ASSOCIATE EDITORS

A. Y. Alanis
University of Guadalajara
MEXICO

M. Akbarzadeh-T
Ferdowsi University
IRAN

P. Benavidez
Univ. Texas, San Antonio
Texas, USA

Y.-L. Chung
National Taiwan Ocean University
TAIWAN

X.-Z. Gao
Helsinki University of Technology
FINLAND

K. El Hindi
King Saud University
SAUDI ARABIA

C.-H. Hsu
Chung Hua University
TAIWAN

J. Hu
Harbin Univ. of Science and Technology
CHINA

M. Joordens
Deakin University
AUSTRALIA

M. Karakose
Firat University
TURKEY

A. Kumar Rav
Indian Inst. of Information Technology
INDIA

W. K. Lai
Tunku Abdul Rahman College
MALAYSIA

D. Li
China Agricultural University
CHINA

H. Marzi
St. Francis Xavier University
CANADA

S. Moein
The University of Arizona
USA

J. Prevost
Univ. Texas, San Antonio
Texas, USA

J. Qiu
Harbin Institute of Technology
CHINA

M. Rajabalinejad
University of Twente
THE NETHERLANDS

K. Tehrani
ESIGEL EC
FRANCE

F. Tian
Chongqing University
CHINA

F. Xu
Hobai University Nanjing
CHINA

Z. Xu
Research Institute of the Ministry of Public Security
CHINA

I. You
Korean Bible University
S. KOREA

C. Zhang
University of Saskatchewan
CANADA

J. Zhao
University of New Haven
USA

EDIToRIAL ADVISoRY BOARD

Chair: C. Nguyen, Chair
Catholic University of America
Washington, DC, USA
Email: nguyen@cua.edu

R. Aliev
Azerbaijan Oil Academy
Baku, AZERBAIJAN

D. Andina
Universidad de Politecnico de Madrid
SPAIN

P. Angelov
University of Lancaster
UK

L. Behera
Indian Institute of Technology
INDIA

Z. Bien
K.A.I.S.T.
Taejon, SOUTH KOREA

G. Bekey
Univ. of Southern California
Los Angeles, CA, USA

H.-C. (Josh) Chao
National Ilan University
Han-Lan, TAIWAN

C. W. de Silva
University of British Columbia
Vancouver, CANADA

Y. Hata
University of Hyogo
JAPAN

A.M. Madni
BEI Technologies, Inc.
Los Angeles, CA, USA

R. Merzouki
Poly tech Lille, FRANCE

G. Muscato
University of Catania
ITALY

S. Nahavandi
Deakin University
AUSTRALIA

G. Parker
Connecticut College
London, CT, USA

D. Ren
China Three Gorges University
CHINA

E. N. Sanchez
CINVESTAV.IPN
Guadalajara, MEXICO

D. Tesar
University of Texas
Austin, TX, USA

E. Tunstel, Jr.
Applied Physics Lab,
JHU Laurel, MD, USA

T. Yamakawa
Kyushu Institute of Technology
Fukuoka, JAPAN

S. Yang
University of Guelph
Guelph, CANADA
Contents

Structure from Motion Using Bio-Inspired Intelligence Algorithm and Conformal Geometric Algebra
Nancy Arana-Daniel, Carlos Villaseñor, Carlos López-Franco, Alma Y. Alanís and Roberto Valencia-Murillo

Erratum: “Verifiable Outsourcing of High-Degree Polynomials and its Application”
Ye, J., Zhou, X., Xu, Z., and Ding, Y.

Comparison of Local Descriptors for Humanoid Robots Localization Using a Visual Bag of Words Approach
Noe G. Aldana-Murillo, Jean-Bernard Hayet and Hector M. Becerra

Application of Multi Agent Systems in Automation of Distributed Energy Management in Micro-grid using MACSimJX
Leo Raju, R. S. Milton and Senthilkumaran Mahadevan

Highly Accurate Recognition of Handwritten Arabic Decimal Numbers Based on a Self-Organizing Maps Approach
Amin Alqudah, Hussein R. Al-Zoubi, Mahmood A. Al-Khassaweneh, and Mohammed Al-Qodah

Synthesis Optimization of Piezo Driven Four Bar Mechanism Using Genetic Algorithm
Laith Sawaqed, Khaled S. Hatamleh, Mohammad A. Jaradat, and Qais Khasawneh

Special Section: Machine Learning and Data Mining for Cyber-Physical Systems

EDITORIAL: Special Issue on Machine Learning and Data Mining for Cyber-Physical Systems
Zheng Xu and Zhiguo Yan

A User Authentication Protocol Combined with Trust Model, Biometrics and ECC for Wireless Sensor Networks
Tao Liu and Gan Huang

The Machine Learning Based Finite Element Analysis on Road Engineering of Built-In Carbon Fiber Heating Wire
Yuhua Peng, Dingyue Chen, Lihao Chen, Jiayu Yu, and Mengjie Bao

Delay-Dependent Stability of Recurrent Neural Networks with Time-Varying Delay
Guobao Zhang, Jing-Jing Xiong, Yongming Huang, Yong Lu, and Ling Wang

A Computable General Equilibrium Model Based Simulation on Water Conservancy Investment
Jun Wang

Optimal Learning Slip Ratio Control for Tractor-Semitrailer Braking in a Turn Based on Fuzzy Logic
Jinsong Dong, Hongwei Zhang, Ronghui Zhang, Xiaohong Jin, and Fang Chen

Kinematic Calibration of Parallel Manipulator for Semi-Physical Simulation System
Dayong Yu
The Virtual Prototype Model Simulation on the Steady-State Machine Performance ...................................................... 581
Huanyu Zhao, Guoqiang Wang, Shuai Wang, Ruipeng Yang, He Tian, and Qiushi Bi

The SLAM Algorithm for Multiple Robots Based on Parameter Estimation ................................................................. 593
MengYuan Chen

The Lateral Conflict Risk Assessment for Low-Altitude Training Airspace Using Weakly Supervised Learning Method .................................................................................................................. 603
Kaijun Xu, Xueting Chen, Yusheng Yao, and Shanshan Li

Simulation and Data Analysis of Energy Recovery Sensing on Parallel Hydraulic Hybrid Crane ......................................... 613
Youquan Chen, Xinrui Liu, Xin Wang, and Jinshi Chen

Intelligent Control for Integrated Guidance and Control Based on Intelligent Characteristic Model ............................................ 623
Jun Zhou, Zhenzhen Ge

NARX Network Based Driver Behavior Analysis and Prediction Using Time-Series Modeling .................................................. 633
Ling Wu, Haoxue Liu, Tong Zhu, and Yueqi Hu

The Data Analyses of Vertical Storage Tank Using Finite Element SOFT Computing .......................................................... 643
Lin Gao, Mingzhen Wang

Association Link Network Based Concept Learning in Patent Corpus .................................................................................. 653
Wei Qin, Xiangfeng Luo

Multi-Phase Oil Tank Recognition for High Resolution Remote Sensing Images ................................................................. 663
Changjiang Liu, Xuling Wu, Bing Mo, and Yi Zhang
Intelligent Automation & Soft Computing - Autosoft is a peer-reviewed journal published by TSI Press, San Antonio, (tsiepress@gmail.com).


Institutional Subscription Rate (print and online): $682
Institutional Subscription Rate (online-only): $499

All current institutional subscriptions include online access for any number of concurrent users across a local area network to the currently available backfile and articles posted online ahead of publication.

Subscriptions purchased at the personal rate are strictly for personal, non-commercial use only. The reselling of personal subscriptions is prohibited. Personal subscriptions must be purchased with a personal cheque or credit card. Proof of personal status may be requested.

Back issues: TSI Press retains a number of years back issue stock of journals. Inquiries can be sent to TSI Press, tsiepress@gmail.com for availability.

Ordering information: Please contact Ms. Sharon Cole (scole48@msn.com) or TSI Press (tsiepress@gmail.com). Checks, PayPal, and credit cards are accepted.

Copyright © 2018 TSI® Press. All rights reserved. No part of this publication may be reproduced, stored, transmitted, or disseminated, in any form, or by any means, without prior written permission from TSI, to whom all requests to reproduce copyright material should be directed, in writing.

Disclaimer: TSI Press makes every effort to ensure the accuracy of all the information (the ‘Content’) contained in its publications. However, TSI Press makes no representations or warranties whatsoever as to the accuracy, completeness or suitability for any purpose of the Content and disclaim all such representations and warranties whether express or implied to the maximum extent permitted by law. Any views expressed in this publication are the views of the authors and are not the views of TSI Press.

TSI Press copyright material for private research use, on the sole basis that requests for such use are referred directly to the TSI Press, tsiepress@gmail.com.

The copyright fee $50.00 exclusive of any charge or fee levied. To order contact TSI Press (tsiepress@gmail.com)

The 2018 US annual subscription price is $526. Airfreight and mailing in the USA by Sundance Print Centers, 20330 Huebner Road, Suite 104, San Antonio, Texas 78258 USA.

Structure from Motion Using Bio-Inspired Intelligence Algorithm and Conformal Geometric Algebra

Nancy Arana-Daniel, Carlos Villaseñor, Carlos López-Franco, Alma Y. Alanís and Roberto Valencia-Murillo

Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara, México

ABSTRACT
Structure from Motion algorithms offer good advantages, such as extract 3D information in monocular systems and structures estimation as shown in Hartley & Zisserman for numerous applications, for instance; augmented reality, autonomous navigation, motion capture, remote sensing and object recognition among others. Nevertheless, this algorithm suffers some weaknesses in precision. In the present work, we extend the proposal in Arana-Daniel, Villaseñor, López-Franco, & Alanís that presents a new strategy using bio-inspired intelligence algorithm and Conformal Geometric Algebra, based in the object mapping paradigm, to overcome the accuracy problem in two-view Structure form motion algorithms. For this instance, we include two new experiments and the inclusion of the circle entity; the circle carries stronger information about its motion than other geometric entities, as we will show.

1. Introduction

The computer vision field has been a very active research topic in last decades, but the problems to solve have proven to be difficult. The research of new techniques and the exploration of new paradigms bring benefits in the development of the field. In the present work, we propose a new strategy to overcome the accuracy problem in Structure from Motion (SfM) algorithms.

SfM algorithms (Hartley & Zisserman, 2003) are a family of computer vision algorithms whose paradigm is based on the relation between the motion and structures, and how this relation is implicit in the images. Using SfM, over two or more images we can calculate structures that represent objects, when we know the motion of the camera or the geometric relation between two or more cameras, as well as we can calculate the motion of known structure.

SfM is based in the epipolar restriction (1.1), where $x_1$ and $x_2$ represent a couple of images of a euclidian point $X$, with respect to two different cameras; $E$ is called the Essential matrix, which relates corresponding points in stereo images, assuming that the cameras satisfy the pinhole camera model.

$$x_2^TEx_1 = 0 \quad (1.1)$$

However, in order to use this useful equation, we have to compute the Essential matrix. This problem is the main difference between SfM variants and has been achieved by a serial of algorithms like Random Sample Consensus (RANSAC) or the eight-points algorithm as is shown in (Hartley & Zisserman, 2003) and (Ma, Soatto, Kosetska, & Sastry, 2004), despite their velocity, they present a serious accuracy problem as we present in the next sections.

In the present time, the Bio-inspired Intelligence Algorithms (BIA) offers good solutions to complex problems, giving competitive times and small errors (Simon, 2013). The principal task in this kind of algorithm is to establish a suitable objective function, this is commonly performing in the linear algebra mathematical framework, but we can obtain good advantages using more general mathematical framework; like Conformal Geometric Algebra (CGA) as we will see in this work.

In this work, we propose a novel technique to overcome the accuracy problem in SfM algorithms with the use of BIA (Hernandez-Vargas, Lopez-Franco, Rangel, Arana-Daniel, & Alanis, 2014) and CGA. With this work, we extend the previous result shown in Arana-Daniel, Villaseñor, López-Franco, & Alanis (2014) to include a projection of the motion estimation error over the image, and in this way to be able to compare the proposal and the classical approach more intuitively and how this can affect the reconstruction of 3D structures. Moreover, we extended the initial proposal to include circles that are rich information geometric entities presented in many real objects. The circle detection is a very spread out research topic (Jiang, 2009; Semeikina & Yurin, 2011).

This paper is organized as follows: Section 2 presents a brief introduction to SfM algorithms, its fundaments, variants, benefits and problems. In Section 3, we present the basis of Geometric Algebra, its representations and operators. Section 4 presents the bio-inspired intelligent algorithm used in order to compute the Essential matrix. Section 5 presents the development of the proposed algorithm. Section 6 shows a comparison between the proposed algorithm and a classic approach with a couple of experiments, while Section 7 is devoted to conclusion and future work.

2. Structure from Motion

As we mentioned in the introduction, if the epipolar restriction holds, then the cameras satisfy the pinhole camera model
Conformal Geometric Algebra is a pseudoeuclidian GA that use a conformal transformation to span the euclidian space, this is denoted by \( \mathbb{R}^m \rightarrow \mathbb{R}^{m+1,1} \). In order to apply it to \( \mathbb{R}^2 \) with basis \( \{ e_1, e_2, e_3 \} \), consider the conformal projection over Minkowski plane with basis \( \{ e_+, e_- \} \) that possess the properties described on (3.1). We can also define a new basis called null basis with (3.2), and finally describe the representation of a tridimensional euclidian point \( x_i \) in CGA constructed over \( \mathbb{R}^{3,3} \) denoted \( G_{4,4} \) in (3.3).

\[
\begin{align*}
\epsilon_+ &= 1, \\
\epsilon_- &= 1, \\
\epsilon_0 &= 0
\end{align*}
\]

\[
\begin{align*}
\epsilon_0 &= 1, \\
\epsilon_+ &= 1, \\
\epsilon_- &= 0
\end{align*}
\]
Table 1. Geometric Entities of CGA.

<table>
<thead>
<tr>
<th>Entity</th>
<th>IPNS</th>
<th>OPNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point</td>
<td>$x = x_0 + \frac{1}{2}x^2e_\omega + e_0$</td>
<td>$x' = s_1 \wedge s_2 \wedge s_3 \wedge s_4$</td>
</tr>
<tr>
<td>Sphere</td>
<td>$s = x_0 + \frac{1}{2}(x - r)^2 e_\omega + e_0$</td>
<td>$s' = s_1 \wedge x_1 \wedge x_2 \wedge s_4$</td>
</tr>
<tr>
<td>Plane</td>
<td>$P = n + de_\omega$</td>
<td>$P' = s_1 \wedge x_1 \wedge x_2 \wedge e_\omega$</td>
</tr>
<tr>
<td>Line</td>
<td>$l = m_i - e_0m_i$</td>
<td>$l' = s_1 \wedge x_1 \wedge x_2 \wedge e_\omega$</td>
</tr>
<tr>
<td>Circle</td>
<td>$C = s_1 \wedge s_2$</td>
<td>$C' = s_1 \wedge x_1 \wedge x_2$</td>
</tr>
</tbody>
</table>

\[ x_c = x_0 + \frac{1}{2}x^2e_\omega + e_0 \quad (3.3) \]

The elements of GA are called multivectors, and for every instance we find two representations; the Inner Product Null Space (IPNS) representation and the Outer Product Null Space (OPNS) representation. The basic geometric entities of CGA are listed in Table 1, where \( r \) represents the radius of a sphere, \( n \) and \( d \) represent the parameters of the Hesse normal form, \( n \) and \( m \) are the Plücker coordinates of a line, \( s_i \) to \( s_4 \) are spheres and \( x_i \) to \( x_4 \) are conformal points.

One of the advantages of CGA is that we are able to use operators over all the geometric objects in CGA. In general, every transformation can be expressed like (3.4), where \( X, X', H \in \mathbb{G}_{n+1} \) and \( \sigma \) is a scalar parameter to ensure the bijection between the null cone and \( \mathbb{R}^7 \).

\[ \sigma X' = HX\bar{H} \quad (3.4) \]

In rigid transformations, the parameter \( \sigma = 1 \). Rotations are represented by the so-called *rotors* defined in (3.5), where \( n \) represents the dual entity of the rotation axis and \( \theta \) is the angle of rotation. Translations are represented with *translators* described in (3.6), where \( t \) is the euclidian translation. Finally we can describe how these operators are used in (3.7).

\[ R = e^{-\frac{1}{2}n} \quad (3.5) \]

\[ T = 1 + \frac{1}{2}te_\omega \quad (3.6) \]

\[ X' = RX\bar{R}, \quad X' = TX\bar{T} \quad (3.7) \]

We can use both equations into a single operator called *motor*, we denote such operator in (3.8). This final equation represents the rigid movement and can be applied on every entity of the algebra.

\[ M = RT, \quad X' = MX\bar{M} \quad (3.8) \]

4. Bio-Inspired Intelligence Algorithm

Bio-inspired intelligence algorithms have been a very important tool for complex problems resolution. Particle Swarm Optimization (PSO) (Eberhart & Shi, 2001) is an evolutionary algorithm based in the observation of bird flocks and insect swarm behaviours that provides high convergence speed. The PSO algorithm consists of an interactive adaptation of a set of multidimensional vectors called particles. These particles communicate information with each other to find a better candidate solution for an objective function.

Nowadays, we can find many variants of the PSO algorithm that intent to alleviate the premature convergence problem in the original PSO. In this work, this new proposed algorithm called Bio-inspired Aging Model Particle Swarm Optimization (BAM-PSO) (Hernandez-Vargas et al., 2014) has been proven to solve the mentioned problem and improve the accuracy results obtained with other PSO variants.

The BAM-PSO algorithm is an evolutionary computation algorithm based in the Aging Leader and Challengers (ALC-PSO) algorithm (Chen et al., 2013). The basic idea in both algorithms is to include an aging process, in the case of BAM-PSO this is extended to each particle in the swarm. ALC-PSO uses a linear aging model. BAM-PSO proposes to use a bio-inspired aging model.

Each particle in the swarm has a velocity \( v_i \) and a position \( x_i \), and they are defined respectively, in Equations (4.1) and (4.2), where \( i \) is the \( i \)-th particle of the swarm, \( j \) is the \( j \)-th element of dimension problem, \( t \) is the iteration counter, \( w \) represents the inertial weight factor (constant), and \( R_j \) are random, normalized and uniformly distributed values, \( c_1 \) and \( c_2 \) represent the social and cognitive parameter respectively, \( x_{ij}(t) \) is the particle \( ij \)-position for the iteration \( t \), \( x_{ij}(t+1) \) is the particle \( ij \)-position for \( t+1 \) iteration, \( v_{ij}(t) \) is the particle \( ij \)-velocity for \( t \) iteration, \( p_{ij}(t) \) represents the local best position for \( ij \)-particle in iteration \( t \), and \( L \) is the leader that holds the best solution found by the swarm at iteration \( t \).

\[ v_{ij}(t+1) = wv_{ij}(t) + c_1R_1(p_{ij}(t) - x_{ij}(t)) + c_2R_2(L_j - x_{ij}(t)) \quad (4.1) \]

\[ x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (4.2) \]

Finally, BAM-PSO includes a measure of premature convergence \( k_j \) in \( j \)-th-dimension defined in (4.3), where \( D \) is the number of the dimension of the problem, \( k_{min} \) is the minimum deviation allowed, and \( \bar{p}_j \) is the mean of all particles in \( j \)-th-dimension.

\[ k_j = \frac{k_{min}}{\sqrt{\sum_{i=1}^{D}(p_{ij}-\bar{p}_j)^2}} \quad (4.3) \]

5. Structure from Motion Using BAM-PSO and CGA

The SfM algorithms based in a bifocal tensor present low accuracy and they are strongly dependent of the exact image correlation. The object mapping approach is the idea of representing objects with a set of geometric entities, and allows compacts and rich information maps. CGA is a suitable framework to represent geometric entities, and recently, mapping algorithms have been developed in this mathematical framework (López-Gonzáles, Arana-Daniel, & Bayro-Corrochano, 2013).

Using the CGA mathematical framework combined with a modern technique of evolutionary computation (BAM-PSO), it is possible to find a suitable SfM algorithm that works with object maps. In order to achieve that, we have to develop the objective function that allows us to find the rigid transformation of an object, represented with various entities of CGA.

To establish a search space for BAM-PSO, we need to define first the parameters in the rigid transformation represented with the \( \bar{M} \) versor described on (5.1), where \( T \) stands for the *translator*, and \( R_x, R_y \) and \( R_z \) are rotors that represent the rotation in the \( x \), \( y \) and \( z \) axis respectively.

\[ M = TR_xR_yR_z \quad (5.1) \]

In order to operate \( M \), let us define \( \theta_x, \theta_y \) and \( \theta_z \) as the rotation angles of \( R_x, R_y \) and \( R_z \) respectively; and \( t_x, t_y \) and \( t_z \) as the euclidian translation on \( T \). Therefore, the search space is defined in (5.2).

\[ \chi = \{ \theta_x, \theta_y, \theta_z, t_x, t_y, t_z \} \quad (5.2) \]
The search space will be bounded for the first three parameters by \([\pi, -\pi]\). Every element and operator of CGA can be represented in a computer with an array of 32 spaces; to get a better management of the computing resources, we can parameterize each geometric entity \(\Omega\). In this case, we can find another virtue of CGA, in the IPNS representation we can easily extract the parameter of almost all the geometric entities as shown in Table 2.

Table 2 shows that in the IPNS representation, the extraction of parameters is easily performed. The case of the point and sphere, is almost trivial extracting a euclidian point and a radius in the case of the sphere. The plane is extracted using the Hesse normal form with four parameters, and for the line is used a normalized Plücker coordinates with six parameters.

In order to extract the parameters of a 3D circle, let us define the circle with a euclidian point \(c\) that represents the circle centre, \(r\) is a radius and for the inclination we can use the parameters of the plane that contain the centre of the circle. In this way, we can obtain \(C \rightarrow \{c, r, n, d\}\) with eight parameters. To get these parameters we use (5.3) published in Hitzer (2005), constructed with three points contained in the circle, these points are solutions to the Equation (5.4), for the plane we expand the circle with \(C \land e_\omega\). An alternative to use these parameters, we can represent the circle like a multi-entity object with three points, this is not a unique representation of the circle, but for the goal to obtain the rigid transformation, it works faster.

The circle entity carries strong information of its motion, i.e. the rigid transformation between two circles (local rotation and translation) has just two possible solutions. To understand this, consider first the case of the point. Between two points there is only one possible translation, but the point itself is ambiguous to rotations, then, for a general rigid transformation between two points, we have infinite solutions. This does not mean that the points are useless for motion estimation, but we need more of them to have a better understanding of the motion. We mean with more points we have a stronger assumption.

The case of the line is similar to the one of the point, for example, between two lines we can always calculate a rotation (with just two solutions), but if we include translations in a general rigid motion, we get infinite possible rigid transformations between them, i.e. the line is ambiguous to translations. The case of spheres is the same of the point, ambiguous with rotations, and the plane is ambiguous to translations.

As we can see, points, spheres, lines and planes have ambiguities for general rigid transformations and if we try to calculate it, we will get infinite solutions. For the BIA, it means that we get infinite equal local minima. This issue is a strong restriction for the proposal, we cannot parallelize it to the entity level, another way to say it is that we cannot optimize the movement of an object with any of these entities independently.

Then, for an object represented with multiple entities, we can argue that with more entities, we get better accuracy. Using points and spheres we are giving strong information about translation and with lines and planes, information about rotation.

Consider again the circle, between two circles, we can easily find the translation between their centres, and for the rotation we have exactly the same solution of the planes that contain those circles. Then, we have two solutions for a general rigid transformation. This part is important for the proposal because we have an entity that contributes with strong information about translation and rotation.

Finally, using the Root of the Mean Squared Error (RMSE), we can obtain a suitable objective function described on (5.5).

\[
\min_x \sum_{i=1}^h \left( \frac{1}{\sum_{i=1}^h [p_i(\Omega^t_i) - p_i(M\Omega^t_{i-1},\bar{M})]^2} \right)
\]

6. Experiments and Results

In order to proof the proposal and to extend the previous results including circles between the geometric entities used, we present two representative experiments that show the quality of the applications that can be developed. In the first instance, we summarize some important particularities of the implementation. We are going to compare the proposed SfM algorithm with a classic two-frame SfM algorithm, both of them need previous data that depends of other algorithms; in the first place, the proposal needs a previous mapping, in the second, Classic SfM needs an exact correspondence algorithm. For a fair comparison, these two processes were performed in a supervised way.

A previous calibration of the camera has been done with Zhang method (Zhang, 2000). The map for the proposed method was constructed using a stereo camera system Bumblebee 2 BB2–08S2 from Point Grey through Triclops Stereo SDK. Both algorithms were programmed in MATLAB platform although it is not the fastest language; we assume the time of comparison is useful. The first experiment was done in an Intel i7-2.600 processor with 8 GB of RAM and the second one in an Intel i7-4,770 processor with 16 GB of RAM. Finally, let us define an error to compare both algorithms in (6.1), where \(x_t\) represents a euclidian point that belongs to a rigid object in the time \(t\) and \(H\) is a homogeneous transformation constructed with the found movement on each algorithm.

\[
E_H = ||x_t - Hx_{t-1}||
\]

6.1. First experiment

Consider the images in Figure 2, we have three different scenes (a), (b) and (c) with a human body, let’s define five objects: 1: Head and trunk, 2: Right arm, 3: Left arm, 4: Right leg and 5: Left leg. In the (d) image we present the map of entities with all
INTELLIGENT AUTOMATION & SOFT COMPUTING

6.2. Second experiment

In this second experiment consider the images in Figure 3 as the input images to our algorithms. Let us define the objects like: 1: Head and trunk; 2: Right arm, 3: Left arm, 4: Thigh, 5: Shin, 6: Front wheel and 7: Back wheel. Figure 4 represents the map for the proposed algorithm, where object 1 is represented with three spheres, objects 2 to 5 are represented with two points and one line, and objects 6 and 7 are represented in the scenes included, where object 1 is represented with three spheres, objects 2 to 5 are represented with two points and one line. It is important to notice how geometric entities represent structures and offer more information than cloud points, but the map is still compact in memory. In Table 3, we show the results.

Note that for all the results on Table 3, the proposed SfM has a significant better accuracy, but the classic SfM wins always in runtime.

Table 3. Results of the First experiment.

<table>
<thead>
<tr>
<th>Object</th>
<th>Proposed SfM</th>
<th>Classic SfM</th>
<th>Proposed SfM</th>
<th>Classic SfM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time (s)</td>
<td>$E_r$ (m)</td>
<td>Time (s)</td>
<td>$E_r$ (m)</td>
</tr>
<tr>
<td>Head and truck</td>
<td>2.945</td>
<td>0.0031</td>
<td>2.1512</td>
<td>1.5112</td>
</tr>
<tr>
<td>Right arm</td>
<td>2.4531</td>
<td>0.124</td>
<td>2.2113</td>
<td>2.0021</td>
</tr>
<tr>
<td>Left arm</td>
<td>4.3458</td>
<td>0.021</td>
<td>2.008</td>
<td>3.1240</td>
</tr>
<tr>
<td>Right leg</td>
<td>4.6521</td>
<td>0.124</td>
<td>1.984</td>
<td>4.5121</td>
</tr>
<tr>
<td>Left leg</td>
<td>5.8412</td>
<td>0.013</td>
<td>2.1542</td>
<td>3.2135</td>
</tr>
</tbody>
</table>

Note: Bold values show the best results of time and error between the classic SfM algorithm and the proposed SfM algorithm.

Figure 2. Input Images and Map for the First Experiment. Images (a), (b), and (c) are the Input Images for the Algorithms and (d) is the Map with Geometric Entities, which Represent the Objects on all the Input Images.

Figure 3. Input Images for the Second Experiment. Note the Presence of Circle Shapes in Several Objects.

Figure 4. Map for the Proposed Algorithm.
7. Conclusion and Future Work

As we have shown in the experimentation, the proposed algorithm presents a novel method to use SfM with the object mapping paradigm. The use of bio-inspired intelligent algorithms over the suitable mathematical framework like CGA, allow us to accomplish an algorithm that overcomes the accuracy problem in Two-frame SfM algorithms. The proposed algorithm exhibits excellent results in the error, but it is slower than the classic one.

The principal improvement in comparison with Arana-Daniel et al. (2014), is the inclusion of the circle entity, as we already said, the circle carries strong information about its motion, because it has information about the change in orientation and translation, this property leads to a function with fewer local minima to optimize. Thanks to this property, we can obtain the result so accurate and fast for the front and back wheel in Table 4. Including this entity in a multi-entities object will improve the performance of the algorithm.

As future work, the present paper can be extended to other Geometric Algebras using the same objective function with new parametrization. We can also parallelize the BAM-PSO algorithm to get better times. In the case of the Classic SfM algorithm, we can also parallelize the singular value decomposition, as is shown in (Luk, 1985), and compare again the algorithms.

Acknowledgements

The authors thank the support of CONACYT México, through Projects CB256769 and CB258068 (Project supported by Fondo Sectorial de Investigación para la Educación).

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Nancy Arana-Daniel received in 2003 an M. S. degree in Computer Science and a Ph.D. in Computer Science in 2007, both from Center of Research and Advanced Studies, CINVESTAV, Unidad Guadalajara, México. She is currently a research fellow at the University of Guadalajara, in the Department of Computer Science México, where she is working in the Laboratory of Intelligent Systems and the Research Center on Control Systems and Artificial Intelligence. Her research interests focus in applications of geometric algebra, machine learning, optimization, pattern recognition and robot navigation.

Table 4. Results of Second Experiment.

<table>
<thead>
<tr>
<th>Object</th>
<th>Propose SfM</th>
<th></th>
<th>Classic SfM</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time (s)</td>
<td>$E_H$ (m)</td>
<td>Time (s)</td>
<td>$E_H$ (m)</td>
</tr>
<tr>
<td>Head and Truck</td>
<td>2.1765</td>
<td>0.2450</td>
<td>1.5857</td>
<td>1.2879</td>
</tr>
<tr>
<td>Right arm</td>
<td>3.6910</td>
<td>0.1544</td>
<td>1.5333</td>
<td>1.5482</td>
</tr>
<tr>
<td>Left arm</td>
<td>2.7572</td>
<td>0.0565</td>
<td>1.6499</td>
<td>2.1046</td>
</tr>
<tr>
<td>Thigh</td>
<td>2.8906</td>
<td>0.0045</td>
<td>1.5291</td>
<td>3.2145</td>
</tr>
<tr>
<td>Shin</td>
<td>3.1729</td>
<td>0.1287</td>
<td>1.2430</td>
<td>2.846</td>
</tr>
<tr>
<td>Front wheel</td>
<td>0.9612</td>
<td>0.0476</td>
<td>1.3456</td>
<td>0.5946</td>
</tr>
<tr>
<td>Back wheel</td>
<td>0.5579</td>
<td>0.1151</td>
<td>1.4691</td>
<td>0.7642</td>
</tr>
</tbody>
</table>

Note: Bold values show the best results of time and error between the classic SfM algorithm and the proposed SfM algorithm.

with one circle, in order to shown the strong assumption of using circles.

Table 4 shows the results of the second experiment. Finally, in Figure 5 we project the error in the second image, to present some intuition in the comparison.

In the same way like the first experiment, the proposed algorithm presents results with better accuracy, and in this case, two of the objects have a better runtime. Actually, these two objects contain a circle, with this we can argue experimentally that the circle has strong information of its motion. Moreover, the classic SfM in six of the cases has a better runtime.

In Figure 5, we have taken randomly one point in every object with the mark “*”, with this we can calculate the result between two images and we re-project the resultant point with “o”. Then the line in both images are the re-projection of the movement estimation error, where is easy to see how our proposal presents a smaller error in comparison than the classic approach.

Figure 4. Object Map for the Second Experiment.

Figure 5. Re-projection of the Movement Estimation Error. (a) Proposed SfM, (b) Classic SfM. The “*” Mark Represents the Correct Point and “o” Mark Represents the Estimated Point with the Respect Algorithm. Note how the Proposal Represents Smaller Errors in Comparison to the Classic Approach.
Carlos Villaseñor is a Ph.D. student in the Computer Sciences Department, in the Centro Universitario de Ciencias Exactas e Ingeniería at the Universidad de Guadalajara in México. He received a Master's degree in Computer Science in 2014 and a B.Sc. degree in Industrial Engineering at the same university. His research focuses in bio-inspired algorithms and applications of geometric algebras.

Carlos López-Franco received a Ph.D. degree in Computer Science in 2007 from the Center of Research and Advanced Studies, CINVESTAV, Mexico. He is currently a professor at the University of Guadalajara, México, Computer Science Department, and member of the Intelligent Systems group. His research interests include geometric algebra, computer vision, robotics and intelligent systems.

Alma Y. Alanís, received a Ph.D. degree in Electrical Engineering from the CINVESTAV-IPN, Guadalajara Campus, Mexico, in 2007. Since 2008 she has been with University of Guadalajara, where she is currently a chair professor in the Department of Computer Science. She is also member of the Mexican National Research System (SNI-2) and member of the Mexican Academy of Sciences. She has published papers in recognized international journals and conferences, besides four international books. Her research interest centers on neural control, backstepping control, block control, and their applications to electrical machines, power systems and robotics.

Roberto Valencia-Murillo is a Ph.D. student in Computer Science Department, at the Centro Universitario de Ciencias Exactas e Ingeniería at Universidad de Guadalajara in México. He has received a Master’s degree in Computer Science in 2013 and a B.Sc. degree in Computer Science at the University of Baja California. His research interests focus in machine learning, robot navigation and geometric methods for robotics.

References


Erratum

http://dx.doi.org/10.1080/10798587.2016.1267239

When the above article was first published online, J. Ye's affiliation was incorrectly displayed as
“Guangxi Key Laboratory of Cryptography and Information Security, Lab of Security Insurance of Cyberspace, China”
instead of
“Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things, Sichuan University of Science & Engineering, Guangxi Key Laboratory of Cryptography and Information Security, Lab of Security Insurance of Cyberspace, China”

This has now been corrected.

Taylor & Francis apologises for this error.
Introduction

Visual memory mimics the human behavior of remembering key visual information when moving in unknown environments, to make the future navigation easier in a topological map. This metric-free methodology has been extensively studied in the context of wheeled mobile robots (Becerra, 2014; Courbon, Mezouar, & Martinet, 2009; Diosi, Segvic, Remazeilles, & Chaumette, 2011). However, much fewer studies are reported in the context of humanoid robots. This kind of robots raises particular challenges that are not considered in the aforementioned works. In particular, because of the bipedal locomotion, sharp accelerations produce blurring effects on the images; also, the robot sway motion produces rotations around the optical axis.

The navigation of a robot based on a visual memory typically implies two distinct stages (Courbon et al., 2009; Diosi et al., 2011). First, the learning stage consists in making the robot build a representation of the initially unknown environment, by means of a set of key images that forms the so-called visual memory. Then, in the autonomous navigation stage, the robot has to reach a location associated to a desired key image by following a visual path. That path is defined by a subset of images from the visual memory that topologically connects the key image that is the most similar to the current robot view to the target image. The autonomous navigation stage may be addressed assuming that the visual path is given, like in Becerra, 2014; Becerra, Sagüés, Mezouar, & Hayet (2014), where the control laws are formulated in terms of a geometric constraint, with nonlinear and fuzzy approaches, respectively.

A visual memory is built by selecting a subset of images from a sequence of images captured in a learning stage. The basic criterion for the selection of a key image in the learning stage is to satisfy a compromise between two aspects; a pair of consecutive key images must share enough visual information (for instance a minimum number of matched interest points) and, at the same time, we do not want too many images. The construction of the visual memory is a problem by itself. In this work, we assume that the visual memory is given and we focus on the problem of localizing a robot within this visual memory. Moreover, we assume that the robot is initially placed at an unknown location with no prior knowledge about this location or about where it was previously. This is known as the kidnapped robot problem (Thrun, Burgard, & Fox, 2006).

Few works have been reported about the navigation of humanoid robots based on a visual memory (Delfín, Becerra, & Arechavaleta, 2014; Ido, Shimizu, Matsumoto, & Ogasawara, 2009). In the aforementioned works, the robot is not initially kidnapped, as in our setup, but instead starts its navigation from a known position. Hence, compared to these works, one main motivation in this paper is to propose a solution to the appearance-based localization problem, where the current image is matched to a known location only by comparing images (Ulrich & Nourbakhsh, 2000). In particular, we address the problem of the localization of humanoid robots using monocular vision only, by determining the key image in a visual memory that is the most similar in appearance to the current view of the robot (input image). Figure 1 presents an overview of the problem. Consider that the visual memory consists of $n$ key images ($I_1^*, I_2^*, \ldots, I_n^*$). The current view $I$ has to be compared (a priori) with the $n$ key images and the method should give us as an output the most similar key image $I_i^*$ within the visual memory, from which the visual path to the target image could be defined.

Since a naive comparison of $I$ with all the visual memory would take too much time, depending on the size of the visual memory, we use a method that compresses the visual memory into a compact, efficient-to-access representation, the visual bag of words VBoW (Sivic & Zisserman, 2003). A bag of words
is a structure that represents an image as a numerical vector, allowing fast images comparisons. In robotics, the VBoW approach has been used in particular for loop-closure detection in Simultaneous Localization and Mapping (SLAM) (Botterill, Mills, & Green, 2011; Galvez-López & Tardos, 2012), where re-visited places have to be recognized to manage the proper closure of loops while building maps.

We want to emphasize that our appearance-based localization is part of a more global visual memory approach, which relies on navigating with visual information extracted directly from images, without the need of costly estimation machinery such as SLAM. The SLAM methods may provide a more accurate localization in terms of metric information, but it has been shown that appearance information is sufficient to navigate (Delfin et al., 2014). In this paper, a qualitative evaluation of the VBoW approach for the visual navigation problem is carried out. In particular, we apply the VBoW approach on real datasets captured by a camera mounted in the head of a small-size humanoid robot. An evaluation and comparison of the performance of different local descriptors are presented. The images taken by the humanoid robot are affected by the sway motion due to its locomotion. They undergo blurring and rotation around the optical axis. Hence, a specific visual vocabulary is proposed to tackle those issues. Figure 2 shows two examples of 640 × 480 pixels images captured from our experimental platform; a NAO humanoid robot, where blur and rotation effects are visible.

This paper is an extension to a previously published conference paper, presented at the Mexican Conference on Pattern Recognition (Aldana-Murillo, Hayet, & Becerra, 2015), it is organized as follows: Section 2 introduces the local descriptors included in our evaluation. Section 3 details the VBoW approach as we implemented it. Section 4 presents the results of the experimental evaluation and Section 5 gives a few conclusions about the approach.

2. Local Descriptors

Local features describe regions of interest of an image through descriptor vectors, defined around special image locations, given by local feature detectors. In the context of image comparison, the idea is that groups of local features should be robust against occlusions and viewpoint changes, in contrast to global methods, both in terms of the features locations and in terms of the local descriptors. Hence, from the existing local detectors/descriptors, we wish to select the best option for the specific task of appearance-based humanoid localization. Hereafter, we introduce the local descriptors selected for a comparative evaluation.

2.1. Real-valued Descriptors

A popular system of combined keypoint detector/descriptor is SURF (Speeded Up Robust Features (Bay, Ess, Tuytelaars, & Van Gool, 2008)). It has good properties of invariance to scale and rotation. SURF keypoints can be computed and compared much faster than their previous competitors. Thus, we only selected SURF as a real-valued descriptor to be compared in our localization framework. The detection uses local extrema of the Hessian matrix and makes an approximation of the Hessian with integral images, to reduce the computation time. The descriptor combines approximate Haar-wavelet responses within the interest point neighborhood and also exploits integral images to increase speed. In our evaluation, the standard implementation of SURF (descriptor vector of dimension 64) included in the OpenCV library is used, with 4 octaves and 2 layers in each octave.

2.2. Binary Descriptors

Binary descriptors represent image features by binary strings instead of floating-point vectors. Thus, the extracted information is very compact, occupies less memory and can be compared faster. Two popular binary descriptors have been selected for our evaluation; Binary Robust Independent
Elementary Features (BRIEF (Calonder, Lepetit, Strecha, & Fua, 2010)) and Oriented FAST and Rotated BRIEF (ORB (Rublee, Rabaud, Konolige, & Bradski, 2011)). Both use variants of FAST (Features from Accelerated Segment Tests (Rosten & Drummond, 2006)) as a detector, i.e. they detect keypoints by comparing the gray levels along a circle of radius 3 to the gray level of the circle center. In average, most pixels can be discarded soon, hence the detection is fast. BRIEF uses the standard FAST keypoints, while ORB uses oFAST keypoints, an improved version of FAST including an adequate orientation component. The BRIEF descriptor is a binary vector of user-choice length where each bit results from an intensity comparison between some pairs of pixels within a patch around the detected keypoints. The patches are previously smoothed with a Gaussian kernel to reduce noise. They do not include information of rotation or scale, so they are hardly invariant to them. This issue can be overcome by using the rotation-aware BRIEF descriptor (ORB) that computes a dominant orientation between the center of the keypoint and the intensity centroid of its patch. The BRIEF comparison pattern is rotated to obtain a descriptor that should not vary when the image is rotated in the plane. In our evaluation, we use oFAST keypoints given by the ORB detection method as implemented in OpenCV along with BRIEF with a patch size of 48 and a descriptor length of 256. The ORB implementation is the one of OpenCV with 256 bits descriptors.

2.3. Color Descriptors

We also evaluate the bag-of-words image comparison approach with color information only. To do so, we use rectangular patches and a color histogram is associated to each patch as a descriptor. We select the HSL (Hue-Saturation-Lightness) color space, because its three components are more natural to interpret and less correlated than in other color spaces. Also, only the H and S channels are used, in order to achieve as much robustness as possible against illumination changes. The color descriptor of each rectangular patch is formed by a two-dimensional histogram of hue and saturation and the length of the descriptor was set experimentally to 64 bits. Three different alternatives are evaluated to select the patches:

- Random patches: A number of $48 \times 64$ patches randomly selected. Hereafter, this option is referred to as Color-Random.
- Uniform grid: A uniform grid of patches covering the image, with patches overlapped by half of their size. This option is referred to as Color-Whole.
- Uniform grid on half of the image: Instead of using the whole image, only the upper half is used. This is because the inferior parts, when taken by the humanoid robot, are mainly projections of the floor and do not discriminate well among possible locations. This option is referred to as Color-Half.

3. Visual Bag of Words for Humanoids Localization

As mentioned above, this work relies on the hierarchical visual bag of words approach (Nister & Stewenius, 2006) to combine the highly descriptive power of local descriptors (see above) with the versatility and robustness of histograms. In Section 3.1, we recall the main characteristics of the work presented in Nister & Stewenius (2006), and then in Section 3.2, we introduce a novel use of the BRIEF descriptor suited within a VBoW approach in the context of humanoid robots navigation.

3.1. Hierarchical Visual Bag of Words Approach

The visual bag of words approach starts by discretizing the local descriptors space in a series of words, which can be considered as clusters in the local descriptors space. Obviously, the nature of this space (histograms, responses to filters) depends on the selected descriptors. Here, we follow the strategy of (Nister & Stewenius, 2006), who performs this discretization step in a hierarchical way. In the set of $n$ acquired key images $I_1, I_2, \ldots, I_n$ forming the visual memory, a pool of D local descriptors is detected, as illustrated in Figure 3 left. The local descriptors can be extracted by any of the methods mentioned above. Given a branch factor $k$, the idea is to form $k$ clusters among the D descriptors by using the kmeans++ algorithm. Then, the sets of descriptors associated to these $k$ clusters are recursively clustered into other $k$ clusters, and so on, up to a maximum depth of $L$ levels, as depicted in Figure 4. The leaves of this tree of recursively refined clusters correspond to the visual words, i.e., the clusters in the local descriptors space.

When handling a new image $I$, $d$ descriptors are extracted, and each of these is associated to a visual word as explained above. This way, we obtain an empirical distribution of the visual words $v(I)$ (see Figure 3, right). Now, the content of $I$ can be compared with any of the key images $I^*_j$ by comparing their words in histograms. Of course, because $n$ may be very high, it is out of question to compare the histogram of $I$ with the $n$ histograms of the key images. That is why an important element in this representation is the notion of inverse dictionary; for each visual word, one stores the list of images from the visual memory containing this word. Then, on a new image, we can easily determine, for each visual word it contains, the list of key images also containing this word. This way, to limit the number of comparisons, we restrain the search for the most similar images to the subset of key images having at least 5 visual words in common with the test image.

For an image $I$, each histogram entry $v_i$ (where $i$ refers to the visual word) is defined as:
\[ v_i = \frac{c(I)}{n} \log \left( \frac{n}{n^i} \right). \]  

Where \( c(I) \) is the total number of descriptors present in \( I \), \( c(I) \) the numbers of descriptors in \( I \) classified as word \( i \), and \( n \) the number of key images where the word \( i \) has been found. The log term allows to weight the frequency of word \( i \) in \( I \) in function of its overall presence; if a word is present everywhere in the database (\( n^i = n \)), then the information of its presence will not help in discriminating among images.

Last, we should choose how to compare histograms. After intensive comparisons made among the most popular metrics for histograms such as dot product, \( \chi^2 \), Bhattacharyya coefficient, \( L_1\text{-norm} \) and \( L_2\text{-norm} \), we have observed that the best results were obtained with the \( \chi^2 \) distance.

Hence, for any new image \( I \), we obtain a score against any image from the visual memory sharing at least 5 words with it, Figure 5 sums up the whole methodology.

### 3.2. A BRIEF-based Vocabulary for Humanoid Robots Localization

We introduce a novel use of the BRIEF descriptor, suited for a VBoW approach in the context of humanoid robots. This is a specific vocabulary that we called BRIEFROT, which deals with the issues generated by the humanoid locomotion. As mentioned above, the humanoid locomotion generates rotated images within a certain rotation angle due to the lateral sway motion, as can be seen in Figure 6. Due to this, the detected interest points may appear rotated relatively to how they have been learned in the visual memory. However, the majority of detected points rotate with a similar angle with respect to their learned appearance. Hence, we apply rotations that span the possible values of this transformation to all the detected BRIEF points and generate a new vocabulary: BRIEFROT.

This comes as an alternative to ORB descriptors that are invariant to rotation, as BRISK or FREAK. However, for these rotational invariant descriptors, each point has an intrinsic orientation that may correspond to a different rotation angle for each descriptor when it is not correctly estimated. This may lead to the non-detection of some of the visual words present in the image. Our proposed BRIEFROT vocabulary rotates all the detected descriptors to a certain fixed value by trying to imitate the real rotation of the points in the images, because of the robot’s sway motion and, therefore, more visual words can be retrieved. Also, we make the computation times lower, by avoiding the computation of the intrinsic orientation.

The humanoid locomotion also generates blurry images. To try to mitigate this problem, patches around the detected points are smoothed with a Gaussian kernel to reduce the difference between descriptors of sharp images and descriptors of blurry images.

The BRIEFROT vocabulary contains three independent internal BRIEF vocabularies, two of which are rotated at a fixed angle; one anti-clockwise, the other clockwise. To create these vocabularies, BRIEF points are extracted and they are rotated by an adequate angle to generate the rotated vocabularies. The third vocabulary is identical to the normal BRIEF. The angle of rotation was obtained experimentally by varying the angle to obtain the largest number of correct results, as will be described in Section 4.3. Formally speaking, we have three vocabularies, generated from the three rotations. Hence, for any new image \( I \), we can define \( z^{(o)}_{\chi^2}(I) \), the score between image \( I \) and the \( i \)-th image of the vocabulary \( o \in \{1, 2, 3\} \). Note that by construction, the key images are the same in all cases; what is different is the set of features contained in the vocabulary. Then, the overall score is simply defined as:

\[ z_{\chi^2}(I) = \max_{o \in \{1, 2, 3\}} z^{(o)}_{\chi^2}(I). \]  

These rotated vocabularies were implemented with the idea of coping with the slight rotation caused by the locomotion of these robotic systems. The rotated vocabularies represent the images of the visual memory as rotated features. When evaluating a new image, we evaluate it under the three vocabularies separately and the winner is the vocabulary that obtains the maximum result.

The idea of using three vocabularies is that if the input image is rotated with respect to any image of the visual memory, then the image should be detected by one of the rotated vocabularies; if the input image is not rotated with respect to any image of the visual memory, then it is detected with the vocabulary without rotation.

### 4. Experimental Evaluation

We evaluated the local descriptors mentioned in Section 2 on 5 distinct datasets. Four of these datasets correspond to indoor environments (CIMAT-NAO-A, CIMAT-NAO-B, Toulouse (Alcantarilla, Stasse, Druon, Bergasa, & Dellaert, 2013) and Bicocca (Bonarini et al., 2006)) and one corresponds to an outdoor environment (New College (Smith, Baldwin, Churchill, Paul, & Newman, 2009)). Table 1 summarizes the datasets.
characteristics. The tests were done with a laptop using Ubuntu 12.04 with 4 Gb of RAM and a 1.30 GHz processor.

### 4.1. Description of the Evaluation Datasets

The CIMAT-NAO-A data-set was acquired with a NAO humanoid robot inside CIMAT. This data-set contains 640 × 480 images of good quality, but also many blurry ones. Some images are affected by rotations introduced by the humanoid locomotion or by lighting changes. We have used 187 images, hand-selected, as a visual memory and 258 images for testing. Figure 7 shows a map of the locations associated to the images of the visual memory for the CIMAT-NAO-A data-set. This metric information was obtained from the robot’s odometry for visualization purposes and it is not used by our localization method. The CIMAT-NAO-B data-set was also captured indoors at CIMAT, with the same humanoid robot. It also contains good quality and blurry 640 × 480 images, but it does not have images with drastic light changes, as in the previous data-set. We have used 94 images as a visual memory and 94 images for testing.

Table 1. Datasets Used for the Evaluation of the Local Descriptors.

<table>
<thead>
<tr>
<th>Data-set</th>
<th>Description</th>
<th>Image size (px×px)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIMAT-NAO-A</td>
<td>Indoors/humanoid</td>
<td>640×480</td>
</tr>
<tr>
<td>CIMAT-NAO-B</td>
<td>Indoors/humanoid</td>
<td>640×480</td>
</tr>
<tr>
<td>Bicocca 2009–02-25b</td>
<td>Indoors/Wheeled robot</td>
<td>320×240</td>
</tr>
<tr>
<td>New College (Smith et al., 2009)</td>
<td>Outdoors/Wheeled robot</td>
<td>384×512</td>
</tr>
<tr>
<td>Toulouse (Alcantarilla et al., 2013)</td>
<td>Indoors/humanoid</td>
<td>572×390</td>
</tr>
</tbody>
</table>

The Toulouse data-set was acquired by the humanoid robot HRP-2 inside a laboratory and it is available online (Alcantarilla et al., 2013). It contains 572 × 390 grayscale images of good quality and very few blurry images. Moreover, compared with CIMAT-NAO-A, the images have almost zero rotation around the optical axis. The Bicocca 2009–02-25b data-set is also available online (Bonarini et al., 2006) and was acquired by a wheeled robot inside a university. The 320 × 240 images have no rotation around the optical axis nor blur. We used 120 images as a visual memory and 120 images for testing. Unlike the three previous datasets that were obtained indoors, the New College data-set was acquired outside the Oxford University by a wheeled robot (Smith et al., 2009), with important light changes. The 384 × 512 images are of good quality with no rotation nor blur. For this data-set, 122 images were chosen as a visual memory and 117 images for testing. The last two datasets were used to compare the visual vocabularies obtained by wheeled robots with the images obtained by humanoids robots, in order to make clearer the acquisition issues arising in the second case.

### 4.2. Evaluation Metrics

Since the goal of this work is to evaluate different descriptors in a VBoW approach, it is critical to define adequate evaluation metrics to assess the quality of the result from our application. We propose two metrics that compare the ground-truth data to the algorithm output and generate a score; the first one is the ranking (according to the bag-of-words algorithm) of the theoretically most similar image:

\[ \mu_1(I) = \text{rank} \left( k(I) \right), \]

(3)
Where $k(I)$ is defined as the ground truth index of the key image associated to $I$. In the best case, the rank of the closest key image in the list of closest images given by our algorithm should be one, so $\mu_1(I) = 1$ means that the retrieval is perfect, whereas higher values correspond to worse evaluations.

The second metric is similar in essence:

$$\mu_2 = \sum_l \frac{z_l(I_{k(I)})}{\sum_l z_l(I_{k(I)})} \text{rank}(l), \quad (4)$$

Where $z_l(I)$ is the similarity score between the key image $l$ from the visual memory and the image $I$. Hence, $z_l(I_{k(I)})$ refers to the similarity score between the key image $l$ and the key image $k(I)$ (the theoretically closest image to ours, according to the ground truth). This metric is “fair” in the sense that it handles the presence of close, similar key images within the data-set; hence, with this metric, the final score integrates weights (normalized by $\sum_l z_l(I_{k(I)})$ to sum to one) from the key images $l$ similar to the closest ground truth image $k(I)$; this ensures that all the closest images are also well ranked.

### 4.3. Parameters Selection

There are three free parameters in the VBoW algorithm; the number of clusters $k$ at each level of the tree, the tree depth $L$, and the measure of similarity. Tests were performed by varying the value of $k$ from $k = 8$ to $k = 10$ and varying the value of $L$ from $L = 4$ to $L = 8$. Also, different similarity measures between histograms were tested; $L_1$-Norm, $L_2$-Norm, $\chi^2$, Bhattacharyya and dot product (note that some of them are distances, other are similarity coefficients). To do the tests, we generated ground truth data, by defining manually the most similar key image to the input image, with which we can define the evaluation metrics $\mu_1$ and $\mu_2$ explained before. By examining the statistics of $\mu_1$ scores, we obtained the best results with $k = 8$, $L = 8$ and the metric $\chi^2$. The data-set used for the parameters selection was the CIMAT-NAO-A, since it is the most challenging data-set for the type of images it contains.

For interest-point feature detection, an important parameter to consider in the whole approach is the number of features to detect and to use to build the word histogram based on the vocabulary. In Figure 8 we show effectiveness results (proportion of test images for which $\mu_1 = 1$) obtained for different features and different numbers of points, for the CIMAT-NAO-A data-set. Based on these results, we chose the number of 500 extracted features and discretized into words, in all the following experiments.

A specific parameter to be adjusted for the BRIEFROT vocabulary is the angle of rotation for the rotated vocabularies. The tests were done by varying the angle from 0 to 20 degrees. In Figure 9, the results are presented for different numbers of interest points, again in the case of the CIMAT-NAO-A data-set. The best results occur when the angle of rotation is 8 degrees. Therefore, the rotation value was set at this value and with 500 interest points. Note that for 750 points, better results were obtained, but we chose the 500 points, because it requires...
classified as correct was 1 (i.e., a perfect match). On the other hand, for the level of confidence $\mu_2$ we choose a threshold of 2.5, so that all tests showing scores below 2.5 were considered as correct. This second level of confidence is looser, but takes into account the similarity between images within the visual memory.

4.4.2. CIMAT-NAO-A data-set

In Table 2, we present results obtained for the CIMAT-NAO-A data-set. In this case, the BRIEF vocabulary obtained the best behavior for both levels of confidence. For the case of $\mu_1$, it has an efficiency of 60.85% (measured as the proportion of correct test cases, i.e. such that $\mu_1 = 1$) and for $\mu_2$ of 75.19% (proportion of correct test cases, i.e. such that $\mu_2 < 2.5$). Also, the ORB vocabulary offered good performance for $\mu_2$ and was the second best for $\mu_1$. The Color-Half vocabulary obtained the worst results.

4.4.3. CIMAT-NAO-B data-set

In Table 3 we present results for the CIMAT-NAO-B data-set. It can be noticed that the SURF vocabulary behaved better than BRIEFROT, but with higher computation times. The times reported were measured from the stage of features extraction to the stage of comparison. ORB, again, behaved well and was the second best for $\mu_1$. The Color-Random vocabulary obtained the worst performance for $\mu_1$, but for $\mu_2$ it was one of the best vocabulary; this means that it tends to put the correct key image in the second rank. Color-Half had the worst results for $\mu_2$.

4.4.4. Toulouse data-set

In the Toulouse data-set, we have images with radial distortion due to the robot cameras and we un-distort them for testing. We do not use color vocabularies, because the images are grayscale.

Figure 10. Performance Evaluation of the Color-Half and Color-Whole for Different Numbers of Partitions of the Image.

Table 2. Percentages of Correct Results for the CIMAT-NAO-A Data-set.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Number of tests</th>
<th>Correct tests $\mu_1$</th>
<th>Effectiveness $\mu_1$ (%)</th>
<th>Correct tests $\mu_2$</th>
<th>Effectiveness $\mu_2$ (%)</th>
<th>Average time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRIEF</td>
<td>258</td>
<td>132</td>
<td>51.16</td>
<td>185</td>
<td>71.71</td>
<td>122.6</td>
</tr>
<tr>
<td>BRIEFROT</td>
<td>258</td>
<td>157</td>
<td>60.85</td>
<td>194</td>
<td>75.19</td>
<td>132.4</td>
</tr>
<tr>
<td>Color-Random</td>
<td>258</td>
<td>110</td>
<td>42.64</td>
<td>117</td>
<td>45.35</td>
<td>129.4</td>
</tr>
<tr>
<td>Color-Half</td>
<td>258</td>
<td>104</td>
<td>40.31</td>
<td>160</td>
<td>62.01</td>
<td>93.1</td>
</tr>
<tr>
<td>Color-Whole</td>
<td>258</td>
<td>110</td>
<td>42.64</td>
<td>162</td>
<td>62.79</td>
<td>101.8</td>
</tr>
<tr>
<td>ORB</td>
<td>258</td>
<td>144</td>
<td>55.81</td>
<td>194</td>
<td>75.19</td>
<td>107.5</td>
</tr>
<tr>
<td>SURF</td>
<td>258</td>
<td>135</td>
<td>52.32</td>
<td>187</td>
<td>72.48</td>
<td>296.5</td>
</tr>
</tbody>
</table>

The underlines correspond to the vocabularies that obtained the best performance for the two proposed metrics.
In Table 4, the results for this data-set, after image un-distortion, are presented. The BRIEFROT vocabulary achieved the best results for both evaluation metrics. The BRIEF vocabulary also obtained good performance, very similar to BRIEFROT. This is, because the images of this data-set have almost zero rotation with respect to the camera optical axis. This can be seen in Figure 11, in which we evaluate different BRIEFROT rotation angles. It can be seen that there is no clear angle value for which optimal results are obtained. This is due to the fact that the zero-degree vocabulary of BRIEFROT can explain all the images by itself. In Figure 12, this is illustrated by a plot showing the frequencies of vocabularies (among the three) having the largest response, on average for this data-set. In Figure 13, we show the same information for the CIMAT-NAO-A data-set. The vocabularies with rotation have more detections for this data-set. Also, the vocabulary rotated by θ degrees has significantly more detections than the vocabulary rotated by -θ degrees. This is, because there are more images in the data-set rotated in that direction.

Tests were also performed on the images with radial distortion. In this case, the effectiveness decreases significantly with respect to the results obtained without radial distortion, as it can be seen in Figure 14. Table 5 shows that the BRIEF vocabulary obtained a higher number of correct results, followed by

### Table 3. Percentages of Correct Results for the CIMAT-NAO-B Data-set.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Number Correct Effectiveness</th>
<th>Correct Effectiveness Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Correct</td>
</tr>
<tr>
<td></td>
<td>of tests</td>
<td>μ₁ μ₁ (%)</td>
</tr>
<tr>
<td>BRIEF</td>
<td>94</td>
<td>63</td>
</tr>
<tr>
<td>BRIEFROT</td>
<td>94</td>
<td>65</td>
</tr>
<tr>
<td>Color-Random</td>
<td>94</td>
<td>62</td>
</tr>
<tr>
<td>Color-Half</td>
<td>94</td>
<td>64</td>
</tr>
<tr>
<td>Color-Whole</td>
<td>94</td>
<td>68</td>
</tr>
<tr>
<td>ORB</td>
<td>94</td>
<td>69</td>
</tr>
<tr>
<td>SURF</td>
<td>94</td>
<td>70</td>
</tr>
</tbody>
</table>

### Table 4. Percentages of Correct Results for the Toulouse Data-set with Rectified Images.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Number Correct Effectiveness</th>
<th>Correct Effectiveness Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Correct</td>
</tr>
<tr>
<td></td>
<td>of tests</td>
<td>μ₁ μ₁ (%)</td>
</tr>
<tr>
<td>BRIEF</td>
<td>770</td>
<td>562</td>
</tr>
<tr>
<td>BRIEFROT</td>
<td>770</td>
<td>564</td>
</tr>
<tr>
<td>ORB</td>
<td>770</td>
<td>554</td>
</tr>
<tr>
<td>SURF</td>
<td>770</td>
<td>536</td>
</tr>
</tbody>
</table>

![Figure 11](image1.png) Effectiveness of BRIEFROT for Different Degrees of Rotation for Toulouse Data-set with Rectified Images.

![Figure 12](image2.png) Frequency that a Vocabulary gets a Minimum Distance for Toulouse Data-set with Rectified Images.
4.4.6. New College data-set

With the New College data-set, in Table 7, the SURF vocabulary obtained the best behavior for both evaluation metrics. In both cases, the BRIEFROT vocabulary obtained a good behavior, close to SURF, but BRIEFROT consumes less than half of the time required by SURF.

As an illustration, we give in Figure 15 and Figure 12 a few examples of test images for the BRIEFROT vocabulary from the views represented in Figure 7 as blue robot locations. In Figure 16, images evaluated with $\mu_1 = 1$ (i.e., perfect match) to have the $I^*_{ij}$ as closest key image are shown; similarly, in Figure 16 we present
We collected the average value of the proportion of highest frequency output considering $\mu_1 = 1$. As seen in Table 8, BRIEF, BRIEFROT and SURF have particularly high repeatability rates.

### 5. Conclusions

The problem of appearance-based localization of humanoid robots is tackled in this paper, which consists in determining the most similar image among a set of previously acquired images (visual memory) to the current robot view. We used a hierarchical visual bag of words (VBoW) approach to achieve this goal. We evaluated and compared the performance of different local descriptors used to feed the VBoW method. Real-valued, binary and color descriptors were compared on real datasets captured by humanoid robots, in particular with a small-size humanoid platform (a NAO robot). We presented a novel use of the BRIEF descriptor suited to the VBoW approach for humanoid robots; BRIEFROT. According to our evaluation, the BRIEFROT vocabulary is very effective in this context, as reliable as SURF to solve the localization problem, but in much less time. We also show that keypoints-based vocabularies performed better than color-based vocabularies.

### Table 7. Percentages of Correct Results for the New College Data-set.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Number of tests</th>
<th>Correct tests $\mu_1$</th>
<th>Effectiveness $\mu_1$ (%)</th>
<th>Correct tests $\mu_2$</th>
<th>Effectiveness $\mu_2$ (%)</th>
<th>Average time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRIEF</td>
<td>117</td>
<td>70</td>
<td>59.83</td>
<td>85</td>
<td>72.65</td>
<td>105.9</td>
</tr>
<tr>
<td>BRIEFROT</td>
<td>117</td>
<td>70</td>
<td>59.83</td>
<td>87</td>
<td>74.36</td>
<td>134.6</td>
</tr>
<tr>
<td>Color-Random</td>
<td>117</td>
<td>48</td>
<td>41.02</td>
<td>69</td>
<td>58.97</td>
<td>110.9</td>
</tr>
<tr>
<td>Color-Half</td>
<td>117</td>
<td>34</td>
<td>29.06</td>
<td>64</td>
<td>54.70</td>
<td>60.4</td>
</tr>
<tr>
<td>Color-Whole</td>
<td>117</td>
<td>44</td>
<td>37.61</td>
<td>74</td>
<td>63.25</td>
<td>110.5</td>
</tr>
<tr>
<td>ORB</td>
<td>117</td>
<td>69</td>
<td>58.97</td>
<td>85</td>
<td>72.65</td>
<td>107.0</td>
</tr>
<tr>
<td>SURF</td>
<td>117</td>
<td>74</td>
<td>63.25</td>
<td>89</td>
<td>76.07</td>
<td>302.3</td>
</tr>
</tbody>
</table>

### Table 8. Repeatability of Visual Vocabularies.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Repeatability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRIEF</td>
<td>100</td>
</tr>
<tr>
<td>BRIEFROT</td>
<td>98</td>
</tr>
<tr>
<td>Color-Random</td>
<td>87.6</td>
</tr>
<tr>
<td>ORB</td>
<td>89.1</td>
</tr>
<tr>
<td>SURF</td>
<td>97.2</td>
</tr>
</tbody>
</table>

images associated by the localization method with the key image $I_{142}$ with $\mu_1 = 1$. In both cases, it can be appreciated that the appearance of the sequence of input images is similar, but with sometimes significant rotations or light effects. Despite that, the localization method gives the correct most similar key image.

#### 4.4.7. Repeatability Analysis

An important aspect of the method is that it has a random component, at the level of the construction of the words tree; in the kmeans++ algorithm, the center of each cluster is chosen randomly. Hence, it is important to evaluate the repeatability of the evaluation results, when generating new vocabularies. We repeated 100 times the building of visual vocabularies, with different features, and tested them for 10 different input images.
In this article, we have assumed that the visual memory is given and the construction of the visual memory is left as future work. However, the construction of the visual memory is critical and may have a great impact on the performance of the localization stage and on the autonomous visual navigation. The visual memory should come with no gaps, i.e., every pair of consecutive key images should have a minimum amount of common visual information. However, we stress that the localization stage is not as dependent on the visual memory quality as the navigation stage, since the localization will give an approximate location whenever some similarity in appearance is detected.

Also as future work, we will implement the method onboard the NAO robot using a larger visual memory. We will explore the combination of visual vocabularies to robustify the localization stage and on the autonomous visual navigation. The visual memory should come with no gaps, i.e., every pair of consecutive key images should have a minimum amount of common visual information. However, we stress that the localization stage is not as dependent on the visual memory quality as the navigation stage, since the localization will give an approximate location whenever some similarity in appearance is detected.

As future work, we will implement the method onboard the NAO robot using a larger visual memory. We will explore the combination of visual vocabularies to robustify the localization results. We also wish to use the localization algorithm in the construction of the visual memory to identify revisited places.

Note

Acknowledgments
The first and third authors were supported in part by CONACYT [grant 220796].

Disclosure statement
No potential conflict of interest was reported by the authors.

Notes on contributors

Noé G. Aldana-Murillo received a degree in Electro-mechanics Engineering from the Tecnológico Nacional de México, campus León; an M.Sc. degree in Optomechatronics from Centro de Investigaciones en Óptica, León, México, in 2011 and 2014, respectively. Since 2015, he is a doctoral student at Centro de Investigación en Matemáticas, CIMAT-Guanajuato, México. His research interests include applications of visual control to robotics.

Jean-Bernard Hayet graduated from Ecole Nationale Supérieure de Techniques Avances in 1999. He received his master’s degree from University Paris VI and his Ph.D. degree from University of Toulouse in 2003. His doctoral dissertation tackled the problem of landmark-based navigation with computer vision in indoor environments and was supervised by Dr. M. Devy and Dr. F. Larase. From 2003 to 2006 he did a post-doctoral stay in visual tracking at University of Liege with Pr. Justus Piater. Since 2007 he is researcher at the Centro de Investigacion en Matematicas, in México, and is focusing his research on probabilistic methods in visual tracking and in handling vision constraints in motion planning and locomotion problems.

Héctor M. Becerra received a degree in Electronics Engineering from the Tecnológico Nacional de México, campus Ciudad Guzmán, a M.Sc. degree in Automatic Control from Cinvestav-Guadalajara, México, and a Ph.D. degree in Systems Engineering and Computer Science from the University of Zaragoza, Spain, in 2003, 2005 and 2011, respectively. He is currently a full researcher at Centro de Investigación en Matemáticas, CIMAT-Guanajuato, México. His research interests include applications of automatic control to robotics, particularly the use of computer vision as main sensory modality for visual control and visual navigation of mobile robots.

References
Application of Multi Agent Systems in Automation of Distributed Energy Management in Micro-grid using MACSimJX

Leo Raju, R. S. Milton and Senthilkumar Mahadevan

Department of Electrical and Electronics Engineering, SSN College of Engineering, Chennai, India; Department of Computer Science and Engineering, SSN College of Engineering, Chennai, India

ABSTRACT

The objective of this paper is to monitor and control a micro-grid model developed in MATLAB-Simulink through Multi Agent System (MAS) for autonomous and distributed energy management. Since MATLAB/Simulink is not compatible with parallel operations of MAS, MAS operating in Java Agent Development Environment (JADE) is linked with MATLAB/Simulink through Multi Agent Control using Simulink with Jade extension (MACSimJX). This allows the micro-grid system designed with Simulink to be controlled by MAS for realizing the advantages of MAS in distributed and decentralized micro-grid systems. JADE agents receive environmental information through Simulink and they coordinate to take best possible action, which is reflected in MATLAB/Simulink simulations. After validation and performance evaluation through dynamic simulations, the operations of the agents at various scenarios are practically verified by using the Arduino microcontroller. These validation and verification moves MAS closer to Smartgrid applications and takes micro-grid automation to a new level.

1. Introduction

The world today is witnessing a gradual transition in the form of generation and distribution of electricity. The liberalization of the energy industry, together with decarbonisation and inclusion of renewable energy sources, as well as ground-breaking Information and Communication Technologies (ICT) developments have huge consequences on the way sectors work. We are moving rapidly towards a more decentralized, more sustainable, and smarter power system. The smart grid paradigm represents a transition towards an intelligent, digitally enhanced, two-way power delivery grid (ME El-hawary, 2014). Micro-grid is a building block of smart grid and it plays a major role in adoption of renewable energy resources like solar and wind. However, their intermittent nature entails new challenges when integrating with grid and impacts the stability of the micro-grid. To meet these challenges, micro-grid should incorporate control strategies to achieve balance between generation and demand. Distributed energy management is necessary for the integration of renewable energy onto the micro-grid. MAS approach is used for distributed decision-making in a decentralised environment to improve responsiveness and stability. Energy management of micro-grid using MAS is discussed in Jun, Junfeng, Jie, & Ngan (2011). An agent-based modelling approach is used to model the micro-grid in Hatzigiargiou (2014). The interaction between individual intelligent decision-makers in the micro-grid are analysed through MAS simulation in Reddy & Veloso (2011). Optimization of micro-grid with intermittent renewable energy resources using MAS is given in detail in Eddy & Gooi (2011). Here, MAS is simulated in JADE for making distributed decisions to improve performance. But, no attempt is made to go beyond JADE simulations. MAS controlling the main operations of micro-grid are discussed in Hatzigiargiou & Dimeas (2005). Fully decentralized approach is considered here, which is not economical. The design and implementation details of MAS in micro-grid energy management are discussed in detail in Pipattanasomporn, Feroze, & Rahman (2009). Hierarchical approach of MAS is considered here, but facilitating MAS for practical implementation is not discussed. Real-time digital simulator is used for real-time operation of MAS on micro-grid is discussed in Logenthiran, Srinivasan, Khambadkone, & Aung (2012). The monitoring and control strategies are not discussed here. Multi-agent based distributed energy management for intelligent micro-grid is discussed in (no reference listed). Micro-grid management and market operations are discussed here, but practical implementation is not discussed. The various trends in micro-grid control are discussed in Olivares et al. (2014). The complete review of micro-grids in MAS perspectives is discussed in Gomez-Sanz, Garcia-Rodriguez, Cuartero-Soler, & Hernandez-Callejo (2014). Distributed online optimal energy management for smart grid is discussed in Liu, Zang, & Yu (2015). Unit commitments and demand side management for social welfare is discussed here, but hardware in loop simulations is not discussed. A detailed review on agent concepts applied to intelligent energy systems is given in Vrba et al. (2014). Coordination approach of micro-grids for resilient operation is discussed in Rivera, Farid, & Youcef-Toumi (2014). The electrical constraints of micro-grid are discussed here, but the details of implementation is not given here. Micro-grid control strategies are discussed in Logenthiran, Naayagi, Woo, Phan, & Abidi (2015). But, real time implementation is not discussed here. Two agents scheduling is discussed in Yin, Cheng, Wu, & Wu (2013). Multiple agent scheduling applications is discussed in Yin (2016). In most of these references, MAS operations are simulated in JADE, but JADE simulations are not adequate for validation. MAS have to be
linked with MATLAB for a comprehensive simulation, which can be validated. But MATLAB does not support parallel operations, which is essential for a decentralized approach of MAS. A novel approach, MACSimJX, is introduced to link MATLAB with MAS to allow system designed with Simulink to be controlled by agents that operate in external program in JADE, a powerful development environment for modelling MAS (Robinson & Peter, 2010). Although many micro-grid research activities involving MAS have been reported, all of them implement MAS in JADE platform and generates the basic simulations, which can only be used for theoretical proof of demonstration. Because of its collaborating and negotiating behaviour, MAS in JADE is not capable of performing real-time constrained control operation of micro-grid. So, there is a need to investigate the linking of MAS in JADE with MATLAB Simulink, which opens MATLAB to agent development environments for modelling real-time hardware systems alongside and interacting with multi-agent systems. This brings MAS closer to practical implementation. So an attempt has been made in this paper to use MACSimJX for autonomous energy management of solar micro-grid, considering the intermittent nature of solar power, randomness of load, dynamic pricing of grid and control of Non-Critical Load (NCL) for demand side management. The rest of the paper is organized as follows: A detailed discussion on multi agent system approach is given in Section 2. MACSimJX is explained in Section 3. Section 4 deals with implementation of dynamic energy management of solar micro-grid using MACSimJX. Case study, simulation and validation are given in Section 5. Practical verification of demand side management of solar micro-grid using Arduino is discussed in Section 6. Conclusion is given in Section 7.

2. Multi Agent Systems Approach

Autonomous actions and coordination are the basic ingredients of any distributed system. The limitations of distributed systems are that they lack run-time adaptive behaviour and require continuous communication. These considerations have motivated the development of approaches to a distributed system based on agents, which provide ways for adaptation and on-going interaction. A Multi Agent System (MAS) is a distributed system consisting of multiple agents, which form a loosely coupled network to work together to solve problems that are beyond their individual capacities. Multi-agents overlay a way to elaborate systems that are decentralized, emergent, and concurrent. MAS have inherent benefits such as extensibility, autonomy and reduction in problem complexity. Agents have certain behaviours and tend to satisfy certain objectives using their resources, skills and services. With a decentralized approach, MAS has its own perception of the environment, goal and agenda and they try to achieve the best for themselves, while behaving strategically. Plug and play adaptability and connection to external grid are seamless in MAS based micro-grid energy management. By nature MAS can be scaled up by adding other agents or by dispersing them in new environment with new resources and capacities. MAS are particularly useful for designing distributed systems requiring autonomy of their entities.

3. Multi Agent Control using Simulink with Jade extension (MACSimJX)

MAS are implemented in JADE framework for energy management of a micro-grid. MATLAB is not compatible with parallel operation, which is an essential characteristic of MAS; it becomes unstable if several processes run in parallel. To overcome this problem, a middleware, MACSimJX, is used as an interface between Simulink models and the agents in JADE, bringing MAS closer to the physical models through hardware in loop simulations. The agent considers the environment values received from the micro-grid devices and takes best possible actions autonomously for dynamic energy management of the solar micro-grid in a distributed environment. MACSimJX has client-server architecture. Client is in the Simulink and the server is at the agent environment. Communication between client and the server is done through named pipes (Mendham, 2005). Configuration information and simulation information are passed through separate pipes allowing the two processes to be run asynchronously. Queries are sent by the client and responded by the server. Simulation signals are passed on to the agent model. The agent environment facilitates coordination and messaging. It assigns the work to various agents for executing the operations. The result of agents operations or the command signals are given back to the agent environment. After receiving the completion information from all the agents, the agent environment sends the output to the Simulink through the windows pipe server. The resulting actions of agents are synchronized with Simulink simulation cycle for hard ware in loop simulations.


A grid-connected micro-grid system containing two solar Photo Voltaic (PV) systems, one in the department of the college premises and the other in a hostel is considered. Each PV system consists of local consumer, a solar PV system and a battery. The Simulink model is developed for the department and hostel solar power, load, and batteries along with grid and diesel (SDP, SHP, LDP, LHP, BD, BH, GRD, and DSL). In the department solar unit Simulink model, the department solar power is given to the department load then the remaining power is given to the hostel load. The further remaining power is given to the department battery, and then to the hostel battery for charging and if anything remaining is finally given to the grid. The hostel solar unit also designed in the similar way. Simulink model of the department solar unit is given in Figure 1. Similarly the hostel solar unit is modelled. In the department load model, the department load first receive power from the department solar unit and then from hostel solar unit. Further requirement of power is taken from department battery, then from hostel battery. If power is still required, it is received from diesel unit and grid unit based on unit price at that point of time. The loads are divided into critical and Non Critical Loads (NCL) so that NCL load shedding can be done before going to the external power resources like diesel or grid. The department load Simulink model is shown in Figure 2. The hostel load is modelled in the similar way. The Simulink model of the department battery is shown in Figure 3. In this model, three types of State Of Charge (SOC) are considered in the battery. Fully charged, SOC is in between fully charged (100%) and fully drained (40%), and SOC is fully drained. In the battery, if lesser the available power, the longer it takes for the battery to fully charge, and vice versa. Similarly, if more power is drawn, it discharges in short time and if less power is drawn it discharges for a long time. The hostel battery is modelled in similar way.
The proposed approach considers a two layer framework; comprising and functional layer and the agent layer. The physical infrastructure of micro-grid is simulated in MATLAB-Simulink in the functional layer and the behaviour parts of the

**Figure 1.** MATLAB/Simulink Model of Department Solar Unit.

**Figure 2.** MATLAB/Simulink Model of Department Load Unit.
agents are done in the agent layer aiming to optimize micro-grid operation. These two layers are linked with MACsimJX, which acts as a gateway to translate semantics from agent world to services world. The functional layer feeds the environmental values and requests services and the agent layer take necessary actions for optimizing the solar micro-grid under dynamic environment. The agent layer interacts with the function layer to test and validate the control strategies. Every component in the solar micro-grid is considered as an agent. All the agents are programmed in Java and made to run in JADE platform in Eclipse background. The multi-threading feature of Java and multiple behaviours per agent leads to heterogeneous concurrency, which is the essence of Multi Agent Systems. All of the smart grid features are implemented by coordinated actions of agents. The agent paradigm of computing has flexibilities, than the object oriented paradigm as they cannot access other agent's methods without its permission. The agents can say no and can negotiate for optimal action. These agents dynamically interact in the run time using the directory facilitator to manage the uncertainties of solar power supply and randomness of the load. They take strategic decision for demand side management, self-healing, dynamic pricing and plug and play. The characteristics of agent based computing such as autonomy, proactive and social are incorporated using inherent features of JADE and its various simple and composite behaviours. The semantic, Agent Communication Language (ACL) is used to communicate and coordinate among agents.

4.1. Agent Relationship

Agents sense the environment and make simple decisions, which are within its capacity. They are called, reflective action. For complex decisions and deliberative actions, it communicates with the control agent and informs the requirements. For example, the department solar agent gives power to the department load requirement if it is having enough power. If there is a shortage of power, it informs the control agent about the power requirements. The control agent gets the status of all the agents from Directory Facilitator (DF), where all the agents are registered with their status and services. It coordinates with other agents and validates the request made by other agents and informs other agents about the best place to get the requirement for optimal energy management in the solar micro-grid. The agent's relationship diagram is shown in Figure 4.
Here, the load agents participate in the system as buyers of energy, while the solar generator agent as sellers of energy. Here the hierarchical approach, which combines centralized and decentralized approach, is used. Every agent autonomously takes local decision, and then through control agent it communicates to other agents for strategic decision. We consider Solar Department Agent (SDA), Load Department Agent (LDA), Battery Department Agent (BDA), Solar Hostel Agent (SHA), Load Hostel Agent (LHA), Battery Hostel Agent (BHA), Grid Agent (GA), Diesel Agent (DA) and Control Agent (CA). Every hour based on the net power availability and the load requirement, the transaction with the grid is made by the CA. When LDA requests power from SDA, SDA gives power to LDA autonomously. If surplus power is available, it is given to BDA. Further excess power is given to BHA and finally to the Grid agent (GA). The validation is done through the control agent. The priority of the operation is mentioned as 1, 2, and 3 in the diagram. If there is not enough power available in SDA, then LDA contacts SHA and receives the available power. If the power is still required, it contacts the BDA and then BHA and finally it communicates with the control agent to do the NCL shedding and finally the post NCL shedding power is received from grid or diesel agent based on the unit price at that point of time. Similarly the LHA contacts SHA for power. SHA makes the autonomous decision to give the power to LHA. If surplus power is available after supplying to LHA, SHA checks the BHA and BDA for charging and further excess power is given to grid. If SHA does not have enough power required by LHA, LHA contacts SDA and gets the power available there. If power is still required, LHA contacts BHA and BDA to get the power available in the battery. The further required power is managed by contacting the control agent, which does the NCL shedding and finally the post NCL shedding power is received from grid or diesel agent after comparing their unit prices. The priority of operation is mentioned as 1, 2, 3, 4, and 5 in the diagram.

## 4.2. Control Strategies for Autonomous Energy Management of a Solar Micro-grid

In the solar micro-grid model, 12 input ports and 25 output ports are considered. Environment variables are given to the input port and the output ports are connected to the switches. All possible conditions are considered for optimal energy management. Based on values of these parameters, an 8 bit number is formed. The first bit is formed by comparing the solar power values of department (SDP) and host load (LDP). If the difference (SDP-LDP) is positive, then the first bit value is 1, if negative the bit value is 0. Similarly the second bit is found by comparing the value of department solar power (SDP) and load (LHP). The third bit is formed by comparing (SDP-LDP) and (LHP-LDP). If the difference is positive, the value is 1 and if negative, the value is 0. The fourth and fifth bits are found by knowing the department battery charging value. If it is fully charged the value is 00 and initially charged, the value is 01. If the charge reaches the cut off value then the value is 10. Similarly the sixth and seventh bits are found for the hostel battery. The eighth bit is found by looking at grid and diesel unit price. If the diesel price is greater, the value is one and if grid price is greater, the value is 0. Thus the 8 bit number is formed. This number is broadcasted to all the agents and by looking at it each agent comes to know the switching operations it is expected to do. JADE does the switching operations in parallel due to multi-threading and reduces the time of operations considerably. The load in department and hostel is divided into critical and non-critical load (CL and NCL). NCL is defined for every hour, which can be shed before going to grid. Along with the 8 agents already mentioned; control agent is also included, which specifically controls the NCL. Two layers of agents are formed: The control layer contains control agent for NCL control and the system layer contains all the other agents for controlling the critical load. Communication between these two layers is established for micro-grid control switching operations. Control layer agent does the switching operation of LDNC, LHN, GRD_LD, GRD_LH, DSL_LD and DSL_LH to control the NCL load for demand side management. Coordination between control agent and system layer agents is formed for controlling NCL. If the power requirement is managed in post NCL shedding, the control agent does not favour the grid even though the system layer agent favouring it. For control agent actions, separate state model is worked out and a 6 bit number is generated based on the conditions as shown in Figure 5.

The 6 bit number generates 64 states, out of which 50 states are active and the remaining are null states. For all the active states the switching operations are identified. Then for every switching action the state numbers are consolidated. The control agent is programmed with consolidated state numbers for each switching operations. The control agent is given a special state to bypass the logic for switching action of external power resources and go for grid irrespective of the unit price, when the diesel power is exhausted. The special status is formed as shown in Figure 6. If the NCL switches of department and hostel are switched on and if the total load power requirements is greater than the generated solar power, the Special State (SS) of Control Agent checks whether the required load power, after exhausting the department and hostel solar power, is greater than the available diesel power. If it is greater, SS of CA is made to 1 and the grid is connected to the loads irrespective of prices of diesel and grid.

### 4.3. Power Transfer Strategy

The power transfer strategies are shown in Figure 7. Here there are two layers namely system and control layers. Various
components of the solar micro-grid agents function in the system layer. The Control agent functions are in the control layer. Here the interaction of the agents for energy management and demand side management is clearly indicated. Solar Hostel Agent (SHA) has the possibility to link with Load Hostel Agent LHA, LDA, BDA, BHA, and GA based on hourly status of the solar power and load. It gives power first to local load and if excess power is available, then it goes to department load. Further excess power is given to its own battery and then to the department battery and finally to the grid. If there is deficiency for LHA, then LHA receives power from SDA, then from BHA and BDA and the further power requirement is received from Grid or Diesel agents. This way it is connected to all these agents. Similarly Department solar agent (SDA) has the possibility to link with LDA, LHA, BDA, BHA and GA. In the figure, department agents are mentioned in the upper part and hostel agents are represented in the lower part. All the possible connections between agents are given in detail. Non critical load shedding is coordinated by control agent before going for power transactions with the grid. The control agent at the control layer is used for NCL shedding at LDA and LHA. The agent operations are done based on the various environmental states, as projected in the 8 bit number, and the output from agent’s operations is sent to activate these switches through Simulink. Thus the interaction among agents takes place to consider all possible options for dynamic energy management under intermittent supply of solar power and varying load.

4.4. Example for Switching Strategies for System State and Control State Number

We consider a case for which the value of SDP = 200 kW, LCDP = 300 kW, LNCDP = 200 kW, SHP = 100 kW, LCHP = 400 kW, LNCHP = 100 kW, DPR = 8, GPR = 10. An eight bit system state number, 111101010 is formed as explained in Section 4.2. This gives the state number as 245. The relevant switch operations for this state are identified as SD_LD, SH_LH, DSL_LD and DSL_LH. There are 8 agents in ATF and the state number 245 is broad-casted for all the agents. The agents look at it and execute the required switching operations. Solar department agent executes the switching operation of SD_LD. Solar hostel agent does the switching operations of SH_LH. Diesel agent does the switching operation of DSL_LH and DSL_LD. For NCL operations, a 6 bit number 101111 is formed based on Figure 2 and it specifies the state as 47. For this state number, the grid agent switches GRD_LD and GRD_LH. The switch control strategy of agents for the system state and also the control state are shown in Figure 8. The special state of control agent is used to switch on the grid after the full diesel power is exhausted. In this case, out of the total load demand of 700 kW, NCL shedding is done for 300 kW and out of the remaining demand of 400 kW, the full capacity of 300 kW is received from DSL and the remaining 100 kW is received from grid through special state of control agent as the special state value is one. Thus the agents corresponding to the switching operations are identified and executed.

5. Case Study

For the department and the hostel, LDPC and LHPC represent critical loads, and LDPNC and LHPNC represent non-critical loads, respectively. The input values given to Simulink model are shown in Table 1. After the coordinated operations of the agents, the simulation output for the above input values is shown in Figure 9. In the 5th hour, both the hostel and the department have 200 kW deficits. Here, the non-critical load of 200 kW in the department and 100 kW in the hostel are shed. Then the required 100 kW is received from the grid as the grid price (GPR) is less than the diesel price (DPR). In the 11th hour, both department and the hostel have a deficit of 300 kW and 400 kW respectively. Here the batteries supply to the corresponding load until it is drained and then the non-critical load of 200 kW in the department and 100 kW in the hostel are shed. Then out of required power of 400 kW, 300 kW is received from diesel as it is economical and the remaining 100 kW is received from grid by activating the special state action of control layer agent.

6. Practical Verification of Demand Side Management of Solar Micro-grid using Arduino

JADE agent operations are verified by linking it with Arduino microcontroller. A Java library called RxTx is used to facilitate serial communication with the Arduino. Initially COM serial port is connected to the Arduino board in the device manager of the operating system. Then a connection is established between the Arduino and PC. Arduino software is ran on a PC to read or write the data available at the connection port through a Java code developed in Eclipse. The environment variables are sensed by using potentiometer where the maximum ranges are fixed for the value 0-1023. For example if hostel solar power is 50 kW, then 1023 represents 50 kW. Similarly all the environment variable values, department solar and load, hostel solar and load, department and hostel battery, grid and diesel values are fixed in eight potentiometers. Then, by varying the potentiometer, these values can be changed within the fixed range. These values are given to the
of the corresponding micro-grid devices. Thus the real time operation of MAS for energy management of the micro-grid is practically verified by using the Arduino microcontroller as shown in the Figure 10. Here we consider in the tenth hour of simulation, both the department and hostel have 100 kW excess after supplying their respective loads. This 200 kW excess power is used to charge the batteries. Here all the agents are active except grid and diesel agent. So, in the figure the last two LEDs representing Grid and Diesel are OFF and all other LEDs are ON. Thus the agent operations are verified using Arduino.

control agent through serial communication and from the control agent all other agents get their environmental values. The agents communicate and coordinate with other agents for strategic decision and the resulting command signals are passed to the actuators for physical action. This is observed through the corresponding LEDs, Digital write of high or low is done on the agent when the ACL communication is active. We consider an 8 bit number, each bit represent a micro-grid device action. Bit value 1 represents ON and 0 represents OFF. These 8 bit output is given through Arduino microcontroller to the corresponding LEDS to show the physical action of the corresponding micro-grid devices. Thus the real time operation of MAS for energy management of the micro-grid is practically verified by using the Arduino microcontroller as shown in the Figure 10. Here we consider in the tenth hour of simulation, both the department and hostel have 100 kW excess after supplying their respective loads. This 200 kW excess power is used to charge the batteries. Here all the agents are active except grid and diesel agent. So, in the figure the last two LEDs representing Grid and Diesel are OFF and all other LEDs are ON. Thus the agent operations are verified using Arduino.

Table 1. Case Study Values (kW).

<table>
<thead>
<tr>
<th>Time</th>
<th>SDP</th>
<th>LCDP</th>
<th>LNC</th>
<th>SHP</th>
<th>LCHP</th>
<th>LNC</th>
<th>DPR</th>
<th>GPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>600</td>
<td>400</td>
<td>0</td>
<td>400</td>
<td>200</td>
<td>0</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>300</td>
<td>500</td>
<td>0</td>
<td>500</td>
<td>300</td>
<td>0</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>300</td>
<td>0</td>
<td>200</td>
<td>300</td>
<td>0</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>500</td>
<td>0</td>
<td>200</td>
<td>400</td>
<td>0</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>300</td>
<td>400</td>
<td>100</td>
<td>400</td>
<td>200</td>
<td>200</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>300</td>
<td>400</td>
<td>0</td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>600</td>
<td>400</td>
<td>0</td>
<td>400</td>
<td>300</td>
<td>0</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>400</td>
<td>500</td>
<td>0</td>
<td>300</td>
<td>300</td>
<td>0</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>300</td>
<td>400</td>
<td>0</td>
<td>300</td>
<td>400</td>
<td>0</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>500</td>
<td>400</td>
<td>0</td>
<td>400</td>
<td>400</td>
<td>0</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>300</td>
<td>200</td>
<td>100</td>
<td>400</td>
<td>100</td>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>
7. Conclusion

Since MAS implementation in JADE framework is not capable of performing real-time constrained control operation of micro-grid, it is linked with solar micro-grid model, developed in MATLAB/Simulink, through MACSimJX. This helps to validate the agent operations through hardware in loop simulations bringing MAS closer to practical implementation. Autonomous energy management and demand side management of solar micro-grid is implemented to achieve the lowest possible cost of power generation under intermittent nature of solar power and randomness of load. The proposed framework explores all possible logical sequences of options, understand the stochastic environment, and select the optimal energy management actions to increase operational efficiency autonomously in a distributed environment. MAS in JADE are implemented with Arduino microcontrollers and the agent operations for environment dynamics are practically verified. Future work will focus on extension to multiple agents integrating diverse renewable generators (solar and wind etc.). Furthermore, this autonomous, distributed energy management of micro-grid can be scaled through big data, data analytics and cloud computing principles for smart grid and industry 4.0 applications.

Disclosure statement

No potential conflict of interest was reported by the authors

Notes on contributors

Leo Raju is an assistant professor of Electrical and Electronics Engineering at SSN College of Engineering, Chennai, India, since 1999. He has 18 years of teaching and 4 years of industry experience. He received a B.E. (EEE) from Government College of Technology, Coimbatore, Bharathiyar University in the year 1994 and an ME in Computer Science and Engineering from Anna University, Chennai, in the year 2005. He is currently pursuing his PhD in Anna University, Chennai. His research interests include multi agent systems, applications of multi agent systems in micro-grid and smart grid, application of reinforcement learning in renewable energy management, multi agent system in cyber physical systems.

R. S. Milton is a professor of Computer Science and Engineering at SSN College of Engineering, India since January 2010. He has been teaching for 26 years. He received a B.E. degree with honors in Electronics and Communication Engineering from Thiagarajar College of Engineering, Madurai and an ME in Computer Science and Engineering from College of Engineering Guindy, and a PhD in Information and Communication Engineering (Rough Set Theory and Relational Learning), from Anna University, Chennai. He worked in Madras Christian College, Chennai, from 1993 to 2010. He is a member of the Machine Learning Research Group of SSN. His research interests include statistical relational learning, natural language processing, multi agent systems and computer science education.

SenthilKumar Mahadevan is an associate professor of Electrical and Electronics Engineering at SSN College of Engineering, since 2004, India. He obtained a B.E. (EEE), and an ME (Applied Electronics), both from University of Madras in the year 1999 and 2000 respectively. He obtained his PhD from Anna University in the year 2014. He has 15 years of teaching experience. His area of specialization includes smart grid, multi agent system, simulation, modelling and control of renewable energy resources (wind and solar), applied electronics, power electronics, power quality and electrical machines. He has worked extensively in the area of Matrix Converter and its application in Wind energy system. He is so passionate in teaching and guiding under graduate projects in Signals and systems and Control.

References


Highly Accurate Recognition of Handwritten Arabic Decimal Numbers
Based on a Self-Organizing Maps Approach

(1,2)Amin Alqudah, (2)Hussein R. Al-Zoubi, (2,3)Mahmood A. Al-Khassaweneh, and (1)Mohammed Al-Qodah

(1)Department of Electrical Engineering, College of Engineering, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia
(2)Computer Engineering Department, Hijjawi Faculty for Engineering Technology, Yarmouk University, Jordan.
(3)Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA.

ABSTRACT
Handwritten numeral recognition is one of the most popular fields of research in automation because it is used in many applications. Indeed, automation has continually received substantial attention from researchers. Therefore, great efforts have been made to devise accurate recognition methods with high recognition ratios. In this paper, we propose a method for integrating the correlation coefficient with a Self-Organizing Maps (SOM)-based technique to recognize offline handwritten Arabic decimal digits. The simulation results show very high recognition rates compared with the rates achieved by other existing methods.

KEY WORDS: Correlation coefficient; reference image; recognition rate; SOM

1 INTRODUCTION
The electronics industry is currently transforming our traditional way of living. People now communicate using the Internet and mobile telephones and use digital devices in almost every aspect of their lives. Nevertheless, many daily activities and tasks still depend on classical methods of communication, such as using paper and a pencil or a pen. In many applications, automatic handwriting recognition can clearly save time and effort (Xiaofang, et al., 2011; Khalil, et al., 2017; Claudio, et al., 2001; Gómez, et al., 2008).

Automatic recognition of handwritten numbers is a field that has attracted many researchers to develop algorithms to increase recognition rates (Abirami and Murugappan, 2011; Ahmed and Moskowitz, 2004; Alhoniemi, 2002; Al-Omari and Al-Jarrah, 2004; Al-Zoubi, et al., 2011; Alqudah, et al., 2012). In this study, we employed the correlation coefficient (Ahmed and Moskowitz, 2004; Sun, et al., 2003) and Self-Organizing Maps (SOMs) (Kaski, et al., 1998; Kohonen, 1997; Kohonen, et al., 1996; Simula and Kangas, 1995) for the offline recognition of handwritten decimal numbers and through an extended number of experiments demonstrated that very high recognition rates can be achieved. The measured features of Arabic numerals include angular span, distance span, horizontal span, and vertical span (Mahmoud, 2008).

In this section, we review the latest proposed methods in the field of numeral and character recognition with a focus on the recognition of handwritten numerals (Milson and Rao, 1976; Narasimhan, et al., 1980). For example, the authors of (Rajput, et al., 2010) proposed a method to recognize handwritten and printed Kannada numerals using crack codes and a Fourier descriptors template. The work of (Plamondon and Srihari, 2000) presented a survey of the most popular algorithms for preprocessing, character and word recognition, signature verification, writer authentication, and handwriting learning tools for online and offline recognition. In (Choudhary, et al., 2011), the authors proposed a handwritten numeral recognition method using a modified Back Propagation Neural Network structure. In (Abirami and Murugappan, 2011), the identification of English numerals was conducted using rule-based classifiers and their sub-classifiers to implement a set of classification rules. This technique involves document vectorization, by which vectors are generated based on characters' shape, density, and
transition features. The overall performance of this technique was observed to be 94%.

The authors of (Sas and Kaczmar, 2012) proposed a recognition method for writer-dependent handwriting. The proposed method continuously segments handwritten words using genetic algorithms. The goal is to improve the recognition accuracy by maximizing the similarity among images of the same character within subsets of images. The authors of (Zhao and Küt, 2011) used unsupervised learning to enhance the performance of supervised learning during Chinese-word segmentation. To integrate the strengths of the two methods, the authors applied various goodness-of-fit measures and were able to achieve an 8.96% reduction in error. In (Liu, et. al., 2011), a fast SOM clustering method was proposed to leverage the performance of text clustering systems while at the same time preserving comparable clustering quality. Unlike most term weighting approaches, which are based on document indexing, the authors of (Ren and Sohrab, 2013) proposed a class-indexing-based term weighting approach for automatic text classification (ATC). The experimental results indicated better classification than that achieved by existing classification methods.

Many methods that are used for the recognition of online and offline handwritten English and Arabic numerals are based on neural networks (Al-Omari and Al-Jarrah, 2004; Goltsvev and Rachkovskij, 2005; Kherallah, et. al., 2008; Kussul and Baidyk, 2004; Lauer, et. al., 2007; Liu and Sako, 2006; Said, et. al., 1999) and support vector machines (Lauer, et. al., 2007; Soltanzadeh and Rahmati, 2004). Hidden Markov models have also been employed in the recognition of offline handwritten numerals (Mahmoud, 2008). Parkins and Nandi applied genetic programming to recognize handwritten digits (Parkins and Nandi, 2004). The authors of (Peng and Xu, 2012) proposed a classifier for pattern recognition called the Twin Mahalanobis distance-based Support Vector Machine (TMSVM), which represents a generalization of Twin Support Vector Machines (TSVMs) and other classification methods. TMSVM constructs two Mahalanobis distance-based kernels to optimize nonparallel hyperplanes using the covariance matrices of two classes of data. Lastly, the authors in (Al-Zoubi, et. al., 2011) proposed a global motion estimation method for the recognition of offline machine-printed Arabic digits.

Herein, we present a fast and accurate algorithm that exhibits very high recognition rates based on SOMs integrated with a correlation coefficient for the offline recognition of handwritten Arabic decimal numbers. For the numerals (1 – 9), we calculate the correlation coefficient between certain reference images generated by a SOM and the image under consideration. Recognition is performed by finding the reference image that has the highest correlation with the target image using the following equation:

\[
Corr = \frac{1}{Ng} \sum_{i=1}^{Ng} \frac{(I_R(i) - mean(I_R))(I_T(i) - mean(I_T))}{\sigma_{I_R} \sigma_{I_T}},
\]

where \( Ng \) is the number of pixels in the image, \( I_R \) and \( I_T \) are the reference and the target images, respectively, and \( \sigma_{I_R} \) and \( \sigma_{I_T} \) are the standard deviations of the reference and the target images, respectively. Moreover, we apply a special procedure for the recognition of the numeral 0, achieving recognition rates close to 100%, as discussed in Section 4.

This paper is organized as follows: a description of the preprocessing stage is presented in Section 2. Section 3 provides an introduction to SOM basics. In Section 4, we describe the proposed approach. SOM-classified images are shown in Section 5. The experimental results are presented in Section 6. Finally, Section 7 concludes the paper.

2 PREPROCESSING

In offline recognition, we obtain a scanned image of the numeral to be recognized. The scanned image (original image) is converted to a gray-level image, which is then converted to a binary image. The binary image is cleaned using opening and closing transforms and through erosion and dilation to smooth the boundaries of the numeral to be recognized, maintain its size, and eliminate any noise. The original dimensions of the images used in this work are (32 × 32) pixels (including the numeral and the background). Because the numerals in images are different sizes, resizing is used to scale the numeral in the image to a standard size. All numerals (1 – 9) are resized to a common size by selecting the value of the nearest point that yields a piecewise-constant interpolant (http://www.mathworks.com) based on the nearest-neighbor interpolation algorithm. Figure 1 (a) shows the number “4” before preprocessing (image size of 32 × 32). Figure 1 (b) shows the number “4” with an image size of (20 × 15) pixels after the preprocessing step. The number has been scaled and shifted to the middle of the image.

The preprocessing step provides images with the same value a better chance to achieve a high correlation. For example, the images in Figures 1 and 2 represent the number “4”. The correlation between the images in Figure 1 (a) and Figure 2 (a) is 0.03, whereas the correlation between the numerals in the same images after shifting and scaling (shown in Figure 1 (b) and Figure 2 (b)) is 0.78. Deformation does not significantly affect the recognition rate. Rather, centering the scaled images provides a good chance for images that show the same numeral to have a high correlation value, which leads to higher recognition rates.
Figure 1. a) Digit “4” located near the bottom of the image (entire image size 32x32). b) Digit “4” centered in the middle of the image (image size 20x15).

Figure 2. a) Digit “4” located near the top of the image (entire image size 32x32). b) Digit “4” centered in the middle of the image (image size 20x15).

Figure 4. Updating the best matching unit (BMU).

3 SELF-ORGANIZING MAPS (SOM’S)

CLASSIFYING handwritten numbers can help us extract useful information that can be used to discover new properties of handwriting and improve recognition rates. A self-organizing map is a type of neural network that is based on unsupervised learning. It performs a nonlinear mapping of data onto a two-dimensional map grid (Kohonen, 1997). SOMs have been used in different engineering applications (Kaski, et. al., 1998; Kohonen, 1997; Kohonen, et. al., 1996; Simula and Kangas, 1995; Alqudah and Al-Zoubi, 2015). A SOM consists of connected neurons, each of which is represented by a $d$-dimensional weight vector, where $d$ is the dimension of the input vectors used for training. Figure 3 shows the SOM structure used in this study. Training the SOM requires updating the best matching unit (BMU) and its neighbors on the map, as shown in Figure 4, where the BMU and its neighbors are pulled toward the input sample, marked ($\times$). The solid lines represent the connections before updating, whereas the dashed lines represent the connections after updating. At the end of training, the neurons will be organized into a structure that classifies the training samples into classes, where each neuron is the center of that class.

Two training algorithms are generally used when training SOM networks: the sequential algorithm and the batch algorithm. Both algorithms provide the same mapping (Kohonen, 1993). In this work, we use the batch algorithm because of its performance in terms of speed. More information about SOM training can be found in many published studies (Alhoniemi, 2002; Kaski, et. al., 1998; Kohonen, 1997; Kohonen, et. al., 1996; Kohonen, 1993; Kohonen, 1991; Simula and Kangas, 1995; Vesanto, et. al., 2000).

4 THE PROPOSED APPROACH

THE proposed recognition approach consists of two stages: obtaining the reference images and recognition. The reference images are obtained by clustering a large number of handwritten numbers into clusters using SOMs and then drawing a reference image for every cluster. This image is created by taking the average of all of the images that belong to the same cluster. Averaging is performed using the following equation:

$$R_{Im}(i, j) = \frac{1}{L} \sum_{n=1}^{L} I_{Im}^{(n)}(i, j)$$

where $R_{Im}$ is the created reference image, $(i,j)$ is the location of the pixel in the image, $I_{Im}^{(n)}$ is the $n^{th}$ image in class $R$, and $L$ is the number of images in class $R$.

We cluster/classify the images of each Arabic decimal number (1–9) into $N$ clusters/classes, yielding a total of $(9 \times N)$ clusters or reference images. Recognition is performed using the correlation coefficient between the image under consideration and all of the reference images.

The image is compared to all of the reference images and placed in the class with the highest correlation. Figure 5 illustrates the first stage of the proposed approach. It is worth mentioning that clustering the images of each decimal number (1–9) into $N$ classes emphasizes the features that characterize how people write decimal numbers.

Figure 3. The hexagonal lattice structure of the SOM used in this study.
Furthermore, having one class for each decimal would hide many features of written decimals, especially uncommon ones. Therefore, having many classes for each number divides the numbers into groups, and uncommon features have a better chance of being shown.

To recognize a certain numeral, the number in the image is first centered and scaled to a $(20 \times 15)$ pixel image, as shown in the previous step. The correlations between this new image and the $(9 \times 9)$ reference images are calculated, yielding $(9 \times N)$ correlation values, one value for each reference image. The number is classified as having the value of the image with the maximum correlation. For the numeral “0”, a special procedure is applied in which the ratio between the height and the width of the tested numeral is calculated. If the ratio is in the range (0.5-2.0) and the ratio of white pixels to black pixels is very small (less than 0.5, ideally zero), then the numeral under test is classified as “0”. Otherwise, the tested numeral undergoes the previously described process. With this method, we are able to achieve very high recognition rates. Preprocessing is applied after the “0” special case test to avoid scaling the numeral (if it is actually “0”).

5 CLASSES

As mentioned in Section IV, $N$ reference images are constructed for each of the 9 numerals (1 – 9) corresponding to $N$ classes. Each reference image is obtained by calculating the average of all of the images that belong to the same class. In this study, we test different values of $N$ (1, 4, 9, 16, 25, and 49).

Table 1 shows 144 reference images at $N = 16$. This table reveals the most common styles that people follow in handwriting Arabic numbers. It is worth mentioning that the images shown in Table 1 are created by classifying and averaging the classes of the
shifted and resized images of each number, which is why fuzzy images are observed, as in the case of decimal 6 class 2. In this study, a large number of people have participated by providing us with real handwritten documents containing words and numbers. The numbers are extracted with an image size of \((32 \times 32)\) pixels. Table 2 shows the 10 Arabic numerals and their equivalents in English numerals. We collect approximately 4,860 samples of each decimal, yielding 48,600 images for all decimal digits \((0 - 9)\). Out of the 48,600 images, 30,000 images are used to train the SOM classifier (3000 images/numeral), and the rest are used for testing.

To obtain a better understanding of the data we are using, the frequency of each of the handwriting styles is calculated. Table 3 shows the frequency of the images in each class for digits \((1 - 9)\) when the number of classes \(N = 16\). From Table 3, it can be observed that the digit “6” has the largest deviation, 1.93, between different handwriting styles, whereas the digit “7” has the smallest deviation, 0.74. Table 4, which is based on information derived from Table 3, shows the images of the most and the least frequent classes. Moreover, the table shows interesting and useful statistics about the most and least frequent classes (the frequencies of these classes are shown in bold font in Table 3).
Table 3. Frequency (%) of images in each class for digits (1 – 9), $N = 16$.

<table>
<thead>
<tr>
<th>Number/Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.73</td>
<td>5.90</td>
<td>6.16</td>
<td>7.19</td>
<td>4.93</td>
<td>7.56</td>
<td>6.96</td>
<td>5.13</td>
<td>6.36</td>
</tr>
<tr>
<td>2</td>
<td>5.03</td>
<td>7.60</td>
<td>5.90</td>
<td>6.93</td>
<td><strong>4.39</strong></td>
<td><strong>2.50</strong></td>
<td>6.26</td>
<td><strong>3.63</strong></td>
<td>8.33</td>
</tr>
<tr>
<td>3</td>
<td>4.63</td>
<td>6.26</td>
<td>8.00</td>
<td>6.56</td>
<td><strong>8.79</strong></td>
<td>6.90</td>
<td>6.03</td>
<td>7.33</td>
<td>7.80</td>
</tr>
<tr>
<td>4</td>
<td>8.16</td>
<td>5.36</td>
<td>7.10</td>
<td>5.33</td>
<td>5.03</td>
<td>5.06</td>
<td>6.10</td>
<td>5.26</td>
<td><strong>2.90</strong></td>
</tr>
<tr>
<td>5</td>
<td>4.83</td>
<td>4.86</td>
<td>5.50</td>
<td>7.63</td>
<td>7.10</td>
<td>8.33</td>
<td>5.86</td>
<td>5.70</td>
<td>5.70</td>
</tr>
<tr>
<td>6</td>
<td>7.43</td>
<td><strong>4.46</strong></td>
<td>5.70</td>
<td><strong>9.03</strong></td>
<td>8.03</td>
<td>2.53</td>
<td><strong>7.06</strong></td>
<td>5.06</td>
<td>7.70</td>
</tr>
<tr>
<td>7</td>
<td><strong>3.90</strong></td>
<td>5.80</td>
<td>7.10</td>
<td>7.86</td>
<td>7.70</td>
<td>6.90</td>
<td>6.06</td>
<td>6.96</td>
<td>8.36</td>
</tr>
<tr>
<td>8</td>
<td>6.33</td>
<td>6.90</td>
<td>7.93</td>
<td>6.13</td>
<td>5.76</td>
<td>4.33</td>
<td>6.96</td>
<td>5.83</td>
<td>6.16</td>
</tr>
<tr>
<td>9</td>
<td>6.80</td>
<td>4.90</td>
<td>5.09</td>
<td>7.13</td>
<td>4.53</td>
<td>6.60</td>
<td>6.46</td>
<td>7.06</td>
<td>5.50</td>
</tr>
<tr>
<td>10</td>
<td>7.43</td>
<td>5.76</td>
<td><strong>4.00</strong></td>
<td><strong>3.63</strong></td>
<td>7.33</td>
<td>7.56</td>
<td>5.93</td>
<td>6.90</td>
<td><strong>8.60</strong></td>
</tr>
<tr>
<td>11</td>
<td>6.36</td>
<td>7.30</td>
<td><strong>8.96</strong></td>
<td>6.66</td>
<td>6.10</td>
<td>5.66</td>
<td>7.03</td>
<td>6.20</td>
<td>4.93</td>
</tr>
<tr>
<td>12</td>
<td>6.00</td>
<td><strong>8.43</strong></td>
<td>6.93</td>
<td>6.73</td>
<td>7.23</td>
<td>5.46</td>
<td>6.73</td>
<td>6.66</td>
<td>6.23</td>
</tr>
<tr>
<td>13</td>
<td>5.83</td>
<td>5.13</td>
<td>5.56</td>
<td>4.93</td>
<td>4.43</td>
<td><strong>9.40</strong></td>
<td>5.53</td>
<td>6.73</td>
<td>5.73</td>
</tr>
<tr>
<td>14</td>
<td><strong>8.63</strong></td>
<td>7.96</td>
<td>5.30</td>
<td>5.73</td>
<td>6.56</td>
<td>7.43</td>
<td>6.90</td>
<td>6.96</td>
<td>6.40</td>
</tr>
<tr>
<td>15</td>
<td>6.36</td>
<td>6.20</td>
<td>6.03</td>
<td>4.30</td>
<td>7.00</td>
<td>7.76</td>
<td><strong>4.20</strong></td>
<td><strong>8.30</strong></td>
<td>4.10</td>
</tr>
<tr>
<td>16</td>
<td>4.50</td>
<td>7.13</td>
<td>4.70</td>
<td>4.16</td>
<td>5.03</td>
<td>5.96</td>
<td>5.86</td>
<td>6.23</td>
<td>5.16</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>6.34</td>
<td>6.05</td>
<td>5.96</td>
<td>6.61</td>
<td>6.33</td>
<td>6.75</td>
<td>6.18</td>
<td>6.44</td>
<td>6.19</td>
</tr>
<tr>
<td><strong>STDEV</strong></td>
<td>1.40</td>
<td>1.19</td>
<td>1.32</td>
<td>1.47</td>
<td>1.41</td>
<td><strong>1.93</strong></td>
<td><strong>0.74</strong></td>
<td>1.12</td>
<td>1.60</td>
</tr>
</tbody>
</table>

Table 4. Most and least frequent classes for digits (1 – 9), $N = 16$.

<table>
<thead>
<tr>
<th>Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Most frequent handwriting class after preprocessing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14th</td>
<td>12th</td>
<td>11th</td>
<td>6th</td>
<td>3rd</td>
<td>13th</td>
<td>6th</td>
<td>15th</td>
<td>10th</td>
<td></td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td>8.63</td>
<td>8.43</td>
<td>8.96</td>
<td>9.03</td>
<td>8.79</td>
<td>9.40</td>
<td>7.06</td>
<td>8.30</td>
<td>8.60</td>
</tr>
<tr>
<td><strong>Average image</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Least frequent handwriting class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th</td>
<td>6th</td>
<td>10th</td>
<td>10th</td>
<td>2nd</td>
<td>2nd</td>
<td>15th</td>
<td>2nd</td>
<td>4th</td>
<td></td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td>3.90</td>
<td>4.46</td>
<td>4.00</td>
<td>3.63</td>
<td>4.39</td>
<td>2.50</td>
<td>4.20</td>
<td>3.63</td>
<td>2.90</td>
</tr>
<tr>
<td><strong>Average image</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6 RESULTS AND DISCUSSION

The approach described in the previous section is first tested using the 30,000 training images; as expected, the recognition rate is 100%, which certifies the sufficiency of the features considered. The same approach is also tested using 1,860 test images for each Arabic decimal number, yielding 18,600 test images. If the numeral in the image is not determined to be “0”, it is centered and scaled to a (20 × 15) image as a preprocessing step to improve recognition. Then, the correlation between the test image (the centered and scaled numeral) and the reference images is calculated. The number value corresponding to the reference image with the highest correlation with the test image is assumed to be correct. As the most
important step, all preprocessed training images are clustered using the SOM technique. The images are clustered into 1, 4, 16, 25, 36, and 49 classes. Each time, the recognition rate is calculated. We observe that the recognition rate increases with the number of classes. This relationship holds true until the number of classes reaches 16, when a recognition rate of 100% is achieved. This recognition rate remains constant until the number of classes reaches 25; after that point, the recognition rate starts to decrease, as shown in Figure 6 and Table 5.

![Figure 6. Average recognition rate vs. number of classes. (The number “0” is recognized based on a special case test. The numbers (1 to 9) are recognized based on SOM).](image)

### Table 5. Effect of the number of classes on recognition rate and testing time, including the special case test for “0”.

<table>
<thead>
<tr>
<th>Number of Classes</th>
<th>Recognition Rate (%)</th>
<th>Time to Test (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90.08</td>
<td>228</td>
</tr>
<tr>
<td>4</td>
<td>91.79</td>
<td>666</td>
</tr>
<tr>
<td>9</td>
<td>96.53</td>
<td>1602</td>
</tr>
<tr>
<td>16</td>
<td>99.99</td>
<td>2322</td>
</tr>
<tr>
<td>25</td>
<td>99.97</td>
<td>3587</td>
</tr>
<tr>
<td>36</td>
<td>92.43</td>
<td>5196</td>
</tr>
<tr>
<td>49</td>
<td>92.59</td>
<td>7009</td>
</tr>
</tbody>
</table>

The recognition rate is at least 90% for any number of classes. This behavior occurs because by increasing the number of classes, we increase the number of hidden features that are observed, which helps to increase the recognition rate. This trend holds true until the number of classes reaches 25, after which the recognition rate begins to decrease. This decrease occurs because having too many classes makes the recognition mechanism very sensitive to minor features, which leads to false recognitions. Furthermore, this degradation in performance could be related to the ratio of the reference image size to the number of images in any class. For example, when the number of classes is 49, the average number of images in each class is approximately 61 images/class (3000/49); the size of the reference image created from these images in this class, after averaging, is (20 × 15 = 300) pixels. This large image size (300 pixels), when compared to the size of the 61 images/classes, suggests that each pixel has only a small chance to be represented. Each pixel has a recognition probability of 61/300, which is minor compared to that observed when the number of classes is 16. In that case, the ratio is approximately 187.5/300. The latter case results in a representation that is approximately 3 times better; therefore, we focus on the case of \( N = 16 \) classes.

#### 6.1 Effect of reducing the number of subclasses used for recognition

We select the case involving 16 classes for further study. We obtain the frequency of selection for each class. The results are shown in Table 6. As shown in the table, it is clear that there are dominant classes for each number. For example, for the decimal number “1”, class 11 is selected more than 99% of the time, whereas for the decimal number “6”, class 3 and class 4 are selected at a similar rate. Therefore, the most and the second most selected class for each number are studied separately. (These classes are indicated in bold in Table 6).

Table 7 shows the most and the second most selected for digits (1 – 9) for the case of \( N = 16 \) classes. The table shows the representative image for each class as well as the selection percentage. If we focus on the most selected class for each number, it is observed that 67.27% of the correct recognitions are derived from this particular class. Furthermore, if we focus on the second most selected class, it is observed that 26.19% of the correct recognitions are derived from this class. Combining both classes and ignoring the other classes would give us a recognition rate of approximately 93.46%. This high rate implies that this system could be made even simpler if we consider only these two classes, in which case we would still have a considerable recognition rate.

#### 6.2 Time requirements

In addition to the recognition accuracy, a system with a high recognition rate should also take into consideration the time requirements for testing. Therefore, we measure the time required to test 1,860 images for each digit. This work is carried out using MATLAB version 7.8 (R2009a) on a computer running Windows Vista (64-bit) with 4 GB of RAM and a processor with a 2-GHz clock rate. Table 5 lists the recorded average recognition rate versus the test time when 1, 4, 9, 16, 25, 36, and 49 classes are used (includes the special test for “0”). As expected, longer test times are needed when more classes are used. Consequently, the tradeoff between accuracy and performance should be determined. In the rest of this
section, the case of \( N = 16 \) classes will be considered. We will discuss methods for further improving the performance of the proposed recognition system.

### 6.3 Case study: Number of classes \( N=16 \)

To obtain a profound understanding of the recognition behavior of our proposed method, we will closely examine the results that we have obtained. Table 8 shows the recognition rate achieved for each digit with 16 classes. At a glance, we can notice the high accuracy of the proposed recognition approach, for which an average recognition ratio of 99.99% is achieved, with a 100% recognition ratio for the digits 0, 2, 3, 4, 6, 7, 8, and 9.

Table 6 presents the correct recognition percentages achieved for each class for digits (1 – 9), \( N = 16 \).

<table>
<thead>
<tr>
<th>Number/Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>62.58</td>
<td>0</td>
<td>0.05</td>
<td>0.05</td>
<td>0</td>
<td>6.13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
<td>20.81</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11.02</td>
<td>0</td>
<td>2.74</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.94</td>
<td>0</td>
<td>79.68</td>
<td>50.81</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.11</td>
<td>0</td>
<td>0.05</td>
<td>49.09</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33.98</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.27</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0.43</td>
<td>0</td>
<td>1.4</td>
<td>0</td>
<td>0.91</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
<td>52.69</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>67.37</td>
<td>48.87</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.66</td>
<td>0.05</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.22</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
<td>32.9</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.38</td>
</tr>
<tr>
<td>10</td>
<td>0.05</td>
<td>0</td>
<td>0.32</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0.65</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>99.03</td>
<td>0</td>
<td>0</td>
<td>1.72</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0.16</td>
<td>0.22</td>
<td>0</td>
<td>3.98</td>
<td>17.53</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>41.34</td>
</tr>
<tr>
<td>13</td>
<td>0.22</td>
<td>16.29</td>
<td>28.66</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0.11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>81.29</td>
<td>0.32</td>
<td>2.53</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
<td>63.12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.86</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>99.94</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>99.94</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 7. Highest and second highest correct recognition percentage for digits (1 – 9), \( N = 16 \).

<table>
<thead>
<tr>
<th>Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>First class of highest correct recognition</td>
<td>11th</td>
<td>1st</td>
<td>7th</td>
<td>7th</td>
<td>3rd</td>
<td>3rd</td>
<td>14th</td>
<td>15th</td>
<td>6th</td>
</tr>
<tr>
<td>Recognition rate (RC1 %)</td>
<td>99.03</td>
<td>62.58</td>
<td>67.37</td>
<td>48.87</td>
<td>79.68</td>
<td>50.81</td>
<td>50.81</td>
<td>81.29</td>
<td>63.12</td>
</tr>
<tr>
<td>Average image</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second class of highest correct recognition (RC2 %)</td>
<td>6th</td>
<td>2nd</td>
<td>13th</td>
<td>5th</td>
<td>12th</td>
<td>4th</td>
<td>2nd</td>
<td>8th</td>
<td>12th</td>
</tr>
<tr>
<td>Recognition rate %</td>
<td>0.43</td>
<td>20.81</td>
<td>28.66</td>
<td>33.98</td>
<td>17.53</td>
<td>49.09</td>
<td>11.02</td>
<td>32.9</td>
<td>41.34</td>
</tr>
<tr>
<td>Average image</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC1+RC2 (%)</td>
<td>99.46</td>
<td>83.39</td>
<td>96.03</td>
<td>82.85</td>
<td>97.21</td>
<td>99.9</td>
<td>92.31</td>
<td>96.02</td>
<td>94.03</td>
</tr>
</tbody>
</table>
Table 8. Recognition rates for each decimal digit, $N = 16$. (Number "0" was recognized based on a special case test).

<table>
<thead>
<tr>
<th>Decimal digit</th>
<th>Correct recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit 0</td>
<td>100</td>
</tr>
<tr>
<td>Digit 1</td>
<td>99.94</td>
</tr>
<tr>
<td>Digit 2</td>
<td>100</td>
</tr>
<tr>
<td>Digit 3</td>
<td>100</td>
</tr>
<tr>
<td>Digit 4</td>
<td>100</td>
</tr>
<tr>
<td>Digit 5</td>
<td>99.94</td>
</tr>
<tr>
<td>Digit 6</td>
<td>100</td>
</tr>
<tr>
<td>Digit 7</td>
<td>100</td>
</tr>
<tr>
<td>Digit 8</td>
<td>100</td>
</tr>
<tr>
<td>Digit 9</td>
<td>100</td>
</tr>
<tr>
<td>Average</td>
<td>99.99</td>
</tr>
</tbody>
</table>

On the other hand, Table 10 shows all of the misrecognized numbers. The table shows the image index, the original image, the actual digital value, the misrecognized value, the misclassified class, and the reference image for that class in each case. The table shows that the digits "1" and "5" have been misrecognized only once each as "0" and "2", respectively.

6.4 System performance when using 16 classes out of 16 vs. when using 2 classes out of 16

A closer look at Table 6 reveals that the majority of the classes achieve a 0% or very poor recognition rate (in other words, Table 6 is sparse). Therefore, in a trial to obtain better performance using our proposed system, we investigate a system using only the most and second most successful classes for each of the decimal digits. That is, rather than using all of the 16 classes, why not use only 2? Table 11 shows the recognition rates and the test times obtained when all of the 16 classes are used and when only the two most
Table 10. The incorrectly recognized numbers.

<table>
<thead>
<tr>
<th>Image Index</th>
<th>Original Image</th>
<th>Preprocessed Image</th>
<th>Actual Digit Value</th>
<th>Classification Based on 16 classes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Misrecognized As</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 11. Comparison of recognition rates and test times when all of the 16 classes are used and when only the two most successful of the 16 classes are used.

<table>
<thead>
<tr>
<th>Decimal digit</th>
<th>2 classes: Recognition Ratio (%)</th>
<th>16 classes: Recognition Ratio (%)</th>
<th>2 classes: Time to Test 1860 images (seconds)</th>
<th>16 classes: Time to Test 1860 images (seconds)</th>
<th>Speedup (T2/T1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit 1</td>
<td>100</td>
<td>99.94</td>
<td>47</td>
<td>398</td>
<td>8.46</td>
</tr>
<tr>
<td>Digit 2</td>
<td>100</td>
<td>100</td>
<td>47</td>
<td>400</td>
<td>8.51</td>
</tr>
<tr>
<td>Digit 3</td>
<td>100</td>
<td>100</td>
<td>50</td>
<td>399</td>
<td>7.98</td>
</tr>
<tr>
<td>Digit 4</td>
<td>99.90</td>
<td>100</td>
<td>48</td>
<td>395</td>
<td>8.22</td>
</tr>
<tr>
<td>Digit 5</td>
<td>100</td>
<td>99.94</td>
<td>48</td>
<td>396</td>
<td>8.25</td>
</tr>
<tr>
<td>Digit 6</td>
<td>100</td>
<td>100</td>
<td>48</td>
<td>399</td>
<td>8.27</td>
</tr>
<tr>
<td>Digit 7</td>
<td>100</td>
<td>100</td>
<td>48</td>
<td>397</td>
<td>8.25</td>
</tr>
<tr>
<td>Digit 8</td>
<td>100</td>
<td>100</td>
<td>47</td>
<td>405</td>
<td>8.61</td>
</tr>
<tr>
<td>Digit 9</td>
<td>100</td>
<td>100</td>
<td>48</td>
<td>403</td>
<td>8.39</td>
</tr>
<tr>
<td>Average</td>
<td>99.99</td>
<td>99.99</td>
<td>47.9</td>
<td>399.1</td>
<td>8.33</td>
</tr>
</tbody>
</table>

Successful of the 16 classes are used. Surprisingly, it can be observed that the average recognition rates are the same in both cases (99.99%). Looking at the recognition rate for each digit individually, other than that for the digit “4”, the recognition rates are either enhanced or remain the same.

Table 12 shows all of the misrecognized numbers based on 2 classes among the 16 classes. The table shows that for the digit “4”, two misrecognized pictures appear; the recognition rate decreases by 0.10% (from 100% to 99.90%).

To explain the changes in the achieved recognition rate when the number of classes is reduced to 2 from 16, let us look at Table 13. The table lists all of the misrecognized numbers when using 2 classes out of 16 and when using all 16 classes. Table 13 shows that the incorrectly assigned classes are removed from the set for the case in which only 2 classes are used.

These results demonstrate why the recognition rates for the digits “1” and “5” increased while the recognition rate for the digit “4” decreased: in the 2-class case, the 11th and 6th classes were selected for the digit “1”, the 7th and 5th classes were selected for the digit “4”, and the 3rd and 12th classes were selected for the digit “5”. To better illustrate this point, let us take the 2nd row of Table 13 as an example. This row clarifies what has happened for the image with Index 1. This image, with an actual digital value 4, was correctly recognized as “4” when the number of classes was 16; Class number 1 of the digit “4” showed the highest correlation, 0.5232. In the 2-class case, Class number 1 of the digit “4” was removed because it was not one of the most successful classes. Therefore, Class number 2 of the digit “7”, which remained in the set, exhibited the next highest correlation, 0.3516, and was selected; the number “4” was misrecognized as the number “7”.

In terms of performance, an average increase in speed of 8.33 is achieved when 2 classes are used rather than 16, as indicated in Table 11, with the same recognition accuracy. The same table also shows that the average test time needed to recognize 1,860 images is 47.9 seconds, which means that an average of 26 milliseconds is needed to recognize a number. Thus, in addition to exhibiting high accuracy, the proposed system is also fast and can be implemented in online applications.

7 CONCLUSIONS AND FUTURE WORK

In this study, we propose a method that integrates the use of correlation coefficients with that of SOMs for the recognition of handwritten Arabic decimal numbers. The proposed recognition approach has three main advantages. First, the method is simple: the proposed approach uses 2 basic methods, the correlation coefficient method and the SOM method,
both of which are known to be simple and fast. Second, the method is highly accurate: with the proposed method, the results are nearly 100% accurate. Third, the method exhibits high performance: the proposed approach performs recognition in only a few milliseconds. In future work, we plan to apply this approach for the recognition of handwritten Arabic and English characters.

ACKNOWLEDGMENT

This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project # 2017/01/7551.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

REFERENCES


NOTES ON CONTRIBUTORS

Amin Alqudah received his M.Sc. and Ph.D. in Electrical Engineering from the University of Colorado, USA in 2005, and from Colorado State University, USA in 2009, respectively. He received his bachelor degree in Communications Engineering from Yarmouk University, Jordan in 1999. Since 2009, he has been working with the Department of Computer Engineering, Hijjawi Faculty for Engineering Technology, Yarmouk University, Jordan. He has been working as an Associate Professor in the Electrical Engineering Department at Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia since Fall 2015. His research interests include Image Processing, Neural Networks, Machine Learning and adaptive signal processing.

Hussein R. Al-Zoubi received his MSE and Ph.D. in Computer Engineering from the University of Alabama in Huntsville, USA in 2004 and 2007, respectively. He received his bachelor degree in Electrical Engineering from Jordan University of Science and Technology. Since 2007, he has been working with the Department of Computer Engineering, Hijjawi Faculty for Engineering Technology, Yarmouk University, Jordan. He is currently a Professor. His research interests include machine vision, pattern recognition, image processing, computer networks and their applications: wireless and wired, security, multimedia, queuing analysis, and high-speed networks.

Mahmood A. Al-Khassaweneh received the B.S. degree in electrical engineering from Jordan University of Science and Technology, Jordan, in 2000, and the M.S. and the Ph.D. degrees in electrical engineering from Michigan State University, USA, in 2003 and 2007, respectively. Since Fall 2007, he has been working with the Department of Computer Engineering, Hijjawi Faculty for Engineering Technology, Yarmouk University, Jordan. He has been working as a Visiting Associate Professor in the Electrical and Computer Engineering Department at Michigan State University, East Lansing, MI, USA, since Fall 2014. His research interests include Image Processing, Computer Vision, Medical Imaging, Data Security and applying signal processing methods to the brain signals.

Mohammed Al-Qodah received his Bachelor and M.Sc. in Electrical Engineering from the Jordan University of Science and Technology, Jordan in 2004 and 2008, respectively. Since 2009, he has been a Lecturer with the Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia. His research interests include information theory, nonlinear and chaos theory and image processing.
Synthesis Optimization of Piezo Driven Four Bar Mechanism Using Genetic Algorithm

Laith Sawaqed¹, Khaled S. Hatamleh¹,², Mohammad A. Jaradat¹,², Qais Khasawneh¹,³

¹Mechanical Engineering Department, Jordan University of Science & Technology, P.O.Box 3030 Irbid, Jordan
kshh@just.edu.jo, majaradat@just.edu.jo, qakhasawneh@just.edu.jo
²Mechanical Engineering Department, American University of Sharjah, P.O. Box (2666) Sharjah, UAE
khatamleh@aus.edu, mjaradat@aus.edu
³Electromechanical Engineering technology department, Institute of Applied Technology, Abu-Dhabi, UAE
Qais.khasawneh@adpoly.ac.ae

ABSTRACT
Over the past few years, there has been a growing demand to develop efficient precision mechanisms for fine moving applications. Therefore, several piezoelectric driven mechanisms have been proposed for such applications. In this work an optimal synthesis of a four-bar mechanism with three PEAs is proposed. Two evolutionary multi-objective Genetic Algorithms (GAs) are formulated and applied; A Genetic Algorithm Synthesis method (GAS) is first used to obtain a synthesis solution for the mechanism regardless of power consumption. Then another Genetic Algorithm Minimum Power Synthesis method (GAMPS) is used to obtain the synthesis solution of minimum power consumption. For that purpose, the study performs simulation investigation of the aforementioned algorithms for each point along sinusoidal and kidney shaped paths of motion. Results show capability of both methods in obtaining a synthesis solution. However, GAMPS outperformed GAS in terms of driving power consumption as it is minimized by 99% ratio.

KEY WORDS: Genetic Algorithm, Four-bar Mechanism, Synthesis optimization.

1 INTRODUCTION
PIEZOELECTRIC actuators (PEAs) provide micro-level fine tuning and precise positioning control. Therefore, PEAs were deployed within several mechatronic system applications that require such positioning accuracy over the last decade. For instance, PEAs were proposed to run hard disk drives and micro-manipulator mechanisms (Nambi, et. al. (2012); Lopez-Martinez and Campo (2003); Krishnan and Saggere (2007); Tan (2001); Liaw, et. al. (2008); Liaw and Shirinzadeh (2009); Lin and Lin (2012)). However, the proposed solutions consist mainly of multi bar mechanisms with coupled links. This in turn, created a challenge in obtaining a valid and optimum synthesis solution.

Zhang, et. al. (1984) presented an atlas of curves that approximates the geometrical solution needed to follow a desired path by a geared five-bar mechanism; each of those curves is associated with a certain gear phase angle. However, the approximated solutions need an optimization process for accurate path following.

Chanekar and Ghosal (2013) introduced a two-stage sequential quadratic based optimization algorithm for synthesis of adjustable planar four-bar mechanisms to solve a continuous path generation problem. The first stage of the algorithm obtains possible driving dyads, which are then passed on to the second stage to obtain the remaining mechanism parameters using least-squares based circle-fitting procedure. The authors used an objective function that is based on the geometry of the four bar mechanism. Synthesis results presented by two numerical examples using the proposed method demonstrated more efficient optimization than literature work as it uses less number of design variables in the search process.
Gogate and Matekar (2012) presented two evolutionary new objective functions for optimum synthesis of mechanisms path generating. The authors used Differential Evolution optimization to assess the performance of the presented new objective functions through several simulations runs over a four-bar mechanism three different prescribed paths of motion. Obtained Results are compared with those generated using the regular tracking error function between followed path and desired path. The presented objective functions are concluded as better option for optimization in accordance to designer interest and problem nature.

Recently, simplicity of implementation and fast convergence speed realized by powerful computers encouraged other researchers to use modern optimization and evolutionary techniques to solve complex mechanisms synthesis such as Diab and Smaili (2017); Yi and Liu (2017). For instance, Cabrera, et. al. (2002) used simulation techniques to achieve minimal error between the desired and actual paths of motion followed by planner mechanism using Genetic Algorithms (GA). Another study presented by Laribi, et. al. (2004) in which a fuzzy logic is proposed to adjust the initial bounding intervals used by GA for path generation of a four-bar mechanism synthesis, simulation results of the Fuzzy GA method against classical GA method show higher accuracy and faster convergence. Cabrera, et. al. (2007) applied an evolutionary GA based technique “POEMA” to optimize a multi-objective synthesis problem of a planar hand robot. The study extended “POEMA” using Pareto-based approach to classify the outcomes and choose the optimal solution. The fast convergence results were encouraging to be used for other mechanism synthesis. Affi, et. al. (2007) considered both mechanism synthesis and motor characteristics while controlling a motor-driven four-bar system using multi-objective optimization approach. Continual efforts in the field, were carried out such as; the work done by Acharyya and Mandal (2009), in which a synthesis of a four-bar mechanism for trajectory following was done using GA, Particle Swarm Optimization Technique (PSO), and Differential Evolution (DE). The latter outperformed GA and PSO. In 2009, Erkaya and Uzmay (2009) were able to minimize deviations of a four-bar mechanism from desired path of motion arising from clearance; where two GA were used to minimize trajectory errors after obtaining direction of joint clearance using minor search space knowledge. Khorshidi, et. al. (2011) presented a multi-objective optimization algorithm for path-generation of four-bar linkage mechanisms. The proposed algorithm used Pareto Genetic algorithm to enhance the performance of a Non-dominant strong Genetic algorithm by setting the radius of the local search neighbourhood at each search step. The presented hybrid algorithm used triple-objective functions at once during optimization namely; minimum tracking error, deviation from 90 degrees angle in addition to maximum angle velocity ratio. Optimization results using the proposed algorithm over four-bar mechanism’s path generation design outperformed other literature algorithms in terms of energy efficiency, computational power, and practical viability.

Recently, El-kribi, et. al. (2013) used Non-dominant Strong Genetic Algorithm to select the optimal motor and optimal inertia distribution of a considered motor driven mechanical system. The proposed algorithm used minimum motor torque and minimum velocity fluctuation as objective functions. Obtained results of the optimal (driving motor/mechanical system properties) proved to be more efficient than electromechanical design strategies. In addition, the algorithm was able to minimize considered system power consumption and velocity fluctuation without the need of sophisticated controllers.

However, the majority of the discussed solutions within the literature did not include the type of driving system for the proposed solution. Although PEAs were deployed within fine tuning applications for high precision mechanisms, they suffer from the nonlinear behaviour that affects motion accuracy and makes it more challenging to find their synthesis solution (Liaw, et. al. (2008); Liaw and Shirinzadeh (2009); Lin and Lin (2012)).

In this paper, unlike the conventional four bar mechanism (Oetomo, et. al. (2006); Sitti (2003); Tari and Su (2011)), due to the lack of a closed form solution for the inverse kinematics of the piezo-driven four bar mechanism with 6 variables, a multi-objective optimization of the mechanical system using PEAs is presented, namely, the four-bar mechanism, as illustrated schematically by Figure 1. Optimal synthesis is achieved based on a genetic algorithm approach. The objective functions are used to minimize both the deviation of the end effector location from the desired position, and the change in the four-bar mechanism links lengths which guarantee reduced power consumption of the PEAs when it moves from one position to another. Figure 1 shows a flexure-based four-bar mechanism driven by three PEAs. Accordingly, the main contribution is to determine the optimal inverse kinematic solution for the piezoelectric actuated four-bar mechanism to move the end effector to a predefined position by using genetic algorithm. The dynamical model of the deployed PEA in the mechanism is assumed to be compensated as in Lin and Yang (2006); Badr and Ali (2010).

The paper will be structured as follows: in Section 2, formulates the analytical model of the four bar mechanism. Section 3 states the dynamic model of the PEA. Section 4 formulates the optimization problem of the considered four bar mechanism. Section 5 illustrates the working principle of the genetic
algorithm and formulates the fitness function. Section 6 shows discussion of the simulation results based on defined fitness function. Finally, Section 7 summarizes the conclusions of this study.

Figure 1: Piezo-driven flexure-based four-bar micro/nano manipulator.

2 MECHANISM MODELING

The main objective is to provide an inverse kinematic solution for the piezoelectric actuated four-bar mechanism using GA. The mechanism end effector follows a desired path within its reachable workspace while considering the geometrical constraints of the mechanism. For this purpose, the four-bar mechanism shown in Figure 1 can be simplified into four links pinned together as shown in Figure 2. The dimensional synthesis model, as shown in KHALAF (2012), can be summarized as follows:

\[ \theta_1 = \tan^{-1}\left(\frac{Y}{X-L}\right) \]

When the mechanism’s end effector is following some desired path, the deformation of the first piezoelectric (PEA1) link \( \Delta L_1 \) and its corresponding angular deflection \( \theta_1 \) are determined geometrically using Eqs. (2-3):

\[ \Delta L_1 = \sqrt{(X-L)^2 + Y^2} - L \]

\[ \theta_1 = \tan^{-1}\left(\frac{Y}{X-L}\right) \]

where \( L \) is the original link length, \( (X, Y) \) are coordinates of the desired end effector location, and \( \theta_1 \) is the orientation of the first link as shown in Figure 2.

Figure 2: Simplified four-bar mechanism.

Since the first link has a unique solution, then the synthesis problem is reduced into solving for the deformations of the second and third (PEA2 & PEA3) links which are referred to as \( (\Delta L_2, \Delta L_3) \) and their corresponding angular deformations \( (\theta_2, \theta_3) \). Based on geometry and parameters shown in Figure 2, the following relations can be determined as:

\[ \gamma = \tan^{-1}\left(\frac{Y}{X}\right) \]

\[ \theta_2 = \alpha \pm \gamma \]

\[ \alpha = \begin{cases} \theta_2 - \gamma, & \theta_2 \geq \gamma \\ \gamma - \theta_2, & \theta_2 < \gamma \end{cases} \]

\[ c\lambda = \frac{(L + \Delta L_2)^2 + (L + \Delta L_3)^2 - (X^2 + Y^2)}{2(L + \Delta L_2)(L + \Delta L_3)} \]

\[ \sin \lambda = \frac{\sqrt{X^2 + Y^2}}{(L + \Delta L_3)} \sin \alpha \]

\[ \lambda = \tan^{-1}\left(\frac{\sin \lambda}{\cos \lambda}\right) \]

\[ \theta_3 = \begin{cases} +(180 - |\theta_3|), & \theta_2 < \gamma \\ -(180 - |\theta_3|), & \theta_2 > \gamma \end{cases} \]

The previous relations are coupled and hard to be solved individually for each parameter. However, they could be solved for an arbitrary numerical solution as introduced by Khalaf (2012), or by enumerative search methods, or by a soft computing methods such as GAs. In this paper, a multi-objective GA is used to solve for the optimal power unknown parameters \( \Delta L_2, \Delta L_3, \theta_2, \) and \( \theta_3 \), as shown in Figure 3. Where \( \zeta_\delta \) is the desired end effector location represented by \((x_d, y_d)\) coordinates.
Figure 3: General block diagram for the proposed approach.

3 MODELLING

The general model describing positioning mechanism using PEA is represented by mass spring damper system with an applied force from the PEA as presented in Lin and Yang (2006); Badr and Ali (2010). The general model developed by Bouc-Wen for the PEA describes the nonlinearity of the system by Lin and Yang (2006):

\[ m\ddot{l} + b\dot{l} + k\ddot{h} = f = k(du - h) + \rho \]

where \( m \) represents the effective mass of the PEA, \( b \) is the damping coefficient, \( k \) is the stiffness, \( f \) the generated force due to the applied voltage to the PEA, \( d \) represents the ratio between the output displacement and applied input voltage \( u \) to the actuator, \( h \) represents the hysteretic nonlinear term for the PEA, \( l \) is the output displacement of the piezoelectric actuated mechanism in its respective dimension, \( \rho = kl_0 \) and \( l_0 \) is the initial displacement of the actuator as the applied voltage is \( u = 0 \); \( a, \beta, \gamma \) are the parameters which characterize the hysteretic loop’s magnitude and shape of the PEA actuator. Finally \( \dot{l}, \ddot{l}, \dot{h} \) and \( \ddot{h} \) are the derivatives of \( l \) and \( h \) with respect to time \( t \), respectively. The modelling parameters using the Bouc-wen model were determined in Lin and Yang (2006).

The modeled hysteresis of the PEA is compensated by a feedforward loop for fine tuning application which mainly focuses on the motion manipulation, based on this approach the system is represented by Lin and Yang (2006) as:

\[ \dot{l} = a \dot{d} - \beta \dot{h} - \gamma \dot{v} \]

Figure 4: PEA feedforward compensator loop (Lin and Yang (2006)).

The controller command is the summation from the feedforward controller output \( u(t) \) and the hysteresis observer output \( Oh(t) \) given as in equation (2)

\[ U(t) = u(t) + Oh(t) \]

(11)

while the augmented closed loop model of the system is given as Lin and Yang (2006)

\[ \dot{X} = AX + Bl_d \]

(12)

where \( x_1 = l \), \( x_2 = \dot{l} \) and \( x_3 = e(t) = l_d - l(t) \). For position fine tuning application a proportional integral (PI) controller was proposed by Lin and Yang (2006) to handle the system hysteresis based on linear matrix inequality approach. Where the augmented closed loop model of the system matrices are given as

\[ A = \begin{bmatrix} 0 & 1 & 0 \\ K_c & B_c & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ u_c \\ 1 \end{bmatrix} \]

(13)

where \( B_c \) and \( K_c \) are the closed loop augmented damping coefficient, and stiffness, respectively.

\[ u_c = \frac{d \times k \times k_p}{m} \]

(14)

where \( k_p \) is the controller proportional term.

4 OPTIMIZATION

Two scenarios will be considered. The first compromises optimization of the four bar mechanism geometrical synthesis, while the second compromises optimization of the four bar mechanism geometrical synthesis along with minimum change in the four-bar mechanism links length which guarantee reduced power consumption. The optimization problem can be formulated as follows (El-kribi, et. al. (2013)):

\[ \min : \ g_j(Z), j = 1, \ldots, n \]

(15)

subject to:

\[ Z \in \mathbb{Z} \]

(16)

where \( g_j \) is the objective function. \( Z \) is a design vector of the four-bar mechanism, which contains all the design variables. \( z_{i_{\min}} \) and \( z_{i_{\max}} \) define the limits of each design variable \( z_i \).

Consider the mono-objective function

\[ g = \sum_{j=1}^{n} \mu_j g_j(Z) \]

(17)

where \( \mu_j \)’s are arbitrary weighting factors, \( g_j (j = 1, \ldots, n) \) are the objective functions to be optimize. The weighting factors can be chosen based on the importance of the objective to the system for a given the application.

5 GENETIC ALGORITHM

GA is a soft computing method that can efficiently determine global minima/maxima of linear or nonlinear problems. GA is highly recommended for problems involving large number of unknowns that are hard to be determined using conventional methods. It is mainly based on the natural selective principle in which the fittest candidates in a population are selected as parents from which a replacement population is generated by mutation and crossover
operations to create better fit candidates often called offsprings (Haupt and Haupt (2003); Holland (1992)).

Figure 5 shows flow of the suggested GA until the stopping criteria is reached, i.e. maximum number of iterations (epochs) is reached or fitness values dropped below a predefined certain threshold. Generally the iterations (epochs) is reached or fitness values dropped offsprings (Haupt and Haupt (2003); Holland (1992)).

operations to create better fit candidates often called offsprings (Haupt and Haupt (2003); Holland (1992)).


g = e^2 = (x_e - x)^2 + (y_e - y)^2 \\
x = (L + \Delta L_2 )\cos(\theta_2) + (L + \Delta L_3 )\cos(\theta_2 - \theta_3) \\
y = (L + \Delta L_2 )\sin(\theta_2) + (L + \Delta L_3 )\sin(\theta_2 - \theta_3)

where the desired end effector location is represented by $(x_d, y_d)$, and the end effector position, obtained from the geometry, is represented by $[(L + \Delta L_2 )\cos(\theta_2) + (L + \Delta L_3 )\cos(\theta_2 - \theta_3) , (L + \Delta L_2 )\sin(\theta_2) + (L + \Delta L_3 )\sin(\theta_2 - \theta_3)]$. $\Delta L_2 , \Delta L_3$ are the next required PEA deformations of links 2 and 3. Finally, $(\theta_2 , \theta_3)$ are the next angular positions of links 2 and 3, as shown in Figure 2. This objective function will be minimized to obtain the design variables defined in (14).

5.2 Genetic Algorithm Minimum Power Synthesis (GAMPS)

Here, the fitness function takes into consideration the fitness function defined in the previous section and the power consumption represented as the difference in PEA deformation when moving the end effector from the current to the following position, mathematically:

\[
g = e^2 = (x_e - x)^2 + (y_e - y)^2 + \nabla \Delta L_2^2 + \nabla \Delta L_3^2 \\
x = (L + \Delta L_2 )\cos(\theta_2) + (L + \Delta L_3 )\cos(\theta_2 - \theta_3) \\
y = (L + \Delta L_2 )\sin(\theta_2) + (L + \Delta L_3 )\sin(\theta_2 - \theta_3) \\
\nabla \Delta L_2 = \Delta L_2 - \Delta L_{2,i} \\
\nabla \Delta L_3 = \Delta L_3 - \Delta L_{3,i}
\]

where, $(\Delta L_{2,i} , \Delta L_{3,i})$ are the current PEA deformations of links 2 and 3, and $(\theta_{2,i} , \theta_{3,i})$ are the current angular positions of links 2 and 3. Similar to the previous section, the objective function will be minimized to obtain the design variables defined in (18) based on not only the geometry error but also the power consumption. Based on the optimization defined above, equal weighting factors are considered.

Table 1 summaries the genetic algorithm options value used in the conducted simulations.

<table>
<thead>
<tr>
<th>GA option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>50</td>
</tr>
<tr>
<td>Crossover fraction</td>
<td>80%</td>
</tr>
<tr>
<td>Elite count</td>
<td>1</td>
</tr>
<tr>
<td>Mutation Rate</td>
<td>1%</td>
</tr>
<tr>
<td>Number of Generations</td>
<td>400</td>
</tr>
</tbody>
</table>

Table 2: The lower and upper bounds of the unknown parameters.
6 RESULTS

SINCE the proposed micro-manipulator is required to manipulate micro-objects or particles along any arbitrary path, two simulation runs were conducted to obtain the mechanism’s minimum power synthesis solution using developed GA. The runs used sinusoidal and kidney-shaped paths of motion. At each run the path was discretized into a series of consecutive points that starts from the mechanism’s unactuated position and ends at the considered path end. There are infinite synthesis solutions for \( Z = \{ \Delta L_2, \Delta L_3, \theta_2, \text{ and } \theta_3 \} \) when moving from a current point to the next point along the path of motion, the proposed GAS searches for an arbitrary synthesis solution using the objective function described by equation (19). On the other hand, GAMPS uses the objective function described by equation (20) to extract the synthesis solution that represents the minimized power consumption.

Obtained changes in length of PEA link 1 along all path points are obtained using equation (2). Links 2 and 3 on the other hand, has many possible solutions. Therefore, GAS and GAMPS will differ unless they coincide by chance. Figure 9 and 10 show obtained changes in lengths of PEA links 2 and 3, namely \( (\Delta L_2, \Delta L_3) \) along all points of defined path. As expected, the figures show how GAS solution has higher values of \( (\Delta L_2, \Delta L_3) \) than those obtained by the GAMPS solution. Figures also show the expected fluctuation in obtained changes of PEA lengths while moving from one point to another using the GAS solution. The GAMPS solution resulted in a smooth mechanism motion along the defined path, unlike the GAS solution which requires higher power rates with rough mechanism motion.

The difference vector between current and next PEA links length is calculated and recorded for all points of motion path as described by the following equation:

\[
\nabla (\Delta L_m) = \Delta L_m - \Delta L_{m-1}, \quad m = 1, 2, 3 \quad (21)
\]

This equation is considered twice, once using the GAS solution, and another using the GAMPS solution. The net change in length of PEA link 1 \( \nabla (\Delta L_1) \), is the actual change required in PEA link 1 when the mechanism moves from one point to another. Differences of change in links 2 and 3 are shown in Figure 11 and 12, respectively. It can be clearly noticed that there is a great enhancement over the required difference in links lengths. \( \nabla (\Delta L_2) \) and \( \nabla (\Delta L_3) \) when GAMPS is used. Summations of all difference values along the defined path of motion for PEA links 2 and 3 are illustrated in Table 3. The table shows how the total amount of absolute values of differences, closely related to the required power, are dramatically smaller for both kidney-shape and sinusoidal paths.

<table>
<thead>
<tr>
<th>Sum of absolute changes in ( (\mu m) )</th>
<th>Kidney-Shape Path</th>
<th>Sinusoidal path</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAS ( \sum_{q=1}^{r} \nabla (\Delta L_2) )</td>
<td>296</td>
<td>0</td>
</tr>
<tr>
<td>GAS ( \sum_{q=1}^{r} \nabla (\Delta L_3) )</td>
<td>256</td>
<td>0</td>
</tr>
<tr>
<td>GAMPS ( \sum_{q=1}^{r} \nabla (\Delta L_2) )</td>
<td>358</td>
<td>0</td>
</tr>
<tr>
<td>GAMPS ( \sum_{q=1}^{r} \nabla (\Delta L_3) )</td>
<td>460</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 8 shows that the fitness value and the average distance between individuals (chromosomes) decrease with the number of iterations represented by the number of generations (50 generations is shown) which indicates the conversion to a solution of the optimal mechanism synthesis. Figure 6 shows four positions of the proposed mechanism along the kidney-shaped path of motion; at each position GAS and GAMPS solutions are indicated, if GAS solution is traced from position 1 to 4, it is clear how the mechanism behaves in a fluctuating type of motion. On the other hand, tracing GAMPS solution over the same figure reveals the smooth type of motion obtained. Path tracking animations are provided for kidney-shaped and sinusoidal paths. The previous discussion for kidney-shaped path is also verified by sinusoidal path as shown by Figure 7.
Figure 7: GAMPS vs. GAS solution snapshots along the sinusoidal path of motion.

Figure 8: Fitness value and average distance vs. number of generations using (a) GAMPS, (b) GAS.

Figure 9: Deformation of PEA2 for kidney-shaped path.

Figure 10: Deformation of PEA3 for kidney-shaped path.

Figure 11: Required gradient in the second link deformation along the kidney-shape path.

Figure 12: Required gradient in the third link deformation along the kidney-shape path.

7 CONCLUSION

AN optimal synthesis of a four-bar mechanism with three PEAs is proposed. The proposed mechanism synthesis for each point along the discretized path is a complex and computationally expensive process if conventional numerical and enumerative methods are used. In addition, these methods will result in multiple solutions that need optimization. Hence, Genetic Algorithm was proposed to obtain a synthesis solution at lower cost and faster
In order to achieve that, GAS and GAMPS were proposed to obtain a general synthesis solution and a minimum power synthesis solution, respectively. Simulation results tested over two different paths showed that GAMPS was able to obtain a solution that has less power consumption than GAS. Furthermore, it was noticed that the transition of mechanism links along the path were much smoother and had fewer fluctuations than GAS. The proposed GA can replace conventional numerical methods to synthesize other forms of multi bar mechanisms as it is easy to implement and has a faster convergence. Such advantages will make this approach more attractive to be deployed within feedback system to provide the required synthesis for the predefined path or set point.

8 REFERENCES
B. El-kribi, et al., 2013. Application of multi-objective genetic algorithms to the mechatronic design of a four bar system with continuous and discrete variables. 61, pp. 68–83.


9 DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

NOTES ON CONTRIBUTORS

Laith Sawaqed, Ph.D. is the Mechanical Engineering-Acting chairperson of Department. Dr. Sawaqed holds a Bachelor’s and Master’s Degrees in Mechanical Engineering - Mechatronics from Jordan University of Science and Technology in 2006, 2008, respectively. He received his Doctorate in Mechanical Engineering from University of Maryland – College Park in 2013. Dr. Sawaqed's primary area of research has focused on Artificial Intelligence, Robotics, and Control Systems.

Khaled S. Hatamleh received his B.Sc. (2002) in Mechanical Engineering from JUST, Jordan. Received his M.Sc. (2006) in Industrial Automation Engineering from Yarmouk University, Jordan, and his Ph.D. (2010) in mechanical engineering from New Mexico State University, NM, USA. Currently he is an Assistant Professor at the American University of Sharjah, and JUST. His research areas are: robotics, UAV parameter estimation, IMU, and control systems.

Mohammad Abdel Kareem Jaradat received the B.Sc. degree from Jordan University of Science & Technology, Jordan, and the M.Sc. and Ph.D. degrees in mechanical engineering from Texas A&M University, College Station, TX, USA. Currently he is an Associate Professor at the American University of Sharjah, and with Jordan University of Science & Technology. During his integrated experience several projects, prototypes and publications are conducted, specialized in the following research areas: robotics, artificial intelligent systems, mechatronics system design, sensor fusion, fault diagnostics, intelligent nano-systems, intelligent control, and embedded control systems.

Qais Khasawneh received his BSc (1999) From JUST and received his MSc (2002) and PhD (2008) from the University of Akron, USA. He has been serving as an assistant professor in the mechanical engineering department/mechatronics since 2009. His research activities involve MEMS, Nanotechnology, renewable energy, and control.
EDITORIAL
Special Issue on Machine Learning and Data Mining for Cyber-Physical Systems
Intelligent Automation & Soft Computing

Zheng Xu and Zhiguo Yan
The third research institute of the ministry of public security, 339 Bisheng Road, 201142, Shanghai, China

RECENTLY, machine learning as well as data mining for cyber-physical systems has become the main focus of research in this area. Challenges imposed by the large scale of CPS Data, the uncertainty related to contradictory and incomplete information, and also, by properties and characteristics of Linked Data represent an interesting domain for emerging machine learning and data mining approaches. The challenges in managing and analyzing data can require fundamentally new techniques and technologies in order to handle the size, complexity, or rate of availability of this data. The cyber-physical system (CPS) has been coming into our view and will be applied in our daily life and the business process management. The emerging CPS must be robust and responsive for its implementation in coordinated, distributed, and connected ways. It is expected that future CPS will far exceed today’s systems on a variety of characteristics, for example, capability, adaptability, resiliency, safety, security, and usability. With the rapid development of computing and sensing technologies, such as ubiquitous wireless sensor networks, the amount of data from dissimilar sensors and social media has increased tremendously.

A new user authentication protocol using trust model, elliptic curve cryptography and biometrics for WSNs is proposed by Liu and Huang (A User Authentication Protocol Combined with Trust Model, Biometrics and ECC for Wireless Sensor Networks). The result of the trust model analysis indicates that the model can improve the model’s ability of withstanding attacks from the malicious nodes. With indoor laboratory test data as the basic parameters, using ABAQUS finite element software simulation, an analysis was carried out by Peng et al. (The Machine Learning Based Finite Element Analysis on Road Engineering of Built-In Carbon Fiber Heating Wire) of the degree that the surface temperature of heating wire, the thermal physical parameters of asphalt concrete, and environmental conditions have influence on the melting effect. Zhang et al. (Delay-Dependent Stability of Recurrent Neural Networks with Time-Varying Delay) investigated the delay-dependent stability problem of continuous recurrent neural networks with time-varying delay. A new and less conservative stability criterion is derived through constructing a new augmented Lyapunov-Krasovskii functional (LKF) and employing the linear matrix inequality method. A regional dynamic CGE model was constructed by Wang (A Computable General Equilibrium Model Based Simulation on Water Conservancy Investment) to simulate and analyze the short-term and long-term influence of water conservancy investment to water conservancy industry itself, other national economy sectors and macro economy, so as to provide scientific proof for policy-making of water conservancy, and coordinated and sustainable development police-making of the whole society. The tire braking model and the dynamic characteristic model of the brake torque with the variable of the controlling air pressure were established by Dong et al. (Optimal Learning Slip Ratio Control for Tractor-Semitrailer Braking in a Turn Based on Fuzzy Logic). A new calibration method is presented by Yu (Kinematic calibration of parallel manipulator for semi-physical simulation system) that takes into consideration all geometrical parameter errors and coordinate transformation errors. A virtual prototype model of the articulated tracked vehicle is established by Zhao et al. (The Virtual prototype model simulation on the steady-state machine performance) based on the multi-body dynamic software RecurDyn. An Iterative Squared-Root Cubature Kalman Filter (ISR-CKF) algorithm proposed by Chen (The SLAM Algorithm for Multiple Robots Based on parameter estimation) is aimed at improving SR-CKF algorithm on simultaneous localization and mapping (SLAM). The AMESim simulation models are established by Chen et al. (Simulation and data analysis of Energy Recovery sensing on Parallel Hydraulic Hybrid Crane) and
analyzed by establishing vehicle dynamics model and referencing to the actual data of the crane and physical hydraulic components, the simulation results are verified by road tests on the experimental prototype. In order to effectively evaluate the safety interval and lateral collision risk in training airspace, TSE error performance was modeled by Xu et al. (The lateral conflict risk Assessment for Low-Altitude Training Airspace Using Weakly Supervised Learning Method). An adaptive integrated guidance and control (IGC) scheme for the homing missile is proposed by Zhou and Ge (Intelligent Control for Integrated Guidance and Control Based on Intelligent Characteristic Model) based on the novel continuous characteristic model and the dynamic surface control technique. The NARX neural network was used by Wu et al. (NARX Network Based Driver Behavior Analysis and Prediction Using Time-Series Modeling) to predict real-time speed with the heart rate regarded as the input variable. The results indicated that familiarity with the experimental route did decrease drivers’ mental stress but resulted in higher speed. Three models of the vertical cylindrical oil storage tank in different sizes, which are commonly used by Gao and Wang (The data Analyses of Vertical Storage Tank USING Finite Element SOFT computing) in practical engineering are established. The dynamic characteristics, sloshing wave height and hydrodynamic pressure of oil tank considering liquid-structure coupling effect are analyzed by using ADINA finite element software, which are compared with the result of the standard method. Qin and Luo (Association Link Network based Concept Learning in Patent Corpus) investigated a challenging problem to automatically construct the patent concept learning model. The model consists of two main processes, which are the acquisition of the initial concept graph and refine process for initial concept graph. A multi-phase oil tank recognition of remote sensing images, namely coarse detection and artificial neural network (ANN) recognition is proposed by Liu et al. (Multi-phase Oil Tank Recognition for High Resolution Remote Sensing Images).

**Acknowledgments.** The guest editors would like to thank Dr. Mo Jamshidi who is the Special Issues Editor of *Intelligent Automation & Soft Computing*. His help and trust is the most important thing for the success of this SI. The guest editors would like to thank the reviewers for their high quality reviews, which provided insightful and constructive feedback to the authors of the papers.

**Zhiguo Yan** received degree certificate of Doctor of Philosophy form Shanghai Jiaotong University in 2008, and accomplished post-doctor research in Fudan University in 2013. He is a vice professor of the third research institute of Ministry of public security. His scholar interests focus on video intelligent analysis, big data techniques, IoT techniques on public security, etc.

**Zheng Xu** was born in Shanghai, China. He received Diploma and Ph.D. degrees from the School of Computing Engineering and Science, Shanghai University, Shanghai, in 2007 and 2012, respectively. He is currently working in the third research institute of ministry of public security and the postdoctoral in Tsinghua University, China. His current research interests include topic detection and tracking, semantic web and web mining. He has authored or co-authored more than 70 publications including IEEE Trans. on Fuzzy Systems, IEEE Trans. on Automation Science and Engineering, IEEE Trans. on Cloud Computing, IEEE Trans. on Emerging Topics in Computing, IEEE Trans. on Systems, Man, and Cybernetics-Part C, etc.
A User Authentication Protocol Combined with the Trust Model, Biometrics and ECC for Wireless Sensor Networks

Tao Liu, Gan Huang*, Ping Zhang
School of Computer and Information, Anhui Polytechnic University, Wuhu, China

ABSTRACT
In this article, a new user authentication protocol using trust model, elliptic curve cryptography and biometrics for WSNs is submitted. The result of the trust model analysis indicates that the model can improve the model's ability of withstanding attacks from the malicious nodes. The results of safety analysis and performance analysis for our proposed user authentication protocol demonstrate that this protocol can be flexible to all sorts of common known attacks and performs similarly or better compared with some active user authentication protocols. It is suitable for WSNs which have a prominent request for the security and the performance.

KEY WORDS: User authentication, Ant colony algorithm, Biometrics, Elliptic curves cryptography (ECC), Trust model, Wireless sensor networks (WSNs)

1 INTRODUCTION
WIRELESS sensor networks, which consists of many sensor nodes, which are limited in computation, storage and energy, can inspect and gather data in the distributed area, then process and transport data to users ultimately. For the characteristics of powerful self-organization and flexible network topology structure of WSN, it has been widely applied in the fields of military affairs, industries, real-time traffic monitoring, measurement of seismic activity, wildlife monitoring and etc. (Chong & Kumar, 2003). The data may be sent to users through queries. Most of the queries are often issued at the points of base stations or gateway (GW) nodes in WSN applications. Nevertheless, there are great needs to access the real-time data from sensor nodes. Therefore, real-time data may no longer be accessed, only through base stations or GW nodes. The data can be accessed from sensor nodes by external users (Li, 2009).

Compared with traditional networks, for the characteristics of dynamic topology, opening channel and etc., the WSN is more vulnerable to be attacked, captured and destroyed. But for military or high-tech applications, the data collected is confidential or valuable. Authentication is the first line of defense for a security system. The characteristics of itself determine the traditional security authentication mechanism, and can't fit very well into their security needs. Therefore, it is necessary to find a suitable security and authentication scheme for sensor networks (Radi, DezFouli, Bakar, 2012).

In 2002, Adrian Perrig et al. (Adrian, Robert, & Tygar, 2002) proposed a kind of security protocols for sensor network, one of secure network encryption protocol (SNEP) using symmetric cryptography to realize the communication of confidentiality, integrity and point-to-point authentication. Two keys used for encryption and authentication of a node to the base station are derived by the same algorithm as the main key shared with the base station. The shared secret key between nodes is temporarily allocated by the base station. WSN user entity authentication of public key system is proposed for the first time in (Watro, Kong, & Cuti, 2004), but the anti-capture property of this protocol is poor. Literature (Malan, Welsh, & Smith, 2004) based on the elliptic curve cryptosystem algorithm of the strong user authentication protocol for the document program has been improved. Wong et al. (Wong, Zheng, & Cao, 2006) put forward a
dynamic user authentication protocol for WSNs by means of the hash function and exclusive-OR mathematical operation. This protocol saves computer loads and decreases the complexity of the computation. However, Das et al. (Das, 2009) stressed that Wong et al.'s protocol didn’t keep a look out replaying attacks and impersonating attacks and an enhanced two-factor user authentication protocol for WSNs is proposed. It not only avoids the replay attacks and stolen-verifier attacks, but also defends against the code-guessing attacks and masquerade attacks. However, it doesn’t defend the off-line code-guessing attacks and node compromising attacks, and it doesn’t prevent the response information from the sense node to the user. In addition, it lacks mutual authentication.

In order to enhance the performance and the security of Das et al.'s protocols, A user authentication protocol for WSNs through employing elliptic curve cryptography is proposed by Yeh et al. (Yeh, Chen, & Liu, 2011) and Choi et al. (Choi, Lee, Kim, 2014). Compared with other cryptographic systems, it can reach the identical security with a key in a rather smaller size. Therefore, ECC is quite suitable for WSNs. Recently; Xue et al. (Xue, Ma, Hong, 2013) proposed a user authentication protocol for WSNs. In this protocol, a temporal credential produced by GW is distributed to the user and the sensor nodes. Through the method, the user, the sensor node and the gateway can achieve the mutual authentication between all of them. Besides, hash function and XOR operations are only used. But it is which only using simple passwords infeasible for WSNs to use the traditional authentication mechanisms, because of the differences lying among various networks. Our proposed user authentication protocol absorbs the merits of the above protocols and increases the trust model to enhance the performance of the above protocols and the accuracy of the authentication between entities.

2 RELATED WORK

2.1 Biometric Authentication

BIOMETRIC authentication technology as a means of identity authentication has become more widely used, such as the fingerprint attendance system, etc. Biometric identity authentication system is based on users' biological physiological structure features and behavior patterns for authentication. Because these biological characteristics are on the physiology and behavior associated with the user directly, biometric keys have some advantages, as follows: (YUAN, JIANG, JIANG, 2010).

(1) They are very difficult to copy with or share.
(2) They are not easy to drop out nor to forget.
(3) They are extremely hard to guess.
(4) They cannot be forged or distributed easily.

Therefore, compared with the traditional identity authentication system based on password, using the latest biometric identification technology of identity authentication system can provide better security (Long, 2015).

2.2 BAN-logic

Formal analysis for security protocols has widely gained attention in the field of security information. BAN logic is one of formal analysis methods for security protocols, which can be used to find protocol vulnerabilities based on formal analysis (QING, 2003).

Ban is a modal logic based on belief. In the process of BAN inference, participants’ belief in a protocol would be changed along with the message exchanged. When BAN is applied, it should idealized the steps, which transform protocol messages to the formal Logic formula of BAN; next, they make rational assumptions according to the specific situation, and reasoning for the assumptions based on logic rules and Idealized protocols; in the end, it gets results whether the reasoning protocols meet with expected goals (HAN, & DING, 2011, Syverson & Oorschot, 1994).

BAN logic tries to answer the following questions:
(1) What the protocols can accomplish for goals?
(2) What is the assumption for the protocols?
(3) Whether there is redundant information for the protocols?

Whether to encrypt the message in the protocols can be passed using clear key without affecting the protocols’ security?

BAN logic discusses abstractly the security for the certification agreement. Therefore, it cannot consider the security defects, which are induced by a specific implementation and the protocol defects, which are induced by the encryption system. On the whole, the BAN logic system makes the following assumptions:

(1) Blocks of cipher text cannot be tampered with, and several blocks of smaller cipher cannot compose a bigger block of cipher text.
(2) Two blocks of cipher text in a message are regarded as two arrive, respectively.
(3) Cipher Key Assumption, always assumes that the encryption system is perfect, that is to say, only the principal of the master key could understand the cipher text messages. A cipher key that does not have the right key cannot decode the cipher texts that are generated by the right cipher key. With the correct key to decrypt the cipher text is to definitely have a clearer meaning, and the wrong key decryption expressly does not make sense.
(4) The cipher text contains plenty of redundant information, which makes the whistleblower being judged whether he used the correct key.
(5) The message contains plenty of redundant information, which makes the principal being judged whether the message comes from itself.
(6) Principal Assumption, BAN logic assumes that the principal of the participation agreement is honest.

(7) Time Assumption, two times in protocol analysis are past-time and current-time. Current-time starts at the beginning of this agreement running stage, and before this is past-time. If a viewpoint is at the beginning of the agreement, so the current-time is established. However, the vice may not set up. Therefore, using time to distinguish can prevent message replay.

As a many-sorted modal logic, BAN logic mainly includes three process objects; principal, keys, and formula. The formula also can be named statement. BAN logic only contains proposition conjunction, which can be represented as a comma. The conjunction meets the exchange law and the combining law. According the rules of BAN logic, P, Q, and R represents the variable of principle; K represents the variable of keys and X, Y represents the variable of formula, respectively. A, B represents the normal principal, and S is the certificate server. $K_a, K_b, K_c$ represents specific shared keys; $K_a^+, K_b^+, K_c^+$ represents specific public keys; $K_a^{-1}, K_b^{-1}, K_c^{-1}$ represents specific private keys; $N_a, N_b, N_c$ represents a temporary variable; h(X) represents the one-way hash function. We would introduce the inference of formal analysis of BAN logic as follows:

The goal of formal analysis tool based on the BAN logic system resolves the following questions:

(1) Whether the authentication protocol is correct;
(2) Whether the goal of the authentication protocol is attained;
(3) Whether the initial of the authentication protocols is appropriated;
(4) Whether the authentication protocol is redundant?

3 TRUST MODEL BASED ON OPTIMIZED ANT COLONY ALGORITHM

At present, a series of studies were launched around the trust model of WSNs. However, the traditional trust model based on the biology algorithm for WSNs (BTRM-WSN) model has some demerits, such as high complexity and inaccurate trust calculation (Pan, Yu, & Yan, 2013, Mármo, & Pérez, 2011). In order to address these issues; a Trust Model Based on Optimized Ant Colony Algorithm [23] is proposed. This model is composed of the pheromone update, the path quality assessment, the trust evaluation and the punishment and reward mechanism. In addition, in order to enhance the accuracy of the global pheromone calculation, when the global pheromone is calculated, the optimal solution retention strategy is introduced into the trust model.

3.1 Principle of the Ant Colony Algorithm

In fact, the ant colony algorithm is a simulation of the real ant colony behavior of the simulated evolutionary algorithm (FENG & CHEN, 2014), and it is a randomized algorithm that is based on the natural world and the ant colony foraging behavior. When the ants are out in action, they release in its path through a special secretion called pheromones, and what’s more, ants can sense the substance to guide the direction of their movement. So the more the ants go through the certain path, the more pheromone is left on the path. Thus the probability of choosing the path is greater when other ants are out for food, then the later ants added the ant pheromone at the same path and this phenomenon forms a positive feedback mechanism, and finally the whole ant colony finds the optimal path.

3.2 Update of the Pheromone

The definition of the pheromone is as follows:

Define 1. The value of the pheromone means the trust between nodes, using $\tau_{ij}(t)$ in time $t$. Initially, $m$ ants are put on $n$ initiate nodes and the first element of the tabuk of every ant sets the original node. In the beginning, the pheromone of every track is equal, that is to say, $\tau_{ij}(0) = \zeta(C)$ is a small constant).

Each ant based on pheromone value on the path determines the next node. The probability of ant $k$ transferring from node $i$ to node $j$ is $p_{ij}^k(t)$ is as follows:

$$p_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}(t)}{\sum_{r \in J_k} \tau_{ir}(t)}, & \text{when } j \in J_k(i), \\ 0, & \text{or not} \end{cases}$$

$I_k(i)$ = {1, 2, ..., $n$} - tabuk symbolizes the set of next nodes ant $k$ can choose. When ant $k$ passes a node, it puts this node into tabuk.$k$.

The update of the pheromone consists of the local pheromone update and global pheromone update. Each ant passing one node every time will carry out a local pheromone update. When an ant transfers from a node $i$ to a node $j$, the side $ij$ will update the value of the pheromone as follows:

$$\tau_{ij}(t + 1) = (1 - \rho)\tau_{ij}(t) + \rho\tau_0$$

$\rho$ is an evaporation factor of the pheromone on the path, enabling the path to forget a bad situation in order to avoid the path into a sub-optimal situation. $\tau_0$ is an initiate value of the pheromone on the path? The role of the local pheromone update is that when the ants have found the pheromone of the selected side reduces, it will turn to choose sides, which are not chosen.

After each iteration the global pheromone of the optimum path all the ants have found will be updated. The global pheromone of the optimal path $ij$ is updated as follows:
\[ \tau_{ij}(t + 1) = (1 - \rho) \tau_{ij}(t) + \rho \Delta \tau_{ij} + \Delta \tau_{ij}', \quad (3) \]
\[ \Delta \tau_{ij} = \sum_{k=1}^{m} \Delta \tau_{ik}^k, \quad (4) \]
\[ \Delta \tau_{ij}^k = \begin{cases} \frac{Q}{L_k} & \text{when ant } k \text{ pass the side } ij \text{ } \text{or not} \\ 0 & \text{otherwise} \end{cases}, \quad (5) \]
\[ L_k \text{ is the length of the path ant } k \text{ travels in this iteration}, \Delta \tau_{ij}^k \text{ is the increasing pheromone of the path } ij \text{ the optimal ant passes}, \sigma \text{ is the number of the optimal ants}, L_{gb} \text{ is global optimal solution}. \]

### 3.3 Path Quality Assessment

Each time the ants return the source node after one iteration they will remember the path they traversed. After that, the source node will evaluate the quality of the path this ant passed. Especially ants hold node lists and pheromone on the path. The calculation of the path quality assessment is as follows:

\[ Q(S_k) = \frac{\tau_k}{\text{Length}(S_k) \cdot \text{PLF}} \cdot A_k, \quad (7) \]
\[ \tau_k \text{ is the average pheromone of the path the ant } k \text{ found}, \text{PLF} \in [0,1], A_k \text{ is the radio of the number of ants selecting the same path with the ant } k \text{. By the above method, ants in the trust model will choose the shorter path as much as possible, and will choose this path many times}. \]

### 3.4 Trust Evaluation

In the networks, every node maintains a set of pheromones, determining which path ants will choose. Nevertheless, pheromone values are often confused with trust values. In order to differentiate the two concepts, \( T_i \) is used to signify the trust of node \( i \). The calculation process is as follows:

\[ T_i = \sum_{j=1}^{\text{[i]}} \sin(\theta_{ij}) / \text{[0]} \cdot \text{[0]}, \quad (8) \]
\[ \text{If(i)} \text{ means the nodes having a way to the node } i. \text{ Each node will maintain a trust list on the neighbor nodes}. \]

### 3.5 NPunishment and Reward Mechanism

For the sake of enhancing the accuracy of the trust model and preventing the attack of the malicious node, “addition increase, multiplication decrease” is carried out after the trust evaluation is finished. The process is as follows:

\[ T_i = \begin{cases} \text{T}_{\max} & \text{upper limit} \\ T_i + \text{step} & \text{When the communication is good, } \tau \cdot T_i \text{ When the communication is not good} \end{cases}, \quad (9) \]
\[ T_w \text{ is a trust threshold. If } T_i > T_w, \text{ the communication is good. Otherwise, the communication is not good}. \]

## 4 THE PROPOSED PROTOCOL

Based on the descriptions and explanations on the above protocol, a new enhanced user authentication protocol integrated with biometrics is proposed in an effort to solve the weaknesses. Before analysis, the notations with their corresponding meanings are summarized in Table 1 first.

Our proposed protocol can be illustrated step by step.

### 4.1 Registration Phase

In this phase, the user \( U \) submits an identity \( ID_U \) and a hash password \( PW_U \) to the GW node in a secure way. Then, the GW node issues a smart card to \( U \). Its description is as follows:

1. The user \( U \) selects his/her identity \( ID_U \) and password \( PW_U \) freely, inputs his/her personal biometrics \( B_U \) (fingerprint, etc.) and computes \( f_U = h(B_U), PW_U = h(PW_U \oplus f_U) \). \( U \) sends \( ID_U \) and \( PW_U \) to GW via a secure channel.

2. Once received, GW computes \( K_U = h(x||y||ID_U) \times P \) and \( W_U = h(ID_U||PW_U||f_U) \oplus K_U \). Then GW saves \( \{W_U, PW_U, h(\cdot)\} \) into a smart card and transmits it to the user \( U \) via a secure channel.

### 4.2 Login Phase

In order to access data from WSNs, \( U \) should provide his/her identity \( ID_U' \), identity \( ID_U' \) and password \( PW_U' \). The smart card goes through the subsequent steps to confirm the justifiability of \( U \).

1. \( U \) inserts his/her smart card into the terminal and inputs personal biometrics \( B_U' \), identity \( ID_U' \) and password \( PW_U' \).

2. The smart card computes \( f_U' = h(B_U') \) and \( PW_U' = h(PW_U' \oplus f_U) \), checks whether \( \overline{PW_U'} = PW_U \). If the equation is incorrect, the smart card refuses the demand. Or else, the smart card chooses randomly a number \( \gamma_U \in Z_q^* \) and calculates:

\[ K_U = W_U \oplus h(ID_U \oplus PW_U' || f_U) \times X = \gamma_U \times P \times X' = \gamma_U \times K_U, \omega \]
\[ = h \left( ID_U \left| ID_{\bar{S}_n} \right| \left| X \right| \left| y \right| \left| T_U \right) \right), \]
\[ \alpha = h(ID_U \left| ID_{\bar{S}_n} \right| \left| X \right| \left| X' \right| \left| T_U \right| \left| \omega \right). \]

where \( T_U \) is the current timestamp of the user \( U \). Finally, \( U \) sends \( M_t = \{ID_U, ID_{\bar{S}_n}, X, T_U, \omega, \alpha \} \) to the gateway GW.

1. \( S_n \) first verifies if \( T^- - T_U \leq \Delta T \) and \( T^- - T_U \leq \Delta T \). If the requirement is reached, the validity of \( \Delta T \) and \( T_U \) can be guaranteed. Then \( S_n \) can pursue the next step. Otherwise, \( S_n \) refuses to carry on.
Table 1. Notations.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>p, q</td>
<td>Two large prime numbers</td>
</tr>
<tr>
<td>U</td>
<td>A user</td>
</tr>
<tr>
<td>ID_{U}</td>
<td>A user’s identity</td>
</tr>
<tr>
<td>ID_{S_n}</td>
<td>The identity of sensor node S_n</td>
</tr>
<tr>
<td>PW_{U}</td>
<td>The password of user U</td>
</tr>
<tr>
<td>B_{U}</td>
<td>Biometric template of user U</td>
</tr>
<tr>
<td>x, y</td>
<td>The master keys of GW node</td>
</tr>
<tr>
<td>h(γ)</td>
<td>An anti-collision one-way hash function</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>⊕</td>
<td>A string XOR operation</td>
</tr>
<tr>
<td>F_p</td>
<td>A finite field</td>
</tr>
<tr>
<td>P</td>
<td>A point on an elliptic curve E with an order n</td>
</tr>
</tbody>
</table>

(2) Then S_n tests whether \( \beta = h(\text{ID}_{U}|X||T_{U}|\alpha||\text{ID}_{S_n}) \). If the equation doesn’t correspond, S_n terminates the authentication. Or else, S_n calculates \( X = h(\text{ID}_{U}|X|\gamma \times X \). Then S_n checks if \( \alpha = h(\text{ID}_{U}|X||T_{U}|X|\alpha||\text{ID}_{S_n}) \). If the check fails, the authentication ends. Otherwise, it produces a random number \( \gamma_{a} \in Z_{q}^* \), computes \( Y = \gamma_{a} \times P \) and gets the current timestamp \( T_{S_n} \) and computes \( \gamma = h(\text{ID}_{U}|X|\alpha||\text{ID}_{S_n}|Y||T_{S_n}) \).

In the end, S_n node sends \( M_3 = \{ T_{S_n}, \gamma, \delta \} \). Once receiving \( M_3 \) at \( T'''' \), GW performs to authenticate sensor S_n as follows:

(1) GW tests whether \( T'''' - T_{S_n} \leq \Delta T \). If the condition is reached, the validity of \( T_{S_n} \) can be committed and S_n can proceed. If not, S_n refuses the request.

(2) GW checks whether \( \gamma = h(\text{ID}_{U}|X||\alpha||\text{ID}_{S_n}|Y||T_{G}) \). If the answer is no, the authentication ends. Otherwise, GW gets the current timestamp \( T_{G} \) and transfer \( M_4 = \{ T_{G} \} \) to S_n.

When getting \( M_4 \) at \( T'''' \), S_n proceeds as follows:

(1) S_n checks whether \( T'''' - T_{G} \leq \Delta T \). If not, S_n terminates the following process. Otherwise, S_n computes \( K_{US} = \gamma_{a} \times X \) and the session key \( SK = h(X||Y||K_{US}) \).

(2) S_n gets the current timestamp \( T_{S_n} \). Then S_n sends \( M_5 = \{ Y, T_{S_n} \} \) to U.

Upon receiving \( M_5 \), U operates the below process.

U checks if \( T'''''' - T_{S_n} \leq \Delta T \). If \( M_5 \) is not fresh, U stops the session; otherwise, U inquires the trust of sensor S_n and checks if \( T_{U}(S_{U}) > T_{W} \). If not, U stops the session; otherwise, U computes \( K_{US} = \gamma_{a} \times Y \) and gets session key SK= \( h(X||Y||K_{US}) \). U and S_n can make sessions.

4.3 Password Update Phase

When the user wants to reset his password, the password update phase happens. The password update phase is explained below.

(1) The user inserts his/her smart card into the terminal and enters the biometric templates \( B_{U} \), the identity \( \text{ID}_{U} \), the old password \( PW_{U} \) and a new one \( PW_{U}' \).

(2) The smart card checks if \( f_{U} = h(B_{U}) \). If the answer is no, U rejects the demand. Otherwise, the smart card computes \( PW_{U} = h(PW_{U} \oplus f_{U}) \). The smart card checks if \( PW_{U}' = PW_{U} \). If the equation doesn’t hold, the password update phase is over. Otherwise, the smart card calculates \( K_{U} = W_{U} \oplus h(\text{ID}_{U}|PW_{U}|f_{U}) \) and replaces \( W_{U} = h(\text{ID}_{U}|PW_{U}|f_{U}) \) and \( W_{U}' = h(\text{ID}_{U}|PW_{U}'|f_{U}) \) and correspondingly.

5 TRUST MODEL ANALYSIS

WITH the aim of assessing the performance and reliability of the proposed trust model, this paper is compared with BTRM-WSN models throughout three simulations using TRMSim-WSN (Már mol, Pérez, 2009). The first set of experiments compares the two models in the
accuracy of the search trusted node. The second set of experiments compares the two models for the average path length for finding the required trusted node. The third set of experiments compares the total energy the two models consume.

### 5.1 TRMSim-WSN

TRMSim-WSN is a Java-based wireless sensor network trust model simulation, which provides a very convenient way to test the trust model. In order to simulate two trust models, some parameters need setting as; Client: 15%, Relay Server: 6%, Radio Range: 10, Min Num Sensors and Max Num Sensors: 100, Num networks: 400, Number executions: 100. Then, the simulator will randomly generate a wireless sensor network based on these parameters. Note that 85% of the nodes are server nodes to provide the service.

### 5.2 Accuracy

The exact rate of the trust model was used to estimate the reliability and safety of the trust model, which uses the ratio of the number of credible sensor nodes the trust model chooses successfully to the total number of treatments. A good trust model is able to have good control over the malicious node attacks. The accuracy of the trust model and the accuracy of the BTRM-WSN were compared in Figure 1. It can be seen over the figure that when the ratio is less than 50 percent of malicious nodes, the two trust models on the accuracy of finding the credible nodes is similar. However, when the ratio of malicious nodes is higher than 50 percent, the proposed model can provide higher accuracy and security than the BTRM-WSN.

### 5.3 Energy Consumption

The wireless sensor network is an energy-limited network, so a designed trust model for the wireless sensor network is required to reduce energy consumption to ensure network security and enlarge the life cycle of the networks. Energy expenditure in the wireless sensor network includes the energy source nodes that are consumed in sending information and the energy consumed by malicious nodes providing malicious service and the energy consumed in order to search trusted nodes in the networks. Figure 2 is a comparison of two trust models in energy consumption. As shown, the energy consumption of the trust model proposed is less than the BTRM-WSN, so the proposed trust model is more suitable for the wireless sensor network.
6 SECURITY ANALYSIS

In this section, the security of our proposed protocol for WSNs will be analyzed using BAN-logic.

6.1 Analysis of the Proposed Protocol Using BAN-logic

In this subsection, BAN-logic (Burrows, Abadi, Needham, 1989) is used to analyze the proposed protocol. First of all, some notations and statements are shown as follows: Where X, Y and K symbolize statements and P signifies a principal.

- (X, Y): conjunction of two statements X and Y.
- \{X\}+: X is encrypted with the key K.
- \{X\}_+: X is decrypted with the key K.
- (X)_+: X is hashed throughout the key K.
- \langle X\rangle: X combined with Y.
- #(X): X is fresh.
- P\equiv X: P believes X and will continue to do something for the rest of the protocol.
- P\triangleright X: P has jurisdiction over X and should be trusted as to X.
- P\sim X: P once conveyed X.

In addition, some main logic postulates of the BAN-logic are described as follows, which will be used in our analysis.

- The message-meaning rule: \[ P \equiv \text{X} \implies P \equiv X \]
- The freshness-conjunction rule: \[ P \equiv #(X) \implies P \equiv X \]
- The nonce-verification rule: \[ P \equiv #(X), P \equiv Q \implies X \]
- The jurisdiction rule: \[ P \equiv Q \triangleright X, P \equiv Q \equiv X \]

Our proposed user authentication protocol needs to meet the following goals.

- Goal 1. U|\equiv (U \leftrightarrow S_n)_K
- Goal 2. U|\equiv S_n|\equiv (U \leftrightarrow S_n)_K
- Goal 3. S_n|\equiv (U \leftrightarrow S_n)_K
- Goal 4. S_n|\equiv U|\equiv (U \leftrightarrow S_n)_K
- Goal 5. U|\equiv (U \leftrightarrow S_n)_K
- Goal 6. U|\equiv S_n|\equiv (U \leftrightarrow S_n)_K
- Goal 7. S_n|\equiv (U \leftrightarrow S_n)_K
- Goal 8. S_n|\equiv U|\equiv (U \leftrightarrow S_n)_K

First of all, it’s necessary to make some assumptions in order to derive our goals.

- A1: U|\equiv #(T_U)
- A2: U|\equiv #(T_U)
- A3: U|\equiv #(T_{Gn})
- A4: U|\equiv \overset{\omega=h(ID_U||ID_Sn||x||y||T_U)}{\rightarrow} GW
- A5: U|\equiv S_n \rightarrow U \leftrightarrow S_n
- A6: U|\equiv S_n \rightarrow U \leftrightarrow S_n
- A7: GW|\equiv #(T_U)
- A8: GW|\equiv GW
- A9: GW|\equiv S_n \overset{\beta=h(ID_U||x||T_U||x||y||ID_Sn)}{\rightarrow} GW
- A10: S_n|\equiv #(T_U)
- A11: S_n|\equiv #(T_G)
- A12: S_n|\equiv #(T_G)
- A13: S_n|\equiv GW \rightarrow U \leftrightarrow S_n
- A14: S_n|\equiv GW \rightarrow U \leftrightarrow S_n
- A15: S_n|\equiv U \rightarrow U \leftrightarrow S_n
- A16: S_n|\equiv U \rightarrow U \leftrightarrow S_n

Second, the proposed protocol needs to be transformed into the idealized form.

-(Msg1. U\rightarrow GW(T_U, U \leftrightarrow S_n, U \leftrightarrow S_n)_{h(ID_U||PWU||T_U)}
-(Msg2. GW\rightarrow S_n, T_U, U \equiv U \leftrightarrow S_n, U \equiv U \leftrightarrow S_n
-(Msg 3. S_n \rightarrow GW: T_{sa}, U \equiv U \leftrightarrow S_n, U \equiv U \leftrightarrow S_n
-(Msg 4. GW\rightarrow S_n, T_{sa}, U \equiv U \leftrightarrow S_n, U \equiv U \leftrightarrow S_n
-(Msg 5. S_n \rightarrow U: T_{sa}, U \equiv U \leftrightarrow S_n, U \equiv U \leftrightarrow S_n

Third, the idealized outline of the proposed protocol is analyzed using the BAN logic. The chief steps are depicted as follows:

- By Msg 1, it is easy to reach the statement as follows:

  \[ 1_{S_U} \rightarrow GW(T_U, U \leftrightarrow S_n, U \leftrightarrow S_n)_{h(ID_U||PWU||T_U)} \]
By A9, $S_1$ and the message-meaning rule it is easy to achieve

$$ S_2: \text{GW} | \equiv U | \sim (T_{10}, U \leftarrow IDU S_n, U \leftrightarrow S_n). $$

By A7, $S_2$ and the freshness-conjunction rule, it is easy to achieve

$$ S_3: \text{GW} | \equiv U | \equiv (U \leftarrow IDU S_n, U \leftrightarrow S_n). $$

By $S_3$ and the belief rule, it is easy to get

$$ S_4: \text{GW} | \equiv U | \equiv (U \leftarrow IDU S_n). $$

By $S_4$ and the conjunction rule, it is easy to achieve

$$ S_5: \text{GW} | \equiv U | \equiv (U \leftarrow IDU S_n). $$

By Ms 2, 3, 4, it is easy to achieve

$$ S_6: S_n \leftarrow (T_{10}, T_{10}, U | \equiv U \leftarrow IDU S_n, U | \equiv U \leftarrow IDU S_n). $$

By $S_6$ and the jurisdiction rule, it is easy to achieve

$$ S_7: \text{GW} | \equiv U | \equiv (U \leftarrow IDU S_n, U | \equiv U \leftarrow IDU S_n). $$

By A13, A14, S6 and the message-meaning rule, it is easy to achieve

$$ S_8: s_n \equiv \text{GW} | \sim (T_{10}, T_{10}, U | \equiv U$$

By A10, A11, A12, S6 and the freshness-conjunction rule, it is easy to achieve

$$ S_9: s_n \equiv \text{GW} | \equiv (U | \equiv U \leftarrow IDU S_n). $$

By $S_9$ and the belief rule, it is easy to achieve

$$ S_{10}: s_n \equiv \text{GW} | \equiv (U | \equiv U \leftarrow IDU S_n). $$

By A14, S10 and the jurisdiction rule, it is easy to attain

$$ S_{11}: s_n \equiv \text{GW} | \equiv (U | \equiv U \leftarrow IDU S_n). $$

By A16, S11 and the jurisdiction rule, it is easy to attain

$$ S_{12}: s_n \equiv \text{GW} | \equiv (U | \equiv U \leftarrow IDU S_n) (\text{Goal 8}). $$

By A13, S12 and the jurisdiction rule, it is easy to attain

$$ S_{13}: s_n \equiv \text{GW} | \equiv (U | \equiv U \leftarrow IDU S_n) (\text{Goal 7}). $$

By A16, S13 and the jurisdiction rule, it is easy to attain

$$ S_{14}: s_n \equiv \text{GW} | \equiv (U | \equiv U \leftarrow IDU S_n). $$

By A14, S14 and the message-meaning rule, it is easy to attain

$$ S_{15}: s_n \equiv \text{GW} | \equiv (U | \equiv U \leftarrow IDU S_n) (\text{Goal 4}). $$

By A16, S15 and the jurisdiction rule, it is easy to attain

$$ S_{16}: s_n \equiv \text{GW} | \equiv (U | \equiv U \leftarrow IDU S_n) (\text{Goal 3}). $$

By A17, S16 and the message-meaning rule, it is easy to attain

$$ S_{17}: s_n \equiv \text{GW} | \equiv (U | \equiv U \leftarrow IDU S_n, U \leftrightarrow S_n). $$

By A8, S17 and the message-meaning rule, it is easy to attain

$$ S_{18}: U | \equiv S_n | \sim (T_{10}, U \leftarrow IDU S_n, U \leftrightarrow S_n). $$

By A3, S18 and the freshness-conjunction rule, it is easy to attain

$$ S_{19}: U | \equiv S_n | \equiv (U \leftarrow IDU S_n) (\text{Goal 6}). $$

By A6, S19 and the jurisdiction rule, it is easy to attain

$$ S_{20}: U | \equiv S_n | \equiv (U \leftarrow IDU S_n) (\text{Goal 5}). $$

By A3, S18 and the freshness-conjunction rule, it is easy to attain

$$ S_{21}: U | \equiv S_n | \equiv (U \leftarrow IDU S_n). $$

Since $SK = h(X||Y||K_{US})$, it is easy to attain

$$ S_{22}: U | \equiv S_n | \equiv (U \leftarrow IDU S_n) (\text{Goal 2}). $$

By A5, S22 and the jurisdiction rule, it is easy to say

$$ S_{23}: U | \equiv S_n | \equiv (U \leftarrow IDU S_n) (\text{Goal 1}). $$

Throughout (Goal 1), (Goal 2), (Goal 3), (Goal 4), (Goal 5), (Goal 6), (Goal 7), (Goal 8), it is indicated that both $U$ and $S_n$ believe that a session key $SK$ and an identity $IDU$ are shared between them.

### 6.2 Other Discussion

**THE OREML.** The proposed user authentication protocol can assist a stolen-verifier attack and many users who suffer from the same login-id attack, man-in-the-middle attack, the session key attack, the stolen-smart card attack and the sensor energy exhausting attack.

1. **Stolen-verifier Attack:** An attacker can do the stolen-verifier attack if the GW nodes stores verifier tables for authentication. However, our proposed protocol doesn’t need any verifier tables. Therefore, it’s impossible that the stolen-verifier attack happens.

2. **Many Users who suffer from the Same Login-id Attack:** In the login process to WSNs. Our proposed protocol requires the user to provide his/her identity $ID_U$, the password $PW_U$, and his/her personal biometrics $B_U$. It is noted that $PW_U$ is in alliance with $f_U$. Though two attackers may have the same identity $ID_U$ and the same password $PW_U$, their biometrics $B_U$ and $PW_U = h(PW_U \oplus f_U)$ are distinct. Therefore, our proposed protocol can defend this attack.

3. **Session Key Attack:** In a session key attack, an attack intercepts the user’s sending
authentication message and constructs a session key with the user. In our proposed protocol, $\alpha = h(ID_U || ID_{Sn} || X || X' || T_{U} || \omega)$ consists of $ID_U$ and $ID_{Sn}$ denoting the user wanting to communicate with the sensor node $Sn$. GW node uses $\delta = h(ID_U || X || X' || T_{U} || Y || T_{Sn})$ denoting GW node has authenticated the status of the user $U$ and the sensor node $Sn$. The session key attack can be avoided by using $\alpha$ and $\delta$.

7 PERFORMANCE ANALYSIS

Table 2 will contrast our proposed protocol with Yeh et al.'s protocol and Xue et al.'s protocol towards the computation cost. First, notations need to be illustrated as follows:

H: The time for calculating the hash function;
X: The time for calculating the bit XOR; E: The time for calculating the bit ECC encryption/decryption. Table 2 of computer cost comparison is as follows:

As shown in Table 2, the total computational cost of three protocols is $8H+8E, 24H+13X$ and $14H+2X+6E$ respectively. Our protocol and Yeh et al.'s protocol have better capability than Xue et al.'s protocol in the $Sn$ and the GW. In WSNs, our proposed protocol and Yeh et al.'s protocol can save energy consumption better and be more effective for the WSNs than Xue et al.'s protocol. The property of our protocol is analogous to that of Yeh et al.'s protocol. However, Yeh et al.'s protocol is flimsy to some common attacks. The proposed protocol in the article can contribute to the safety of the user authentication protocol. Therefore, the proposed protocol has better performance than other protocols.

Table 2. Computation Cost.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>User</th>
<th>Sensor</th>
<th>Gateway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeh et al.'s protocol</td>
<td>H+2E</td>
<td>3H+2E</td>
<td>4H+4E</td>
</tr>
<tr>
<td>Xue et al.'s protocol</td>
<td>7H+5E</td>
<td>7H+5X</td>
<td>10H+3X</td>
</tr>
<tr>
<td>Our proposed protocol</td>
<td>6H+2X+3E</td>
<td>5H+3E</td>
<td>3H</td>
</tr>
</tbody>
</table>

8 CONCLUSION

In this paper, a novel user authentication based on a trust model, biometrics and ECC is proposed in the WSN. Throughout the trust model analysis, the model shows higher accuracy and less energy consumption. Analysis throughout the BAN logic indicates the proposed user authentication protocol can attain higher security requirements as to user authentication in WSNs. Moreover, it can resist a variety of attacks. Performance evaluation indicates that the proposed protocol has better performance than Xue et al.'s protocol and Yeh et al.'s protocol. In the future, the current work towards the WSNs will be hereby extended.

9 ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 61300170, No. 61501005), the University Provincial Natural Science Foundation of Anhui Province (No.KJ2016A057), the Natural Science Foundation of Anhui Province under Grant (No. 1608085MF147).

10 REFERENCES


11 NOTES ON CONTRIBUTORS

Tao Liu is a professor. She received B.S. degree in Computer Application Technology from HeFei University of Technology, China in 2004. She was a visiting scholar at the University of Science and Technology of China in 2012. Her main
research interests include machine learning, computer network and Information security. Email:liutao@ahpu.edu.cn

Gan Huang received M.S. degree from the School of Computer and Information at Anhui Polytechnic University, China, in 2016. He is now studying at the Department of Electrical and Computer Engineering, Sungkyunkwan University for a Ph.D. degree. In his master’s study, he worked on wireless sensor networks. His current research interests are in edge computing and software-defined networking. In this paper, he is corresponding author. Email:ahczhg@skku.edu

Ping Zhang received the B.S. degree in mathematics from Anhui Normal University, China, in 2004, the M.S. degree in statistics from Southeast University, China, in 2007, and the Ph.D. degree in the School of Information Science and Engineering from Southeast University, China, in 2015. His research interests include sensor network localization, network data mining, and statistical signal processing.
The Machine Learning based Finite Element Analysis on Road Engineering of Built-in Carbon Fiber Heating Wire

Yuhua Peng\textsuperscript{a}, Dingyue Chen\textsuperscript{a}, Lihao Chen\textsuperscript{b}, Jiayu Yu\textsuperscript{b} and Mengjie Bao\textsuperscript{a}

\textsuperscript{a}Chang'an University, Xi'an, China
\textsuperscript{b}Northwest University, Xi'an, China

\section*{ABSTRACT}
For the study of the effect of deicing with carbon fiber heating wire in the bridge pavement structure, through built-in carbon fiber heating wire in the bridge pavement structure, experimental studies were carried out indoor on the effects of thermal conductivity in different embedding positions, layout spacing and the installs power of carbon fiber heating wire. With indoor laboratory test data as the basic parameters, using ABAQUS finite element software simulation, an analysis was carried out of the degree that the surface temperature of the heating wire, the thermal physical parameters of asphalt concrete, and environmental conditions have influence on the melting effect. With a 50 m bridge between the left Shanggaoqiao tunnel exports with 2\# Dabaozhai tunnel entrance in Ma Zhao Highway as the test section, this paper introduces the carbon fiber heating wire grooving laying process and construction methods.

\section*{1 INTRODUCTION}
In the winter, the roads in most areas of China will be frozen, because of snowfall, freezing rain and some other natural climate, these situations would bring serious impacts on roads unimpeded and traffic safety. Because the highway bridge and tunnel portal pavement, which in cold and high altitude areas, above natural ford or canyon, the wind is strong, bridge decks and pavements could freeze easily, and those sections of highway are high incidence of a freeze disaster. In order to keep the roads unimpeded and safe, traffic management departments mobilize a lot of manpower and material resources to participate in resisting freezing and keeping roads unimpeded, maybe that would arrive at getting effective results, but with a high cost.

There is little research about bridge pavement deicing in our country, the research achieved doesn’t give satisfaction to the people, for example: using industrial salt and deicing chemicals, which could bring enormous negative effects on the environment, and material composition used to try and lower the freezing point of bridge deck asphalt up-layer or use flexible pavement form self-break-ice pavement structure, these methods won’t supply us great satisfaction in practice.

There is little research about the heating wire systems applied to bridge pavement structure deicing in China. (Guo Zhengu, et. al. 2012) researched how to apply heating wires to road deicing, and introduced the construction technology detailed. (Yang Jie, Li Hai-tao etc. 2009) explored how to use electro thermal methods to clear road snow, and researched the technical index of heating wires. Li (Li Yanfeng, et. al. 2008) and (WU Haiqin 2005) (Li Yanfeng, et. al. 2006) who are at Beijing University of Technology has carried on the theoretical analysis about some key problems of heating wire deicing, and did some experiments indoor, researched the system’s pavement efficiency, the circuit control technology, and economy.

There are various kings of material heating wires. Compared with other materials, carbon fiber has the feature of low density, heat-resistant, abrasion resistance, anti-corrosion, low heat expansion coefficient, fever quickly, high electro thermal conversion efficiency, can adjust to heating temperature and energy by wires’ length and voltage, in the conditions that current load’s area is the same. Carbon fiber’s intensity is 6–10 times higher than metal wires’. And it’s not easy to break off, with a long life cycle.
This article is aimed at researching the method of arranging carbon fiber heating wires in bridge deck construction, determine the layout’s distance, pavement power, layout’s location, analyze concrete’s performance parameter, which is used in experiments by ABAQUS, simulate bituminous concrete’s thermophysical properties, heating wires’ heat, environmental conditions and some different influence factors, analyze the heating rate’s change in bridge pavement when influence factors changed, come up with specific construction methods and rely on Mliuwan-Zhaotong highway at last.

2 CARBON FIBER HEATING WIRES DEICING PAVEMENT EXPERIMENT

The typical pavement form on a highway bridge pavement is; 4cm up-layer + 6cm middle-layer + 10cm concrete pavement, an experiment on indoor mold corresponding thickness test piece was done respectively. Consider the convenience of construction, and the main inbuilt places of carbon fiber heating wires are; between up-layer and middle-layer (as shown in Figure 1a); between middle-layer and concrete pavement (as shown in Figure 1b); in concrete pavement (as shown in Figure 1c), 3 ways.

The gap of concrete’s bar-mat reinforcement is 10cm usually, when concrete is not congeal, the need to strap heating wires and bar-mat reinforcement, could explore influence of heating wires’ inbuilt place to bridge pavement deicing.

2.1 Carbon Fiber Heating Wires’ Inbuilt Place

When doing the experiment, set the space of heating wires to 10cm, set the power of heating wires to 35W/m, set temperature to -5°C, no wind, place the test piece in PJT horizontally (Yang Fei. 2014). Keep the bottom of test piece superterranean, the upper surface is a heat delivery surface. When test piece’s temperature achieve -5°C, start experiment with electricity (Shi Yiping, Zhou Yurong. 2006), the voltage loaded on heating wires is 65V, keep galvanical for 2 hours, the data of the test piece surface’s temperature changes over time by temperature data logger.

According to Figure 2, we can see that there is a big difference in the surface temperature of the specimen-heating wires located in 3 different depths after 2 hours. When the heat wires are placed in the steel bar network within the concrete layer, the surface temperature remains -5°C after the power is on for 2 hours, the temperature is low and the effect of heat is poor (Zhang Huiyu, Zou Lingdeng. 2011). When the heating wires are placed between the middle-layers and the concrete layer, a better effect can be obtained. When the heating wires are placed between the upper-layer and middle-layer and the surface temperature reaches -4°C after 2 hours with the power on, the highest temperature and best effects are obvious and the surface temperature increases with the increasing of the powering time.

In order to prevent the stability in high temperature from injury resulting from high temperature in the surface layer of the asphalt, we should constrict the temperature of the heating wires within a rational area.

2.2 Carbon Fiber Heating Wires’ Burying Spacing

To obtain directly the effects about the spacing of the heating wires on heat conduction, setting the spacing to 5 cm, 8 cm and 10 cm respectively indoors, the specimens in the junction of up-layer and middle-layers and the power to 35W/m. The heating wires and environmental temperature are – 5°C, to process circuit and load voltage 65 V. In addition, keeping galvanical for 2 hours and setting the interval of temperature data logger for 5 minutes. Finally, collecting the surface temperature data of heating wires respectively, using three kinds of spacing. Figure 3 shows the surface temperature of heating wires with three kinds of spacing.

According to Figure 3, under the same heating time, the smaller the heating wires are spaced, the higher the wires’ surface temperature in the same station, the better the effect about the temperature sensitivity (Chen Depeng, et. al. 2007). However, when the burying area is the same, the smaller the heating wires are spaced, the longer the heating wires are used. Hence, considering the construction condition, economic cost and temperature effect, the spacing of heating wires using 10 cm is relatively appropriate.

2.3 The Pavement Power of Carbon Fiber Heating Wire

In order to determine the influence of power for the heating system on surface temperature of the specimen, we choose 6 different types of power: 200W/m², 400W/m², 600W/m², 800 W/m², 1000W/m², 1200W/m², the spacing of the heating wires is 10cm, the laying place is located between the upper and median layer with the environment temperature is -5°C under windless conditions. The input voltages corresponding to the 6 powers are listed in the Table 1.

When the temperature of the specimen decreases to -5°C, we put on the circuit and load the voltage of the heating wires according to Table 1. Collecting the temperature data of the specimen’s surface temperature based on the spacing of 300s after 1 hour and the data of the surface temperature of the specimens can be finally obtained.

From Figure 4 we know that with increasing the power, the surface final temperature of the heating wires increases constantly. When the power is larger than 800W/m², the final temperature of the specimen’s surface obtains an obvious increase-up to 3.4°C, whereas, when the power is 200~800W/m², there is no
obvious difference in the final temperature of the specimen’s surface (Zhang Renyi, GU Qiangkang, 2012), fluctuating within -2--1°C. After the power is on for 1 hour, the temperature increases gradually corresponding to 400–800W/m², the temperature increases rapidly when the power is larger than 800W/m². At the same time, there is a big difference in the stable temperature that the heating wires can reach under different power shown in Table 2 below.

According to Table 2, with increasing pavement power, the surface temperature of the heating wires increases simultaneously. When the power is larger than 600W/m², the surface temperature of the cable has already over 45°C, which may influence stability in high temperature of the asphalt mixture.

Thus the operating cost will be reduced meanwhile we should make sure that the system can melt ice and snow in time (Gao Yiping, 1999). Considering the factor that the temperature of heating wires should not be too high, we choose power of the heating system as 400W/m² under the same situation as the experiment.

3 DEICING OF CARBON FIBER HEATING WIRE SIMULATION BY ABAQUAS

To analyze the heat convey process of bridge pavement deicing with carbon fiber heating wires, utilizing the ABAQUIS finite element software to simulate the effect of melting ice of heating wires
Table 1. Pavement Power and Corresponding Input Voltage.

<table>
<thead>
<tr>
<th>Pavement power (W/m²)</th>
<th>200</th>
<th>400</th>
<th>600</th>
<th>800</th>
<th>1000</th>
<th>1200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing voltage (V)</td>
<td>37</td>
<td>42.7</td>
<td>52.3</td>
<td>60.4</td>
<td>74</td>
<td>80</td>
</tr>
</tbody>
</table>

Figure 4. Test Piece’s Surface Temperature Rise Curve in the 6 Pavement Power.

Table 2. The Heating Wires’ Surface Temperature Changes with Different Pavement Power.

<table>
<thead>
<tr>
<th>Pavement power (W/m²)</th>
<th>200</th>
<th>400</th>
<th>600</th>
<th>800</th>
<th>1000</th>
<th>1200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature of cable’s surface (°C)</td>
<td>18.6</td>
<td>38.7</td>
<td>58.9</td>
<td>79.0</td>
<td>99.1</td>
<td>118.3</td>
</tr>
</tbody>
</table>

under the different condition and comparing with experiment data indoor (Gao Qing, et. al. 2007). Last, considering the result that to analyze the concrete thermal parameters and simulate the asphalt concrete thermal physical performance parameters, the temperature of heating wires and environment condition as the basis choosing carbon fiber.

3.1 Concrete Thermal Parameters Analysis

Initially, setting the concrete thermal parameters by ABAQUS, and utilizing a trial and error method to determine the test coefficient of thermal conductivity of the mixture indoor is done. The basic step of a trial and error method; assuming reasonably a coefficient of thermal conductivity of the concrete and simulating the heating process of concrete slab (Tang Zuquan, et. al. 2001), then comparing the concrete slab temperature field distribution calculated by conductivity coefficient and the experimental results. If the simulation value is smaller than experimental value, conductivity coefficient needs to be increased, antisense versa. The course is over until one is close to the other.

(a) up-layer trial
(b) middle-layer trial
1) Test plan

When locating the heating wires in the junction of the upper-layer and middle-layers, setting the space of heating wires to 10cm and the pavement power to 400W/m², the coefficient of thermal conductivity of the asphalt surface is measured. However, the heating wires in the inner of the cement pavement, 4 cm from the bottom of the middle-layers, the space of heating wires to 10cm and the pavement power to 400W/m² for the coefficient of the thermal conductivity of the cement concrete is measured.

2) Trial process

When calculating the temperature raising of the concrete slab, selecting a different value of coefficient of thermal conductivity and specific volume heat (Liu Hongmei. 2006) is done. By comparing the simulated values with ABAQUS and test values indoor, concerning the concrete surface temperature distribution and temperature curve that shows the temperature rising locating in the minimum temperature of the concrete surface and keeping the error within 1°C, the coefficient of thermal conductivity and specific volume heat of the test specimen is determined. Figure 5 shows the comparison curve graph that represents simulated values and test values about the thermal physical performance parameters in different layers.

Figure 5 indicates that the simulated value is near to the test. According to the result of the test indoor, the thermal physical performance parameters of the concrete in different layers are determined, for example in Table 3.

3.2 The Study of Bridge Deck Melting Ice Influencing Factors

Influence of asphalt concrete thermo-physical performance parameters
In order to study the influence of asphalt concrete thermo-physical performance parameters on the bridge deck surface temperature, taking asphalt concrete density, thermal conductivity, specific heat capacity value (Yuhui Wu, 2010), the heating line depth of 4cm, environmental and pavement structure initial temperature of -5°C, the heating line epidermis temperature of 40°C and windless conditions run 4h, simulating the influence of the heating-up rate through changing parameters.

a) The influence of the asphalt concrete density on the surface warming effect.

Concrete density can be drawn through laboratory test measured calculations. All kinds of concrete mix, materials and density are different, prevailing by the measured results. Take the upper-layer of asphalt concrete density being 2200Kg/m³, 2300Kg/m³, 2400Kg/m³, 2500Kg/m³ and 2600Kg/m³. After power-4h, the curve of concrete slab surface warming rates corresponding different concrete densities was shown in Figure 6.

As can be seen from Figure 6, with the increase of the density of asphalt concrete, the warming rate decreased. Therefore, to reduce the density of asphalt mixture can improve the road performance of the asphalt mixture.

b) The influence of asphalt concrete thermal conductivity on the surface warming effect.

Thermal conductivity refers to under the heat transfer conditions, the heat flux inside the object generated at the role of the unit temperature gradient (Qianwen Zhang, et. al. 2015), characterization of the size of the object thermal capacity and the unit is \( W/(m^2\cdot ^\circ C) \). Factors affecting the thermal conductivity of asphalt concrete mainly include aggregate, porosity, surface structure of the bridge deck, whetstone ratio, moisture and heat conductive filler, etc., in which the aggregate has played a decisive role to the thermal properties of asphalt concrete. Therefore, to obtain a
thermal conductivity of specific asphalt concrete, you will need the aid of experimental testing. The thermal conductivity of concrete refers to literature (Sherif Yehia, et al. 1999) obtaining a range of values, and then were confirmed by comparison between laboratory tests and the results obtained and error method.

c) The study of temperature changing in the thickness direction.

In 2.1, the warming condition of the heating line buried in a different position measured by the test, the test measured the heating line located between the middle-layer and the upper-layer, which heated better, so in this position the simulate heating line temperature is changing in the thickness direction.

Buried depth of 4cm (distance from the surface of the model), the heating line placement pitches 10cm, ambient temperature and the initial temperature of the pavement structure set to -5°C, the heating line skin temperature 40°C, no wind, run 4h, research shows the bridge surface temperature changes when the heating line cloth depth changes.

Figure 7 shows the final temperature cloud model, the highest temperature area close to the heating line and the lowest temperature area away from the heating line.

As seen from Figure 7, the temperature field is axisymmetric when shown the buried depth of 4cm. Take the models upper left corner as a coordinate origin, to observe the final temperature value of a measuring point at intervals of 0.03m and found a difference between the surface maximum temperature and the lowest temperature only 1.5°C, which the surface uniformity is better.

![Figure 7. The Final Temperature Cloud Model](image)

Thus, taking into account the construction continuity of the pavement structure layer, warming rate, uniformity and other factors (Xie Ping, et al. 1996), the heating line is buried at a depth of 4cm at best. Similarly, when the pavement structure is three layers, the heating line should be located between the upper-layer and middle-layer; if the pavement structure is two layers, the heating line should be located between the upper-layer and under-layer.

4 THE ENGINEERING APPLICATION

4.1 Project Summary

SELECT the road of Maliuwan-Zhaotong highway, which is easy to freeze between the exit of Shang Gaoqiao tunnel and Da Baozhai 2# tunnel’s exit (K43+064.50 ~ K43+114.50) in the left roadway (Sherif Yehia, et al. 2000), which the length is 50m, being the test section. In the K43+071 and K43+101 installing bridge expansion joints in two places are done. The need to empty out the space of the one meters, actual length of the pavement area is 48 meters. This project uses the carbon fiber in the bridge deck structure internally, after being electrified there will form a thermal heating element, the heat will be transferred to the bridge through asphalt upper-layer, the temperature of the bridge deck will be higher than the freezing point.

4.2 The Construction Technical Requirements

At present a mature way of the carbon fiber wire hotline laid mainly is divided into the following two kinds; the first one is made from carbon fiber heating grid, the second one is groove. The advantages and disadvantages of the two kinds of the laying methods a comparative analysis is shown in Table 4.

Due to the surface layer being thinner, the existence of the heat grid will affect the bonding between the layers directly, easy to cause the asphalt pavement traction, upheaval and shorten the service life of bridge deck pavement. By means of the grooving way, the carbon fiber wire heating wires are embedded in the middle-layer, not affecting the upper-layer’s pavement and roller compaction construction, and the impact on the bond strength between layers is small. Therefore, this project has chosen the grooving way to lay and fix carbon fiber heating wires.

When using grooving the groove’s depth and width are chosen according to the standard of the carbon fiber wires heating wire. The projects main grooves are divided into three categories; the first one is across the driveway, used in the embedded carbon fiber heating wires (Zenewitz. J. A.. 1997); the second slots are to embed the K type joint connected with heating
Table 4. Groove Way Compared with Fever and the Advantages and Disadvantages of the Grille.

<table>
<thead>
<tr>
<th>Comparison method</th>
<th>Pavement method</th>
<th>Dapping method</th>
<th>Heating grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixing method</td>
<td>in the surface of the middle-layer, making the heating wire in grooves</td>
<td>Spray tack coat, fix the steel nail on middle-layer</td>
<td></td>
</tr>
<tr>
<td>The construction problems that may occur</td>
<td>Debris in the trench may abrade the heating wires</td>
<td>The heating grid will be rolled up and the heating wire will be crushed when the paver make the follow-up completion operation</td>
<td></td>
</tr>
<tr>
<td>The influence on pavement</td>
<td>Grooving destroys the middle−layer and is nothing wrong with the up-layer</td>
<td>The heating grid has certain thickness, both its rigidity and strength are different from original pavement. Damage may be emerged after traffic service</td>
<td></td>
</tr>
<tr>
<td>Difficulty Level of the construction</td>
<td>Unspecific type, worksite layout need put the heating wires in order. It has more connectors comparing with the normal wire</td>
<td>Have been fixed well and been specific; easy construction</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8. A Single Set of Carbon Fiber Cable Layout Diagram.

wires cold and hot wires, which are in the left of passing lane; the third one is used to connect the cold wire and it’s joint. Should choosing the appropriate grooving machines when constructing, its working performance needs to satisfy the following conditions: (a) The first slots are about 5 mm wide, 10 mm deep. A single slot is 15 m long, it’s groove spacing is 10 cm, all of the bridge decks should be grooved, it’s corner part should be rounded; (b) The second slots are about 80 mm wide, 25 mm deep, a single slot is 1 m long; (c) The third slots are 30 mm wide, 25 mm deep, a single slot is about 8.5 cm long.

In the end of the groove, the groove should be cleaned up then lay the hair heating wires in the first kind of the slots. Five bars of the wire should be precast with a K type connector together in actual construction, unified arrangement. Each slot when necessary, the SBS modified asphalt will be send for the heating wires, which are bonded in the groove. K type of the connectors are placed in the second kind of slot, the connecting wires are placed in the third kind of slot and should be supported well. A single set of carbon fiber wire layout diagram is shown in Figure 8.

When the whole heating wires system is laid, a electrify test is done for 5~10 minutes. The normal operation indicates that the installation and use of the heating system is right. On the contrary, if we found the cable damaged, we should rework immediately.

4.3 The Construction Scheme

1) The first step should be lineation when we make grooves, the following step is grooving. The carbon fiber wire must be in the embedded groove (Cress. M. D.,. 1995), and it can’t be shown on the surface of the middle-layer. The wire installation location should avoid the bridge expansion joints. Before grooving the entire section, we should try to groove and make first of all, and be clear that what grooving machines and the installation technology of carbon fiber should be selected.

2) Installing cables should not be installed in the construction environment at 5 degrees below zero or less, when the environmental temperature decreases,
the cable will get hard and then is not good to be installed, and heat cable appropriately (Rebala S. R., Estakhri C. K., 1995). The heating wire placed in the groove should be tried to avoid bump phenomenon of projection, and to prevent the heating wire being rolled up when the paver is working. If the asphalt of the under-layer is damaged at the corners of the installation the carbon fiber wire can’t be fixed, and steel nails are used to fix it.

5 CONCLUSION

FOR carbon fiber heating wire in a bridge deck structure and the use of three kinds of layout ways makes the indoor experiment results show that heating wires, which are between the middle-layer and upper-layer, will generate the highest calorific efficiency. Melting ice time is too long when the heating wires are between the middle-layer and cement concrete pavement. When heating wires are located inside the paving layer, there is almost no change for the temperature on the surface. Therefore, when the pavement structure is of three layers, the layout of the position should be located on the place between the upper-layer and middle-layer. If the pavement structure is of two layers, the heating wires should be located on the place between the two layers.

The smaller the heating wire spacing is, the faster the heating rate is and shorter time is needed for bridge deck melting ice and melting snow. Take the heating wires’ costs and the amount of work for the fixing heating wire into consideration comprehensively. The heating wires’ spacing is 10cm reasonably in actual construction. The influence of thermo-physical performance parameters, the temperature of the heating wires, environmental conditions of the asphalt on the temperature and rise of melting of the ice bridge pavement is analyzed by finite element software. The main conclusions are as follows: (a) with the increase of asphalt concrete density, heating rate reduces. (b) With the increase of asphalt concrete thermal conductivity, heating rate on the surface grows exponentially. (c) With the increase of specific heat capacity of asphalt concrete, heating rate on the surface decreases constantly. (d) With the increase of the heat transfer coefficient, heating rate on the surface of the bridge pavement increases rapidly. (e) With the increase of heat exchange coefficient on the surface, heating rate of the deck surface decreases.

6 ACKNOWLEDGMENT

THE transportation projects of Yunnan Provincial Department (Yunnan transportation projects 2013 (B) 03) are gratefully acknowledged by the authors. The authors are also grateful to the anonymous reviewers for their critical review and valuable suggestions for improving the manuscript.

7 DISCLOSURE STATEMENT

NO potential conflict of interest was reported by the authors.

8 REFERENCES


9 NOTES ON CONTRIBUTORS

Yuhua Peng is a professor at Chang'an University, Xi'an, China. Dr. Peng's research interests include traffic and transportation engineering, computer simulation and optimizational design. His research interests are in the deicing pavement. He has published some articles in core Journals.

Dingyue Chen is a teacher at Chang'an University, Xi'an, China. Dr. Chen research interests include vehicles and traffic engineering, optimizational control method and computer simulation. He has published some articles in international Journals.

Lihao Chen is a master's student of Northwest University, Xi'an, China.

Jiayu Yu is a master's student of Northwest University, Xi'an, China.

Mengjie Bao is a master's student of Chang'an University, Xi'an, China.
Delay-dependent Stability of Recurrent Neural Networks with Time-varying Delay

Guobao Zhang\textsuperscript{a,b}, Jing-Jing Xiong\textsuperscript{a,b}, Yongming Huang\textsuperscript{a,b}, Yong Lu\textsuperscript{a,b,c}, and Ling Wang\textsuperscript{a,b}

\textsuperscript{a}School of Automation, Southeast University, Nanjing 210096, China
\textsuperscript{b}Key Laboratory of Measurement and Control of CSE (School of Automation, Southeast University), Ministry of Education, Nanjing 210096, China
\textsuperscript{c}Earthquake Administration of Jiangsu Province, Nanjing 210014, China

ABSTRACT
This paper investigates the delay-dependent stability problem of recurrent neural networks with time-varying delay. A new and less conservative stability criterion is derived through constructing a new augmented Lyapunov-Krasovskii functional (LKF) and employing the linear matrix inequality method. A new augmented LKF that considers more information of the slope of neuron activation functions is developed for further reducing the conservatism of stability results. To deal with the derivative of the LKF, several commonly used techniques, including the integral inequality, reciprocally convex combination, and free-weighting matrix method, are applied. Moreover, it is found that the obtained stability criterion has a lower computational burden than some recent existing ones. Finally, two numerical examples are considered to demonstrate the effectiveness of the presented stability results.

KEY WORDS: Neural networks, Lyapunov-Krasovskii functional, Stability criterion, Time-varying delay, Lower computational burden

1 INTRODUCTION
DELAY-DEPENDENT stability criteria for neural networks with constant time or time-varying delays have received considerable attention in recent years (Zhang, et al. 2013; Li, et al. 2013; Kwon, et al. 2013; Chen & Zheng, 2013; Zhang, et al. 2014; Zhang, et al. 2014; Ge, et al. 2014; Wang, et al. 2015). On the one hand, the reason is that neural networks have received considerable attention due to their extensive applications, such as optimization (Nissinen, et al. 1999), automatic control (Benallegue & Meddah 2001), and others. On the other hand, delay-dependent stability results, which consider the information of time delays, are less conservative than the delay-independent stability results, especially when the time delays are small. Since the time delays are frequently encountered in electronic implementations of neural networks due to the finite switching speed of amplifiers and the inherent communication time between neurons, which may cause hidden oscillations, divergence, chaos, instability, or other poor performance behaviors (Zhang, et al. 2014; Ge, et al. 2014). Therefore, it is more effective to solve the delay-dependent stability problem of neural networks with time-varying delay from two perspectives of less conservatism and lower computational burden.

combination (Liu, 2013; Yang & Zhang, 2014; Farnam, et al. 2016), and their combinations (Zhang, et al. 2014). For the derivative of LKF, it is necessary to estimate the derivative for deriving stability criteria in terms of linear matrix inequalities (LMIs). It is usually difficult to deal with the integral terms of the derivative of LKF such that the derivative is enlarged. Although numerous techniques have been developed for estimating the time derivative of LKF, there still exists room for further study. It can be summarized as follows: 1) For the augmented LKF introduced in Kwon, et al. (2013), Zheng, et al. (2010), Kwon, et al. (2013), Rakkiyappan, et al. (2016), and Yang, et al. (2017), the delayed state derivative terms including $\dot{y}(t-\tau(t))$ and $\dot{y}(t-\tau_m)$ are contained in the derivative of LKF. However, if these terms are replaced by the delayed neural networks, it can be found that the time-varying delay is doubled or the delay interval $[0, \tau_m]$ imperceptibly becomes $[0, 2\tau_m]$. This means that the calculation results listed in Kwon, et al. (2013), Zheng, et al. (2010), Kwon, et al. (2013), Rakkiyappan, et al. (2016) and Yang, et al. (2017) are doubled. Simultaneously, derived activation function $g(y(t-2\tau(t)))$ and $g(y(t-\tau(t)) - \tau_m))$ are non-existent for the considered neural networks. 2) For the delay-decomposition LKF introduced in Ge, et al. (2014) and Zeng, et al. (2011), and the delay-partitioning LKF introduced in Wang, et al. (2015) and Lakshmanan, et al. (2013), the larger the number of subintervals is, there is less conservatism of stability results. However, the conservative reduction trends to be in-apparent as the increasing of the number of time delay subintervals, which lead to a large computational burden (Zhang, et al. 2014). 3) For Leibniz-Newton formula, many free weighting matrices are often needed to be introduced in the derived stability criteria and leads to significant increases in the computational burden. In Zhang, et al. (2014), the Leibniz-Newton formula was used to estimate the derivative of LKF, which caused a large number of matrix variables. Thus, how to further reduce the conservatism and computational burden of the stability results may be more attractive.

This paper further investigates the stability condition for continuous recurrent neural networks with time-varying delay by constructing a newly augmented LKF and employing the LMI method. This paper aims to derive a new and less delay-dependent stability criterion for recurrent neural networks with time-varying delay, while reducing the computational burden via less useful matrix variables. The contributions and improvements are summarized as follows: 1) Unlike the augmented LKF in Kwon, et al. (2013), Zheng, et al. (2010), Kwon, et al. (2013), Rakkiyappan, et al. (2016), and Yang, et al. (2017), a newly augmented LKF is constructed to improve the results, where the delayed state derivative terms including $\dot{y}(t-\tau(t))$ and $\dot{y}(t-\tau_m)$ do not appear in the derived stability criterion. 2) The delay-decomposition or delay-partitioning ideas introduced in Ge, et al. (2014), Zeng, et al. (2011), Wang, et al. (2015) and Lakshmanan, et al. (2013) are not used. Instead, by considering more information of the delayed state and neuron activation functions as augmented elements, a newly augmented LKF is developed. 3) The commonly used techniques, including integral inequality, reciprocally convex combination, and free-weighting matrix method, are applied in the estimation of the derivative of the constructed LKF. Those improvements lead to the obtained stability results having lower computational burden. Finally, two numerical examples are given to verify the effectiveness of the proposed stability criterion and its improvements over the recent existing ones.

The rest of this paper is organized as follows: Section 2 gives the problem formulation. Section 3 presents the new and less conservative delay-dependent stability criterion. In Section 4, two numerical examples are given to verify the effectiveness of the proposed stability criterion. Finally, the conclusion is made in Section 5.

Throughout this paper, the superscript $T$ means the transpose of a matrix, $R^n$ denotes the n-dimensional Euclidean space, $R^{m \times n}$ denotes the set of all $m \times n$ real matrices, $P > 0 \ (\geq 0)$ means that $P$ is a real symmetric and positive-definite matrix, $[\cdot]$ denotes the absolute value, $Y_{ij}$ represents the element in row $i$ and column $j$ of matrix $Y$, diag$[\cdots]$ denotes a block-diagonal matrix, symmetric term in a symmetric matrix is denoted by *.

2 PROBLEM FORMULATION

CONSIDER the following recurrent neural network with time-varying delay:

$$\dot{x}(t) = -C x(t) + A f(x(t)) + B f(x(t-\tau(t))) + J \cdot x(t),$$

where $x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T \in R^n$ is the state vector. $f(\cdot) = [f_1(\cdot), f_2(\cdot), \ldots, f_n(\cdot)]^T$ denotes the activation functions. $C = \text{diag}[c_1, c_2, \ldots, c_n]$ is a diagonal matrix with $c_i > 0$. $A$ and $B$ are the connection matrices. $J = [J_1, J_2, \ldots, J_p]^T$ is an external constant input vector. $\tau(t)$ is time-varying delay and satisfies

$$0 \leq \tau(t) \leq \tau_m, \quad \dot{\tau}(t) \leq \mu,$$

where $\tau_m$ and $\mu$ are constants.

Assume that the activation function satisfies
along the state trajectories of (3) yields,

\[ \quad \sigma_i^- \leq \frac{f_i(a) - f_i(b)}{a - b} \leq \sigma_i^+ , \quad \forall a, b \in \mathbb{R}, a \neq b, i = 1, 2, \ldots, n, \] (3)

where \( \sigma_i^- \) and \( \sigma_i^+ \) are known real constants, which can be positive, zero, and negative.

Suppose there exists an equilibrium point \( x^* \) for the neural network (1), one can shift the equilibrium point of (1) to the origin by changing variables \( y(t) = x(t) - x^* \),

\[ g(y(t)) = f(y(t) + x^*) - f(x^*), \] (4)

Then, (1) is rewritten as

\[ y(t) = -Cy(t) + Ag(y(t)) + Bg(y(t - \tau(t))) , \] (5)

where \( y(t) = [y_1(t), y_2(t), \ldots, y_n(t)]^T \),

\[ g(y(t)) = [g_1(y_1(t)), g_2(y_2(t)), \ldots, g_n(y_n(t))]^T. \]

In addition, it is easily obtained from (3) that

\[ \sigma_i^- \leq \frac{g_i(a) - g_i(b)}{a - b} \leq \sigma_i^+ , \quad \forall a, b \in \mathbb{R}, a \neq b, i = 1, 2, \ldots, n, \] (6)

Let \( a \neq 0 \) and \( b = 0 \), then

\[ \sigma_i^- \leq \frac{g_i(a)}{a} \leq \sigma_i^+ , \quad \forall a \neq 0, \quad i = 1, 2, \ldots, n, \] (7)

and

\[ \quad [g_i(a) - \sigma_i^- a][g_i(a) - \sigma_i^+ a] \leq 0, \quad i = 1, 2, \ldots, n \] (8)

This paper aims to derive a new and less conservative delay-dependent stability criterion guaranteeing that the delayed neural network (5) is globally asymptotically stable, while reducing the computational burden. To obtain the main results, the following lemmas are introduced.

**Lemma 1** (He, et al. 2016). For any constant positive-definite matrix \( Q \in \mathbb{R}^{m \times m} \) and \( \beta \leq s \leq \alpha \), the following inequalities hold:

\[ (\alpha - \beta) \int_b^a x^T(s)Qx(s)ds \geq \int_b^a x^T(s)ds^T \begin{bmatrix} \int_b^a Qx(s)ds \end{bmatrix} , \] (9)

\[ -\frac{(\alpha - \beta)^2}{2} \int_b^a x^T(s)Qx(s)ds \leq \begin{bmatrix} \int_b^a x^T(s)ds \end{bmatrix}^T \begin{bmatrix} \int_b^a Qx(s)ds \end{bmatrix}, \] (10)

**Lemma 2** (Wang, et al. 2015). For any vectors \( h_1, h_2 \) with appropriate dimensions, scalars \( a > 0, \quad b > 0 \) and \( a + b = 1 \), if there exist matrices \( M \in \mathbb{R}^{n \times n} \), \( \beta \leq s \leq \alpha \), and any matrix \( S \) with appropriate dimensions, and the following inequalities hold:

\[ \begin{bmatrix} M & S \\ * & M \end{bmatrix} \geq 0, \] (11)

\[ \begin{bmatrix} h_1^T Mh_1 - \frac{1}{b} h_2^T Mh_2 \leq \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}^T \begin{bmatrix} M & S \\ * & M \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} , \]

**3 STABILITY CRITERION**

This section discusses the stability of (5) and derives a delay-dependent stability criterion by employing newly augmented LKF and above mentioned lemmas.

**Theorem 1.** For given \( \tau_m, a > 0 \) and \( \beta \), and diagonal matrices \( L_1 = \text{diag}[\sigma_1, \sigma_2, \ldots, \sigma_n] \), \( L_2 = \text{diag}[^{\sigma_1^+}, ^{\sigma_2^+}, \ldots, \sigma_n^+] \), the neural network (5) with (6) and a time-varying delay satisfying condition (2) is globally asymptotically stable if there exist real matrices

\[ \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ * & M_{22} & M_{23} \\ * & * & M_{33} \end{bmatrix} > 0, \begin{bmatrix} P_{11} & P_{12} \\ * & P_{22} \end{bmatrix} > 0, \]

\[ G = \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ * & G_{22} & G_{23} \\ * & * & G_{33} \end{bmatrix} > 0, \quad K_1 > 0, \]

and positive diagonal matrices

\[ T_i (i = 1, 2, 3), \quad H_1 = \text{diag}[h_1, h_2, \ldots, h_n], \quad H_2 = \text{diag}[h_1, h_2, \ldots, h_n], \]

and any matrix \( N_i \in \mathbb{R}^{n \times n} (i = 1, 2, \ldots, 11) \), with appropriate dimensions, then the following LMIs hold:

\[ \Sigma < 0, \] (12)

\[ \begin{bmatrix} G & S \\ * & G \end{bmatrix} \geq 0, \] (13)

Calculating the time derivatives of \( V_i(t) \), \( i = 1, 2, \ldots, 5 \) along the state trajectories of (5) yields

\[ V_i(t) = 2 \begin{bmatrix} y(t) \\ \int_{t-\tau_1}^{t} y(s)ds + \int_{t-\tau_1}^{t-\tau_2} y(s)ds \end{bmatrix}^T \begin{bmatrix} \int_{t-\tau_1}^{t} g(y(s))ds + \int_{t-\tau_1}^{t-\tau_2} g(y(s))ds \end{bmatrix} \]
By applying Lemma 2, one can obtain

\[ \dot{V}_2(t) \leq \begin{bmatrix} y(t) \\ g(y(t)) \end{bmatrix}^T \begin{bmatrix} P_{11} & P_{12} \\ * & P_{22} \end{bmatrix} \begin{bmatrix} y(t) \\ g(y(t)) \end{bmatrix} - (1-\mu) \begin{bmatrix} y(t) \\ g(y(t)) \end{bmatrix}^T \begin{bmatrix} P_{11} & P_{12} \\ * & P_{22} \end{bmatrix} \begin{bmatrix} y(t) \\ g(y(t)) \end{bmatrix} \]

\[ \dot{V}_3(t) = 2[\dot{g}(y(t)) - L_1 y(t)]^T H_1 \dot{y}(t) \]

\[ + 2[L_2 y(t) - g(y(t))]^T H_1 y(t), \]

\[ \dot{V}_4(t) = \tau_m^2 \begin{bmatrix} y(t) \\ g(y(t)) \end{bmatrix}^T \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ * & G_{22} & G_{23} \\ * & * & G_{33} \end{bmatrix} \begin{bmatrix} y(t) \\ g(y(t)) \end{bmatrix} \]

\[ - \tau_m \int_{t-\tau_m}^t \begin{bmatrix} y(s) \\ g(y(s)) \end{bmatrix}^T \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ * & G_{22} & G_{23} \\ * & * & G_{33} \end{bmatrix} \begin{bmatrix} y(s) \\ g(y(s)) \end{bmatrix} ds \]

Since \( \begin{bmatrix} G & S \\ * & G \end{bmatrix} \geq 0 \), by applying Lemma 2, one can obtain

\[ \dot{V}_4(t) \leq \tau_m^2 \begin{bmatrix} y(t) \\ g(y(t)) \end{bmatrix}^T \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ * & G_{22} & G_{23} \\ * & * & G_{33} \end{bmatrix} \begin{bmatrix} y(t) \\ g(y(t)) \end{bmatrix} \]

\[ - \tau_m \int_{t-\tau_m}^t \begin{bmatrix} y(s) \\ g(y(s)) \end{bmatrix}^T \begin{bmatrix} G & S \\ * & G \end{bmatrix} \begin{bmatrix} y(s) \\ g(y(s)) \end{bmatrix} ds \]

By using Lemma 1, one can obtain

\[ \dot{V}_5(t) = (\tau_m^2 / 2)^2 \ddot{y}(t)^T K_1 \ddot{y}(t) \]

\[ - \left( \tau_m^2 / 2 \right) \int_{t-\tau_m}^t \ddot{y}(u) K_1 \ddot{y}(u) du ds \]

\[ \leq \left( \tau_m^2 / 2 \right) \ddot{y}(t)^T K_1 \ddot{y}(t) \]

\[ = \left( \tau_m^2 / 2 \right) \ddot{y}(t)^T K_1 \ddot{y}(t) \]
According to (6)-(8), it implies that, for any diagonal matrices \( T_i = \text{diag}\{t_{i1}, t_{i2}, \ldots, t_{im}\} > 0 \), \( i = 1, 2, \ldots, 5 \), the following inequality holds:

\[
0 \leq -2\big[ g(y(t)) - L_1 y(t) \big]^T T_1 \big[ g(y(t)) - L_2 y(t) \big] \\
-2\big[ g(y(t - \tau(t))) - L_1 y(t - \tau(t)) \big]^T T_2 \big[ g(y(t - \tau(t))) - L_2 y(t - \tau(t)) \big] \\
-2\big[ g(y(t - \tau_m)) - L_1 y(t - \tau_m) \big]^T T_3 \big[ g(y(t - \tau_m)) - L_2 y(t - \tau_m) \big] \\
-2\big[ g(y(t)) - g(y(t - \tau(t))) \big]^T T_4 \big[ g(y(t)) - g(y(t - \tau(t))) \big] \\
-2\big[ g(y(t)) - g(y(t - \tau(t))) \big]^T T_5 \big[ g(y(t)) - g(y(t - \tau(t))) \big] \\
-2\big[ g(y(t)) - g(y(t - \tau(t))) \big]^T T_6 \big[ g(y(t)) - g(y(t - \tau(t))) \big] \\
-2\big[ g(y(t)) - g(y(t - \tau(t))) \big]^T T_7 \big[ g(y(t)) - g(y(t - \tau(t))) \big].
\]

(21)

By combination of the concerned neural network (5) and free-weighting matrix method, the following equation is true for any matrices \( N_i \) (\( i = 1, 2, \ldots, 11 \))

\[
0 = 2\big[ y^T(t) N_1 + y^T(t - \tau(t)) N_2 + y^T(t - \tau_m) N_3 \\
+ g^T(y(t)) N_4 + g^T(y(t - \tau(t))) N_5 \\
+ g^T(y(t - \tau_m)) N_6 + \left( \int_{-\tau(t)}^{t} y(s)ds \right)^T N_7 \\
+ \left( \int_{-\tau_m}^{t} g(y(s))ds \right)^T N_8 + \left( \int_{-\tau(t)}^{t} g(y(s))ds \right)^T N_9 \\
+ \left( \int_{-\tau_m}^{t} g(y(s))ds \right)^T N_{10} + y^T(t) N_{11} \big]
\]

\[
\times [-c y(t) + A g(y(t)) + B g(y(t - \tau(t))) - \dot{y}(t)].
\]

(22)

Finally, by combining (15)-(22), one can derive that

\[
\dot{V}(t) \leq \dot{\xi}^T(t) \Sigma \dot{\xi}(t).
\]

(23)

If \( \Sigma < 0 \), then \( \dot{V}(t) < 0 \) for \( \forall \dot{\xi}(t) \neq 0 \), the concerned system (5) is globally asymptotically stable. This completes the proof.

Remark 1: Recently, the extended reciprocally convex combination approach to reduce the conservatism of the stability criteria for recurrent neural networks with time-varying delays is used in Wang, et al. (2015). Motivated by this idea, the approach is applied in this paper, which is shown in (19) and has potential to yield less conservative condition.

Remark 2: From the perspective of a computational burden, the stability criterion obtained in this paper has lower computational burden than the criteria obtained by the delay-decomposition LKF in Ge, et al. (2014) and Zeng, et al. (2011) and delay-partitioning LKF in Wang, et al. (2015) and Lakshmanan, et al. (2013), for the reason that the number of decision matrix variables is greatly reduced. Similarly, compared with the recent constructions of augmented LKF in Kwon, et al. (2013), Zheng, et al. (2010), Kwon, et al. (2013), Rakkiyappan, et al. (2016), and Yang, et al. (2017), the main difference is that the delayed state derivative terms are not considered in the constructed LKF in this paper, which greatly reduces the dimensions and computational burden of the stability criterion. Moreover, in Kwon, et al. (2013), Zheng, et al. (2010), Kwon, et al. (2013), and Rakkiyappan, et al. (2016), the free-weighting matrix method was not employed when they established the delay-dependent stability criteria. Therefore, the computational burden of the delay-dependent stability criterion obtained in this paper is lower.

Remark 3: In Theorem 1, the free-weighting matrix method used in (22) plays an important role in reducing the conservatism of stability criterion via incorporating the system model (5), in which many free weighting matrices are involved. Therefore, the obtained criterion will have better results than the existing ones.

Remark 4: The stability criteria derived in this paper give the matrix variables to be determined and the LMI-based constraint conditions for ensuring the stability of delayed neural networks; one can use the feasp function in MATLAB/LMI toolbox to solve those variables from the corresponding criterion.

4 NUMERICAL EXAMPLES

In this section, two numerical examples are considered to verify the effectiveness of the obtained stability criterion. The main objective is to derive an acceptable maximum upper bound (AMUB) on time delays such that delayed neural networks are globally asymptotically stable. Meanwhile, the larger the AMUB is, there is less conservatism of the corresponding stability criterion.

Example 1: Consider the following 4-neuron delayed neural network (5) with the parameters (Zhang, et al. 2014)

\[
C = \text{diag}\{1.2769, 0.6231, 0.9230, 0.4480\}
\]
First case: $\sigma_i^+ = 0$, $i = 1,2,...,4$. The comparison results on the AMUB of time-varying delay via the different methods presented in recent works (Zhang & Han, 2011; Ge, et al. 2014; Wang, et al. 2015; Kwon, et al. 2013; Kwon, et al. 2013; Zeng, et al. 2015; Liu, et al. 2015; Zhang, et al. 2014; Zhang, et al. 2016; Zhang, et al. 2017; Yang, et al. 2017) are listed in Table 1. From Table 1, it can be seen that Theorem 1 provides larger AMUBs of time-varying delay than the existing results in the literatures, especially when $\mu \geq 0.5$. When $\mu = 0.9$, the AMUBs are 0.3602, 0.3654 and 0.3767 in Zhang, et al. (2014). Applying Theorem 1, the AMUB is 1.6203, which is much better than those results in Zhang, He, Jiang, Wu, and Wu (2014).

When $\mu = 0.5$, the AMUB of time-varying delay is obtained by Theorem 1, $\tau_m = 2.7335$, and the initial values are randomly chosen as $[2.5,-5,2]^T$, the simulation result is shown in Figure 1. Obviously, the concerned neural network is globally asymptotically stable.

When $\mu = 0.9$, the AMUB of time-varying delay is obtained by Theorem 1, $\tau_m = 3.6220$, and the initial values are randomly chosen as $[0.2,0.5,-0.5,0.2]^T$, the simulation result is shown in Figure 2. Obviously, the concerned neural network is globally asymptotically stable.

Second case: $\sigma_1^- = -0.4$, $\sigma_2^- = 0.1$, $\sigma_3^- = 0$, $\sigma_4^- = -0.3$. The AMUBs of time-varying delay for various $\mu$ obtained by Theorem 1 and the methods presented in Zhang, et al. (2014) and Liu, et al. (2015) are listed in Table 2. From Table 2, it can be found that the proposed method in Theorem 1 significantly enhances the feasible region of stability criterion compared to those results in Zhang, et al. (2014) and Liu, et al. (2015), especially when $\mu \geq 0.5$. When $\mu = 0.9$, the AMUBs are 0.3602, 0.3654 and 0.3767 in Zhang, et al. (2014). Applying Theorem 1, the AMUB is 1.6203, which is much better than those results in Zhang, He, Jiang, Wu, and Wu (2014).

When $\mu = 0.9$, the AMUB of time-varying delay is obtained by Theorem 1, $\tau_m = 2.7335$, and the initial values are randomly chosen as $[2.5,-5,2]^T$, the simulation result is shown in Figure 1. Obviously, the concerned neural network is globally asymptotically stable.

When $\mu = 0.5$, the AMUB of time-varying delay is obtained by Theorem 1, $\tau_m = 3.6220$, and the initial values are randomly chosen as $[0.2,0.5,-0.5,0.2]^T$, the simulation result is shown in Figure 2. Obviously, the concerned neural network is globally asymptotically stable.
Inspired by the application given in Wang, Liu, Shan and Zhang (2015), the application is utilized to further verify the effectiveness of the stability criteria obtained in this paper. When the time delay is 
\[ \tau(t) = 1.7637 + 0.3536 \sin(0.4260t) + 0.7725 \cos(0.5\sqrt{t}) \], 
\[ \tau_m = 2.8853 \], \[ \mu = 0.8324 \], and the initial values are \([-0.5, -0.2, 0.1, 0.3]^T\), the simulation result is shown in Figure 4. Obviously, the concerned neural network is globally asymptotically stable. Furthermore, compared with the result reported in Wang, Liu, Shan and Zhang (2015) \( (\tau_m = 2.2477, \mu = 0.8324) \), the acceptable upper bound of time-varying delay in this paper is better.

**Example 2.** Consider the following 2-neuron delayed neural network (5) with the parameters (Zhang, et al. 2014)
\[
\begin{align*}
C &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \\
A &= \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \\
B &= \begin{bmatrix} 0.88 & 1 \\ 1 & 1 \end{bmatrix}, \\
\sigma_1 &= 0.4, \quad \sigma_2 &= 0.8, \quad \sigma_i = 0, \quad i = 1, 2.
\end{align*}
\]
When \( \mu = 0.9 \), the AUMBs \( \tau_m \) of \( \tau(t) \) obtained by Theorem 1 and different methods presented in Zhang, et al. (2013), Zhang, et al. (2014), Ge, et al. (2014), He, et al. (2016), Wang, et al. (2017), Kwon, et al. (2013) and Rakkiyappan, et al. (2016) are listed in Table 3. Seen from Table 3, it is obviously found that the result calculated based on the criterion in this paper is less conservative than the existing ones.

When \( \mu = 0.9 \), \( \tau_m = 2.9158 \), and the initial values are randomly chosen as \([-0.1, 0.1]^T\), the simulation result is shown in Figure 5. Obviously, the concerned neural network is globally asymptotically stable. All above, the obtained stability criterion in this paper is much effective and less conservative than most of the existing results in the literatures.

<table>
<thead>
<tr>
<th>Methods</th>
<th>( \mu = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang, et al. 2013</td>
<td>1.6375</td>
</tr>
<tr>
<td>Ge, et al. 2014</td>
<td>1.9562</td>
</tr>
<tr>
<td>Zhang, et al. 2014</td>
<td>1.9603</td>
</tr>
<tr>
<td>He, et al. 2016</td>
<td>2.2201</td>
</tr>
<tr>
<td>Wang, et al. 2017</td>
<td>2.3582</td>
</tr>
<tr>
<td>Kwon, et al. 2013</td>
<td>2.8339</td>
</tr>
<tr>
<td>Rakkiyappan, et al. 2016</td>
<td>2.8881</td>
</tr>
<tr>
<td>Theorem 1</td>
<td>2.9158</td>
</tr>
</tbody>
</table>
CONCLUSION

This paper has studied the delay-dependent stability for continuous recurrent neural networks with time-varying delay and lower computational burden. A new and less conservative delay-dependent stability criterion has been established by constructing a newly augmented LKF and employing LMI method. It has been proven that the obtained stability criterion has a lower computational burden. The new LKF that considers more information of the slope of the activation functions has been further developed. Finally, two numerical examples have been considered to demonstrate the effectiveness of the proposed stability criterion.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation (nos. 61503081, 61473079), Jiangsu Province Natural Science Foundation (nos. BK20140649, BE2016805, BE2014728) and Science for Earthquake Resilience (no. XH171001).

REFERENCES


8 NOTES ON CONTRIBUTORS

Guobao Zhang received bachelor’s degree from Zhejiang University in 1986. Later received a Master’s degree and Doctor’s degree from Southeast University in 1991 and 1998, respectively, China. During 1999 and 2001, he has been a senior visiting scholar in Hong Kong Polytechnic University. Presently, he is a professor at School of Automation, Southeast University in China. His research interests include complex networks and intelligent control.


**Jing-Jing Xiong** received B.S. degree from the Institute of Electrical and Electronic Engineering from Changchun University of Technology, Changchun, China, in 2012, and M.S. degree from the College of Mechanical and Electrical Engineering from China Jiliang University, Hangzhou, China, in 2015. He is currently pursuing the Ph.D. degree from the School of Automation from Southeast University. His current research interests include tracking control of quadrotor UAV, stability analysis and synchronization of delayed neural networks, sliding mode control, and adaptive control.

**Yongming Huang** was born in April 1982. He is a Dr. of Automation and now a lecture of College of Automation in Southeast University. He mainly engaged in speech recognition, speech emotion recognition, affective computing and robot for the elderly. He has published 4 first-author SCI papers and 15 EI retrieval papers.

**Yong Lu** was born in January 1973. He received a Master’s degree from University of Science and Technology of China in 2004. In 2005, he was employed as a senior engineer in Seismological Bureau of Jiangsu Province. Since participating in the work, he has been engaged in the erection, maintenance, research and development of seismic instruments.

**Ling Wang** is currently working towards a Master’s degree at the Department of Automation, Southeast University, Nanjing, Jiangsu, China. Her research interests are pattern recognition and intelligent control.

**APPENDIX**

$S$ with appropriate dimensions, then the following LMIs hold:

\[
\Sigma < 0,
\begin{bmatrix}
G & S \\
* & G
\end{bmatrix} \geq 0,
\]

Where

\[
\Phi_{11} = 2M_{12} + P_{11} + \tau_m^2 G_{11} - G_{33} - \tau_m^2 K_1 - 2L_1^T (T_1 + T_4)L_2 - 2N_1 C, \\
\Phi_{12} = G_{33} - S_{33} + 2L_1^T T_4 L_2 - C^T N_2^T, \\
\Phi_{13} = -M_{12} + S_{33} - C^T N_2^T, \\
\Phi_{14} = M_{13} + P_{12} + \tau_m^2 G_{12} + (L_1^T + L_2^T)(T_1 + T_4) + N_1 A - C^T N_4^T, \\
\Phi_{15} = -(L_1^T + L_2^T)T_4 + N_1 B - C^T N_5^T, \\
\Phi_{16} = -M_{13} - C^T N_6^T, \\
\Phi_{17} = M_{22} - G_{13} - \tau_m K_1 - C^T N_7^T, \\
\Phi_{18} = M_{22} - S_{31} + \tau_m K_1 - C^T N_8^T, \\
\Phi_{19} = M_{23} - G_{23} - C^T N_9^T, \\
\Phi_{1,10} = M_{23} - S_{32} - C^T N_{10}, \\
\Phi_{1,11} = M_{11} - L_1^T H_1 + L_2^T H_2 + \tau_m^2 G_{13} - N_1 - C^T N_{11}, \\
\Phi_{22} = -(1 - \mu)P_{11} - 2G_{33} + 2S_{33} - 2L_1^T (T_2 + T_4 + T_5)L_2, \\
\Phi_{23} = S_{33} + G_{33} + 2L_1^T T_3 L_2, \\
\Phi_{24} = -(L_1^T + L_2^T)T_4 + N_2 A, \\
\Phi_{25} = -(1 - \mu)P_{12} + (L_1^T + L_2^T)(T_2 + T_4 + T_5) + N_2 B, \\
\Phi_{26} = -(L_1^T + L_2^T)T_5, \\
\Phi_{27} = G_{15}^T - S_{13}^T, \\
\Phi_{28} = S_{31}^T - G_{13}^T, \\
\Phi_{29} = G_{25}^T - S_{23}^T, \\
\Phi_{2,10} = S_{32} - G_{23}^T, \\
\Phi_{2,11} = -N_2, \\
\Phi_{33} = -G_{33} - 2L_1^T (T_5 + T_3)L_2, \\
\Phi_{34} = N_3 A, \\
\Phi_{35} = -(L_1^T + L_2^T)T_5 + N_3 B, \\
\Phi_{36} = (L_1^T + L_2^T)(T_3 + T_5), \\
\Phi_{37} = -M_{22} + S_{13}^T, \\
\Phi_{38} = -M_{22} + G_{13}^T, \\
\Phi_{39} = -M_{23} + S_{23}^T.
\[ \Phi_{3,10} = -M_{23} + G_{23}, \]
\[ \Phi_{3,11} = -N_3, \]
\[ \Phi_{44} = P_{22} + r_m^2 G_{22} - 2T_1 - 2T_4 + 2N_4 A, \]
\[ \Phi_{45} = 2T_4 + N_4 B + A^T N_7^T, \]
\[ \Phi_{46} = A^T N_7^T, \]
\[ \Phi_{47} = M_{23}^T + A^T N_7^T, \]
\[ \Phi_{48} = M_{23}^T + A^T N_8^T, \]
\[ \Phi_{49} = M_{33} + A^T N_9^T, \]
\[ \Phi_{4,10} = M_{33} + A^T N_{10}^T, \]
\[ \Phi_{4,11} = H_1 - H_2 + r_m^2 G_{23} - N_4 + A^T N_{11}^T, \]
\[ \Phi_{55} = -(1-\mu)P_{22} - 2T_2 - 2T_4 - 2T_5 + 2N_5 B, \]
\[ \Phi_{56} = 2T_3 + B^T N_6^T, \]
\[ \Phi_{57} = B^T N_6^T, \]
\[ \Phi_{58} = B^T N_8^T, \]
\[ \Phi_{59} = B^T N_9^T, \]
\[ \Phi_{5,10} = B^T N_{10}^T, \]
\[ \Phi_{5,11} = -N_5 + B^T N_{11}^T, \]
\[ \Phi_{66} = -2T_3 - 2T_5, \]
\[ \Phi_{67} = -M_{23}, \]
\[ \Phi_{68} = -M_{23}, \]
\[ \Phi_{69} = -M_{33}, \]
\[ \Phi_{6,10} = -M_{33}, \]
\[ \Phi_{6,11} = -N_6, \]
\[ \Phi_{77} = -G_{11} - K_1, \]
\[ \Phi_{78} = -S_{11} - K_1, \]
\[ \Phi_{79} = -G_{12}, \]
\[ \Phi_{7,10} = -S_{12}, \]
\[ \Phi_{7,11} = M_{12}^T - N_7, \]
\[ \Phi_{88} = -G_{11} - K_1, \]
\[ \Phi_{89} = -S_{21}, \]
\[ \Phi_{8,10} = -G_{12}, \]
\[ \Phi_{8,11} = M_{12}^T - N_8, \]
\[ \Phi_{99} = -G_{22}, \]
\[ \Phi_{9,10} = -S_{22}, \]
\[ \Phi_{9,11} = M_{13}^T - N_9, \]
\[ \Phi_{10,10} = -G_{22}, \]
\[ \Phi_{10,11} = M_{13}^T - N_{10}, \]
\[ \Phi_{11,11} = r_m^2 G_{33} + (r_m^2/2)^2 K_1 - 2N_{11}. \]

\[ V(t) = V_1(t) + V_2(t) + V_3(t) + V_4(t) + V_5(t), \quad (14) \]

Where

\[ V_1(t) = \begin{bmatrix} y(t) \\ \int_{t-r_m}^t y(s) \, ds \end{bmatrix}^T \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ * & M_{22} & M_{23} \\ * & * & M_{33} \end{bmatrix} \begin{bmatrix} y(t) \\ \int_{t-r_m}^t g(y(s)) \, ds \end{bmatrix}, \]

\[ V_2(t) = \int_{t-r_m}^t \begin{bmatrix} y(t) \\ \int_{t-r_m}^t g(y(s)) \, ds \end{bmatrix} \begin{bmatrix} y(t) \\ \int_{t-r_m}^t g(y(s)) \, ds \end{bmatrix} \, ds, \]

\[ V_3(t) = 2 \sum_{i=1}^n \left[ h_{b_i} (g_i(s) - \sigma_i s) \right. \]
\[ + h_{b_i} (\sigma_i s - g_i(s)) \right. \left. \right] ds, \]

\[ V_4(t) = \tau_m \int_{t-r_m}^t \int_{t-r_m}^t \begin{bmatrix} y(u) \\ \int_{t-r_m}^t g(y(u)) \, du \end{bmatrix} \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ * & G_{22} & G_{23} \\ * & * & G_{33} \end{bmatrix} \begin{bmatrix} y(u) \\ \int_{t-r_m}^t g(y(u)) \, du \end{bmatrix} \, du, \]

\[ V_5(t) = \left( \tau_m^2 / 2 \right) \int_{t-r_m}^t \int_{t-r_m}^t \begin{bmatrix} y^T(v) K_1 y(v) \end{bmatrix} \, dv \, du, \]

\[ \xi^T(t) = [y^T(t), y^T(t-\tau(t)), y^T(t-\tau_m), \] \[ \times g^T(y(t)), g^T(y(t-\tau(t))), \]
\[ \times g^T(y(t-\tau_m)), \int_{t-r_m}^t y^T(s) \, ds, \]
\[ \times \int_{t-r_m}^t y^T(s) \, ds \int_{t-r_m}^t g^T(y(s)) \, ds, \]
\[ \times \int_{t-r_m}^t g^T(y(s)) \, ds, \]
\[ \times y^T(t)]. \]

**Proof.** Consider the following candidate for the LKF:
A Computable General Equilibrium Model based Simulation on the Water Conservancy Investment

JUN WANG
Zhejiang University of Water Resources & Electric Power, Hangzhou, China

ABSTRACT

The water conservancy industry is one of the oldest fundamental industries in human history with high attention all the time, which is a vital factor to national well-being and people’s livelihood. The “Five Water Governance” is one of important strategies as a breakthrough to force transforming and upgrading for ecological water conservancy and sustainable development. A regional dynamic CGE model was constructed to simulate and analyze the short-term and long-term influence of the water conservancy investment to water conservancy industry itself, other national economy sectors and macro economy, so as to provide scientific proof for policy-making of water conservancy, and coordinated and sustainable development police-making of the whole society.

KEY WORDS: Computable general equilibrium, Five water governance, Water conservancy

1 INTRODUCTION

THE water conservancy industry is always one of the fundamental and oldest industries of economy society in all ages, which aims at “taking advantages and getting rid of disadvantages of water” and is closely linked to the economy, people’s livelihood, and even national security. In view of the special role, all governments have paid great attention to the water conservancy industry and the topic has appeared many times in central number one documents. Investment of water conservancy will definitely promote the economy (W. Yi-jia and F. Mei-lan 2006, L. CHEN 2009, L. Lei 2011) and support the development of society (W. Huang, X. Zhang, et al. 2011). Literature about benefits of investment of water conservancy are numerous, but they are usually based on water conservancy itself or correlation analyses to a part of the national economy, which means we can only see the trees but not the forest to the role of water conservancy in the whole economy.

In view of industry linkage, investment of water conservancy will promote production of upstream industries, such as steel, cement and machinery, and will facilitate the development of downstream industries such as agriculture, electricity and water transport. Water conservancy constructions may be used for flood control and disaster reduction, farmland water conservancy, reclamation, water resource optimization, and water ecological protection. That is to say, different construction purposes will lead to different social and economic benefits. In view of capital flow, increment of water conservancy investment will crowd out investment non-beneficial industries to some extent, and governmental, public and fundamental investment of water conservancy will crowd out private capital. Therefore, some industries will benefit from water conservancy investment directly or indirectly, and some will not. So, only to analyze the whole framework of the national economy may get comprehensive, objective and accurate results and provide references and basis for scientific decision making.

Some scholars have analyzed the investment influence of water conservancy with input-output table or social accounting matrix, SAM for short, in the framework of macro economy. (L. Xiuli and C. Xikang 2003) studied the shadow prices of industrial and productive water for water price reform with the IO table of the nine major river basins in China. (T. Wen-jin, X. Xiao-wei, et al. 2011, W. Tang, X. Xu, et al. 2012) evaluated the backward economic effects of different kinds of water conservancy investment and the pull effect of large-scale investment in water conservancy based on SAM multiplier analysis. It can only roughly analyze linear influence of water conservancy investment. It cannot describe
endogenous behaviors and macro performance of economy society, and cannot simulate and forecast mid-term and long term trends of economy on the condition of external shocks and policy changes.

It is a specialty and advantage of the Computable General Equilibrium model, CGE for short, to simulate external shocks and policy changes, which may help to make in-depth and comprehensive analysis of linkages among industries and different term effects of all parts of the national economy. For example, (H. Gan and M. Cheng 2012) simulated different investment portfolios of 4 trillion to survey optimized financial investment structures. CGE analysis can provide scientific proof for policy making, evaluation, and implementation based on SAM by simulating and predicting the economy system running with abstracted equations from the structure and running mode of the economy system, regional simulation model from behavior instantiation of economic agents, and external shock variables from quantification of economic environment changes. (Z. Jing, H. Xiao-li, et al. 2013) evaluated investment benefits of water supply with the CGE model. (M. D. Azdan and M. S. M.A. 2001), (Z. Yong, W. Jin-feng, et al. 2008), (Z. Na, J. Yang-wen, et al. 2014) and some other literatures simulated and analyzed policies of the water resource, such as water price and water right. (C. Wen, X. Hao, et al. 2012). (T. OKUDA and C. NI 2010) and some other literatures paid attention on evaluation of water environment policies.

Zhejiang province is famous for its abundant resources, which is a land of milk and honey, with eight major river systems and over 30 lakes with at least 1 million cubic meters volume. The government always attaches great importance to water conservancy construction. The Five Water Governance was proposed as a development strategy at the 4th session of 13th provincial Party committee, which is regarded as a breakthrough to force and promote the transformation and upgrading. The amount of investment grew from 27,287 billion at 2011 to 55 billion at 2015, standing first in China. In the whole 12th five-year period, the total investment of water conservancy was more than 200 billion, with 93% increment than 11th five-year period and average 19% year growth rate. In the 13th five-year period, eight key projects with more than 300 billion are listed in the development planning. Long term and sustained investment of water conservancy must bring many influences to the economy and society. Therefore, it has great theoretic and practical significance for CGE modeling and analyzing on investment of water conservancy, taking Zhejiang province as an example.

2 REGIONAL SAM CONSTRUCTION OF WATER CONSERVANCY

2.1 Regional Macro SAM Construction

SAM can reflect complex relationships of the whole economy system much more completely, comprehensively and consistently. It is mainly based on IO table and kinds of statistical year books, and is the fundamental of the macro economy analysis with SAM multiplier or CGE. Its structure is familiar to the IO table with column accounts and row accounts as the framework of the data organization to describe information of income, expenditure, transfer payment, transaction cost and so on for those corresponding accounts such as household, government, enterprise etc. The account system of SAM is the premise of IO table, which is main data source meanwhile, as shown in Table 1.

The cross cell of the row account and the column account means expenditure of the row account to the column account. According to the above SAM framework, we get the following macro SAM of Zhejiang province shown in Table 2.

2.2 Regional Water Conservancy Micro SAM Construction

(1) Water Conservancy Account

Accounts of SAM can be customized as required with great flexibility, but usually based on the IO accounting system. There are usually 122 departments, 42 departments IO tables based on related rules of national economic accounting classification. Accounts of SAM are usually accounts or accounts combination of IO table. As the main data source, the grain size of IO table decides the data consistence of SAM to IO table.

In view of availability and authority of the national statistical year book, the account scale of the national statistical bureau is often regarded as the accounting standard, though there are some rules and studies of item definitions of the water conservancy industry. But there is no account of water conservancy according to national statistical scale, which means all data of related items of water conservancy should be extracted from the current statistical account and merged to a new account as the water conservancy. According to Classification of National Economic Industries (GB/T4754-2011), the water conservancy industry is usually regarded as a collection of “Water Resource Manufacture of Special Purpose Machinery”, “Production and Supply of Water”, “Construction of Water Conservancy and Inland Port Engineering”, “Service in Irrigation” and “Service in Fishery” of “Service in Agriculture”, in the national statistical yearbook. Unfortunately, limited to issue IO tables, some data are unavailable, as Table 3 shows.
### Table 1. General Framework of Regional SAM

<table>
<thead>
<tr>
<th>Activities</th>
<th>Factors</th>
<th>Households</th>
<th>Enterprises</th>
<th>Government</th>
<th>Rest Of The World</th>
<th>Saving-Investment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodities</td>
<td>Labor</td>
<td>intermediate inputs</td>
<td>private consumption</td>
<td>government consumption</td>
<td>call out</td>
<td>exports</td>
<td>investment</td>
</tr>
<tr>
<td>Labor</td>
<td>Capital</td>
<td>labor payment</td>
<td>capital returns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factors</td>
<td>Households</td>
<td>labor income</td>
<td>labor income</td>
<td>surplus</td>
<td>transfers</td>
<td>foreign transfers</td>
<td>Household Income</td>
</tr>
<tr>
<td>Enterprises</td>
<td>Capital</td>
<td>capital income</td>
<td>capital income</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government</td>
<td>Other Regions</td>
<td>production taxes, tariffs</td>
<td>direct taxes</td>
<td>direct taxes</td>
<td>transfers</td>
<td></td>
<td>Government Income</td>
</tr>
<tr>
<td>ROW</td>
<td>Abroad</td>
<td>call in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saving-Investment</td>
<td>Other Regions</td>
<td>imports</td>
<td>capital income</td>
<td>payment abroad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>Total Input</td>
<td>Factor Expenditures</td>
<td>Factor Expenditures</td>
<td>Household Expenditures</td>
<td>Enterprise Expenditures</td>
<td>Government Expenditures</td>
</tr>
</tbody>
</table>

### Table 2. Macro SAM of Zhejiang Province in 2012 (100 million RMB)

<table>
<thead>
<tr>
<th>Act</th>
<th>Fac</th>
<th>Hou</th>
<th>Ent</th>
<th>Gov</th>
<th>Other Regions</th>
<th>Abroad</th>
<th>S-I</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Com</td>
<td>Lab</td>
<td>80171.5</td>
<td>12072.0</td>
<td>3743.8</td>
<td>19450.0</td>
<td>14716.7</td>
<td>14688.3</td>
<td>144842.3</td>
</tr>
<tr>
<td>Fac</td>
<td>Cap</td>
<td>14775.8</td>
<td>15782.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hou</td>
<td>14775.8</td>
<td>0.2</td>
<td>-202282.5</td>
<td>311.0</td>
<td>22583.1</td>
<td></td>
<td></td>
<td>38657.7</td>
</tr>
<tr>
<td>Ent</td>
<td>15781.9</td>
<td>178.9</td>
<td>537.0</td>
<td>0</td>
<td>-11.0</td>
<td></td>
<td></td>
<td>16627.8</td>
</tr>
<tr>
<td>Gov</td>
<td>5502.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6207.4</td>
</tr>
<tr>
<td>Other Regions</td>
<td>22599.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19450.0</td>
</tr>
<tr>
<td>Abroad</td>
<td>6010.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6010.9</td>
</tr>
<tr>
<td>S-I</td>
<td>26406.8</td>
<td>213835.3</td>
<td>2147.7</td>
<td></td>
<td>-235369.0</td>
<td></td>
<td></td>
<td>11538.8</td>
</tr>
<tr>
<td>Total</td>
<td>144842.3</td>
<td>14775.8</td>
<td>15782.1</td>
<td>38657.7</td>
<td>16627.8</td>
<td>6207.4</td>
<td>19450.0</td>
<td>6010.9</td>
</tr>
</tbody>
</table>
Table 3. Availability from IO Table of Water Conservancy Sub-industries

<table>
<thead>
<tr>
<th>Sub-industries of Water Conservancy</th>
<th>Statistical Account (GB/T4754-2011)</th>
<th>Corresponding Secondary Directory</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service in Agriculture and Water Conservancy (SAWC for short)</td>
<td>Service in Irrigation (Code: 0512)</td>
<td>Service in Support of Agriculture (Code: 05)</td>
<td>No</td>
</tr>
<tr>
<td>Production and Supply of Water (PSW for short)</td>
<td>Production and Supply of Water (Code: 46)</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Construction of Water Conservancy (CWC for short)</td>
<td>Construction of Water Conservancy and Inland Port Engineering (Code: 482)</td>
<td>Civil Engineering (Code: 48)</td>
<td>No</td>
</tr>
<tr>
<td>Management of Water Conservancy (MWC for short)</td>
<td>Management of Water Conservancy (Code: 76)</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

For data that is unavailable in the IO table, proportional approximation for evaluation based on other available statistical reports is often adopted. The production information of sub-industries of the Service in Support of Agriculture is still unavailable in all kinds of regular statistical reports. The second national census provides a good data source for proportional approximation. The proportional factor of sub-industry can be defined as the output value proportion of it to its parent item, where the statistical information of the SAWC and the CWC can be found. For example, define the proportional factor of the SAWC as the output value proportion of the sum of Irrigation and Fishery to Agriculture. The statistical information of Construction of Water Conservancy can be found in the China Industry Statistical Yearbook and China Construction Statistical Yearbook as an independent item. These statistical reports are always nationwide, so only nationwide proportional factors can be defined by this method, which is only the average level in China. In view of massive rivers and water conservancy facilities in Zhejiang province, with great development in water conservancy, a compromised method is used.

\[ \alpha_v = \frac{v}{V} + \frac{31 \times i}{I} \quad (1) \]

In the above equation, “v” is the nationwide output value of one sub-industry, “V” is the nationwide output value of its parent industry, “i” is the investment of water conservancy in Zhejiang and “I” is the total investment of water conservancy in China. Therefore, the proportional factor, \( \alpha_v \), reflects both the industry proportion relationship and regional industry development.

(2) SAM Construction

Given \( \alpha_v = t \), a parent account A can be split into two children accounts, \( A^1, A^2 \), as Figure 1 shows. According to statistical reports such as the China Industry Statistical Yearbook, China Construction Statistical Yearbook, China Water Conservancy Statistical Yearbook and China Rural Statistical Yearbook, we get the SAM of water conservancy in Zhejiang with operations of account split and immerge, shown in Table 4.

3 REGIONAL CGE MODEL

THE LHR model proposed by (H. Lofgren, R. L. Harris, et al. 2002) is an international used and easy extended CGE model with many characters of a developed country. The key equations of price block, trade block, institution block and system constraint block are listed as follows:

3.1 Trade Block

This block defines the market optimization in the process of production and sales. Commodities from abroad, local region and other regions usually compete in local markets, and enterprises often optimize their sales of outputs among export, local sales and domestic sales.

For one commodity, given local outputs defined as QD in quality and PDD in price, other regions outputs defined as QI in quality and PI in price, imports defined as QM in quality and PM in price, they are incomplete substitutive of each other. Consumers have to optimize their purchase of different kinds of commodities to maximize their utility under the constraint of income, which can be described as the following equations.
\[ QQ_c = \alpha^Q \cdot \left( \delta^Q QM_c^{\rho^Q} + \delta^Q QI_c^{\rho^Q} + (1 - \delta^Q - \delta^{Q^2}) QD_c^{\rho^Q} \right)^{-1}, \tag{2} \]

\[ \frac{QM_c}{QD_c} = \left( \frac{PDD_c}{PM_c} \cdot \frac{\delta^Q}{1 - \delta^Q - \delta^{Q^2}} \right)^{1 + \rho^Q}, \tag{3} \]

\[ \frac{QI_c}{QD_c} = \left( \frac{PDD_c}{PI_c} \cdot \frac{\delta^{Q^2}}{1 - \delta^Q - \delta^{Q^2}} \right)^{1 + \rho^Q}. \tag{4} \]

\[ QX_c = \alpha^Q \left( \delta^{Q^2} QE_c^{\rho^Q} + \delta^{Q^2} QO_c^{\rho^Q} + (1 - \delta^Q - \delta^{Q^2}) QD_c^{\rho^Q} \right)^{-1}, \tag{5} \]

\[ \frac{QE_c}{QD_c} = \left( \frac{PE_c}{PDS_c} \cdot \frac{1 - \delta^Q - \delta^{Q^2}}{\delta^Q} \right)^{1 - \rho^Q}, \tag{6} \]

\[ \frac{QO_c}{QD_c} = \left( \frac{PO_c}{PDS_c} \cdot \frac{1 - \delta^Q - \delta^{Q^2}}{\delta^{Q^2}} \right)^{1 - \rho^Q}. \tag{7} \]

Figure 1. The Diagram of Account Split in IO Table

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>B</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

For the outputs of one commodity, it can be sold abroad, QE in quality and PE in price, locally, QDD in quality and PDS in price, and domestic, QO in quality and PO in price. To maximize the profit, enterprises will optimize their sales strategy, which can be described as the following equations.

### 3.2 Price Block

Given no trade barrier in the domestic market, commodities can flow freely, so the price of one commodity is the sum of original price and trade cost as equations (8 and 9) show. In the process of trade optimization, no new value is created, so identical equations (10 and 11) are established.

### 3.3 System Constraint Block

For the region as a whole, given RSAV representing regional saving or regional trade balance, the balance of region can be defined as equation 13 and equation 14.

### 3.4 Dynamic Block

Some influence of investment of water conservancy will not appear in a short time, which needs a long process of conduction and diffusion, when the dynamic CGE model is required. The variations of saving accumulation and labor growth are main powers of the dynamic of the economy system for the static CGE model. The dynamic recursive equations of capital and labor are shown in the following equations, given the depreciation rate and labor growth rate are \( r \) and \( \delta \), CAP and LAB are capital and labor, \( t \) is the simulation term, QFS is the quality of factors and QINV is the inventory.
$PI_c = pwi_e + \sum_{e'} PQ_{ee'} \cdot i_{ee'} \cdot i_{ee'}$ \quad(8)

$PO_c = pwo_e + \sum_{e'} PQ_{ee'} \cdot i_{ee'}$ \quad(9)

$PQ_{ee'} \cdot (1 - iq_{ee'}) \cdot QQ_{ee'} = PDD_{ee'} \cdot QD_{ee'} + PM_{ee'} \cdot QM_{ee'} + PI_{ee'} \cdot QI_{ee'} \quad(10)$

$PX_e \cdot QX_e = PDS_{ee'} \cdot QD_{ee'} + PE_{ee'} \cdot QE_{ee'} + PO_{ee'} \cdot QO_{ee'} \quad(11)

$\sum_{e'} (pwi_{ee'} \cdot QI_{ee'}) = \sum_{e'} (pwo_{ee'} \cdot QQ_{ee'}) + RSAV \quad(12)$

$QFS(CAP)_{t+1} = (1 - \delta) \cdot QFS(CAP)_t + QINV \quad(13)$

$QFS(LAB)_{t+1} = QFS(LAB) \cdot (1 + r) \quad(14)$

4 SIMULATION AND ANALYSIS OF INVESTMENT OF WATER CONSERVANCY

4.1 Simulation Scenarios

The investment of water conservancy in Zhejiang province has increased rapidly every year always at a two-digit rate in recent years. It was up to 55 billion in 2015, as the top one in China. It will keep a high rate increment in the 13th five year period, which will be 300 billion totally, that is 60 billion a year. Therefore, we simulated the CGE model with the investment information, with a perfect competition market assumption and small country assumption as well.

4.2 Simulation Results Analysis

According to the simulation results, the investment increment of water conservancy will promote directly the development of water conservancy, which is proven by the simulated outputs, showing a 45.17% increase. Meanwhile, the outputs will change with 0.83 percent increments for the secondary industry, 2.68 percent decline for the primary industry and 1.3 percent decline for the tertiary industry. It is well known that the investments of the water conservancy are usually in fundamental construction, therefore, the secondary industry, tightly related to fundamental construction, has obviously positive correlation with the investment of water conservancy. However, the primary industry and tertiary industry have slightly negative correlation with it.

As for the view of the whole economy, we get the following conclusion that the increment of water conservancy investment will promote the social production, based on the static simulation results of 0.73 percent increment of social outputs, 0.06 percent increment of the nominal GDP, 0.08 percent increment of the real GDP, 0.32 percent increment of total imports and 4.08 percent increment of total investment. The variation of GDP in the dynamic simulation with 5 terms is shown in Figure 2.

[Figure 2. Dynamic Simulation of GDP]

It can be concluded that the GDP of Zhejiang province is rising yearly for the influent of water conservancy investment, and the growth rate of real GDP is always higher than that of nominal GDP, which means a higher quantity of economy. So, the investment of water conservancy mainly in form of “Five Water Governance” should promote the regional economy development significantly in both short-term and long-term rising yearly. Dynamic simulations of detailed items of the economy are shown in Figure 3 and Figure 4.

The imports and exports of Zhejiang province are rising yearly under the influence of water conservancy investment, and the total outputs is still in a increment trend as imports and exports. However, social consumption is constrained due to the increment of investment, with a totally controllable and gentle trend.
Table 4. Micro SAM of Water Conservancy in Zhejiang province in 2012 (100 million RMB)

<table>
<thead>
<tr>
<th></th>
<th>Water Conservancy</th>
<th>Primary Industry</th>
<th>Secondary Industry</th>
<th>Tertiary Industry</th>
<th>Lab</th>
<th>Cap</th>
<th>Hou</th>
<th>Enterprises</th>
<th>Government</th>
<th>Other Regions</th>
<th>Abroad</th>
<th>S-I</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water Conservancy</td>
<td>40.91</td>
<td>2.79</td>
<td>114.45</td>
<td>42.34</td>
<td>79.83</td>
<td>39.07</td>
<td>7.21</td>
<td>17.24</td>
<td>1035.24</td>
<td>1379.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Industry</td>
<td>3.00</td>
<td>129.17</td>
<td>1623.43</td>
<td>383.25</td>
<td>1164.53</td>
<td>5.81</td>
<td>165.05</td>
<td>18.07</td>
<td>280.78</td>
<td>3773.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary Industry</td>
<td>737.08</td>
<td>613.37</td>
<td>54860.66</td>
<td>4533.54</td>
<td>5862.67</td>
<td>16052.44</td>
<td>12618.50</td>
<td>11363.78</td>
<td>106642.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tertiary Industry</td>
<td>269.51</td>
<td>230.07</td>
<td>8452.69</td>
<td>8135.23</td>
<td>4964.94</td>
<td>3698.96</td>
<td>2062.91</td>
<td>2008.51</td>
<td>33048.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>152.38</td>
<td>1594.48</td>
<td>6288.08</td>
<td>6740.82</td>
<td>14775.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38657.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cap</td>
<td>103.67</td>
<td>132.35</td>
<td>7877.05</td>
<td>7669.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15782.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14775.76</td>
<td>0.22</td>
<td>202262.48</td>
<td>311.05</td>
<td>225833.15</td>
<td>38657.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterprises</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15781.87</td>
<td>4.91</td>
<td></td>
<td>841.03</td>
<td>16627.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government</td>
<td>44.82</td>
<td>-70.20</td>
<td>3523.37</td>
<td>2004.58</td>
<td>178.93</td>
<td>536.97</td>
<td></td>
<td>-11.02</td>
<td>6207.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Regions</td>
<td>18.17</td>
<td>997.87</td>
<td>18885.07</td>
<td>2698.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3149.52</td>
<td>19450.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abroad</td>
<td>9.54</td>
<td>143.20</td>
<td>5017.23</td>
<td>840.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6010.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26406.80</td>
<td>218353.32</td>
<td>2147.65</td>
<td>-235368.99</td>
<td>11538.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1379.06</td>
<td>3773.09</td>
<td>106642.03</td>
<td>33048.15</td>
<td>14775.76</td>
<td>15781.87</td>
<td>16627.81</td>
<td>19450.04</td>
<td>6010.88</td>
<td>11538.79</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Since there is an independent water conservancy account, corresponding sub-industries of water conservancy are excluded from primary industry, secondary industry and tertiary industry in this paper.
5 CONCLUSION

THE "Five Water Governance" is one of the main contents and essential breakthroughs in order to improve eco-environmental quantity overall, forcing economic transforming and upgrading by enhancing water ecological treatment and ecological water conservancy construction, constructing beautiful Zhejiang and creating a good life for residents in Zhejiang. It is proven in the static and dynamic simulations of CGE in this paper. As one main content of the “Five Water Governance”, water conservancy investment cannot only promote the development of the water conservancy industry significantly, but also be beneficial to short-term and long-term development of society and economy of Zhejiang province, with increments of total outputs, GDP, imports and exports. Though there will be some negative impacts in the agriculture industry, tertiary industry and consumption due to the investment increment of water conservancy, but totally it is controllable and encouragement policies about them are suggested to keep coordinated development of all industries and steady economy development.

6 ACKNOWLEDGEMENT

THE work was supported by Humanity and Social Science Foundation of P. R. China for Young Scholars of Education Ministry, China (Grant No. 13YJC630164, 14YJC630109).

7 REFERENCES

Hans Lofgren, Rebecca Lee Harris, et al. (2002) "A Standard Computable General Equilibrium (CGE) Model in GAMS."


Jun WANG received Ph.D. degree in Management Science and Engineering at School of Information Management & Engineering, Shanghai University of Finance & Economics. He works at Zhejiang University of Water Resources & Electric Power in China as an associate professor. His research interests mainly include industry management science, policy simulation and analysis.
Optimal Learning Slip Ratio Control for Tractor-semitrailer Braking in a Turn based on Fuzzy Logic

Jinsong Donga, Hongwei Zhanga, Ronghui Zhangb,*, Xiaohong Jinc and Fang Chenb

d Key Laboratory of Operation Safety Technology on Transport Vehicles, Research Institute of Highway Ministry of Transport, Beijing 100088, China.  
b Xinjiang Laboratory of Perception & Control Technology for IOT, Xinjiang Nor-West Star Information Technology Co., Ltd, Urumqi 830011, China.  
c Department of Transportation, School of Mines, China University of Mining and Technology, Xuzhou 221000, China.  
d Xinjiang Communications Construction Group Co., Ltd, Urumqi 830016, China.

ABSTRACT
The research on braking performance a of tractor-semitrailer is a hard and difficult point in the field of vehicle reliability and safety technology. In this paper, the tire braking model and the dynamic characteristic model of the brake torque with the variable of the controlling air pressure were established. We also established a nonlinear kinematic model of the tractor-semitrailer when it brakes on a curve. The parameters and variables of the model were measured and determined by the road experiment test. The optimal control strategy for the tractor-semitrailer based on the optimal slipping ratio was proposed. Then the PID controller and the fuzzy controller were designed respectively. Simulation results show that the reasonable control strategy can significantly improve the braking directional stability when a tractor-semitrailer runs on a curving road. The research results provide technical references for improving the lateral stability when a tractor-semitrailer brakes on a curve, and it also provides a technical reference for the road traffic safety.

KEY WORDS: Optimal learning control, Fuzzy logic algorithm, Computing simulation, Intelligent system, Vehicle dynamics, Tractor-Semitrailer

1 INTRODUCTION
WITH increasing the investment scale for highway construction and the development of freight vehicle technology in China, the road freight transport has turned into a stage of rapid development and taken an important position in the integrated transport service system. Road freight transport has become a major part in economic development. As one of the main carrying tools in the road transportation system, the semi-trailer train plays an important role. At the same time, because of its vehicle performance and safety performance, it guarantees the safety and smoothness for the personnel, goods, vehicles and other traffic participants in the cargo transport process. Research results have indicated that the braking system modeling and control strategy of the semi-trailer train are important factors that affect the safety of the braking system. Bad performance or failure of the braking system could directly lead to traffic accidents.

Hubert L. and Dietz from the German Stuttgart Automobile Research Institute conducted a study on the lateral stability of the combination vehicle in 1937 (Ahn, et al., 2017; Guan, et al., 2002; Thanpattranon, et al., 2016). In the 1960s, many scholars began to study the braking performance of freight cars and articulated semi-trailers (Chieh et al., 2000; Chadwick et al., 2014; Goel et al., 1983; Khalaji et al., 2016; Zhang et al., 2017). In the 1990s, the advanced control technology was widely used for improving the braking performance or the vehicle operating stability. In recent years, studies on braking stability of the combination vehicle have made great progress, and the research results focus on three aspects; folding, rollover and ABS. For literature (Gaspar et al., 2005; Li et al., 2017; Li et al., 2007; Wang et al., 2005; You
et al., 2015; Zhang et al., 2016), an accurate 15 degrees of freedom simulation model was established by methods of Lagrange, which is combined with the Extended Kalman Filter to realize the measurement of the longitudinal slip ratio, lateral acceleration of the traction vehicle and yaw rate velocity. For literature (Kress et al., 2009; Huang et al., 2010; Huang et al., 2017; You et al., 2014; Zhang et al., 2017), it establishes the 14 degrees of freedom model to evaluate the roll angle and the semi-trailer train lateral acceleration. Based on that, it develops the linear active control system to prevent the rollover occurrence. For literature (Abdur et al., 2016; Bao et al., 2016; Ozdalyan et al., 2008), it establishes the mathematical models of bike and trailer automobile train by methods of Lagrange equation and Newton's mechanics. For literature (Demir et al., 2014; Zhang et al., 2016), it establishes the semi-trailer train model with 8 degrees of freedom, which carries on the simulation analysis by means of Matlab/Simulink and validates the accuracy of the established model by steering and braking tests. For literature (Ding et al., 2014; Shuwen et al., 2014), it establishes the braking simulation model of the semi-trailer train with Matlab/Simulink software and analyses the effect that braking deceleration has on the load transferring during the brake process, as well as the effect that load transferring has on the wheel lock sequence, and analyzes the effect of ABS arrangement form on the braking performance.

Literature has shown that the research works in this field focuses on mechanics and kinematics with the braking performance of the combination vehicle, but there is less research works on the semi-trailer train braking performance on a curve, and it is not enough due to limitations by the technical conditions. This paper establishes a semi-trailer train tire model and nonlinear dynamics model when braking on a curve. We measure and determine the model parameters and variables through road tests, and put forward the strategy method based on the braking control on a curve with optimum slip rate. The research results can give a useful reference for the brakes design and the semi-trailer train selective assembly.

The structure of this paper is as follows: Section 2 describes braking in a turn modeling of the semi-trailer train. Section 3 checks the model parameter, Section 4 designs the braking control laws based on optimal slip ratio, and we give a conclusion in Section 5. Finally, the last section concludes acknowledgments.

2 CORNERING BRAKING MODELING

2.1 Tire Model

In order to accurately describe the mechanical characteristics of a tire with power and movement, and to carry out unified analysis, and establish the tire mechanics model, the Society of Automotive Engineers (SAE) has worked out a standard tire movement coordinate system. This coordinate system is defined as a three-dimensional right-handed coordinate system under the normal direction coordinate.

If a tire is free from the lateral force, when the driver steps on the brake pedal, the brake torque, through the transmission of power, will act on the wheel and make the wheel decelerate. Meanwhile a brake force will be generated in the contact area between the tire and ground. When the tire tread is getting into the ground-contact area and the fore-end distorts under the drawing force, the sidewall will also correspondingly have shear deformation.

Commonly, we used slip the ratio, which indicates the length of the sliding distance, or the degree of braking during the increase of the brake torque. The slip ratio is defined as the ratio of the difference between the wheel translator velocity and the wheel tangential velocity with respect to the wheel translator velocity. Shown as Eq(1).

\[ s = \frac{u_a - \omega R}{u_a} \times 100\% \]  (1)

In the formula,
- \( s \)—longitudinal slip ratio of wheel;
- \( u_a \)—wheel translatory velocity;
- \( \omega \)—wheel angular velocity;
- \( R \)—rolling radius of wheel.

2.2 Brake Torque Modeling

The air brake is a widely used braking mode in road freight transport combination vehicle. In order to control the service brake of a semi-trailer train, it needs the power conversion device to convert the input air pressure in the brake chamber to mechanical action that realizes the braking function. The brake chamber in the braking system of semi-trailer train and the brake drum assembly are coordinated to achieve the above functions.

The braking force generated by each wheel in the pneumatic braking system can be expressed as (Chou, Y. C., & Nakajima, M., 2016; Jia et al., 2015; Rajamani et al., 2011) and shown as Eq (2).

\[ M_p = P_r A_o \eta_B F \left( \frac{I}{2 l_p} \right) R. \]  (2)

Where:
- \( A_o \)—active area of brake chamber;
- \( P_r \)—pressure of brake chamber;
- \( \eta_B \)—braking efficiency of brake chamber, brake shoe and brake mechanism;
- \( BF \)—braking ratio of brake drum (Braking ratio is the friction force generated by the friction surface of the brake drum and the braking force of the brake shoe);
- \( R \)—radius of brake drum;
2.3 Cornering Braking Model

A semi-trailer train is a complex multi-body dynamic system. This paper mainly studies the cornering braking stability of the semi-trailer train. Based on the analysis of the lateral force, longitudinal force and its motion, the mathematical model of the semi-trailer train cornering braking was established. In order to accurately describe the force and the semi-trailer train movement in the process of cornering braking, a detailed description of the semi-trailer train movement condition was required. So, it needs to introduce more freedom degrees, which greatly increases the complexity and the analysis difficulty.

In order to facilitate the research work, based on the analysis of the key factors, this paper simplifies the following aspects when establishing the cornering braking model of semi-trailer train:

1) The semi-trailer train is regarded as two rigid bodies, which realizes the restraint of the mutual movement via the coupling of the traction pin and the traction seat.

2) The movement with two degrees of suspension roll and pitching in the process of semi-trailer train cornering braking is ignored, and only the longitudinal with lateral load transfer caused by cornering braking was taken into account.

3) It was assumed that the tire characteristics were consistent with the Brush model, regardless of the influence from the aligning torque generated by the angle of roll and the side slip angle.

4) It considers the rolling resistance and air resistance of the vehicle, but ignores the resistance moment caused by the rotational energy.

In order to accurately describe the motion state of the vehicle, a coordinate system must be set up. The setting of a coordinate system directly affects the complexity of the vehicle motion description, so it should be set according to the vehicle motion characteristics. Based on refs. (Deng et al., 2016; Dixit et al., 2016; Tseng et al., 2007), this paper establishes a fixed coordinate system on the ground as well as the tractor and semi-trailer coordinate systems. Among them, the traction vehicle and semi-trailer coordinate system take each centroid as the coordinate’s origin. The vehicles forward direction is the X axis and the Y axis points to the driver right side. The Z axis is vertical down according to the right-hand rule.

In the semi-trailer train, the front-axle of the towing tractor has a single tire on each side; the rear axle of the towing tractor and semi-trailer are dual-axis with two tires on each side. In order to facilitate the mechanical analysis and control strategy research, it simplifies the towing tractor and semi-trailer with dual-axis, and two tires on each side to the type of single axle and single tire. The parameters of the vehicle can refer to the standard definition of reference (Cao, X., & Zhu, D., 2015; Gao et al., 2016; Yin et al.2017; Zhang et al., 2017).

It analyzes the movement condition and the force situation of the semi-trailer train when braking on a curve, and gets the motion equation of the towing tractor and the semi-trailer.

The motion equation of towing tractor:

\[
\begin{align*}
    m_1(\ddot{u}_1 - v_1r_1) & = \sum F_{x1} \\
    m_1(\dot{v}_1 + u_1r_1) & = \sum F_{y1} \\
    I_{r1}\ddot{r}_1 & = \sum M_{z1}
\end{align*}
\]

The motion equation of semi-trailer:

\[
\begin{align*}
    m_1(\ddot{u}_2 - v_2r_2) & = \sum F_{x2} \\
    m_1(\dot{v}_2 + u_2r_2) & = \sum F_{y2} \\
    I_{r2}\ddot{r}_2 & = \sum M_{z2}
\end{align*}
\]

3 MODEL PARAMETER CHECK

THE model parameter is the link between the theoretical model and the real vehicle and it is also one of the necessary conditions to realize the mutual authentication between the vehicle test and the theoretical model. Therefore, the model parameter confirmation is an extremely important link in simulation. The authors’ measure variables related to the model like the tire parameter, rolling resistance, air resistance coefficient and brake torque parameter through road tests to analyze and determine the performance index of the tractor-semi trailer cornering braking. And validate the accuracy of model through the prototype comparison test under the same condition (Abdurrahim et al., 2016; Huynh et al., 2017; Wang et al., 2015; Xu et al., 2016; Zhang et al., 2017).

The straight braking test of the semi-trailer train is conducted in the test field for road traffic of the Ministry of Communications. The combination vehicle performance tester VTRS1 developed for the Ministry of Communications’ Western Transportation Construction Project; “Researches on Technical Conditions and Test Methods for Safe Operation of Freight Vehicles” is applied. The test equipment mainly consists of the wheel speed module, GPS module, pedal force module and the embedded host as well as the necessary auxiliary equipment installation. Figure 1 is the testing experiment platform.

Comparisons between test results and simulations results are shown in Figure 2. The comparative analysis between the test data and the simulation data indicates that the maximum error of the braking distance is 3.31%. Besides, the maximum error of the wheel angular velocity before it first reaches the speed zero is 1.14 rad/s and the maximum error of the
vehicle deceleration is 0.48 m/s\(^2\). Therefore, there is no significant difference between the test data and the simulation data.

Figure 1. Testing Experiment Platform.

Figure 2. Comparison between Test and Simulation Braking Distances.

4 BRAKING CONTROL BASED ON OPTIMAL SLIP RATIO

4.1 PID Control based on Slip Ratio

In order to improve the comprehensive performance of the semi-trailer train, the slip ratio needs to be effectively controlled as is the key parameter. Under the condition that the wheel rolling radius of the wheel is fixed, the wheel sliding rate is the function of vehicle running speed and wheel angular velocity, and the slip rate can also be regarded as the function of the brake torque, the braking force in the contact area between tire and ground as well as the vehicle speed. And then it can control the brake torque through the output pressure of the brake chamber. This paper puts forward the research work of PID control and fuzzy control based on the optimal slip ratio.

As the object of this paper, the simulation model of tractor-semitrailer is complex and nonlinear. One advantage of the PID control is being simple and more stable, which makes up the discontinuity of traditional logic threshold braking control. However, the conventional PID control cannot guarantee the stability and reliability of the braking control under the non-linear characteristics of vehicles. The fuzzy control is based on the fuzzy set theory, fuzzy linguistic variables and fuzzy logic reasoning. It is independent of the controlled object mathematical model. It has a strict theory foundation and easy engineering practice. As we know, it is difficult to describe and control the model, even theoretically good simulation, because the tire is a strong non-linear body, and the braking condition is complicated and changeable. As a result, the implementation process may not be satisfactory. The fuzzy control is suitable for this variable condition, especially, the nonlinear characteristics of tractor-semitrailer, so the fuzzy controller for the braking system is a useful attempt. Therefore, the optimization of the braking slip rate based on fuzzy control is studied, and the results show that the fuzzy control has better control effect.

4.2 Fuzzy Control based on Slip Ratio

Although there is the conventional PID control remedy the non-consistency of the traditional logic threshold braking control and the certain stability, but under the situation of the vehicle nonlinear characteristics and road condition changes, the conventional PID control cannot guarantee the stability and reliability of the brake control. Fuzzy control does not depend on the mathematical model of the controlled object, but with a rigorous theoretical basis, it is easy to put into engineering practice. It is an important method of intelligent control engineering research. For the road surface with low adhesion coefficient, the thesis carries out the research on the cornering braking control of semi-trailer train based on fuzzy theory. Figure 3 shows the comparison of the left front wheel slip ratio for towing tractor controlled by PID and the logical threshold. Figure 4 is the basic principles of fuzzy control. According to the design process of the fuzzy controller, the final fuzzy output obtained through fuzzy process. Fuzzy logic reasoning and accurate calculation is shown in Figure 5. In this text, the vehicle test speed is about 50 km/hour.

Figure 3. Comparison of the Left Front Wheel Slip Ratio for Towing Tractor Controlled by the PID and the Logical Threshold.
Figure 4. Basic Principles of Fuzzy Control.

Figure 5. Operation Curve of a Fuzzy Controller.

From the simulation results of the braking distance, when wheel speed and the speed of the towing tractor in the cornering braking of semi-trailer train under PID control and fuzzy control with low adhesion coefficient. Figure 6 and 7 are the simulation results of the braking distance, wheel speed and the speed of the towing tractor in the cornering braking of a semi-trailer train under PID control and fuzzy control with low adhesion coefficient (=0.2). So, it can be seen that the braking distance of the PID control is less than the fuzzy control, reducing to 154.4m from 162.9m. The main reason is that under the control of PID, the folding occurs when the semi-trailer train brakes, so that the drag of semi-trailer train increases. On the road with low adhesion coefficient, PID parameters adopt the same coefficient with high adhesion coefficient road. Due to changes in road parameters, the control fails and the wheels of semi-trailer train lock when braking. Influenced by the lateral force, braking jack-knifing occurs, and the articulation angle of the semi-trailer train is up to around 64deg when the braking terminates. Under the fuzzy control, the change of semi-trailer train’s articulation angle is relatively smooth and the locking of the wheels does not occur, indicating that fuzzy control has good control effect on the low adhesion coefficient road and can adapt to different road adhesion coefficients. Compared with the PID controller, it has better adaptability to pavement.

Through the system test and the simulation analysis research, in Table 1, comparison of brake control methods for semi-trailer train in cornering braking is obtained. According to the table, compared with the logic threshold control and PID control, the fuzzy control has improved significantly in consistency, nonlinearity and has a certain stability as well as strong adaptability to different attachment coefficients. At the same time, Table 1 presents a detailed description of the advantages and disadvantages of the logic threshold value, which PID control and fuzzy control set. PID control and fuzzy control are relatively good, but there are also some problems under the limited condition. A follow-up study can combine the two methods to form the fuzzy PID control according to the different operating conditions, to make full use of their advantages and ultimately improves the control effect on cornering braking of semi-trailer train.
5 CONCLUSIONS

As one of the main carrying tools for the road freight transportation system, a semi-trailer train plays an important role. Compared with the semi-trailer, the coupling action between the towing tractor and the trailer makes the semi-trailer train have a more complex dynamic system. Improper structural parameters, operation parameters and control parameters will lead to folding, spin, sideslip and other non-stability phenomenon of a semi-trailer train in the braking process. During the cornering braking of semi-trailer train, the non-stability phenomenon may cause serious traffic accidents easily.

Based on the theory of vehicle kinematics and kinetics, this paper establishes the tire braking model, dynamic characteristics model of brake torque that changes with control pressure, nonlinear motion model of the cornering braking for a semi-trailer train and measures, as well as identifies model parameters and variables through road tests to validate the model consistency and the test parameters. Finally, it proposes an optimal control strategy based on the optimal slip ratio for a semi-trailer on the curving road, and it designs the PID controller and fuzzy controller respectively. Besides, this paper makes a detailed analysis of the braking performance of the controllers under the conditions with different adhesion coefficients through comparison. The research results provide a reference for improving the direction stability of a semi-trailer train’s cornering breaking, which also provides a reference for road transportation safety (Zhang et al., 2017; Sun et al., 2018).

6 ACKNOWLEDGMENTS

This paper is partially supported by the Soft Science Research Project of Ministry of Communications (Grant no. 2015312223320), and Major Science and Technology Project funded by the Ministry of Communications (2011318223006). The first author would like to thanks all the editors and anonymous reviewers for the valuable comments in improving the quality and presentation of the paper.

7 DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

8 REFERENCES


<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Logic Threshold Control</th>
<th>Optimal Slip Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistency</td>
<td>Bad</td>
<td>Good, Mediocre</td>
</tr>
<tr>
<td>Nonlinearity</td>
<td>Bad</td>
<td>Mediocre, Good</td>
</tr>
<tr>
<td>Stability</td>
<td>Bad</td>
<td>Mediocre, Good</td>
</tr>
<tr>
<td>Portability</td>
<td>Bad</td>
<td>Mediocre, Good</td>
</tr>
<tr>
<td>Reliability</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Cost price</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>


9  NOTES ON CONTRIBUTORS

**Jinsong Dong** was born in Hebei, China. He received B.Sc. degree, M.Sc. degree and Ph.D. degree in Vehicle Operation Engineering at College of Transportation, Jilin University, Changchun, China, in 2003, 2006, 2010. Currently he works as senior engineer at the Research Institute of Highway Ministry of Transport. His current research interests include vehicle dynamic control simulation and freight vehicles performance evaluation.

**Hongwei Zhang** was born in Shandong, China. He received B.Sc. degree and M.Sc. degree in Vehicle Operation Engineering at the College of Transportation, Jilin University, Changchun, China, in 1986, 1989. He is an excellent senior engineer in the Research Institute of Highway Ministry of Transport. His current research interests include vehicle dynamic control simulation and freight vehicles performance evaluation.

**Ronghui Zhang** was born in Jiangxi, China. He received B.Sc. degree in Automation Engineering at College of Telecommunication Information Engineering, Hebei University, Baoding, China, in 2003. He received M.Sc. degree in Vehicle Operation Engineering at the College of Transportation, Jilin University, Changchun, China, 2006. He received Ph.D. degree in Mechanical & Electrical Engineering from Changchun Institute of Optics, Fine Mechanics and Physics, the Chinese Academy of Sciences, Changchun, China, in 2009. Currently he is a research fellow with the Guangdong Key Laboratory of Intelligent Transportation System, School of Engineering, Sun Yat-sen University, Guangzhou, Guangdong, China. He has been a visiting research fellow in Xinjiang Laboratory of Perception & Control Technology for IOT, Xinjiang Nor-West Star Information Technology Co., Ltd, China. His current research interests include computer vision, intelligent control and ITS.

**Xiaohong Jin** was born in Jiangsu, China. She received B.Sc. degree, M.Sc. degree and Ph.D. degree in Transportation Engineering at the College of Transportation, Jilin University, Changchun, China, in 2003, 2006, and 2009. Currently she works as a lecturer at Department of Transportation Specialty, China University of Mining and Technology. Her current research interests include computer vision, intelligent control and ITS.

**Fang Chen** was born in Xinjiang, China. She received B.Sc. degree and M.Sc. degree in Transportation Engineering, Xinjiang Agriculture University, Urumqi, China, in 2012, 2015. Currently she works as an engineer at Xinjiang Communication Construction Group CO., LTD. Her current research interests include intelligent control and ITS.
Kinematic Calibration of a Parallel Manipulator for a Semi-physical Simulation System

Dayong Yu
School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, China

ABSTRACT
In the application of a semi-physical simulation system of a space docking mechanism, the simulation precision is determined by pose accuracy of the parallel manipulator. In order to improve pose accuracy, an effective kinematic calibration method is presented to enable the full set of kinematic parameter errors to be estimated by measuring the docking mechanism’s poses. A new calibration model that takes into account geometrical parameter errors and coordinates transformation errors is derived by using a differential geometry method. Based on the calibration model, an iterative least square algorithm is utilized to calculate the above errors. Simulation and experimental results show the calibration method can obviously improve pose accuracy.

KEYWORDS: Kinematic calibration, Error modeling, Error compensation, Parameter identification, Parallel manipulator

1 INTRODUCTION
Over the last five decades, parallel manipulators have been extensively researched for their some advantages of high rigidity, high accuracy, and high load-carrying capacity, and their general applications in motion simulators, parallel robots, and parallel machine tools etc. (Yao, et.al. 2016) (Yang, et.al. 2014) (Yang, et.al. 2012) (Sun, et.al. 2016) (Bernhard, et.al. 2001). Parallel manipulators can be used to simulate the relative motion between target spacecraft and tracking spacecraft in a semi-physical simulation system of a space docking mechanism by installing a space docking mechanism to the mobile platform. A semi-physical simulation system of a space docking mechanism has been constructed in the Aerospace System Engineering Shanghai in the 921 Manned Space project with Harbin Institute of Technology.

One of the important issues in applying the semi-physical simulation system of a space docking mechanism is pose accuracy of the parallel manipulator. The investigation of pose accuracy for parallel manipulators is mainly divided into two parts. One is accuracy analysis (Wang & Ehmann, 2002) (Ryu & Cha, 2003), and the other is kinematic calibration (Verner, et.al. 2003) (Li, et.al. 2017) (Bai & Teo, 2003). Accuracy analysis is defined as statistical analysis of the effects of manufacturing errors and installing errors on pose errors of parallel manipulators. Sensitivity analysis is used to modify the relevant error parameters to achieve the desired pose accuracy. In the aspect of accuracy analysis, an error model of a Stewart platform was built using D-H matrix theory, considering the manufacturing errors, installing errors and measuring errors. The error analysis results showed that the maximum errors for Stewart platforms occurred in the edge of the work space (Wang & Masory, 1993). A pose error formula of the moving platform was established by differentiating the inverse equation of Stewart platforms. The influence of the length errors of the drive rod and the position error of the hinge on the pose of the moving platform was analyzed (Ropponen & Arai, 1995). For the Stewart platform, an error model for pose accuracy synthesis analysis was established, and an analytic solution for the accuracy analysis of individual components has been given in this article. A pose error model of parallel manipulators was set up, and the pose error spatial distributions in a given work space were obtained by using a Sobol sequence based quasi-Monte Carlo method. This approach can significantly enhance evaluative precision of pose errors and reduce computational burden. According to the above information, the comparison of various design schemes in the stage can be carried on and the best scheme can be found through the accuracy analysis and accuracy synthesis (Li, et.al. 2015). It can also
calculate the influence of an original error on the pose accuracy. Thus, the key segments in parallel manipulators can be discovered, and the key points and directions to improve pose accuracy can be clearly defined. It can provide accurate and reliable information in order to improve design quality and design level of parallel manipulators.

Compared with accuracy analysis, kinematic calibration is the effective technique to solve the pose accuracy problem in the case that manufacturing and installing accuracy cannot be guaranteed. Pose accuracy can be improved without changing the hardware mechanism. A large number of investigations show that kinematic calibration is the most economical and feasible technologies to improve pose accuracy. The basic principle of kinematic calibration is that an error function between measured poses and calculated poses by using the kinematic model is constructed based on constraint and pose error observability. The kinematic parameters are identified by linear or nonlinear least square technique, and the inverse model parameters in the controller are corrected by the identification results, so as to achieve the goal of accuracy compensation. In general, kinematic calibration includes error modeling, pose measurement, error parameter identification and error compensation (Daney, 2003) (Zhuang & Roth, 1995). According to the different pose measurement technologies, calibration methods can usually be divided into external calibration (Takeda, et.al. 2004) (Renaud, et.al. 2006) (Tian, et.al. 2016) and self-calibration (Patel & Ehmann, 2000) (Ren, et.al. 2009). When the pose of parallel manipulators is completely measured, the inverse kinematic model can be used to solve the calibration problem. Since the pose information is partially measured, the calibration problem can only be solved by using the forward kinematic model. The calculation of external calibration based on complete pose measurement is more accurate, and the kinematic parameter identification performance is better because the inverse kinematics of parallel manipulators has analytic solutions. The forward kinematics of the parallel manipulators can usually obtain numerical solutions, so the calibration performance based on partial pose is relatively weak (Rauf, et.al. 2006). In order to ensure the calibration results, most of the external calibration based on the complete pose is used to identify all independent parameters of the model. Kinematic calibration of Stewart platforms was simulated and analyzed by using calibration technology of serial manipulators. The results show that the errors are reduced by an order of magnitude after kinematic calibration, but did not consider the influence of measurement error on calibration results (Masory, et.al. 1993). A simple calibration method was presented to identify a fixed leg length by changing other leg lengths. Kinematic parameters of each leg can be estimated separately, thereby the amount of calculation is reduced greatly. But this method has decreased the measurable work space (Zhuang & Yan, 1998). A calibration of parallel robots was proposed by using a tilt sensor. The effect of the tilt sensor noise and pose measurement number on kinematic calibration was simulated, but this method can only identify some parameters (Besnard & Khalil, 1999).

A kinematic calibration technique considering geometrical parameter errors of parallel manipulators, transformation errors between the docking coordinate frame and the mobile coordinate frame, and transformation errors between the fixed coordinate frame and the world coordinate frame is presented in order to solve simulation precision for a semi-physical simulation system of a spacecraft docking mechanism. In Section 1, the investigation on accuracy analysis and kinematic calibration is outlined, and its importance is pointed out. The article is organized as follows: The inverse kinematics and pose error model are given in Section 2. A kinematic calibration model containing the above errors is established by using the matrix differential theory, and an error parameter identification algorithm is developed based on iterative least squares in Section 3. In Section 4, the feasibility for the kinematic calibration technique is verified by computer simulation, and some experimental results are presented to further confirm the effectiveness of this technology. Section 5 concludes this paper.

2 KINEMATIC MODEL OF THE PARALLEL MANIPULATOR

The parallel manipulator with a docking mechanism installed to its mobile platform is composed of six identical electro-hydraulic actuators connecting a mobile platform to a base platform with Hooke joints in parallel shown in Figure 1. The parallel manipulator can be simplified as a model shown in Figure 2(a). Two coordinate frames are established for the parallel manipulator. The fixed coordinate frame \{B\} is located in the center of the base platform. The mobile coordinate frame \{A\} is located in the center of the mobile platform. \(l_i\) \((i=1,…,6)\) is the \(i\)th actuator length, \(a_i\) \((i=1,…,6)\) represents upper joint coordinates in the mobile coordinate frame \{A\}, \(b_i\) \((i=1,…,6)\) is the lower joint coordinates in the fixed coordinate frame \{B\}. The docking coordinate frame \{D\} is attached to the docking mechanism, and the world coordinate frame \{W\} is a unified reference frame as shown in Figure 2(b).

The homogeneous transformation matrix is used to express the pose of the coordinate frame \{D\} with respect to the world coordinate frame \{W\}, namely

\[
T_{d}^w = T_{d}^a T_{a}^f T_{f}^w
\]  

(1)
2.1 Inverse Kinematics

The inverse kinematics can deal with the actuator lengths corresponding to a set of given poses of parallel manipulators and the kinematic parameters. The actuator lengths are calculated in the form of closed form solutions through analytic method.

The vector loop in Figure 2(a) can be expressed as

\[ l_i = R \cdot a_i + t - b_i \quad i = 1, 2, \ldots, 6 \]  

(2)

The actuator length can be determined by calculating the norm of Eq. (2).

\[ |l_i| = \|R \cdot a_i + t - b_i\| \quad i = 1, 2, \ldots, 6 \]  

(3)

![Figure 1. The Parallel Manipulator Constructed for the Space Docking Motion Simulation.](image)

Figure 2. Schematic Diagram of a Semi-physical Simulation System.

2.2 Error Model

A pose error model of the parallel manipulator will relate the geometrical errors resulting from the manufacturing and assembly of the mobile platform and the base platform to the pose error of its mobile platform. To derive the pose error model, Eq. (2) is performed. The following differentiation is given below

\[ \delta L_i \cdot u_i + L_i \delta u_i = \delta R \cdot a_i + R \delta a_i + \delta t - \delta b_i \]  

(4)

Where \( L_i \) is the actuator length and \( u_i \) is the unit vector of the actuator. Because the rotation matrix \( R \) is orthogonal, it can be expressed as

\[ \delta R = \Omega R \]  

(5)

Where \( \Omega \) is a \( 3 \times 3 \) skew symmetric matrix whose nonzero element represents the orientation error \( \delta \omega \) of the mobile coordinate frame \( \{ A \} \).

Substituting Eq. (5) into Eq. (4) produces

\[ \delta L_i \cdot u_i + L_i \delta u_i = \delta \omega \times R a_i + R \delta a_i + \delta t - \delta b_i \]  

(6)

![Figure 2(b). The Relationships among all Transformation Matrices](image)

By multiplying \( u_i^T \) on both sides of Eq. (6), it can be represented as

\[ \delta L_i = \left[ u_i^T \left( R a_i \times u_i \right) \right] \left[ \delta t \delta \omega \right]^T + \left[ u_i^T R - u_i^T \right] \left[ \delta a_i \delta b_i \right]^T \]  

(7)

Therefore, the pose error model of the parallel manipulator can be expressed as

\[ E = J_j J_i e_g \]  

(8)

Where \( E \) is the pose error vector, \( E = \left[ \delta t \delta \omega \right]^T \), \( e_g \) is geometric parameter errors of the parallel manipulator, \( e_g = \left[ \delta L_1 \delta a_1 \delta b_1 \cdots \delta L_6 \delta a_6 \delta b_6 \right]^T \).

\[ J_j = \left[ \begin{array}{c} u_1^T \left( R a_1 \times u_1 \right) \\ \vdots \\ u_6^T \left( R a_6 \times u_6 \right) \end{array} \right] \]  

(9)
\[
J_s = \begin{bmatrix}
1 & -u_i^T R & u_i^T & \cdots & 0 \\
\vdots & \vdots & \vdots \\
0 & \cdots & 1 & -u_i^T R & u_i^T \\
\end{bmatrix}
\] (10)

Eq. (8) shows the actuator length errors, the upper joint position errors and the lower joint position errors play an important role in pose accuracy decreasing.

3 KINEMATIC CALIBRATION OF THE PARALLEL MANIPULATOR

As described in section 2, \( T_B^a \) in Eq. (1) is imprecise, because of manufacturing and assembly errors of the parallel manipulator. For transformation matrices \( T_B^m \) and \( T_D^d \), the actual values are not provided due to the immeasurability of origins of the mobile coordinate frame \( \{A\} \) and the fixed coordinate frame \( \{B\} \). In order to achieve their precise values, a kinematic calibration technology is established considering not only geometric parameter errors of the parallel manipulator but also the transformation errors for \( T_B^m \) and \( T_D^d \). By using the differential transformation relationship, Eq. (1) can be rewritten as

\[
T_B^m + dT_B^m(T_B^a + dT_B^a)(T_D^d + dT_D^d)
\] (11)

Removing the parentheses on the right side of Eq. (11) and neglecting the high order terms, it can be represented as

\[
dT = T \delta T
\] (13)

where \( \delta T \) is written as

\[
\delta T = \begin{bmatrix}
0 & -\delta z & \delta y & dx \\
\delta z & 0 & -\delta x & dy \\
-\delta y & \delta x & 0 & dz \\
0 & 0 & 0 & 0
\end{bmatrix}
\] (14)

where \( dx, dy, \) and \( dz \) are position errors, \( \delta x, \delta y, \) and \( \delta z \) are the orientation errors.

Substituting Eq. (13) into Eq. (12), the formula can be obtained as

\[
\delta T_B^m = U^{-1} \delta T_B^m U + \delta T_B^m \delta T_D^d + \delta T_D^d
\] (16)

It is noted that \( U \) and \( \delta T_B^d \) can be in the form of

\[
U = \begin{bmatrix}
n_u & o_u & a_u & r_u \\
0 & 0 & 0 & 1
\end{bmatrix}
\] (17)

\[
\delta T_B^m = \begin{bmatrix}
0 & -\delta z & \delta y & dx \\
\delta z & 0 & -\delta x & dy \\
-\delta y & \delta x & 0 & dz \\
0 & 0 & 0 & 0
\end{bmatrix}
\] (19)

\[
\delta T_D^d = \begin{bmatrix}
0 & \delta z_a & \delta y_a & dx_a \\
\delta z_a & 0 & -\delta x_a & dy_a \\
-\delta y_a & \delta x_a & 0 & dz_a \\
0 & 0 & 0 & 0
\end{bmatrix}
\] (20)

\[
\delta T_D^d = \begin{bmatrix}
0 & -\delta z_d & \delta y_d & dx_d \\
\delta z_d & 0 & -\delta x_d & dy_d \\
-\delta y_d & \delta x_d & 0 & dz_d \\
0 & 0 & 0 & 0
\end{bmatrix}
\] (21)

Letting \( \delta_b = [dx_a, dy_a, dz_a]^T \) and \( \delta_a = [dx_d, dy_d, dz_d]^T \), the triple product of matrix \( U^{-1} \delta T_B^m U \) is expressed as

\[
\delta T_B^m \delta T_D^d = T_B^a \left( T_B^m T_D^d \right)^{-1} \delta T_B^m \delta T_D^d + \left( T_D^d \right)^{-1} \delta T_B^m \delta T_D^d + \delta T_D^d
\] (15)

By letting \( U = T_B^m T_D^d \) and \( V = T_D^d \), Eq. (15) can be simplified as
Similarly, letting $\mathbf{d}_i = [dx_i \ dy_i \ dz_i]^T$ and $
abla = [\delta x, \delta y, \delta z]^T$, the triple product of matrix $V^{-1}\delta \mathbf{T}_a^B V$ is written as

$$V^{-1}\delta \mathbf{T}_a^B V = \begin{bmatrix} 0 & -\delta_x \cdot \mathbf{a}_v & 0 \\ \delta_x \cdot \mathbf{a}_v & 0 & 0 \\ -\delta_x \cdot \mathbf{a}_v & 0 & 0 \end{bmatrix}$$

(23)

By using Eq. (16), (19), (22), (23) and (24), the following equations are established

$$dx = \delta_x \cdot (r_x \times n_v) + d_x \cdot n_v + \delta_x \cdot (r_x \times n_v) + dx_d$$

$$dy = \delta_y \cdot (r_y \times o_v) + d_y \cdot o_v + \delta_y \cdot (r_y \times o_v) + dy_d$$

$$dz = \delta_z \cdot (r_z \times a_v) + d_z \cdot a_v + \delta_z \cdot (r_z \times a_v) + dz_d$$

$$\delta x = \delta_x \cdot n_v + \delta_x \cdot n_v + \delta z_d$$

$$\delta y = \delta_y \cdot o_v + \delta_y \cdot o_v + \delta y_d$$

$$\delta z = \delta_z \cdot a_v + \delta_z \cdot a_v + \delta z_d$$

(25)

Letting $\mathbf{R}_n = [n_v \ o_v \ a_v]^T$, $\mathbf{R}_r = [n_r \ o_r \ a_r]^T$,

$$\mathbf{C}_x = [r_x \times n_v \ r_x \times o_v \ r_x \times a_v]^T$$

$$\mathbf{C}_y = [r_y \times o_v \ r_y \times n_v \ r_y \times a_v]^T$$

$$\delta = [\delta x \ \delta y \ \delta z]^T$$

$$\mathbf{d}_j = [dx_j \ dy_j \ dz_j]^T$$

$$\delta_j = [\delta x_j \ \delta y_j \ \delta z_j]^T$$

Eq. (25) is written in matrix form as

$$\begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix} = \begin{bmatrix} R_n & C_v & R_r & C_r & I_{3 \times 3} & O_{3 \times 3} \\ O_{3 \times 3} & R_r & O_{3 \times 3} & R_r & O_{3 \times 3} & I_{3 \times 3} \end{bmatrix} \begin{bmatrix} \delta_x \\ \delta_y \\ \delta_z \end{bmatrix}$$

(26)

Where $I$ is the unit matrix and $O$ is the zero matrix.

Substituting Eq. (8) into Eq. (26), it can be rewritten as

$$\begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix} = \begin{bmatrix} R_n & C_v & R_r & C_r & I_{3 \times 3} & O_{3 \times 3} \\ O_{3 \times 3} & R_r & O_{3 \times 3} & R_r & O_{3 \times 3} & I_{3 \times 3} \end{bmatrix} \begin{bmatrix} \delta_x \\ \delta_y \\ \delta_z \end{bmatrix}$$

(27)

The calibration model given in Eq. (27) is further simplified as

$$\mathbf{x} = \mathbf{J} \mathbf{y}$$

(28)

where $\mathbf{x} = [d \delta]^T$ is a pose error vector and $\mathbf{y} = [d_b \delta_b \ e_b \ d_s \delta_s]^T$ is a parameter error vector. The calibration model can be used to identify the geometrical parameter errors of the parallel manipulator and the transformation errors of $T_w^B$ and $T_p^B$.

There are 54 parameter errors to be estimated as shown in Eq. (28), while the equations are 6. Therefore, at least 9 sets of pose measurement are required to identify 54 parameter errors. Given $n$ sets of pose error vector, $\delta \times n$ calibration equations can be obtained and written in matrix form

$$\bar{\mathbf{x}} = \bar{\mathbf{J}} \mathbf{y}$$

(29)

where $\bar{\mathbf{x}} = [x_1 \ x_2 \ ... \ x_n]^T$ is pose error matrix, $\bar{\mathbf{J}} = [\mathbf{J}_1 \ \mathbf{J}_2 \ \cdot \cdot \cdot \ \mathbf{J}_n]^T$ is parameter error identification matrix. The least squares solution for $\mathbf{y}$ is

$$\mathbf{y} = (\bar{\mathbf{J}}^T \bar{\mathbf{J}})^{-1} \bar{\mathbf{J}}^T \bar{\mathbf{x}}$$

(30)

In order to ensure that the least square solution is closer to its truth value, successive iteration algorithm can be used to identify the parameter error. The parameter errors are substituted as the correction of geometrical parameters of the parallel manipulator and
transformation matrix of the coordinate. The above calculation procedure is repeated until the parameter errors are satisfied the following formula

\[-\varepsilon \leq y^{(i)} - y^{(i-1)} \leq \varepsilon\]  \hspace{1cm} (31)

where \(\varepsilon\) is error tolerance.

4 COMPUTER SIMULATIONS AND EXPERIMENTAL RESULTS

The feasibility for kinematic calibration method presented in the paper was verified by computer simulations. The following hypothesis was made:

1. The transformation matrix, \(T^D_B\), between the docking coordinate frame \(\{D\}\) and the mobile coordinate frame \(\{A\}\) is considered as an identity matrix before kinematic calibration. Similarly, the transformation matrix, \(T^W_B\), between the fixed coordinate frame \(\{B\}\) and the world coordinate frame \(\{W\}\) is also an identity matrix before kinematic calibration.

2. A set of predefined transformation errors was listed in Table 1. These errors were appended the three transformation matrices, namely, \(T^D_B\), \(T^W_A\), and \(T^W_B\). The computation results were used as pose actual value. The transformation matrices generated by computer simulation were taken as pose nominal value and a set of actuator length was given in Table 2.

The computer simulation results of the kinematic calibration are shown in Table 3 and Figure 3. Since the kinematic calibration is an iterative computation process, the relationship between the pose error of the parallel manipulator and the number of iterations in the computer simulation and shown in Figure 3. In order to evaluate the effect of the kinematic calibration, the following two error parameters are defined: RMSPE and RMSOE, which are the root mean square position and orientation errors. From the computer simulation results above, it can be seen that the pose error of the parallel manipulator decreases gradually while the number of iteration increases, and the expected effect of kinematic calibration is obtained. The set of geometrical parameter errors and transformation errors can be identified through least squares algorithm.

Figure 3. The Relationship between Pose Errors and Number of Iteration in a Simulation.

<table>
<thead>
<tr>
<th>Table 1. Predefined Transformation Errors (m or rad)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error matrix</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>(T^D_B)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(T^W_A)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(T^W_B)</td>
</tr>
</tbody>
</table>
Table 2. Parallel Manipulator Actuator Lengths of Simulation (m or rad).

<table>
<thead>
<tr>
<th>No.</th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4$</th>
<th>$l_5$</th>
<th>$l_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8102</td>
<td>-0.0995</td>
<td>-0.1615</td>
<td>0.1087</td>
<td>0.3626</td>
<td>0.0029</td>
</tr>
<tr>
<td>2</td>
<td>-0.4840</td>
<td>0.2078</td>
<td>0.7086</td>
<td>-0.2386</td>
<td>0.0263</td>
<td>0.2194</td>
</tr>
<tr>
<td>3</td>
<td>0.1923</td>
<td>0.5255</td>
<td>-0.7958</td>
<td>-0.3154</td>
<td>-0.3114</td>
<td>-0.0745</td>
</tr>
<tr>
<td>4</td>
<td>-0.0252</td>
<td>0.7593</td>
<td>-0.2648</td>
<td>-0.5076</td>
<td>0.1803</td>
<td>-0.2046</td>
</tr>
<tr>
<td>5</td>
<td>0.7043</td>
<td>0.4288</td>
<td>0.5637</td>
<td>0.2584</td>
<td>0.3541</td>
<td>-0.3250</td>
</tr>
<tr>
<td>6</td>
<td>0.4718</td>
<td>-0.5827</td>
<td>-0.8822</td>
<td>-0.0575</td>
<td>-0.5030</td>
<td>-0.3210</td>
</tr>
<tr>
<td>7</td>
<td>-0.0784</td>
<td>-0.1697</td>
<td>-0.6500</td>
<td>0.4522</td>
<td>0.1898</td>
<td>0.1908</td>
</tr>
<tr>
<td>8</td>
<td>-0.8667</td>
<td>0.7838</td>
<td>-0.5350</td>
<td>-0.0356</td>
<td>-0.1262</td>
<td>-0.2065</td>
</tr>
<tr>
<td>9</td>
<td>0.5785</td>
<td>0.7594</td>
<td>-0.5423</td>
<td>-0.0852</td>
<td>0.3474</td>
<td>0.0436</td>
</tr>
</tbody>
</table>

Table 3. Iterative Error Variation in Simulation (m or rad).

<table>
<thead>
<tr>
<th>No.</th>
<th>RMSPE</th>
<th>RMSOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0087</td>
<td>0.0151</td>
</tr>
<tr>
<td>2</td>
<td>0.0052</td>
<td>0.0104</td>
</tr>
<tr>
<td>3</td>
<td>0.0021</td>
<td>0.0036</td>
</tr>
<tr>
<td>4</td>
<td>0.0015</td>
<td>0.0018</td>
</tr>
<tr>
<td>5</td>
<td>0.0011</td>
<td>0.0016</td>
</tr>
<tr>
<td>6</td>
<td>0.0008</td>
<td>0.0013</td>
</tr>
<tr>
<td>7</td>
<td>0.0005</td>
<td>0.0007</td>
</tr>
<tr>
<td>8</td>
<td>0.0004</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

The kinematic calibration results of transformation matrices $T_{B_1}^W$, $T_{A_1}^B$ and $T_{D_1}^A$ are given in Table 4. From the simulation results above, the position errors of transformation matrices $T_{B_1}^W$, $T_{A_1}^B$ and $T_{D_1}^A$ can be calibrated accurately. It is shown that the orientation errors can be calibrated by the calculation of rotation matrices for transformation matrices $T_{B_1}^W$, $T_{A_1}^B$ and $T_{D_1}^A$.

The goal of the experiments is to explore the feasibility of the kinematic calibration technique proposed in Section 3 and verified through experimental studies. Two sets of pose measurement data of the 6-DOF electro-hydraulic parallel manipulator constructed in the Aerospace System Engineering Shanghai were collected by using a laser tracking measurement system. This data is fed into the kinematic calibration algorithm to provide the experimental verification. The experimental results are shown in Table 5. The first set of pose measurement data is used in the kinematic calibration. It is obvious that the position errors are less than 1 mm, and the orientation errors are not more than 0.1 degree.

Since pose measurement data are inevitably contaminated by noise, more pose measurements are needed in order to improve the identification of parameter errors in the view of least squares. The relationship between the pose accuracy and the number of pose measurement is illustrated in Figure 4. As expected better pose accuracy is achieved as the number of pose measurement increases. However, very little pose accuracy can be obtained once the number of pose measurement exceeds 30 pose measurements. It should be emphasized that the parameter errors cannot be completely eliminated due to the presence of pose measurement errors.

Figure 4. The Relationship between Pose Accuracy and the Number of Pose Measurements.

5 CONCLUSIONS

The kinematic calibration strategy based on the laser tracking measurement method and a more complete pose error model including geometrical parameter errors of the parallel manipulator and transformation errors of the coordinate frame of the semi-physical simulation system have been proposed. The numerical values of the above errors are identified using iterative least squares algorithm. The kinematic calibration results presented in the article showed to improve the 6-DOF parallel manipulator for a semi-physical simulation system of a space docking mechanism pose accuracy to well below 1 mm and 0.1 degree.

6 DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.
### Table 4. Simulated Calibration Results.

<table>
<thead>
<tr>
<th>Error Matrix</th>
<th>dx</th>
<th>dy</th>
<th>dz</th>
<th>δx</th>
<th>δy</th>
<th>δz</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{a}^w$</td>
<td>0.0199</td>
<td>0.0171</td>
<td>-0.0089</td>
<td>0.0301</td>
<td>0.0198</td>
<td>0.0081</td>
</tr>
<tr>
<td></td>
<td>-0.0089</td>
<td>-0.0009</td>
<td>-0.0067</td>
<td>-0.0029</td>
<td>-0.0114</td>
<td>-0.0060</td>
</tr>
<tr>
<td></td>
<td>-0.0054</td>
<td>-0.0111</td>
<td>-0.0009</td>
<td>0.0079</td>
<td>-0.0091</td>
<td>-0.0152</td>
</tr>
<tr>
<td></td>
<td>0.0017</td>
<td>-0.0027</td>
<td>0.0041</td>
<td>-0.0071</td>
<td>0.0036</td>
<td>-0.0034</td>
</tr>
<tr>
<td></td>
<td>-0.0067</td>
<td>-0.0061</td>
<td>0.0072</td>
<td>-0.0020</td>
<td>0.0043</td>
<td>0.0059</td>
</tr>
<tr>
<td>$T_{a}^d$</td>
<td>-0.0023</td>
<td>0.0035</td>
<td>-0.0039</td>
<td>0.0139</td>
<td>-0.0042</td>
<td>-0.0129</td>
</tr>
<tr>
<td></td>
<td>0.0029</td>
<td>0.0038</td>
<td>-0.0045</td>
<td>0.0060</td>
<td>0.0027</td>
<td>-0.0147</td>
</tr>
<tr>
<td></td>
<td>0.0040</td>
<td>0.0042</td>
<td>0.0069</td>
<td>-0.0089</td>
<td>-0.0016</td>
<td>0.0036</td>
</tr>
<tr>
<td></td>
<td>0.0036</td>
<td>-0.0004</td>
<td>-0.0047</td>
<td>0.0108</td>
<td>-0.0141</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td>-0.0073</td>
<td>0.0012</td>
<td>0.0051</td>
<td>0.0039</td>
<td>-0.0150</td>
<td>-0.0151</td>
</tr>
<tr>
<td>$T_{a}^b$</td>
<td>0.0179</td>
<td>0.0151</td>
<td>-0.0049</td>
<td>0.0398</td>
<td>0.0509</td>
<td>0.0019</td>
</tr>
</tbody>
</table>

### Table 5. Experimental Calibration Results.

<table>
<thead>
<tr>
<th>Error Matrix</th>
<th>dx</th>
<th>dy</th>
<th>dz</th>
<th>δx</th>
<th>δy</th>
<th>δz</th>
</tr>
</thead>
<tbody>
<tr>
<td>before</td>
<td>after</td>
<td>before</td>
<td>after</td>
<td>before</td>
<td>after</td>
<td>before</td>
</tr>
<tr>
<td>$T_{a}^w$</td>
<td>0.0162</td>
<td>0.0008</td>
<td>0.0090</td>
<td>0.0004</td>
<td>-0.0137</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>0.0002</td>
<td>-0.0009</td>
<td>0.0049</td>
<td>-0.0001</td>
<td>0.0041</td>
<td>-0.0009</td>
</tr>
<tr>
<td></td>
<td>0.0003</td>
<td>-0.0004</td>
<td>0.0039</td>
<td>-0.0008</td>
<td>0.0016</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>0.0002</td>
<td>0.0009</td>
<td>-0.0010</td>
<td>0.0001</td>
<td>0.0012</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>-0.0008</td>
<td>0.0013</td>
<td>-0.0001</td>
<td>0.0001</td>
<td>-0.0004</td>
</tr>
<tr>
<td></td>
<td>0.0002</td>
<td>0.0002</td>
<td>-0.0011</td>
<td>-0.0006</td>
<td>-0.0050</td>
<td>-0.0002</td>
</tr>
<tr>
<td></td>
<td>-0.0026</td>
<td>-0.0001</td>
<td>0.0035</td>
<td>-0.0007</td>
<td>-0.0027</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>0.0044</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.0008</td>
<td>0.0048</td>
<td>-0.0001</td>
</tr>
<tr>
<td></td>
<td>-0.0041</td>
<td>0.0007</td>
<td>0.0031</td>
<td>-0.0007</td>
<td>0.0036</td>
<td>-0.0008</td>
</tr>
<tr>
<td></td>
<td>0.0004</td>
<td>-0.0006</td>
<td>-0.0011</td>
<td>0.0009</td>
<td>-0.0049</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>0.0040</td>
<td>-0.0001</td>
<td>0.0046</td>
<td>0.0004</td>
<td>-0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>$T_{a}^d$</td>
<td>0.0036</td>
<td>0.0008</td>
<td>-0.0047</td>
<td>0.0009</td>
<td>-0.0008</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>0.0013</td>
<td>-0.0009</td>
<td>0.0045</td>
<td>0.0006</td>
<td>0.0027</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>-0.0024</td>
<td>0.0005</td>
<td>0.0021</td>
<td>-0.0009</td>
<td>-0.0001</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>-0.0012</td>
<td>0.0006</td>
<td>0.0015</td>
<td>-0.0003</td>
<td>0.0002</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>-0.0016</td>
<td>-0.0004</td>
<td>-0.0005</td>
<td>0.0001</td>
<td>0.0028</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>-0.0019</td>
<td>-0.0005</td>
<td>-0.0033</td>
<td>-0.0004</td>
<td>0.0009</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>0.0045</td>
<td>-0.0001</td>
<td>0.0034</td>
<td>0.0003</td>
<td>-0.0037</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0047</td>
<td>-0.0003</td>
<td>-0.0039</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>0.0020</td>
<td>0.0008</td>
<td>-0.0007</td>
<td>0.0009</td>
<td>0.0016</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>0.0038</td>
<td>-0.0001</td>
<td>-0.0025</td>
<td>-0.0003</td>
<td>0.0050</td>
<td>-0.0008</td>
</tr>
<tr>
<td>$T_{a}^b$</td>
<td>0.007</td>
<td>0.0008</td>
<td>0.0003</td>
<td>0.0022</td>
<td>0.0128</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
REFERENCES


Dayong Yu received Ph.D. degree in Mechanical Engineering at the Harbin Institute of Technology. He works at the School of Mechanical Engineering in the University of Shanghai for Science and Technology in China. He is an associate professor at this University. His research interests are in kinematic calibration and accuracy analysis of parallel manipulators.
The Virtual Prototype Model Simulation on the Steady-state Machine Performance

Huanyu Zhao, Guoqiang Wang, Shuai Wang, Ruipeng Yang, He Tian, Qiushi Bi

College of Mechanical Science and Engineering, Jilin University, Changchun, P.R. China

ABSTRACT
Articulated tracked vehicles have high mobility and steering performance. The unique structure of articulated tracked vehicles can avoid the subsidence of tracks caused by high traction from instantaneous braking and steering. In order to improve the accuracy of the steady-state steering of the articulated tracked vehicle, the velocity of both sides of the track and the deflection angle of the articulated point need to match better to achieve the purpose of steering accurately and reduce energy consumption and wear of components. In this study, a virtual prototype model of the articulated tracked vehicle is established based on the multi-body dynamic software RecurDyn. The trend of the driving torque and power of each track changes as the velocity difference of two sides of the tracks and the traveling trajectory of the mass center of the front vehicle change in a specific condition are obtained by the experiment. The experimental results are compared and verified with the results obtained from the virtual prototype simulation. The change law of driving power in the steady-state steering process on the horizontal firm ground as changing the velocity difference of two sides of the tracks, the theoretical steering radius, and the ground friction is obtained by the virtual prototype model simulation analysis. The steering inaccuracy and track slip rate are used as indexes in evaluating the steady-state steering performance of the articulated tracked vehicle. The research provides references for the study of steady-state steering performance of articulated tracked vehicles.

KEY WORDS: Articulated tracked vehicles, Steady-state steering performance, Virtual prototype model simulation, Steering inaccuracy, Slip rate

1 INTRODUCTION
ARTICULATED tracked vehicles have the advantages of high mobility and good steering performance (Fijalkowski, 2003), which contribute to widespread use in the military, agriculture, forestry, and other fields. The unique structure of articulated tracked vehicles can avoid the subsidence of tracks caused by high traction from instantaneous braking and steering (Nuttall, 1964). In addition, the risk of rollover in the running process, as heavy wheeled vehicles (Edgar et al., 2011), is avoided. The steering of articulated tracked vehicles differs from that of four-tracked (Watanabe et al., 1995) or multi-tracked vehicles (Zongwei et al., 2013) because the articulated component is comprised of two sets of crawlers in a series together. Articulated tracked vehicles have two main forms. The first form is a wagon-type articulation of tracked vehicles, such as towing vehicles, that enables a tracked trailer to drive through a drawbar (Alhimdani, 1982; Basher, 2012). The other is composed of two powerful vehicles that use the moment applied on the articulated component between the two sets of crawlers to make each crawler deflect from the other to achieve steering. Articulated crawlers have a certain inaccuracy in the steady-state steering, namely, understeer and oversteer. Inaccurate steering can cause the crawler to excessively slip or skip in the steering process, resulting in excessive wearing and reduced life of parts and increased energy loss. Therefore, conducting a study on the steady-state steering performance of an articulated crawler is necessary.
Many scholars have conducted a study on articulated tracked vehicles. Sasaki et al. (Sasaki et al., 1991) described the articulated tracked vehicle design and its unique features and operating characteristics. They also presented the experimental data of the articulated tracked vehicle that drives on different angles and slopes through hydraulic control. On the basis of single-tracked vehicles (Kitano & Jyozaki, 1976; Kitano & Kuma, 1977) and the steering performance of coupled-tracked vehicles (non-articulated) (Kitano et al., 1981) in the steady and unsteady-state transition process, Watanabe et al. (Watanabe & Kitano, 1986) constructed a mathematical model of the unsteady process of articulated crawler vehicles on horizontal ground and performed an experimental demonstration with a proportional model. They concluded that articulated tracked vehicles require lesser driving torque and track slip rate than single-tracked and non-articulated double-tracked vehicles.

Establishing a theoretical model about the running process of articulated crawlers is complex. The interaction between the track shoe and the firm ground (Wong, & Chiang, 2001) or the soft ground (Wong, 2009; Al-Milli, 2010) is also complex. With the continuous development of virtual prototyping technology, the virtual prototype model is widely used to perform a simulation analysis of the driving performance of tracked vehicles (Janarthanan et al., 2012; Choi et al., 1998; Lee et al., 1998; Wang et al., 2014). Wong (Wong, 1992; Wong, 1992) analyzed the influence of important parameters, such as joint articulation configuration, suspension characteristics, initial track tension, track width, and center of gravity location, on the mobility of tracked vehicles by using the computer simulation model. The experiment showed that the simulation model plays an important role in the optimization design of tracked vehicles. The efficiency of the simulation analysis of the steady-state steering performance of an articulated crawler will be greatly improved by constructing an accurate articulated crawler virtual prototype model.

It can be seen from the above research, the articulated tracked vehicle was not driven by the motor alone on each track and only an articulated device was used to deflect the articulated crawler. For a motor-operated articulated crawler, the velocity difference of two sides of the tracks, the theoretical steering radius, and the ground friction coefficient can all have a considerable impact on the steering performance of the articulated crawler. In this study, each track of the articulated crawler is driven by a motor. The influence of three important factors on the steering performance of articulated tracked vehicle is analyzed.

In this paper, the accuracy of the virtual prototype model is experimentally validated through an experimental prototype. The accurate virtual prototype model is utilized to perform a simulation analysis on the driving power of the crawler and the trajectory change of the front vehicle when the articulated tracked vehicle is moving in steady-state on the horizontal firm ground. The simulation included changing the velocity difference, the theoretical steering radius, and the ground friction coefficient. The simulation results provide a reference for the study on the steering control of articulated tracked vehicles.

2 ESTABLISHMENT OF VIRTUAL PROTOTYPE MODEL

FIGURE 1 shows the virtual prototype model of an articulated crawler established in the low-speed module Track_LM of multi-body dynamic software RecurDyn. The main parameters of the model are shown in Table 1. Rear vehicles are indicated by "••". The maximum velocity of the designed articulated crawler is 0.2 m/s. In the virtual prototype model, the front and rear vehicles are connected by two articulated frames, which are operated by a linear actuator to deflect the front and rear vehicles around the articulated point.

Figure 1. Virtual Prototype Model of the Articulated Crawler: 1 Drive motor, 2 Rear vehicle, 3 Articulated frame, 4 Front vehicle, and 5 Linear actuator.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass: G=G’</td>
<td>450 kg</td>
</tr>
<tr>
<td>Ground contact length: L=L’</td>
<td>1200 mm</td>
</tr>
<tr>
<td>Ground contact pressure: P=P’</td>
<td>6.125 kPa</td>
</tr>
<tr>
<td>Track gauge: B=B’</td>
<td>930 mm</td>
</tr>
<tr>
<td>Width of link pad: b=b’</td>
<td>300 mm</td>
</tr>
<tr>
<td>Pitch of chain link: p=p’</td>
<td>102 mm</td>
</tr>
<tr>
<td>Pitch radius of sprocket: r=r’</td>
<td>152 mm</td>
</tr>
<tr>
<td>Tooth number of sprocket: n=n’</td>
<td>21</td>
</tr>
<tr>
<td>Distance from articulation point to mass center: l=l’</td>
<td>786 mm</td>
</tr>
<tr>
<td>Number of track rollers: n=n’</td>
<td>2</td>
</tr>
<tr>
<td>Number of chain links: n=n’</td>
<td>27</td>
</tr>
<tr>
<td>Deflection angle range</td>
<td>±20°</td>
</tr>
</tbody>
</table>

3 EXPERIMENTAL VERIFICATION

GIVEN that conducting a theoretical verification of the veracity of the virtual prototype is unpractical, verifying the accuracy of the virtual prototype model...
by experiment is necessary. Figure 2 illustrates the experimental prototype of the articulated crawler. The crawler is operated by motors with a rated power and speed of 0.75 kW and 1250 rpm respectively. The reduction ratio of the reducer is 100. The frequency converters are used to change frequency between the power supply and motors. The motor speed is changed by controlling the output frequency of the frequency converters through the computer, and then the velocities of both sides of the crawler are controlled. The stabilized power supply is used to supply 24V DC voltage to the linear actuator. The articulated steering part relies on the push and pull of the linear actuator to adjust the steering angle and direction. A trajectory drawing device depicts the trajectory of the mass center of the front vehicle of the articulated crawler.

3.1 Comparison of trajectories

The articulated crawler runs on the firm ground at a velocity of 0.15 m/s for 30 s at the beginning. Then, the velocity of the left side (inside) of the articulated crawler decreases to 0.13 m/s, whereas the velocity of the right side (outside) of the articulated crawler increases to 0.17 m/s. Meanwhile, the linear actuator elongates at a speed of 5 mm/s to push the articulated point to turn 20°, thereby enabling the crawler to turn left until it runs steadily. Figures 3(a) and (b) show the comparison of the simulation and experimental trajectories of the mass center of the front vehicle of the articulated crawler. A certain lateral slip occurs in the running process of the crawler, causing the second path to be slightly outward compared to the first path in the simulation result. In the experimental results, the deviation of the two trajectory paths is not obvious because of the error of the trajectory drawing device. The deviation of the two paths is small, and the lines of the delineation are rough. However, the consistency of the comparison results is acceptable based on the results of the overall trajectory process.

3.2 Comparison of Driving Torque and Power of Simulation and Experimental Results

Figures 4 and 5 show the comparison of the simulation and experimental results of the average driving torque and the power of the articulated crawler in steady-state at 20° articulated point steering and velocity differences of two sides of the tracks at 0, 0.02, 0.04, and 0.06 m/s, respectively. In Figure 4, the driving torque of the same-side tracks is similar. The driving torque of the left and right side tracks decreases and increases, respectively, as the velocity difference increases. The change trend of the driving power of each track is similar to that of the driving torque, as shown in Figure 5. The simulation and experimental results in Figures 4 and 5 are similar, and the overall change trends are also similar; thus, the consistency is acceptable.
The virtual prototype model can describe the running performance of the articulated crawler. Therefore, the steering performance of the articulated crawler can be analyzed and studied.

4 RESULTS AND DISCUSSIONS

4.1 Steering Inaccuracy and Slip Rate

In the steering process of the articulated crawler, each track is affected by the lateral force and existing slips or skips (Zongwei et al., 2013). Figure 6 shows the diagram of the steering motion of the articulated crawler. Each track has a respective velocity instantaneous center $O_i$ ($i=1, 2, 3, 4$) on the ground plane, which is relative to the geometric center of the track–terrain interface $O_i$ that produces longitudinal $D_i$ and lateral offset $A_i$. $O_i$ is the theoretical steering center and $O_s$ is the actual steering center of the articulated tracked vehicle. The distance between $O_i$ and $O_s$, and the mass center of the front vehicle or rear vehicle ($C_1$ or $C_2$) are the actual steering radius $R_s$ and theoretical steering radius $R_l$ respectively. The steering inaccuracy is expressed as

$$\delta_R = \frac{(R_s - R_l)}{R_l} \times 100\%$$  \hspace{1cm} (1)

Where the theoretical steering radius $R_l$ is expressed as

$$R_l = \frac{l}{\tan(\alpha/2)}$$  \hspace{1cm} (2)

Where $\alpha$ is the deflection angle of the articulated point. When $R_s > R_l$, $\delta_R > 0$, that is, understeer; when $R_s < R_l$, $\delta_R < 0$, that is, oversteer; when $\delta_R$ approaches 0, the steering trajectory is ideal. Therefore, the steering accuracy of the articulated crawler can be evaluated by the steering inaccuracy $\delta_R$.

The skip occurs when the relative velocity (winding velocity) $v_t$ of the track is greater than the entrainment velocity (translational velocity) $v_e$, and the slip occurs when the relative velocity $v_t$ is less than the entrainment velocity $v_e$ (Zongwei et al., 2013). Thus, the slip rate $\delta_S$ can be introduced to quantitatively describe the extent to which the track–terrain interface is skipping or slipping. It is defined as

$$\delta_S = \frac{(v_t - v_e)}{v_e} \times 100\%$$  \hspace{1cm} (3)

When $v_t > v_e$, $\delta_S > 0$, skip occurs in the track; when $v_t < v_e$, $\delta_S < 0$, slip occurs in the track; when $\delta_S$ approaches 0, the running efficiency of track is the highest. Therefore, the running stability of the tracks can be evaluated by the slip rate $\delta_S$.

4.2 Effect of Velocity Difference and Theoretical Steering Radius

To maintain the running velocity $v_e$ of the mass center of the articulated crawler, in the overall running process, in the case of going straight, the relationship between the inner $v_1$ and outer track velocities $v_2$ is $v_1 = v_2 = v_e$; whereas in the case of steering, the inner track velocity $v_1$ is decreased and the outside track velocity $v_2$ is increased to form a velocity difference $\Delta v=v_2-v_1$. The velocity of both sides of the tracks is shown in Table 2. The relationship between the deflection angle of the linear actuator pushing the articulated point and the theoretical steering radius is when $\alpha = 20^\circ$, $R_l = 4.47$ m; when $\alpha = 15^\circ$, $R_l = 6.00$ m; when $\alpha = 10^\circ$, $R_l = 9.01$ m; when $\alpha = 5^\circ$, $R_l = 18.05$ m. The ground friction coefficient is 0.7.
Table 2. Velocity of Both Sides of the Articulated Crawler.

<table>
<thead>
<tr>
<th></th>
<th>Velocity of center of mass: $v_1$ (m/s)</th>
<th>Velocity of inside track: $v_1$ (m/s)</th>
<th>Velocity of outside track: $v_2$ (m/s)</th>
<th>Velocity difference: $\Delta v$ (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.14</td>
<td>0.16</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.13</td>
<td>0.17</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.12</td>
<td>0.18</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.11</td>
<td>0.19</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

Figure 7 shows the effect of velocity difference and theoretical steering radius on the driving power of each track. At the same $\Delta v$, the driving power of each track decreases with the increase of $R_L$. At the same $R_L$, the driving power of the left and right side tracks decreases and increases, respectively, with the increase of $\Delta v$. In addition, as the $R_L$ increases, the driving power of both sides of the crawler slows down with the trend of $\Delta v$.

Figure 8 shows the effect of the velocity difference and the theoretical steering radius on the total driving power of the articulated crawler. The change trend of the total driving power as the $\Delta v$ increases at different $R_L$ is the same, decreasing first and then increasing, and is relatively low at $\Delta v$ of 0.02–0.06 m/s. In the same $\Delta v$ change interval, the variation range of the total driving power increases as the $R_L$ increases.

Figure 9 shows the effect of the velocity difference on the trajectory of the mass center of the front vehicle under different theoretical steering radii. $R_S$ decreases with the increase of $\Delta v$ at the same $R_L$. In the same $\Delta v$ change interval, the range of the mass center of the front vehicle $R_S$ increases as the $R_L$ increases.
tracks on the right side $\delta_S$ increases with an increase of $\Delta \nu$; from $\delta_S<0$ it gradually increases to $\delta_S>0$, that is, from slip to skip, and the slope of $\delta_S$ curve gradually decreases. At the same $\Delta \nu$, the $\delta_S$ of the left side tracks decreases with the increase of $R_L$. The $\delta_S$ of the right tracks increases with the increase of $R_L$.

Figure 11 shows the variation of the slip rate of each track with the velocity difference under different theoretical steering radii. The change trend of $\delta_S$ of both sides of the articulated crawler is similar. At the same $R_L$, the tracks on the left side $\delta_S<0$, that is, slip occurs; $\delta_S$ decreases with the increase of $\Delta \nu$; and the slope of $\delta_S$ curve gradually increases. Moreover, the
Figures 10 and 11 show that when $\delta R = 0$, the corresponding velocity difference of the four theoretical steering radii is 0.05, 0.04, 0.03, and 0.02 m/s. The corresponding slip rate of the left track is approximately −15%, and the slip rate of the right track is approximately −2.5%. Thus, the slip rate is low. Figure 8 shows that each velocity difference is between 0.02 and 0.06 m/s, and the total driving power is low. Thus, the steady-state steering performance of the articulated crawler can be improved by the corresponding velocity difference of each theoretical steering radius.

### 4.3 Effect of Velocity Difference and Ground Friction Coefficient

In the case of the theoretical steering radius is 4.47 m, the effect of the ground friction coefficient and the velocity difference on the steady-state steering performance of the articulated crawler is analyzed. Figure 12 shows the effect of the velocity difference under different ground friction coefficients for each track driving power. At the same $\Delta v$, the driving power of each track increases as the ground friction coefficient $\mu$ increases. At the same $\mu$, the driving power of the left and right side tracks decreases and increases, respectively, with the increase of $\Delta v$. In addition, as $\mu$ decreases, the trend of driving power for both sides of the tracks with $\Delta v$ slowed down.

![Figure 12. Effect of Velocity Difference and Ground Friction Coefficient on Driving Power.](image)

Figure 13 shows the effect of velocity difference and ground friction coefficient on the total driving power of the articulated crawler. The trend of the total driving power of the articulated crawler changes as the $\Delta v$ increases at different $\mu$, which decreases first and then increases, and is relatively low at $\Delta v$ of 0.02–0.06 m/s. In the same $\Delta v$ change interval, the variation range of the total driving power increases as the $\mu$ increases.

![Figure 13. Effect of Velocity Difference and Ground Friction Coefficient on Total Driving Power.](image)

Figure 14 shows the effect of the velocity difference on the trajectory of the mass center of the front vehicle under different ground friction coefficients. $R_S$ decreases as the $\mu$ increases at the same $\Delta v$. In the same $\mu$, the mass center of the front vehicle $R_S$ decreases as the $\Delta v$ increases.
Figure 14. Effect of Velocity Difference and Ground Friction Coefficient on the Steering Trajectory.

Figure 15 shows the variation of the steering inaccuracy of the articulated crawler with the velocity difference under different ground friction coefficients. The change trend of the $\delta_R$ of the mass center of the front vehicle is the same as the $\Delta v$ increases, and the $\delta_R$ of the front vehicle gradually becomes understeer to oversteer as the $\Delta v$ increase. Moreover, the slope of $\delta_R$ curves gradually decreases. When $\delta_R=0$, the corresponding $\Delta v$ decreases as the $\mu$ increases. That is, if $\mu=0.3$, then $\Delta v=0.06$ m/s; if $\mu=0.5$, then $\Delta v=0.055$ m/s; if $\mu=0.7$, then $\Delta v=0.05$ m/s.

Figure 15. Effect of Velocity Difference and Ground Friction Coefficient on the Steering Inaccuracy.

Figure 16 shows the variation of the slip rate of each track with the velocity difference under different ground friction coefficients. The change trend for both sides of the tracks $\delta_S$ is similar. At the same $\mu$, the tracks on the left side $\delta_S<0$, that is, slip occurs, $\delta_S$ decreases as $\Delta v$ increases, and the slope of $\delta_S$ curve gradually increases. Moreover, the tracks on the right side $\delta_S$ increases as $\Delta v$ gradually increases from $\delta_S<0$ to $\delta_S>0$, that is, from slip to skip, the slope of $\delta_S$ curve gradually decreases. At the same $\Delta v$, the $\delta_S$ of the tracks in the left side increases as the $\mu$ increases. The $\delta_S$ of the tracks on the right side increases as the $\mu$ increases when the slip occurs, and $\delta_S$ decreases as the $\mu$ increases when the skip occurs. Therefore, the curve crosses when $\delta_S$ is approximately 0.

Figure 16. Effect of Velocity Difference and Ground Friction Coefficient on the Slip Rate.
between −15% and −25%, and on the right side is between 0 and −2.5%. Figure 13 shows that the velocity difference is between 0.02 and 0.06 m/s, and the total driving power is low. Thus the steady-state steering performance of the articulated crawler can be improved by the corresponding velocity difference of each ground friction coefficient.

### 4.4 Comprehensive Effect of Velocity Difference, Theoretical Steering Radius, and Ground Friction Coefficient on Steering Performance

The effect of velocity difference, theoretical steering radius, and ground friction coefficient on the steady-state steering performance of the articulated crawler is comprehensively analyzed. Figures 17 (a) and (b) show the effect of velocity difference and ground friction coefficient on the steady-state steering inaccuracy of the articulated crawler in the case of theoretical steering radius of 18.05 and 9.01 m respectively. Figures 15 and 17 show that high \( R_L \) and \( \mu \) results in the low \( \Delta v \), which is required for stabilizing the articulated crawler \( \delta_R \) at 0. For relative \( \mu, R_L \) has a great effect on \( \delta_R \) under the same \( \Delta v \).

**Figure 17.** Effect of the Theoretical Steering Radius and Ground Friction Coefficient under Different Velocity Difference on Steering Inaccuracy.

Figures 18 and 19 show the effects of velocity difference and ground friction coefficient on each track slip rate in the case of theoretical steering radius of 18.05 and 9.01 m respectively. Figures 16, 18, and 19 show that as \( R_L \) and \( \mu \) changes, the tracks on the left side slip under the same \( \Delta v \), that is, \( \delta_R < 0 \). A high \( R_L \) and a low \( \mu \) result in the decreased slip rate of the tracks on the left side under the same \( \Delta v \). When the tracks on the right side slip, that is \( \delta_R > 0 \), the \( \delta_R \) increases as the \( R_L \) and \( \mu \) decreases under the same \( \Delta v \).

**Figure 18.** Effect of Ground Friction Coefficient and Velocity Difference under \( R_L = 18.05 \) m.
Figure 19. Effect of Ground Friction Coefficient and Velocity Difference under $R_L = 9.01$ m.

Figures 17, 18, and 19 show that when $\delta R = 0$, the velocity differences corresponding to the three $\mu$ when $R_L = 18.05$ m are 0.03, 0.025, and 0.02 m/s; and the velocity differences corresponding to the three $\mu$ when $R_L = 9.01$ m are 0.04, 0.035, and 0.03 m/s. The corresponding slip rate of the tracks on the left side is between $-15\%$ and $-25\%$, and the slip rate of track on the right side is between 0 and $-2.5\%$.

5 CONCLUSIONS

ON the basis of the proposed virtual prototype simulation of RecurDyn and experimental verification methods, this research studied the impact of velocity difference, theoretical steering radius, and ground friction coefficient on the driving power of the articulated crawler, as well as the change of the trajectory of the front vehicle when the crawler moves in a steady-state on the horizontal firm road surface. The steady-state steering performance of an articulated crawler was likewise evaluated under different circumstances through the steering inaccuracy and the slip rate of each track.

Steering inaccuracy is used to evaluate the deviation between the actual and theoretical steering radii when the articulated crawler is in steady-state steering. When the theoretical steering radius and ground friction coefficient are high, the velocity difference which is required for stabilizing the steering inaccuracy at zero is low. The theoretical steering radius has a great effect on the steering inaccuracy under the same velocity difference.

The slip rate of the crawler is used to evaluate the slip or skip of the track during the steady-state steering of the articulated crawler. For the inside track, the slip usually occurs when changing the velocity difference, the theoretical steering radius, and the ground friction coefficient. For the outside track, both slip or skip may occur as the change of the velocity difference, the theoretical steering radius, and the ground friction coefficient. Under the corresponding velocity difference which improves the steering accuracy and reduces the driving power, the slip rate of both sides of the tracks is in a certain range.

When the theoretical steering radii and ground friction coefficients are different, the steering trajectory of the articulated crawler vehicle in the steady-state steering can be controlled effectively by selecting the appropriate velocity difference and conducting real-time monitoring of the crawler slip rate. Moreover, it can improve steering accuracy, reduce energy consumption, and improve operational efficiency. The research on the steady-state steering performance of articulated crawlers can be a reference for the study on steering control and trajectory optimization of articulated tracked vehicles.

6 ACKNOWLEDGMENTS

THIS work was supported by the Collaborative Innovation Center of Major Machine Manufacturing in Liaoning and Shanxi Provincial Key Scientific and Technological Projects (MJ2014-02).
7 REFERENCES


8 NOTES ON CONTRIBUTORS

Huanyu Zhao received B.E. degree from Jilin University, Changchun, China, in 2009. He is currently working towards a Ph.D. degree in the College of Mechanical Science and Engineering of Jilin University in China. His research interests include the dynamics of tracked vehicle and computer simulation.

Guoqiang Wang received Ph.D. degree in Agricultural Engineering from Jilin University of Technology (Now Jilin University), Changchun, China, in 1994. He has been a professor since 1997 in the School of Mechanical Science and Engineering, Jilin University, Changchun, China. He went to Technische Universität Bergakademie Freiberg, Germany, as a visiting scholar from 1997 to 1998 and University of Victoria, Canada, as a visiting scholar from 2003 to 2004. His research interests include modern design theories and methods (MDTM), reliability, dynamics, vibration and mechanical failure diagnosis.
Shuai Wang received B.S. degree from Changchun University of Science and Technology, Changchun, China, in 2009. He is currently working towards a Ph.D. degree in the College of Mechanical Science and Engineering of Jilin University in China. His major area is modern mechanical design theory and method.

Ruipeng Yang received B.E. degree from Taiyuan University of Science and Technology, Taiyuan, China, in 2016. He is currently working towards a M.S. degree in the College of Mechanical Science and Engineering of Jilin University in China. His research interests include machine vision, electromechanical joint simulation and dynamic simulation.

He Tian received B.E. degree from Jilin University, Changchun, China, in 2015. He is currently working towards a M.E. degree in the College of Mechanical Science and Engineering, Jilin University in China. His research interests include the intelligent control and data acquisition.

Qiushi Bi received the M.E. degree from Jilin University, Changchun, China, in 2014. He is currently working towards a Ph.D. degree in Mechanical Engineering in the College of Mechanical Science and Engineering, Jilin University, Changchun, China. His research interests include mechanical design theory, mining machinery, DEM-FEM simulation and mechanism optimization.
The SLAM Algorithm for Multiple Robots based on Parameter estimation

MengYuan Chen¹ ²
¹. Key Lab of Electric Drive and Control of Anhui Province, Anhui Polytechnic University, Wuhu 241000, China
². Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230027, China

ABSTRACT
With the increasing number of feature points of a map, the dimension of systematic observation is added gradually, which leads to the deviation of the volume points from the desired trajectory and significant errors on the state estimation. An Iterative Squared-Root Cubature Kalman Filter (ISR-CKF) algorithm proposed is aimed at improving the SR-CKF algorithm on the simultaneous localization and mapping (SLAM). By introducing the method of iterative updating, the sample points are re-determined by the estimated value and the square root factor, which keeps the distortion small in the highly nonlinear environment and improves the precision further. A robust tracking Square Root Cubature Kalman Filter algorithm (STF-SRCKF-SLAM) is proposed to solve the problem of reduced accuracy in the condition of state change on the SLAM. The algorithm is predicted according to the kinematic model and observation model of the mobile robot at first, and then the algorithm updates itself by spreading the square root of the error covariance matrix directly, which greatly reduces the computational complexity. At the same time, the time-varying fading factor is introduced in the process of forecasting and updating, and the corresponding weight of the data is adjusted in real time to improve the accuracy of multi-robot localization. The results of simulation shows that the algorithm can improve the accuracy of multi-robot pose effectively.

KEY WORDS: Iteration, Simultaneous localization and mapping, Sampling, Multi-mobile robot, Square Root Cubature Kalman Filter, Strong tracking

1 INTRODUCTION
SIMULTANEOUS Localization and Mapping (SLAM) can be expressed as follows: Via gathering information from the sensors, a multi-mobile robot is capable of creating the feature map in the unknown environment, and modifying the position and orientation continuously (Ullah, et.al. 2017) (LIU, et.al. 2011). It is a hot topic for mobile robots whether to motion autonomously in the field of SLAM.

The mainstream algorithm applied in the field of SLAM is the classical Extended Kalman Filter (EKF), which is proposed by Smith and Cheeseman, and its essence is first-order Taylor series expansion about the nonlinear system model and observation model, which is linearized by means of the standard Kalman Filter (Smith, et.al. 1987) (Smith, et.al. 1990). On the basis of EKF-SLAM, many scholars have put forward a new improved algorithm. Iterative Extended Kalman Filter (IEKF) proposed relies on the combination of environment feature iteration and EKF algorithm, which corrects the nonlinear error of the robot and improves the accuracy of localization (Qiang, et.al. 2013). An anti-interference EKF-SLAM algorithm, based on the detection of external interference by contrast and errors of expansion positioning, is put forward to improve the robustness of the algorithm (Tai, et.al. 2012). The higher-order term is ignored in the process of linearization, and so, it is not easy to achieve the desired results under circumstance of precisely positioning. The Jacobian matrix calculated in the linearization procedure increases the complexity of computation, which is not applicable to a broad environment as well. Multi-robot coordinated control is a hotspot of robot research in recent years, and there is a broad applying prospect in industrial, military, aerospace and so on (WANG, et.al. 2015) (ZHU, et.al. 2015).
Most of the robot systems require that robots can be aware of surroundings and capable of locating itself when exploring the environment. The multi-robots observe mutually and share information by exchanging messages, which reduces the dependency on the external environment and enhances the perception of a single robot, so it can obtain more accurate positioning information than a single robot does, which is named as an approach of cooperative localization for multiple robots. The co-localization study was originally focused on land-based robots, and later developed into underwater robots, parallel structures composed of homogeneous robots (Ling, et.al. 2007), and master-slave structures composed of heterogeneous robots (Bailey, et.al. 2011). In addition, transferring information from the superior quality sensor information to the inferior quality sensor information can obtain overall precision of the system to be best.

In recent years, the popular algorithm Unscented Kalman Filter (UKF) was proposed (Ouyang, et.al. 2014) (Lin, et.al. 2013). The sampling points with different weights (Sigma points) are chosen to be transferred with the non-linear function, and the statistical characteristics of the random variables are obtained by UT.

Under the framework of the Kalman Filter, UKF reduces the errors caused by the linearization of nonlinear equations and achieves the second-order positioning accuracy under the same conditions. In literature (Merwe, et.al. 2004), Merwe and Wan develop this theory effectively by deriving mathematically the square root of covariance matrix instead of transferring covariance matrix, which can be well applied to the SLAM (Zhao, et.al. 2011). In literature (Wang, et.al. 2014), based on the square root UKF, the sampling strategy of the Sigma point is changed and a single-row sampling algorithm is proposed, but UKF may lead to non-definite covariance matrices in the process of filtering, which affects filter performance.

In 2009, Arasaratnam and Haykin proposed the Cubature Kalman Filter (CKF), which provides a new method for state estimation of the nonlinear system. Under the rules of volume in CKF, a set of the corresponding weight of the volume point was selected into the non-linear function (Bian, et.al. 2017) (Zhao, et.al. 2017) (Arasaratnam, et.al. 2009), and statistical properties can be obtained by processing the weighted distribution of random variables. Compared with UKF, CKF also reduces the errors caused by the linearization of nonlinear equations and this algorithm has been recognized by more and more scholars for its low computational complexity, high numerical precision and strong filter stability, which makes it effective to solve various valuation problems. Square Root Cubature Kalman Filter (SR-CKF) is put forward, which can enhance the stability of system state (Kang, et.al. 2017). However, if the system state changes suddenly, SR-CKF positioning has a poor performance.

A Strong Tracking Filter (STF) is proposed, and time-varying fading factor is introduced when recursively updating, which achieves a dynamic adjustment of the gain matrix, and the improved SLAM is still capable of highly tracking even if the system changes the state abruptly, but at the same time it greatly increased the computational complexity (Wang, et.al. 2013).

Therefore, Iterative Square Root Cubature Kalman Filter (ISR-CKF) is presented in this paper, which introduces Gauss-Newton iterative method on the basis of SR-CKF and draws the advantage of SR-CKF; transferring the matrix information with sampling the square root of the covariance to reduce the truncation error on SLAM, increasing the iterative process to facilitate the full use of measurement information and reducing the SR-CKF algorithm in a highly nonlinear environment of the system state estimation error. Compared with the previous UKF and SR-CKF, ISR-CKF on SLAM can improve the accuracy of the robot pose estimation effectively by simulation software MATLAB. In this paper, SR-CKF combined with the theory of strong tracking filter and Strong Tracking Filter Square Root Cubature Kalman Filter SLAM algorithm (STF-SRCKF-SLAM) is proposed, which is to ensure that the system is positive and a strong tracking filter can dynamically adjust the weight of the corresponding data to improve the accuracy of the system. Experiments show that the STF-SRCKF-SLAM algorithm has a good performance on estimation accuracy and it can keep strong tracking ability under the circumstance of the mutation state, which can be used for simultaneous localization and mapping in large areas.

2 THE DESCRIPTION OF SIMULTANEOUS LOCALIZATION AND MAPPING (SLAM) BASED ON BAYESIAN ESTIMATION

The essence of SLAM is to estimate the pose of the robot and feature map by the movement information inside the mobile robot and the observation information collected from the sensors.

Define the following variables at this time of K: $S_k = [S_k^t, M_t^r]^T$ represents the robot system time vector, of which $S_k^t$ represents the robot pose information. $M_t^r = [m_1^t, m_2^t, \ldots, m_r^t]$ signifies K-time map feature vector and $n_z$ denotes the system state vector dimension. $U_k = [u_1, u_2, \ldots, u_k]^T$ embodies the kinematic information of the robot at the time of k and $n_u$ symbolizes the vector control vector dimension of the robot. $Z_k = [z_1, z_2, \ldots, z_k]^T$ Represents the
observation vector of the robot $k$ and $n_z$ represents the observation vector of the robot at the time of $k$.

The motion model of the robot is $S'_k = f(S'_{k-1}, u_k) + W_k$. The robot observation model is $z_k = h(S'_k) + V_k$, of which $W_k$ obey the Gaussian distribution and $V_k$ obey the Gaussian distribution (YANG, et.al. 2014).

SLAM can be expressed by

$$p(S'_{k}, M'_{k} | Z_{1:k}, U_{1:k})$$

(1)

Namely the known controlled quantity $U_{1:k}$ and observations $Z_{1:k}$ estimate the feature map $M'_{k}$ and the state vector $S'_{k}$ of the joint probability distribution at the moment. (1) can be obtained at each moment to meet the normal distribution of the optimal solution $\hat{S}'_k$ and its covariance $P_k$, which is the key to SLAM.ISR-CKF proposed in this paper can estimate the pose of the robot and the feature map through the known information and its core is to calculate the conditional probability density by volume transformation (2) and realize Bayesian filtering under Gaussian domain.

$$I = \int f(y)N(y; \mu, \sum) dy 
\approx \frac{1}{2n_y} \sum f(\sqrt{\sum} \xi_i + \mu)$$

(2)

$\xi_i$ is expressed as a set of perfectly ortho-symmetric volume points to each other, where $\xi_i$ is $2(n_z + n_x)$ columns.

$$\xi = \sqrt{n_z + n_x},
\begin{bmatrix}
1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0
\end{bmatrix}$$

(3)

Based on the principles of the Bayesian filtering, the ISR-CKF algorithm is divided into two steps, which are prediction and renewal to achieve prior probability distribution and posterior probability distribution at the time of $k$.

First, a prior probability distribution is obtained by using the posterior probability distribution at time $k-1$ and the quantity of motion control at time $k$:

$$p(S_{k-1}, Z_{1:k}, U_{1:k}) = \int p(S_{k-1}, Z_{1:k}, U_{1:k}) \cdot p(S_k | Z_{1:k}, U_{1:k}) dS_{k-1}$$

(4)

Second, the posterior probability distribution is updated with the $k$-time observation and the prior probability distribution:

$$p(S_k | Z_{1:k}, U_{1:k}) = \eta p(z_k | S_k) \cdot p(S_k | Z_{1:k-1}, U_{1:k})$$

(5)

$\eta$ is a constant

## 3 ITERATIVE SQUARE ROOT CUBATURE KALMAN FILTER ON SLAM (ISR-CKF-SLAM)

THE SR-CKF algorithm is based on the third-order spherical-phase volume rule in the framework of the Kalman Filter and the mean value of the non-linear function and the square root of the covariance are obtained by weighting the set of points. The covariance matrix square root (based on QR decomposition) instead of covariance matrix is used to solve the filter divergence, which is caused by rounding error. The specific algorithm in prediction process is described as follows:

1. **Forecast stage:**
   - Calculate the volume points:
     $$x_{k-1}^j = L_{k-1}^d \xi_j + S_{k-1}^d$$
     (6)
   - $j$ is the volume point number and its value is $1, 2, 3, \cdots, 2(n_z + n_x)$, $x_{k-1}^j$ contains information about position, feature points and motion control at the time of $k-1$. Referring from the information about the state $S_{k-1}$ and the motion $u_k$, which is expanded to Gaussian noise variation, $L_{k-1}^d$ and $S_{k-1}^d$ can be obtained from (7)

$$S_{k-1}^d = \begin{bmatrix} S_{k-1} \ u_k \end{bmatrix},
L_{k-1}^d = \begin{bmatrix} L_{k-1} & 0 \\
0 & \sqrt{Q_k} \end{bmatrix}$$

(7)

2. **Calculate the priori estimates of each volume point and take volume point $j$ as an example:**

After each volume point has been propagated through the nonlinear motion equation, it can get the information about the position of the robot at $K$-1 time, the motion at $K$ time, and predicts the position of the robot at $K$ time.

3. **Estimation on position state of the robot and predictions for square root factor**

According to the volume transformation (2)

$$S_{k|k-1} = \frac{1}{2(n_z + n_x)} \sum_{j=1}^{2(n_z + n_x)} x_{k|k-1}^j$$

(9)

The square root factor $C_{k|k-1}$, which is for the update phase, can be obtained after the QR decomposition of the error vector.

$$C_{k|k-1} = \frac{1}{\sqrt{Q_k}} \begin{bmatrix} S_{k|k-1} - S_{k|k-1} \ u_{k|k-1} - u_{k|k-1} \ \cdots \ x_{k|k-1}^{(2(n_z + n_x))} - x_{k|k-1}^{(2(n_z + n_x))} \end{bmatrix}$$

(10)
\[
\begin{bmatrix} Q & R \end{bmatrix} = qr \begin{bmatrix} A_{k|k-1} \end{bmatrix}^T, C_{k|k-1} = R^T
\] (11)

Although the SR-CKF-SLAM algorithm can improve the positional accuracy and the system stability of the mobile robot, the dimension of the system observation will be added with the feature points on the map increasing gradually, which leads to the deviation of volume points from the ideal trajectory and therefore results in a large error in the state estimation. In this paper, the iterative method is used to reconstruct the sampling points in the update stage with the usage of the estimated value and the square root factor. The statistic properties of the system are obtained by volume transformation, and then the system is combined with the new observations estimated in the prediction phase to improve the system. The specific update phase algorithm is described as follows:

(2) Update phase
Take the feature point of number i as an example and suppose the iterative initial values are \( S_{k|i-1} \) and \( C_{k|i-1} \) respectively. The lth iterative robot position and square root factor are \( S_{k|i-1}^{(l)} \), \( C_{k|i-1}^{(l)} \). The observed value of time k is \( z_k^i \) and its observation model is

\[
\begin{align*}
D_{k|i-1}^{(l)} &= C_{k|i-1}^{(l)} S_{j}^i + S_{k|i-1}^{(l)}
\end{align*}
\] (13)

Calculate the lth iterative Kalman gain:

\[
\begin{align*}
B_{k|i-1}^{(l)} &= \frac{1}{2n} \sum_{j=1}^{2n} z_{k|i-1}^{(l)}
\end{align*}
\] (15)

\[
\begin{align*}
\begin{bmatrix}
A_{k|i-1}^{(l)}
\end{bmatrix} = B_{k|i-1}^{(l)}^T,
\end{align*}
\] (18)

Kalman gain:

\[
\begin{align*}
W_{k|i}^{(l)} &= \begin{bmatrix} P_{k|i-1}^{(l)} \end{bmatrix} \begin{bmatrix} P_{k|i-1}^{(l)} \end{bmatrix}^{-1}
\end{align*}
\] (19)

(3) The robot pose \( S_{k|i-1}^{(l+1)} \) and the square root factor \( C_{k|i-1}^{(l+1)} \) are calculated when iterating l+1 times.

- Set the condition of stopping iteration
- \( l = L_{\text{max}} \), (22)
- \( L_{\text{max}} \) is the maximum number of iterations, which is a fixed constant set in advance (22)
- Stop iteration and update each data:

\[
\begin{align*}
S_{k|i} &= S_{k|i-1}^{(l)}
\end{align*}
\] (23)

\[
\begin{align*}
C_{k|i} &= C_{k|i-1}^{(l)}
\end{align*}
\] (24)

Update the position

(25)

Repeat equations (12) to (25) when multiple feature points are observed

4 AN IMPROVED MULTI-ROBOT COOPERATIVE LOCATION BASED ON SQUARE ROOT CUBATURE KALMAN FILTER

DURING the movement of the robot, the position of \( R_j \) is observed by the robot \( R_i \) at time k, by means of obtaining the relative azimuth between them with the external sensors. As is shown in Fig. 1, \( \theta_i \) and \( \theta_j \) are the two robots' direction of movement respectively, and \( \phi_j \) is the relative azimuth angle of the robot \( R_a \) to the robot \( R_b \), equation can be expressed as:

\[
\phi_j = \arctan \left[ \frac{y_j^i - y_j^i}{x_j^i - x_j^i} \right] - \theta_j^i
\] (26)

The general form of observation model can be obtained:

\[
\begin{align*}
z_{k,i}^{ob} = h(S_{k|i}, S_{k|i}^+) + V_k
\end{align*}
\] (27)

where \( z_{k,i} \in R^{n_x} \) is the observation matrix of the system with -dimensional vector at time k, and is the observation matrix whose variance is and the matrix is subject to the Gaussian distribution of \( N(0, R_{k|i}) \).
Although the SR-CKF-SLAM algorithm can improve the positional accuracy and the system stability of the mobile robot, the dimension of the system observation will be added with the feature points on the map increasing gradually, which leads to the deviation of volume points from the ideal trajectory and therefore results in a large error in the state estimation. In this paper, a Strong Tracking Square Root Cubature Kalman Filter SLAM (STF-SRCKF-SLAM) algorithm is proposed to solve the divergence of the numerical calculation when the state is abruptly changed. The algorithm proposed can also correct the corresponding estimation bias, guarantee the tracking performance and improve the pose accuracy.

5 STRONG TRACKING FILTER SQUARE ROOT CUBATURE KALMAN FILTER SLAM (STF-SRCKF-SLAM)

5.1 Strong Tracking Filtering Algorithm

AIMING at the problem of filter divergence and numerical instability caused by model uncertainty or system mutation, the strong tracking filter algorithm improves the stability and accuracy of the system in map building by real-time online adjustment, increases the weight of new data, and weakens the influence of old data precision.

The time-varying fading factor in the strong tracking filter can be calculated as follows:

\[
V_k = \begin{cases} 
  e_k^T e_k, & k = 1 \\
  \rho V_{k-1} + e_k^T e_k, & k \geq 2 
\end{cases},
\]

(30)

where \( V_k \) is the residual covariance matrix, \( \rho \) is the forgetting factor, which is range from 0.95 to 0.98. \( e_k \) is the output residual sequence, which can be calculated by (5)

\[
e_k = x_k - \hat{x}_k,
\]

(31)

\[
N_k = V_k - H_k Q_{k-1} H_k^T - \beta R_k,
\]

(32)

\[
M_k = H_k F_{k|k-1} P_{k-1|k-1} F_{k|k-1}^T H_k^T,
\]

(33)

where \( F_{k|k-1} \) and \( H_k \) are the first-order partial derivative matrix of the state equation and the measurement equation to the state variable. \( \beta \) is a weakening factor whose value can be determined according to system change.

Time-varying fading factor can be obtained from (34):

\[
\]
\[
\lambda_k = \begin{cases} 
\lambda_0, & \lambda_0 > 1, \\
1, & \lambda_0 \leq 1,
\end{cases} 
\quad \tilde{\lambda}_k = \frac{\text{tr} \left[ N_k \right]}{\text{tr} \left[ M_k \right]}
\] (34)

Where \( \text{tr}(\cdot) \) indicates the trace of the matrix.

### 5.2 Tracking Iterative Square Root Cubature Kalman Filtering Algorithm

According to the prediction stage and the update stage, the state prediction covariance matrix \( P_{x,k}^{*} \) is obtained, the output prediction covariance matrix \( P_{z,k}^{*} \), the cross-covariance matrix \( P_{x,z,k}^{*} \) without introducing the fading factor. Equation can be obtained as follows:

\[
\lambda_k = \begin{cases} 
\lambda_0, & \lambda_0 > 1, \\
1, & \lambda_0 \leq 1,
\end{cases} 
\quad \tilde{\lambda}_k = \frac{\text{tr} \left[ N_k \right]}{\text{tr} \left[ M_k \right]}
\] (35)

\[
P_{x,k}^{*} = C_{x,k}^{*} \left( C_{x,k}^{*} \right)^T
\] (36)

\[
P_{z,k}^{*} = Z_{z,k}^{*} \left( Z_{z,k}^{*} \right)^T
\] (37)

Plug (10) - (12) into (9), equations can be obtained:

\[
N_k = V_k - Z_{z,k}^{*} \left( C_{x,k}^{*} \right)^T Q_{x,k} \left( C_{x,k}^{*} \right)^T Z_{z,k}^{*} - \beta R_k
\] (38)

\[
M_k = Z_{z,k}^{*} \left( Z_{z,k}^{*} \right)^T - V_k + N_k
\] (39)

\[
\lambda_k = \begin{cases} 
\lambda_0, & \lambda_0 > 1, \\
1, & \lambda_0 \leq 1,
\end{cases} 
\quad \tilde{\lambda}_k = \frac{\text{tr} \left[ V_k - Z_{z,k}^{*} \left( C_{x,k}^{*} \right)^T Q_{x,k} \left( C_{x,k}^{*} \right)^T Z_{z,k}^{*} - \beta R_k \right]}{\text{tr} \left[ Z_{z,k}^{*} \left( Z_{z,k}^{*} \right)^T - V_k + N_k \right]}
\] (40)

When the time-varying fading factor \( \lambda_k \) is introduced, the square root factor of the new state covariance prediction can be obtained as follows:

\[
C_{k+1} = \text{Tria} \left( \sqrt{P_{x,k}^{*} \cdot X_{k+1}^{*}}, C_{Q,k} \cdot 1 \right)
\] (42)

Where \( \text{Tria}(\cdot) \) denotes the triangular matrix obtained by QR decomposition, and \( C_{Q,k} \) is the square root factor of the variance \( Q_k \) of the motion noise.

After introducing the fading factor, the square root \( C_{k+1}^{*} \) of the prediction error variance matrix is calculated when updating Fig 1.

### 6 Simulation and Analysis of the Algorithm

The experiment to the accuracy of SLAM is applied in the Matlab7.0 software environment on the computer with the frequency of 3.4GHz, Core-i3 dual-core processor and 4G memory. On the open-source simulation platform provided by Australian scholar Tim Bailey, UKF-SLAM, SR-CKF-SLAM and ISR-CKF-SLAM are simulated respectively and they are also analyzed on the precision difference.

In the simulation experiment, the motion equation of the robot is:

\[
m_k = \begin{bmatrix} m_{x,k} \\
 m_{y,k} \\
 m_{\theta,k}
\end{bmatrix} = \begin{bmatrix} m_{x,k-1} + T \sin(\theta_{k-1}) \sin(\theta_k) \\
 m_{y,k-1} + T \sin(\theta_{k-1}) \cos(\theta_k) \\
 m_{\theta,k-1} + \beta \end{bmatrix}
\] (43)

where \( m_k \) represents the robot position, \( T \) represents internal sampling interval of the system, \( v_k \) represents the speed of the robot, \( \alpha_k \) represents the rotation angle, and \( L \) represents the wheelbase.

The robot observation equation can be obtained in the simulation experiment:

\[
z_i = \begin{bmatrix} l_i \\
 \beta_i
\end{bmatrix} = \begin{bmatrix} \sqrt{(m_{x,i} - m_{x,k})^2 + (y_{i} - m_{y,k})^2} \\
 \arctan \left( \frac{m_{y,i} - m_{y,k}}{m_{x,i} - m_{x,k}} \right)
\end{bmatrix}
\] (44)

\( l_k \) represents the distance between the robot and the observed feature points of the map, \( \beta_i \) represents the angle between the robot and the observed feature points of the map, and \( (m_{x,i}, m_{y,i}) \) is the position of the map's feature point observed by the robot.

The experimental hypothesis of the map environment is 250 × 200 outdoor environment with 17 identified path points and 35 map feature points. ISR-CKF-SLAM sets the number of iterations to five. The simulation results of the robot's trajectory are shown in Fig. 3, Fig. 4 and Fig. 5. The dotted line represents the ideal trajectory of the robot, the solid line represents the actual trajectory of the robot and the asterisk represents the map feature points. The simulation parameters can be seen in Table 1.
Table 1. Simulation Parameters.

<table>
<thead>
<tr>
<th>Simulation parameters</th>
<th>Value</th>
<th>Simulation parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed of the robot</td>
<td>3m/s</td>
<td>Wheel gauge</td>
<td>4m</td>
</tr>
<tr>
<td>Maximum steering angle</td>
<td>±30º</td>
<td>Control noise from speed</td>
<td>0.3m/s</td>
</tr>
<tr>
<td>Maximum steering angular speed</td>
<td>±20º</td>
<td>Observation noise from distance</td>
<td>0.1m</td>
</tr>
</tbody>
</table>

6.1 Simulation Algorithm of the Iterative Square Root Cubature Kalman Filter

Figure 3. UKF-SLAM.

Figure 4. SR-CKF-SLAM.

Figure 5. ISR-CKF-SLAM.

It is estimated that the whole path of the robot will take 240 seconds, and the average value will be selected after several experiments. In this paper, the first 180s of the simulation results are chosen to analyze.

In the initial stage of travel, the running trajectories deviate little from the ideal trajectories of the three algorithms. However, with the increasing number of the feature points, the trajectories of the three algorithms have different degrees of deviation from the ideal trajectories, but the ISRs of the three algorithms are different from the ideal trajectories. Compared with UKF-SLAM and SR-CKF-SLAM, ISR-CKF-SLAM is more suitable for the ideal trajectory, which means the pose estimation is more accurate.

From Fig 6, it is shown that the evaluated error of ISR-CKF-SLAM in X direction, Y direction and angle is always within a stable range. Compared with UKF-SLAM and SR-CKF-SLAM, ISR-CKF-SLAM maintains a high accuracy.

Performance differences of these three kinds of algorithms can be expressed by statistical data.

Compared with SR-CKF-SLAM, the error in X direction of ISR-CKF-SLAM is reduced by 46.8%, the error in Y direction of ISR-CKF-SLAM is reduced by 13.2% and the error in angle is reduced by 46.6% from the data in Table 2.
Table 2. Comparison in the Statistical Error of the Algorithm.

<table>
<thead>
<tr>
<th>SLAM algorithm</th>
<th>The average of the absolute value in the X direction's evaluated error /m</th>
<th>The average of the absolute value in the Y direction's evaluated error /m</th>
<th>The average of the absolute value in the angle error /rad</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKF</td>
<td>0.5773</td>
<td>0.5621</td>
<td>0.0348</td>
</tr>
<tr>
<td>SR-CKF-SLAM</td>
<td>0.4372</td>
<td>0.2436</td>
<td>0.0234</td>
</tr>
<tr>
<td>ISR-CKF-SLAM</td>
<td>0.2323</td>
<td>0.2113</td>
<td>0.0125</td>
</tr>
</tbody>
</table>

6.2 Simulation of Multi-robot Cooperative Localization on the Strong Tracking Filter Square Root Cubature Kalman Filter

The proposed STF-SRCKF-SLAM algorithm is simulated in a single-robot and a multi-robot cooperative environment and is compared with the average of 30 experimental results of UKF-SLAM and SR-CKF-SLAM, the simulation results of the robot's trajectory are shown in Fig. 7 and Fig. 8, in which "dotted line" represents the ideal trajectory of the robot, "solid line" represents the actual trajectory of the robot, "*" indicates the map feature point and "+" indicates the position of the feature points.

Fig. 7 and Fig. 8 correspond to the experimental results of STF-SRCKF-SLAM on single-robot and multi-robot respectively. Compared with STF-SRCKF-SLAM on a multi-robot, STF-SRCKF-SLAM on a single-robot has a better performance on the actual route than the ideal route, which means that the accuracy of location estimation is higher and the map created is more accurate.

6.2.1 Error Analysis

The comparison curves of X-axis, Y-axis and angle estimation errors are shown in the same environment when a multi-robot is mapping on the CKF-SLAM, SRCKF-SLAM and STF-SRCKF-SLAM, where t represents the experimental time.

As is shown in Fig. 9, the best stability of the three algorithms is the STC-SRCKF-SLAM algorithm, then the SRCKF-SLAM algorithm, and the CKF-SLAM algorithm has the worst performance of them. The error of the three algorithms in the X-axis, Y-axis and angular direction of the specific values can be viewed in Table 2.
From Table 3, the error of the CKF-SLAM is the biggest, and the error of SRCKF-SLAM is smaller than that of the CKF-SLAM, because the square root factor is introduced to guarantee the semi-positive definite of the covariance matrix, which can restrain the truncation error of the computer. The fading factor is introduced to adjust the gain online on the STF-SRCKF-SLAM and the effect of the algorithm can be seen in Fig.9 that the accuracy of the algorithm is high and the error is small, and it is not easy to generate divergence of numerical filtering, which verifies validity and accuracy of the STF-SRCKF-SLAM algorithm.

6.2.2 Root Mean Square Error and Run Time Analysis.

Root mean square error of the formula is;

\[
RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{x}_t)^2 + (y_t - \hat{y}_t)^2}, \tag{45}
\]

where T is the running time, and \((x_t, y_t)\) and \((\hat{x}_t, \hat{y}_t)\) are the actual position and the estimated position of the robot respectively at time k.

The RMSE and run-time of the three SLAM algorithms are showed in Table 4, and it can be seen from this table that the RMSE value of the STC-SRCKF-SLAM algorithm is the smallest, the estimation accuracy is the highest, and meanwhile the running time of STC-SRCKF-SLAM is the shortest.

Table 4. Experimental Statistics of the Three SLAM Algorithms.

<table>
<thead>
<tr>
<th>SLAM algorithm</th>
<th>Root mean square error /m</th>
<th>operation hours /s</th>
</tr>
</thead>
<tbody>
<tr>
<td>STC-SRCKF-SLAM</td>
<td>5.5880</td>
<td>159.5438</td>
</tr>
<tr>
<td>SRCKF-SLAM</td>
<td>6.5906</td>
<td>161.6320</td>
</tr>
<tr>
<td>CKF-SLAM</td>
<td>7.7630</td>
<td>163.7232</td>
</tr>
</tbody>
</table>

7 CONCLUSION

In this paper, an Iterative Squared Root Cubature Kalman Filter SLAM (ISR-CKF-SLAM) is proposed. The SR-CKF algorithm is improved by the Gauss-Newton iterative method. In this method, the latest updated information is efficiently used, which reduces the influence of initial error and linearization error on state estimation, and makes the posterior probability distribution closer to the true value. Compared with the UKF-SLAM and SR-CKF-SLAM, the ISR-CKF-SLAM is more accurate and more stable with the MATLAB simulation, which verifies the validity of the ISR-CKF-SLAM algorithm. A cooperative localization approach for multiple robots on the Strong Tracking Filter Square Root Cubature Kalman Filter SLAM proposed combines the advantages of the SRCKF and STF. On one hand, the integration of square root strategy can ensure the symmetry of the covariance matrix and effectively reduce the divergence caused by the system mutation filter. On the other hand, the weight of the corresponding data can be adjusted adaptively, and the accuracy of the system state estimation can be improved effectively by introducing the fading factor that can adjust the filter gain matrix in real time. It is the key problem to improve the real-time performance of the algorithm and reduce the influence of sensor observation error on the accuracy of the SLAM algorithm in large environment.

8 ACKNOWLEDGMENTS

This work was supported by 2018 Natural Science Foundation of Anhui, China (1808085QF215), 2018 Foundation for Distinguished Young Talents in Higher Education of Anhui, China (gxqyZD2018050) and Anhui Key Research and Development Programs (Foreign Scientific and Technological Cooperation, 1804b06020375).

9 REFERENCES


10 NOTES ON CONTRIBUTORS

Chen Mengyuan was born in 1984, and is an associate professor at the Anhui Polytechnic University. His main research filters include optimization of sensor information fusion and so on.
The Lateral Conflict Risk Assessment for Low-altitude Training Airspace using Weakly Supervised Learning Method

Kaijun Xu¹, Xueting Chen², Yusheng Yao¹, Shanshan Li¹

¹Institute of Flight Technology, Civil Aviation Flight University of China, Guanghan Sichuan 618307, China
²Institute of Foreign Languages, Civil Aviation Flight University of China, Guanghan Sichuan 618307, China

ABSTRACT

The lateral conflict risk assessment of low-altitude training airspace strategic planning, which is based on the TSE errors has always been a difficult task for training flight research. In order to effectively evaluate the safety interval and lateral collision risk in training airspace, in this paper, TSE error performance using a weakly supervised learning method was modelled. First, the lateral probability density function of TSE is given by using a multidimensional random variable covariance matrix, and the risk model of a training flight lateral collision based on TSE error is established. The lateral conflict risk in specific training airspace is analyzed, and then the lateral collision model is built. Through the quantification of the risk probability of lateral collisions, the security level of a specific airspace is evaluated. The analysis of the examples shows that for normal training flight in a variety of 4D flight track data, the lateral collision risk in specific training airspace is $5.43744 	imes 10^{-13}$, the conflict risk meets the requirement of safety target level of international civil aviation organization.

KEY WORDS: Air traffic control, Lateral conflict risk, Low-altitude complex flight, Standard safety interval, TSE

1 INTRODUCTION

CHINA is one of the fastest growing aviation industries in the world. At present, China is witnessing a strong demand in its air transport. As the development expands, it is estimated that by the end of the "Thirteenth-Five-year plan", the transport airports will reach 270 or so. There will be approximately 4600 cargo airplanes and 5,000 general planes, and a total of nearly 10,000 airplanes will be realized in the civil aviation fleet. This undoubtedly brings new opportunities and challenges to China’s civil aviation pilots training. With the rapid development of China’s civil aviation industry, the exposure of aircraft used in the pilots training in low-altitude complex flight conditions, combined with a large number of training hours and strict safety flight requirements, aircraft lateral conflict risk research has become particularly important.

The aircraft flight safety under complex conditions is always of intriguing interest as well as difficulty in international aviation research. In addition, the flight accident rate of aircraft under complex conditions is ten times more than that of a normal flight. The difficulty of realizing a safe flight in a low-altitude complex flight condition is that it is subjected to complex environmental factors such as; diverse terrain and obstacles, extreme weather and so on. Moreover, the high intensity and concentration of various training aircraft within a limited airspace leads to a more frequent occurrence of lateral collision risk for aircraft in the same altitude. Along with the development of air traffic management techniques and flight training techniques, the lateral error of aircraft, in particular the total systematic error (TSE) associated with the aircraft’s ability to maintain track and navigation positioning errors in training flights, is increased in importance for investigation. How training aircraft’s TSE error performance can ensure flight safety and achieve training results has become the hotspot for current air traffic control and flight training research.

At present, a lot of research work has been done on TSE lateral collision. British scholar (Reich P.G, 1966) put forward the aircraft longitudinal, lateral, vertical...
collision risk model, which is the early research of this field. (Russell A, 1996) (Heinz, 1996) of the NASA Ames Research Center look into accounts of the influence of various uncertainties and assumed, according to the central limit theorem, the position error of the track is a probabilistic ellipsoid. In the lateral conflict probability analysis, (Yang, et.al. 1997) analyzed the error distribution of each stage. Russell established a model of the probability-based position prediction error, and integrated the conflict domain to obtain the collision probability. (Hu, et.al. 2005) took the influence of airborne wind and other uncertain factors on flight into dynamic equation. (Yang, et.al. 2004) used the probability estimation method of aircraft intention information in the track prediction model. Chinese scholars have also carried out relevant researches. (Li, et.al. 2004) made preliminary research on the Reich model of colliding risk of parallel route and conducted modelling and relevant analysis. (Chen, et.al. 2002) (Wang, et.al. 2004) used a probability-based collision detection algorithm to analyze the collision probability based on Brownian motion, and obtained the effect of various factors on the probability of conflict. On this basis, the algorithm is extended to three dimensions, and the flow of practical application is given. (Zhao, et.al. 1998) studied the frequency of a dangerous collision of aircraft on two cross-routes, and proposed the concept of dangerous collision/collision zone. However, the research on the collision risk of TSE errors, especially in training flight, is still in the initial exploration stage. Previous studies only consider the effects of one or both of the track definition error (PDE), flight technology error (FTE) and navigation system error (NSE) (ICAO DOC 9689) (Zhao, et.al. 2016) (Zhu, et.al. 2014).

This paper is organized as follows. Section 2 presents an overview of collision risk modelling approaches. The risk of a lateral conflict under the low-altitude complex training flight condition and the TSE localization error is modelled and discussed in Section 3. Firstly, the TSE localization errors in the flight path data are analyzed, and the distribution function of the lateral localization error is given by the covariance matrix of the 3D random variables. In the specific analysis of the composition of TSE error, the probability of lateral overlap of the two planes in the training flight and the lateral conflict risk in the airspace of the low-altitude training are obtained in Section 4. The validity of the lateral flight conflict model for the risk assessment in low-altitude training airspace strategic planning is verified by examples. Finally, results and conclusions are given in Section 5.

2 RISK MODELLING APPROACHES

2.1 Overview of Risk Modelling Approaches

ONE of the principal matters of concern in the daily operation of civil aviation is prevention of conflicts between aircraft, either while airborne or on the ground, which might escalate to collision. Although aircraft collisions have actually been very rare events, contributing to a very small proportion of the total fatalities, they have always caused relatively strong impact mainly due to relatively large number of fatalities per single event and occasionally the complete destruction of the aircraft involved.

The main driving force for developing risk methods/models during the 1960s was the need for increasing airspace capacity over the Atlantic through decreasing aircraft separation minima. In general, separating aircraft using space and time separation standards (minima) has prevented conflicts and collisions. However, due to the reduction of this separation, in order to increase capacity and thus cope with growing air transport demand, assessment of the risk of conflicts and collisions under such conditions has been investigated using several important methods/models. The methods/models were expected to show if a reduction of separation and spacing between flight tracks would be sufficiently safe, i.e., determine the appropriate spacing between tracks guaranteeing a given level of safety.

• The Reich–Marks model was developed in the early 1960s by the Royal Aircraft Establishment, UK (Reich, 1966). It is based on the assumption that there are random deviations of both aircraft positions and speeds from the expected. The model was developed to estimate the collision risk for flights over the North Atlantic and consequently to specify appropriate separation rules for the flight trajectories (Shortle et al., 2004). The model computed the probability of aircraft proximity and the conditional probability of collision given the proximity (Machol, 1995; FAA/EUROCONTROL, 1998).

• The Machol–Reich model was developed after the ICAO had established the NAT SPG (North Atlantic System Planning Group) in 1966 with the idea of developing the Reich-Marks model as a workable tool, as well as to increase of airspace capacity. Consequently, the ICAO NAT SPG adopted the threshold for risk of collision of two aircraft due to the loss of planned separation (Machol, 1975, 1995).

• Intersection models belong to the simplest collision risk models. They are based on the assumptions that aircraft follow pre-determined crossing trajectories at constant speeds. The probability of a collision at the crossing point is computed using the intensities of traffic flows on
each trajectory, aircraft speeds, and airways geometry (Siddiquee, 1973; Schmidt, 1977; Hsu, 1981; Geisinger, 1985; Barnett, 2000).

- Geometric conflict models are similar to intersection models. In these models (developed in 1990s) the speed of any two aircraft is constant, but their initial three-dimensional positions are random. The conflict occurs when two aircraft are closer than the prescribed separation rules (Paielli and Erzberger, 1997, 1999; Irvine, 2002; Alam et al., 2009; Chaloulos et al., 2010; Oscar et al., 2014).

- The generalized Reich model was developed by removing restrictive assumptions from the Reich model based on the fact that Reich model does not adequately cover certain real air traffic situations. Such a generalized collision model was developed during 1990s and has been in use as part of the TOPAZ (Traffic Optimization and Perturbation AnalyZer) methodology (Blom et al., 1998, 2003; Shortle et al., 2004; Blom and Bakker, 2002).

Collision risk methods/models have gradually been developed from Marks, Reich and Machol to the latest versions used in the TOPAZ methodology. The main purpose has always remained to support decision-making processes during system planning and development, through evaluation of the risk and safety of the proposed changes (either in the existing or the new system).

2.2 Risk Modelling Approach Accepted by ICAO

The International Civil Aviation Organization (ICAO) has developed the Collision Risk Model (CRM) as a mathematical tool used in predicting the risk of mid-air collision. During development of CRM ICAO adopted the Reich (1966) and Hsu (1981) formulae (ICAO, 1998, 2002, 2009) and further defined a unified framework for derivation of collision risk models, called the Rice formula (Mehadhebi and Lazaud, 2004; ICAO, 2009). From the Rice formula it is possible to derive the Reich and the Hsu formulae (Mehadhebi and Lazaud, 2004).

ICAO has adopted the CRM model as a crucial part of the Airspace Planning Methodology for determination of separation criteria (ICAO, 1998). The main purpose of the methodology is determination of separation minima based on collision risk modelling. Calculated collision probability based on CRM is used for comparison with the reference system (if it exists) or for evaluation of the system risk against a threshold, the so-called Target Level of Safety (TLS).

According to ICAO Circular 319 (2009), “the purpose of collision risk models in the context of the determination of separation minima is to model the chain of events leading a pair of initially separated aircraft to a collision”. ICAO CRM calculates probability of collision as the lateral or vertical overlap probability, given probability density functions of position errors at a given moment (where position error depend on path definition error, flight technical error, navigation system error and surveillance error) (Mehadhebi and Lazaud, 2004; Fujita, 2009). However, Mehadhebi (2007) pointed out that CRM’s from ICAO Doc. 9689 (1998) are not sufficient for modelling all situations, especially operational errors.

3 FLIGHT CHARACTERISTICS AND LATERAL CONFLICT MODEL USING WEAKLY SUPERVISED LEARNING METHOD

At present, in monitoring flight training, radar monitoring technology is widely used as the core scheme, and the collision detection is carried out for the low-altitude objective safety factors between the aircraft, the aircraft and the terrain, the aircraft and the meteorology. Whether the total system error (TSE), associated with the aircraft’s ability to maintain track and navigation location errors in training flight, can ensure flight safety and achieve training effectiveness has become a hot topic in field of air traffic control and flight training.

In addition, civil aviation flight training has its unique characteristics. Pilots should perform the training with primary, intermediate or advanced trainer aircraft, in designated training airspace and at the designated training altitude, and in accordance with the pre-established training flight plan. Training airspace and height are all restricted to low-altitude, while subjects for training are all types of trainer aircraft, which are heterogeneous aircraft because of their diverse flight performances. Furthermore, pilots’ lack of relevant flight skills and experiences brings many uncertainties.

![Figure 1. Lateral TSE Composition.](image)

3.1 Analysis of Lateral Conflict Factors

Aircraft lateral position error is related to the aircraft’s ability to maintain track and navigation positioning errors and other factors, represented by the total system error (TSE). TSE contains three aspects of error, namely, track definition error (PDE), flight
technology error (FTE) and navigation system error (NSE), as shown in Figure 1.

Since the three errors are vector-distributed in space, TSE is the vector sum of these three errors, and the standard deviation of TSE is the square root of the sum of squares of the standard deviations of these three errors

\[ TSE = \sqrt{PDE^2 + FTE^2 + NSE^2}, \]

(1)

Because of the influence of a variety of factors, aircraft flying in the training airspace may be unable to fly along the scheduled training routes, thus leads to yaw. Among the many factors that affect collision risk, the collision risk is a function of its TSE error, and the acceptability of the risk depends on the safety target level (TLS) of the specific airspace. When the TLS is determined, the minimum TSE error criterion for a particular airspace is determined. The training mission must meet its minimum standards to ensure flight safety.

3.2 TSE Lateral Conflict Model

The flight positioning error caused by PDE, FTE, and NSE error performance is one of the main reasons for the yawing of the training aircraft flying under the TSE error performance environment. Lateral yaw occurs most frequently, but is most likely to be overlooked by pilots. As the flight time increases, if the two planes flying in the specific training space at the same altitude don’t adjust their lateral yaw, and after a certain period of time, the lateral separation between them will be less than the horizontal safe separation, and the risk of a lateral collision will occur, as shown in Figure 2.

Supposing the following conditions are satisfied for the lateral collision model; two aircraft fly in the same training space and at the same altitude; the TSE error performance of the two aircraft is independent, that is, they are independent in the lateral, longitudinal and vertical directions; the yawing speed of the aircraft is independent of its position, and without considering the impact of ground control on the risk of collision.

\[ f(X_{PDE}, X_{FTE}, X_{NSE}) = \frac{1}{(2\pi)^{3/2} [\det(C)]^{1/2}} \exp\left[-\frac{1}{2} (X - \mu)^T C^{-1} (X - \mu)\right] \]

(5)

The probability density function of the three-dimensional random variable \( M \) is

\[ X = \begin{bmatrix} X_{PDE} \\ X_{FTE} \\ X_{NSE} \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \]

(6)

\( \det(C) \) is the determinant of \( C \), \( C^{-1} \) is the inverse matrix of \( C \).

This paper only considers the lateral error caused by the total system error (TSE). If the lateral yawing distance caused by TSE is more than half of the lateral spacing standard \( A \), the two aircraft will overlap laterally. Calculation shows the probability of lateral overlap of two aircraft:
According to Bibliography (Zhao, et.al. 2016), it can be known that the lateral conflict risk of two aircraft is

\[ C_{\text{risk}} = 2NP \]  

(8)

Where \( N \) is the number of studied aircraft in the low-altitude airspace.

At this time, the aircraft’s risk of lateral collision in low-altitude airspace is

\[ C_{\text{risk}} = 2NP \int \int \int f(X_{\text{PDE}}, X_{\text{FTE}}, X_{\text{NSE}}) \, dX_{\text{PDE}}dX_{\text{FTE}}dX_{\text{NSE}} \]  

(9)

In the TSE components, PDE can be reduced by high-precision measurements. Without compromising the generality of the PDE, constraints and control of the PDE are mainly realized through measurements of the coordinates of key points, such as obstacles around airport terminal area and entrances of runway center line, which is achieved by using high-precision measurement equipment based on WGS-84 system before designing the flight program. If PDE has been measured, low-level flight of two aircraft lateral error discussion can usually ignore PDE. Thus, TSE is controlled by the flight technology error FTE and the navigation system error NES.

When ignoring PDE, FTE and NSE should be constrained respectively to control the TSE.

The same assumption is made; \( X_{\text{FTE}}, X_{\text{NSE}} \) obey normal distribution.

\[ X_{\text{FTE}} \sim N(\mu_2, \sigma_2^2) \]
\[ X_{\text{NSE}} \sim N(\mu_3, \sigma_3^2) \]  

(10)

In general, the lateral positioning errors caused by \( X_{\text{FTE}}, X_{\text{NSE}} \) are interrelated and the correlation coefficient \( \rho_{23} \) represents the correlation coefficient between the flight technology error FTE and the navigation system error NES.

The covariance matrix of the two-dimensional random variable \( M' \) is

\[ C = \begin{bmatrix} \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix} \]  

(11)

\[ M' = (X_{\text{FTE}}, X_{\text{NSE}}) \]  

(12)

\[ C^{-1} \] is the inverse matrix of \( C \)

\[ \rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{E(XY)-E(X)E(Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \]  

(14)

The probability density function of the two-dimensional random variable \( M' \) is

\[ f(X_{\text{FTE}}, X_{\text{NSE}}) = \frac{1}{2\pi|\text{det}(C)|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(X - \mu)^T C^{-1}(X - \mu)\right] \]  

(15)

The probability of lateral overlap of two aircraft is

\[ P = \int \int f(X_{\text{FTE}}, X_{\text{NSE}}) \, dX_{\text{FTE}}dX_{\text{NSE}} \]  

(16)

Similarly, the lateral conflict risk of two aircraft is

\[ C_{\text{risk}} = 2NP \]  

(17)

where \( N \) is the studied number of aircraft in the low-altitude airspace.

At this time the lateral conflict risk of airplanes in low-altitude airspace is

\[ C_{\text{risk}} = 2NP \int \int f(X_{\text{FTE}}, X_{\text{NSE}}) \, dX_{\text{FTE}}dX_{\text{NSE}} \]  

(18)

4 ANALYSIS OF THE FLIGHT LATERAL COLLISIONS BASED ON TSE ERROR

It analyzes the ADS-B/GPS flight data of the track of circle procedure and eight character procedures of a Cessna 172G1000 primary trainer aircraft in local training. (As shown in Fig. 3, 4 & 5).
Figure 3. Cessna 172G1000 Primary Training Aircraft.

Figure 4. ADS-B Display Information of the Track of Circle and Eight Character Procedures.

Figure 5. GPS Track Point Flight Data of the Track of Circle Procedure and Eight Character Procedure.

Through the long-term training flight observation, the statistical analysis of the 1254 measurements of the track of circle procedure and 854 measurements of eight character procedures as shown in Fig. 6 & 7.

Figure 6. Circle Procedure Fly Track in 4D.

\[ X_{FTE} \text{ and } X_{NSE} \text{ obeys normal distribution, and the error distribution is as follows.} \]
\[ X_{FTE} \sim N(7.1684, 4.9102^2) \]
\[ X_{NSE} \sim N(0, 4^2) \]  
\[ (19) \]

Correlation coefficient \( \rho_{23} = 0.5 \)

The covariance matrix of the two-dimensional random variable \( \mathbf{M}' \) is
\[ C = \begin{bmatrix} 24.1101 & 9.8204 \\ 9.8204 & 16 \end{bmatrix} \]  
\[ (20) \]

\[ \mathbf{M}' = (X_{FTE}, X_{NSE}), \]  
\[ (21) \]

\[ C^{-1} = \begin{bmatrix} 0.0553 & -0.0339 \\ -0.0339 & 0.0833 \end{bmatrix} \]  
\[ (22) \]

\[ \det(C) = 289.3213 \]  
\[ (23) \]

The probability density function of the two-dimensional random variable \( \mathbf{M}' \) is (as shown in Fig. 8)
\[ f(X_{FTE}, X_{NSE}) = \frac{1}{2\pi[289.3213]^{1/2}} \exp\left[-\frac{1}{2}(X - \mu)^T C^{-1}(X - \mu)\right] \]  
\[ (24) \]

Figure 7. Eight Characters Procedure Fly Track in 4D.

Figure 8. FTE, NSE Probability Density Function of Long-term Training Flight of a Flight Training Institution.
In the local training of a flight training institution, there are 32 aircraft in one flight brigade to perform training tasks, and the number of aircraft \( N = 16 \). Clearance time interval controlled by ground controllers is one minute to one and a half minutes, and the ground monitoring training terminal area is \( 30 \times 30 \text{km}^2 \). Cruise speed for Cessna 172 is \( 80 \text{kt} \), and safety separation for two aircraft is \( 2.5 \text{km} \). When a single aircraft separation is less than \( 1.25 \text{km} \), it can be regarded as violating the minimum safe separation.

Then the probability of lateral collision of two aircraft is
\[
P = \int_{20.25}^{24} \int_{23.5}^{26} f(X_{FTE}, X_{NSE}) dX_{FTE} dX_{NSE} = 1.6992 \times 10^{-15} \quad (25)
\]

The lateral conflict risk of two aircraft is
\[
C_{\text{risk}} = 2NP = 2N \int \int_{-\infty}^{\infty} f(X_{FTE}, X_{NSE}) dX_{FTE} dX_{NSE} = 5.43744 \times 10^{-14} \quad (26)
\]

Comparing the risk value of the two-plane collision risk based on the TSE error with the predetermined safety target level (the ICAO target safety level is \( 5.0 \times 10^{-9} \)), it can be obtained that the current flight training safety level, which is under the lateral safety clearance standards and the existing TSE performance, meets the pre-defined safety target level requirements.

5 CONCLUSIONS

Under low-altitude complex training flight, the total system error TSE has seriously affected keeping safe separation and the reliability and safety of flight situation monitoring, and has become one of the main reasons for accidents.

This paper discusses flight lateral conflict risk assessment for low-altitude training airspace using a weakly supervised learning method by applying TSE error performance. First, a risk model of flight lateral conflict, based on TSE error, is established by covariance matrix of multidimensional random variables, and through calculating the collision risk of the planes with same directions in the same training space is obtained. Comparing it with ICAO’s predetermined safety target level, a safety assessment of the training airspace can be carried out. Then, in the analysis of examples, the collision risk probability derived from the TSE models is compared with the ICAO’s predetermined safety objective, which further improves the training airspace safety assessment. This method can make a scientific assessment of the safety level of the low-altitude training airspace or busy terminal area. It can also verify whether FTE, NSE error level meet the safety requirements.

Since the current study on FTE data statistics and reasons is still incomplete, more in-depth research is required in this field. In conclusion, the lateral collision models established in this paper lay a solid foundation for further study of collision risk at low-altitude complex flight conditions.

6 ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China Civil Aviation Joint Foundation (GrantNo.U1533127), and Open Foundation of Science and Technology Department of Sichuan Province (GrantNo.2015GZ0307). Work partially supported by the Open Foundation of Civil Aviation Flight University of China (GrantNo.J2014-03) and Foundation of Scientific research innovation team in Civil Aviation Flight University of China (GrantNo.JG2016-26).

7 REFERENCES


Chun-jin Li, Ying-xun Wang. The mathematic model of the collision risk in parallel airway. The Air Traffic Control Research Center of Beihang University.


8 NOTES AND PHOTOS OF CONTRIBUTORS

**Kaijun Xu** received Ph.D. degree in Traffic Information Engineering and Control at the School of Information Science and Technology, Southwest Jiaotong University. He works at the Institute of Flight Technology, Civil Aviation Flight University of China (CAFUC). He also works at Beihang University as Post-doctoral. He is an associate professor in CAFUC and his research interests are cognitive affective computing, air traffic management, modern navigation technology. He has published more than thirty papers in international Journals, core Journals in China and international meeting, more than twenty papers are indexed by SCI and EI.

**Xueting Chen** received Master’s degree in English Language and Literature at Sichuan University. She works at the School of Foreign Languages, Civil Aviation Flight University of China. She has published articles concerning the English learning motives of pilot students and the plays of Tennessee Williams.

E-mail: cxtcheng@126.com, TEL: 15883645685

**Yusheng Yao** is a master’s degree candidate at the Institute of Flight Technology, Civil Aviation Flight University of China. His research interest is theory and method of modern navigation.

E-mail: y_s_yao@163.com, TEL:18227109562

**Shanshan Li** is a master’s degree candidate at the Institute of Flight Technology, Civil Aviation Flight University of China. Her research interest is theory and method of modern navigation.

E-mail: feilin_li@163.com, TEL:15283829022
Simulation and Data Analysis of Energy Recovery Sensing on a Parallel Hydraulic Hybrid Crane

Youquan Chen, Xinhui Liu, Xin Wang, Jinshi Chen

School of Mechanical Science and Engineering, Jilin University, Changchun 130022, China

ABSTRACT
In order to study the braking energy regeneration characteristics of the Front-mounted Parallel Hydraulic Hybrid Crane (FPHHC), the AMESim simulation models are established and analyzed by establishing the vehicle dynamics model and referencing to the actual data of the crane and physical hydraulic components, the simulation results are verified by road tests on the experimental prototype. The experiment results basically match with the simulation results. In the vehicle braking process, the hydraulic hybrid system of the experimental prototype can effectively recycle the vehicle braking energy, the energy recovery rate is up to 50.84%, and the energy-saving effect is obvious. The performance in terms of the vehicle acceleration and braking has obviously been improved, then the prototype crane, which can show that this system has a certain practical significance in saving energy.

KEY WORDS: AMESim simulation, Braking energy recovery, Hydraulic hybrid, Road tests

1. INTRODUCTION
In recent years, with the rapid development of worldwide industrial production, transportation and other industries, the globalization of market competition is increasingly fierce, the problems of energy shortage, environmental pollution, and so on, make people think about a vehicle's fuel economy and emissions, Clarke P and Yang, et. al. (2017) has proposed that it makes the automobile industry head to the direction from the type of traditional pure fuel to the type of clean energy eventually. Simon B, et. al. (2007) and Li Jun-cheng et. al. (2011) consider that the hydraulic hybrid technology is an important branch of hybrid species; Offer G J et. al. (2010), Liu Xin-hui (2010), Zeng Zi-wei (2012) and Fu Song-biao et. al. (2010) think that for the hydraulic hybrid's high energy utilization ratio and large power density, environmental protection and other advantages, Zheng, et. al. (2012) have proposed that it is especially suitable for the need of frequent start-stop on large-tonnage truck, engineering vehicles, city buses and other heavy vehicles. According to the energy flow and power flow configuration relations of the driving system, Liu Tao, et. al. (2010) has proposed that the hybrid powertrain can be divided into three kinds of series, parallel and mixed type. Among them, vehicles can be independently driven by the engine or the hydraulic secondary components of the parallel hybrid system, Sun, et. al. (2009, 2010) and Zhang Qing-yong, et. al. (2011) have proposed that making it convenient to refit vehicles based on existing models.

Considering the characteristics of the hydraulic hybrid vehicle configuration synthetically, based on the FPHHC, the vehicle model is established by AMESim software and the simulation analysis is done to study the effect of energy recovery and utilization, and the validity of the simulation results is verified on experimental prototype by road test.

2. WORKING PRINCIPLE AND MODEL OF PARALLEL HYDRAULIC HYBRID VEHICLE

2.1. Composition and Working Principle of Biaxial Parallel Hydraulic Hybrid System
The BIAXIAL parallel hydraulic hybrid system is composed of the traditional internal combustion engine system and hydraulic secondary component, which are connected together by a torque coupler. J.E. Naranjo, et. al. (2012) has proposed that the traditional internal combustion engine system of components include the engine, clutch, gearbox, drive axle, the accelerator pedal and brake pedal. The hybrid system is composed of the hydraulic pump/motor, hydraulic
accumulator, torque coupler and the corresponding hybrid controller and so on. Based on the traditional rear-mounted parallel hybrid vehicle (RPHHV) structure (as shown in Figure 1), the front-mounted parallel hybrid vehicle layout mode (FPHHV) is put forward (as shown in Figure 2), the torque coupling device is put between the engine and the gearbox, the secondary component speed range can more match the range of engine speed, that can make the engine work more in the high efficient area, and then improve the fuel economy and power performance of the vehicle.

**Figure 1.** Structure Diagram of RPHHV.

**Figure 2.** Structure Diagram of FPHHV.

FU, and Zhang, et.al. (2010, 2111) have proposed that the key technology of a hydraulic hybrid system is the interconversion between pump and hydraulic motor by adjusting the work quadrant of the hydraulic secondary component. In the stage of vehicle accelerating, the engine maintains the idle running, while the hydraulic pump/motor works in the condition of motor, the hydraulic accumulator release energy by a hydraulic secondary component, the hydraulic hybrid system drives the vehicle as the main source of power. When the hydraulic accumulator releases the energy completely or the vehicle achieves a certain speed, the vehicles engine begins to work, at the same time, the hydraulic hybrid system separates from the vehicle transmission system and the vehicle is not driven.

### 2.2 Mathematical Model of the Working Condition of Vehicle Driving and Braking

The vehicle is driven by driving forces and various resistance forces in the driving conditions. Ding Zuo-wo, Zhao Dong-biao. (2007) have proposed that the driving forces mainly include the forces provided by the engine and the hydraulic hybrid system. Resistance mainly includes; rolling resistance, air resistance, slope resistance and acceleration resistance. In the process of driving, the dynamic equation of the hydraulic hybrid vehicles is the same with the traditional vehicle, assuming that the vehicle traveling track is a straight line, and then the balance equation in the process of driving is as follows:

\[ F_{dr} - F_f = m \frac{dv}{dt} \]  \hspace{1cm} (1)

Where $F_{dr}$ is the vehicle driving force (N); $F_f$ is the total resistance in the process of vehicle travelling (N); $V$ is vehicle instantaneous velocity ($m/s$).

\[ F_f = F_{air} + F_r + F_g \]  \hspace{1cm} (2)

Where, $F_{air}$ is air resistance (N); $F_r$ is rolling resistance (N); $F_g$ is slope resistance (N). Rolling resistance is determined by the total mass of the vehicle $m$ (kg) and the coefficient of wheel rolling friction $f$, usually take $f = 0.02$.

\[ F_r = mg \cdot f \]  \hspace{1cm} (3)

The air resistance equation is

\[ F_{air} = C_d A \frac{v^2}{21.15} \]  \hspace{1cm} (4)

Where, $C_d$ is air drag coefficient; $A$ is the area of vehicle windward. The slope resistance equation is

\[ F_g = mg \cdot \sin \alpha \]  \hspace{1cm} (5)

The acceleration resistance equation is

\[ F_a = m \cdot \frac{dv}{dt} \]  \hspace{1cm} (6)

The vehicle driving force equation is

\[ F_{dr} = F_e + F_{hb/M} \]  \hspace{1cm} (7)

Where, $F_e$ is the driving force provided by engine (N); $F_{hb/M}$ is the driving force regenerated by hydraulic system (N).

The driving force provided by engine is

\[ F_e = \frac{T_e i_o \eta_i}{r} \]  \hspace{1cm} (8)

Where, $T_e$ is engine torque (Nm); $\eta_i$ is engine transmission efficiency. Then the total force balance equation in the working condition of driving is

\[ \frac{T_e i_o \eta_i}{r} + F_{hb/M} - C_d A \frac{v^2}{21.15} - mgf - mg \sin \alpha = m \frac{dv}{dt} \]  \hspace{1cm} (9)

The main types of forces of the vehicle in braking condition are; air drag, rolling resistance, gradient resistance, engine anti-drag resistance and wheel
braking force. Then the force equation of equilibrium in the braking condition of vehicle is

$$ F_b + F_{air} + F_s + F_g + F_{eb} = m \frac{dv}{dt} $$  \hspace{1cm} (10)

Where, \( F_b \) is wheel braking force (N), \( F_{air} \) is engine anti-drag resistance (N). Whereby, the wheel braking force includes the hydraulic regenerative braking force provided by the hydraulic hybrid system, and the friction braking force of the original vehicle provided by its braking system, whereby, the wheel braking force equation is

$$ F_b = F_{hb/p} + F_s $$  \hspace{1cm} (11)

Where, \( F_{hb/p} \) is hydraulic regenerative braking force; \( F_s \) is friction braking force.

The engine anti-drag resistance equation (Yu, et.al. 2009) is

$$ F_{eb} = \frac{I_e \omega_e^2}{r^2 \eta_e} \frac{dv}{dt} $$  \hspace{1cm} (12)

Where, \( I_e \) is rotational inertia of engine flywheel (kg\cdotm²). Then the total force balance equation in the working condition of braking is

$$ F_{hb/p} + F_s = \frac{C_v A_v^2}{21.15} \cdot mgf + mg \sin \alpha + \frac{I_e \omega_e^2}{r \eta_e} \frac{dv}{dt} - m \frac{dv}{dt} $$  \hspace{1cm} (13)

2.3. Main Parameters Design of the Hydraulic System

The displacement of hydraulic pump/motor is

$$ V_s = \frac{X_p}{X_p^{max}} V_{s^{max}} $$  \hspace{1cm} (14)

Where, \( V_s \) is the displacement of hydraulic pump/motor (m³/rad); \( X_p^{max} \) is the maximal one-way displacement of the variable hydraulic cylinder (m); \( V_{s^{max}} \) is hydraulic pump/motor maximum displacement (m³/rad).

When the hydraulic pump/motor is operated in the hydraulic pump state, its driving force comes from the load. In the process of braking, the braking torque of hydraulic pump/motor is

$$ T_p = \frac{P V_s}{2\pi} $$  \hspace{1cm} (15)

Where, \( P \) is the pressure of hydraulic accumulator (MPa).

The drive balance equation of hydraulic pump/motor in pump working conditions is

$$ T_{p/m} = \frac{F \cdot r}{i_b \eta p} $$  \hspace{1cm} (16)

Where, \( T_{p/m} \) is output torque of the hydraulic pump/motor (Nm); \( \eta_p \) is system efficiency; \( i_b \) is speed ratio of torque coupler, it should ensure that the secondary component runs in its specified speed range and as much as possible works in the efficient area in the braking and driving process;

$$ i_b = \frac{0.377 r 0 p m}{v_{avg}} $$, \( i_b \) is the gear ratio of the main reducer, \( r \) is wheel rolling radius (m). Then when the hydraulic pump/motor works in pump working condition, the wheel braking force is

$$ F_{hb/p} = \frac{T_p i_p M i_g i_m \eta p \eta M}{r \eta m} $$  \hspace{1cm} (17)

When the hydraulic pump/motor works during the motor working condition, the power comes from the oil pressure in accumulator. In the process of driving, the output torque of hydraulic pump/motor is

$$ T_M = \frac{P V_s}{2} $$  \hspace{1cm} (18)

Then in the working condition of hydraulic motor, the driving force of wheel is

$$ F_{hb/M} = \frac{T_M i_p M i_g i_m \eta m \eta M}{r} $$  \hspace{1cm} (19)

Assuming that the vehicle travels on a flat and straight road, when the vehicle brakes, the energy equilibrium equation is

$$ E_r = E_{fr} + E_{acc} + E_{air-res} $$  \hspace{1cm} (20)

Where, \( E_r \) is vehicle kinetic energy (J); \( E_{fr} \) is the loss of energy when the wheel overcome the rolling friction (J); \( E_{acc} \) is the energy of the accumulator recycling (J); \( E_{air-res} \) is the loss of energy when the vehicle overcomes the air resistance (J).

$$ E_r = \frac{1}{2} m (v_0^2 - v_1^2) $$  \hspace{1cm} (21)

Where, \( v_0 \) and \( v_1 \) respectively correspond to the speed of the vehicle at the beginning of the brake and brake speed after time \( t \).

$$ E_{fr} = f G s = f m g s $$  \hspace{1cm} (22)

Where, \( S \) is the displacement in the process of vehicles braking (m).
\[ \sum E_{\text{ac}} = -\int_{v_0}^{v_1} p dV = \frac{p V_1}{n-1} (V_2 - V_1) + \frac{p V_2}{n-1} \left[ \frac{V_2^\gamma}{p_2} - 1 \right] \]  
(23)

Where, \( p_1 \) is the minimum working pressure (MPa); \( p_2 \) is the maximum working pressure (MPa); \( V_1 \) is the gas volume when working pressure is \( p_1 \), \( V_2 \) is the gas volume when working pressure is \( p_2 \), \( n \) is air polytropic exponent (Ding, et al. 2007)

\[ E_{\text{air-res}} = \int_0^t \frac{C_D A ds}{21.5} = \frac{C_D A}{21.5 \times 4a} \left[ v_0^2 - (v_0 - \alpha t)^4 \right] \]  
(24)

Where: \( t \) is braking time (s); \( \alpha \) is vehicle braking deceleration (m/s²).

The recovery rate of braking energy is defined as

\[ \varepsilon_r = \frac{\sum E_{\text{ac}}}{\sum E_r} = \frac{\sum E_{\text{ac}}}{n-1} \]  
(25)

The reuse rate of braking energy is defined as

\[ \varepsilon_u = \frac{\frac{1}{2} m v_2^2}{\frac{1}{2} m v_0^2} = \frac{v_2^2}{v_0^2} \]  
(26)

Where, \( v_2 \) is the highest speed that the vehicle can reach when it is driven forward along the hydraulic hybrid system after the recovery of energy is completed.

### 3. AMESIM SIMULATION MODE OF THE CRANE

In order to verify the fuel economy of the front-mounted parallel hydraulic hybrid crane under complex conditions, Fu Yong-ling, Qi Xiao-ye (2011) have proposed that the vehicles backward simulation, according to the model is set up in the AMESim simulation environment. The drive cycle is inputted by means of IFP Drive. In the vehicle driving cycle, the drive and the brake torque needed by the vehicle is calculated by a different gear, and the control signal is inputted to the engine, braking system of the original vehicle and hydraulic hybrid system through the controller. According to the principle of the system and the models of main components, as shown in Figure 3, the vehicle backward simulation model of AMESim is established, which mainly includes the control module, the switch gear and hydraulic system module. Some specific parameters of the simulation model are listed in Table 1.

### 4. ANALYSIS OF SIMULATION RESULTS

An integrated UDDS driving cycle is chosen and used as a simulation condition, which is put forward by the environmental protection agency (EPA), and used first for testing the fuel economy of medium and heavy trucks in the United States. The simulation time is 1500s; the simulation step size is 0.01s. When a vehicle is in a driving condition of the UDDS, the change of vehicle speed is shown in Figure 4, you can clearly see that whether driving or braking of the vehicle are all very frequent.

The pressure of the hydraulic accumulator and displacement change of the secondary component is shown respectively in Figure 5 and Figure 6. Under the simulation condition of the UDDS, the hydraulic hybrid system can effectively recycle and reuse the vehicle’s braking energy, and the hydraulic secondary components quickly and effectively switch between the working conditions of pump and motor.

As seen from Figure 7, in a complete UDDS condition, the output power of a traditional vehicle engine is generally greater than the one of hybrid vehicle engine. A hybrid vehicle starts and stops frequently on a city road, by regulation of the hydraulic secondary component, the hybrid system realizes the recycling and release of power, which can effectively reduce the power output and fuel consumption of the engine.

As shown in Figure 8, in the phase from 344s to 360s, while the vehicle starts from static, the hydraulic hybrid system has the effect of auxiliary driving, which makes the engine effectively avoid the inefficient zone of low speed and large torque.

In order to clearly verify the ascension of a vehicles dynamic performance, climbing ability and starting acceleration capability, the vehicle is simulated and analyzed respectively. The slope of simulation is set at 35%, the accumulator outlet pressure is set as 30MPa; that is, and the accumulator has been filled and has reached the maximum pressure acquiescently. As seen in Figure 9, the speed of the hydraulic hybrid vehicle has improved significantly and the hydraulic hybrid vehicle can climb while the traditional vehicle’s speed is 0 all the time and it can’t climb. From Figure 10, it can clearly be seen that the hybrid vehicle can accelerate faster than the traditional vehicle; it is almost twice as fast as the traditional vehicle. The simulation results show that the front-mounted parallel hybrid structure can significantly improve the vehicle’s dynamic performance. The comparison of the simulation results of the engine power and fuel consumption of the prototype vehicle, which does not add the hydraulic regenerative system and FPHHV, is shown in Figure 11(a) and (b).
Figure 3. Simulation Model of the FPHHC Based on AMESim.

Table 1. Selection and Parameters of the Main Sub-model of a Simulation Model.

<table>
<thead>
<tr>
<th>Components name</th>
<th>Parameters name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple-disc clutch</td>
<td>Maximum friction torque(Nm)</td>
<td>104</td>
</tr>
<tr>
<td>Safety valve</td>
<td>Opening pressure(MPa)</td>
<td>30</td>
</tr>
<tr>
<td>Hydraulic control one-way valve</td>
<td>Opening pressure(MPa)</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Pilot ratio</td>
<td>3.5</td>
</tr>
<tr>
<td>Hydraulic accumulator</td>
<td>Air polytropic exponent</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>Oil temperature (℃)</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Density(kg/m3)</td>
<td>850</td>
</tr>
<tr>
<td>Hydraulic oil</td>
<td>Modulus of volume elasticity(MPa)</td>
<td>1700</td>
</tr>
<tr>
<td></td>
<td>Absolute viscosity(cP)</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Wind speed(m/s)</td>
<td>0</td>
</tr>
<tr>
<td>Driving environment</td>
<td>Air density(kg/m3)</td>
<td>1.205</td>
</tr>
<tr>
<td></td>
<td>Environment temperature (℃)</td>
<td>25</td>
</tr>
</tbody>
</table>

Figure 4. Vehicle Speed.

Figure 5. Accumulator Pressure.
As seen from the figure, in an effective work circle, the engine of the prototype vehicle works at 72824.8 kJ, and the engine of the FPHHV works at 62510.3 kJ, the energy-saving rate is 14.163%. In a full UDDS condition with the range of 11 km, the fuel consumption of the prototype vehicle is 2.95L, while the FPHHV is 2.34L, and the oil-saving percentage is about 21%.

5. EXPERIMENTS AND RESULTS ANALYSIS

5.1. Experimental Prototype and the Schematic Diagram of the FPHHC

In order to verify the effectiveness of the hybrid configuration and the precision of the simulation model, in this article, an experimental prototype was developed based on a type of domestic crane and the actual road test and experiment were carried out. The overall appearance of the experimental prototype and the layout of hydraulic hybrid system are shown in Figure 12 and Figure 13; the experimental prototype structure is shown respectively in Figure 14.
5.2. Test and Result Analysis of the Experimental Prototype

Considering the road conditions and the actual situation of the vehicle, in this paper, the corresponding performance experiments of the experimental prototype were conducted under the conditions of a flat and climbing road. A series of experimental curves were analyzed through the way of a real vehicle test.

1) Test and result analysis of energy recovery efficiency

When the crane was accelerated to a certain speed, the crane was braked by the fixed displacement of the hydraulic secondary component (A4VG). The crane kinetic energy was calculated based on the initial speed of the start braking; the hydraulic energy under the same accumulator pressure change was calculated according to the AMESim equivalent model. As is shown in Figure 15, the crane starts to brake at the initial speed of 16.687 km/h, when the speed gets to 1.833 km/h, the accumulator pressure begins to decline. The charging pressure of accumulator is 7 MPa, the accumulator volume is 100L, braking time is about 18.8s, and braking distance is about 47m. When the accumulator pressure begins to decline, the revolving speed of the secondary component is about 154 rpm. Then the rate of energy recovery is 50.84%. Braking energy recovery efficiency varies depending on the pump, the control current is different. Considering the influence of the hydraulic pump being noisy and oil make-up insufficiently, the control current was given at 550mA (max: 650mA), the recovery efficiency is on the average above 50%, so the recovery efficiency is higher relatively.

2) Test and result analysis of energy recycling efficiency

Utilize the hydraulic system to brake the crane, and then the braking energy is stored in the accumulator and used for starting again. The energy regeneration efficiency is calculated based respectively on the vehicle speed at the time of braking and the speed at which the vehicle is started again. As is shown in Figure 16, wide fluctuations in the accumulator pressure proves that there is braking energy regeneration and is reused effectively. According to 16 MPa of the initial pressure of accumulator to calculate the energy recycling efficiency, when relied completely on the hydraulic system to brake and drive, the energy recycling efficiency was about 13.1%. If the accumulator pressure is not lost during the waiting of the vehicle, the energy regeneration efficiency can be further improved.

3) Test and result analysis of acceleration performance

The experimental prototype is respectively under the condition of only relying on the original vehicle transmission system and the aid of the hybrid system, by the same throttle opening. The vehicle is accelerated from the speed of 0 km/h to the speed of 30 km/h, and the time and speed was recorded and cleared up as shown in Figure 17.

As seen from the test curve the experimental prototype shifted gear twice in the process of being accelerated to the speed of 30 km/h. When only on the original vehicle transmission system, the experimental prototype took 27.2 seconds to accelerate to target speed. And under the auxiliary working condition of hybrid system, it took only 20.4 seconds for the experimental prototype to achieve the target speed, so that the vehicle's start time was significantly shortened.

4) Test and result analysis of climbing performance

In the vehicle climbing performance test, the road slope is 2.8%, and is chosen to test. By judging the maximum climbing gear of the experimental prototype in the test slope, the vehicle climbing performance could be evaluated. In the form of engine power, with the 8th gear climbing, the vehicle speed and engine speed decreased significantly, as shown in Figure 18a), the vehicle cannot fulfill speeding up and climbing; while in the hybrid system, a prototype could smoothly start and finish climbing, as shown in Figure 18(b). This shows, with the aid of a hybrid system, the power performance of experimental prototype is improved significantly.

5) Test and result analysis of fuel economy

The vehicle's fuel economy is closely related with its running condition. A specific road segment is chosen as a test road segment. Under the original transmission structure and the mode with a hybrid system, the experimental prototype was tested separately. In order to ensure that the experimental results are more representative, and to minimize the impact of traffic factors on the experimental results, the same test period is selected and used, and multiple tests are conducted. The vehicle speed and the average fuel consumption curves are shown in Figure 19. The fuel consumption recorder is installed on the experimental prototype, which is used to measure and record the vehicle's fuel consumption.
Figure 14. The Structure Diagram of the FPHHC Experimental Prototype.

Figure 15. System Parameters Change in the Process of Energy Recovery.

Figure 16. System Parameter Changes in the Process of Energy Utilization.

Figure 17. Speed Curves of the Vehicle is Accelerated from Idle State to the Speed of 30 Km/h.
6. CONCLUSIONS

THROUGH the establishment of AMESim simulation model combined with the experimental prototype, the braking energy recovery characteristics of front-mounted parallel hydraulic hybrid crane (FPHHC) are analyzed. The theoretical basis for modification of the experimental prototype is provided through the simulation model, and the correctness of the theory model is verified by the road test of hydraulic hybrid experimental prototype, and the simulation model, which is established earlier in this article and the alignment, is good. The comparative analysis of two kinds of vehicles shows that, in the process of vehicle braking, the hydraulic hybrid system of the FPHHC can effectively recycle the braking energy of a vehicle, and the energy-saving effect is obvious. In the aspects of vehicle acceleration and braking performance, the FPHHC has an obvious improvement more than the original vehicle, which shows that the system has a certain practical significance in the aspect of energy conservation. But as a result of the leakage of cartridge valve components, the efficiency of energy regeneration efficiency is affected, so the issue should be payed attention to later. In addition, this paper’s analysis method and the experimental results lay the foundation and has a certain reference value for the development and optimization of late hybrid vehicles.

6.1. Acknowledgment

This work was supported by National Natural Science Foundation Project of P. R. China (Grant No. 51405183); the Fundamental Research Funds for the Central Universities Fund of P. R. China and Special research found of doctoral program of Ministry of Education of P. R. China (Grant No.20130061120036), which I am here with acknowledge and sincere thanks.

7. DISCLOSURE STATEMENT

NO potential conflict of interest was reported by the authors.

8. REFERENCES


9. NOTES AND PHOTOS OF CONTRIBUTORS

Youquan Chen is a Ph.D. Candidate in Mechanical and Electrical Engineering Dept. at the School of Mechanical Science and Engineering, Jilin University, Changchun, China. His research interests are in fluid transmission and control. He has published more than ten papers in international Journals, core Journals in China and international meeting, eight of which are indexed by EI and more than five papers are indexed by CPCI-S.

Xinhui Liu received Ph.D. degree in Computational Mechanics at the Department of Mechanics, Jilin University of Technology, Changchun, China. He works at the School of Mechanical Science and Engineering, Jilin University, Changchun, China as an associate dean, professor, and doctoral supervisor, His research interests are in fluid transmission and electro-hydraulic control. He has published at home and abroad more than 70 research papers, and applied for 16 patents.

Xin Wang received Ph.D. degree in Mechanical Engineering at the School of Mechanical and Electrical Engineering, at Harbin Institute of Technology, Harbin, China. He works at the School of Mechanical Science and Engineering, Jilin University, Changchun, China as an associate professor, His research interests are in fluid transmission and electro-hydraulic control. He has published many research articles in international Journals; core Journals in China and international meeting.

Jinshi Chen received Ph.D. degree in Mechanical Engineering at the School of Mechanical and Electrical Engineering, Jilin University of Technology. He works at the School of Mechanical Science and Engineering, Jilin University, Changchun, China as an associate professor, His research interests are in fluid transmission and electro-hydraulic control. He has published many research articles in in international Journals; core Journals in China and international meetings.
Intelligent Control for Integrated Guidance and Control based on the Intelligent Characteristic Model

Jun Zhou, Zhenzhen Ge

Institute of Precision Guidance and Control, Northwestern Polytechnical University, Xi’an, China

ABSTRACT

In this paper, an adaptive integrated guidance and control (IGC) scheme for the homing missile is proposed based on the novel continuous characteristic model and the dynamic surface control technique. The novel continuous characteristic model is first proposed in the presence of unknown model coefficients and uncertainties. Then, the dynamic surface control technique is applied to the continuous characteristic model. The proposed IGC scheme guarantees the line-of-sight angular rates converge to an arbitrarily small neighborhood of zero and all the closed-loop signals to be semi-globally uniformly ultimately bounded, which is proven using the Lyapunov stability theory. Finally, the effectiveness of the adaptive IGC scheme is demonstrated using nonlinear numerical simulations for the maneuvering target.

KEY WORDS: Integrated guidance and control, Intelligent control, Continuous characteristic model, Adaptive control, Dynamic surface control, Homing missile

1 INTRODUCTION

TRACTICAL missiles by virtue of their mission objectives, demand extremely accurate performance from their guidance and control components. The traditional design method of the missile guidance and control system is to design the guidance and control subsystems separately and then integrate them. As it ignores the coupling and time lag existing inevitably in the guidance and control subsystems, guidance and control subsystems fail to work synergistically, and the performance of overall missile system is not fully exploited (Yan & Ji, 2012). Considering these drawbacks in timescale separation design, integrated guidance and control (IGC) has been a hot topic of research in recent years (Wang & Wang, 2014). The IGC views the guidance subsystem and control subsystem as a whole and directly generates the fin deflection commands. Until now, a variety of approaches have been employed to design IGC laws, for example, feedback linearization (Menon et al., 2004), state dependent Riccati equation (Vaddi et al., 2009), active disturbance rejection control (Zhang & Wu, 2016), model predictive static programming (Padhi et al., 2014), sliding mode control (Song & Zhang, 2016), backstepping (Yan et al., 2014) and dynamic surface control (DSC) (Hou et al., 2013; Wang et al., 2015).

It is noted that most of the existing relevant literature on IGC designs assumed the involved model coefficients related to the missile-target range $R$ or the relative velocity along the line of sight (LOS) $\dot{R}$ were known and accessible, and hence, all state variables were available to the control law. However, this assumption is quite restrictive and unreasonable in many applications, especially for the missile equipped with the infrared seeker or passive radar seeker. Furthermore, attack and sideslip angles are usually regarded as part of state variables in many works, whereas they are difficult to measure. Thus, only a subset of state variables is measurable and available to the control law. In this work, we mainly focus on the IGC design when all the involved model coefficients, unmatched uncertainties and states except for the system output are unknown and unavailable to the IGC law. Therefore, a second-order continuous characteristic model is first presented to reconstruct the original system based on the principle that the original system and its continuous characteristic model should have the same output when their inputs are equivalent.

A characteristic model has the following features. 1) It should be equivalent to its practical plant in output for the same input. 2) It compresses high-order terms, uncertainties and disturbances into several parameters without losing any information. In the
1990s, an integrated and practical all-coefficient adaptive control theory and method based on characteristic models were proposed (Wu et al., 2007). This control theory is independent of the original system model and has already been applied to many fields successfully (Zhang & Hu, 2012; Huang & Zhang, 2015). However, it is denoted by a slow time-varying second-order difference equation and requires that the sample time should be small enough to make the coefficients of the difference equation slow time-varying. The model mismatch and the truncation error between the discrete controller and the continuous system are inevitable. To eliminate these shortcomings, a continuous characteristic model is presented in this work. As the novel characteristic model has no truncation error and is not related to the sample time, its characteristic parameters vary much faster. On the account of this fact, the continuous characteristic model is incorporated with the robust adaptive dynamic surface control technique. To avoid the problem of explosion of complexity of the backstepping method (Dimirovski et al., 2006), the DSC technique was proposed by introducing a first-order low pass filter at each design step. It has been extensively researched in many applications ranging from linear systems, nonlinear systems and stochastic systems (Wang & Lin, 2010; Kendrick et al., 2013; Song et al., 2014).

In this paper, we consider the IGC design for the homing missile based on the novel continuous characteristic model and the adaptive dynamic surface control approach in the presence of unknown model coefficients, uncertainties and intermediate state variables. The rest of this paper is organized as follows: In Section 2, the integrated guidance and control design for the pitch channel is introduced, for example, the original pitch channel dynamics model description, continuous characteristic model, design objective, IGC law design procedure and stability analysis. In Section 3, the integrated guidance and control design for the yaw and roll channels are presented. Simulations on the nonlinear missile model are conducted, and conclusions are drawn in Section 4 and 5, respectively.

2 INTEGRATED GUIDANCE AND CONTROL DESIGN FOR THE PITCH CHANNEL

2.1 Continuous Characteristic Model of the Pitch Channel

The pitch channel model for the missile integrated guidance and control loop is given by (Liang et al., 2015).

\[
\begin{align*}
\dot{x}_1 &= a_{11} x_1 + a_{12} x_2 + \Delta_1 \\
\dot{x}_2 &= a_{22} x_2 + x_3 + \Delta_2 \\
\dot{x}_3 &= a_{32} x_2 + b_3 u + \Delta_3 \\
y &= x_1
\end{align*}
\]

Where \( x_1 = \dot{\epsilon} \), \( x_2 = \alpha \) and \( x_3 = \omega_z \) are the system states; \( u = \delta_z \) and \( y = \dot{\epsilon} \) are the system input and output; \( \Delta_1, \Delta_2 \) and \( \Delta_3 \) are the system uncertainties; \( b_3 = QSLm^\alpha_z / J_z \), \( a_{11} = -2\dot{\alpha} / R \), \( a_{12} = -QSc^\alpha_y / mR \), \( a_{22} = -QSc^\alpha_y / mV \) and \( a_{32} = QSLm^\alpha_z / J_z \) are the model coefficients; \( m \) is the mass of the missile; \( V \) is the velocity of the missile; \( S \) and \( L \) are the reference area and length; \( R \) is the missile-target range; \( \varepsilon \) is the elevation angle of LOS; \( \alpha \) is the attack angle; \( \omega_z \) is the body-axis pitch rate; \( Q = \rho V^2 / 2 \) is the dynamic pressure; \( \rho \) is the air density; \( \delta_z \) is the elevator deflection; \( J_z \) is the pitch moment of inertia; \( c_y^\alpha \) is the partial derivative of lift force coefficient with respect to \( \alpha \); \( m^\alpha_y \) and \( m^\delta_z \) are the partial derivatives of pitching moment coefficient with respect to \( \alpha \) and \( \delta_z \), respectively.

Remark 1. In general, \( c_y^\alpha \) and \( m^\delta_z \) are usually positive and negative constants, respectively. Hence, there exist two constants \( \bar{a} < 0 \) and \( \bar{b} < 0 \) such that \( a_{12} \leq \bar{a} < 0 \) and \( b_3 \leq \bar{b} < 0 \).

So far, a great variety of literature deals with IGC designs on the condition that all the involved model coefficients are known and available, whilst only the uncertainties are unknown. Actually, for many missiles equipped with the infrared seeker or passive radar seeker, the model coefficients related to \( R \) or \( \dot{R} \) are unknown and unavailable. So, the restrictive assumption in previous works is unreasonable in many cases. As a result, in this paper, it is assumed that the model coefficients and uncertainties of model (1) satisfy the following assumption instead of the restrictive one.

Assumption 1. All the involved model coefficients and uncertainties of model (1) are unknown bounded time-varying functions with bounded derivatives, and all the bounds are also unknown.

At present, in some literature, the system states are considered to be known and available. But the attack and sideslip angles are difficult to measure. To solve the problem, the model states are supposed to meet the following assumption.

Assumption 2. Only the system input \( u \) and output \( y \) are available for measurement.

Due to Assumption 1 and Assumption 2, it is of great interest to conduct an IGC design without a state observer and the designed adaptive law should be
much simple. In order to achieve this objective, a novel continuous characteristic model denoted by a second-order continuous time-varying differential equation is proposed. For the same input, the continuous characteristic model has the same output as that of its original model. As the proposed IGC design only uses the system output, it is able to substitute the continuous characteristic model for its original model in the design process. The novel continuous characteristic model of original model (1) is presented as follows.

According to (1), the first-order derivative of output is

$$\dot{x}_1 = a_{11}x_1 + a_{12}x_2 + \Delta_1$$  \hspace{1cm} (2)

Then, the third-order derivative of output is expressed as

$$\ddot{x}_1 = a_{11}\ddot{x}_1 + 2a_{12}\dot{x}_2 + a_{11}x_1 + a_{12}x_2 + \ddot{\Delta}_1$$  \hspace{1cm} (3)

From (1), we have

$$\dddot{x}_1 = a_{11}\dddot{x}_1 + a_{11}x_1 + a_{12}\ddot{x}_2 + a_{12}x_2 + \dddot{\Delta}_1 + a_{12}\{a_{22}\dddot{x}_1 + \dddot{\Delta}_1\}$$  \hspace{1cm} (5)

According to (2), it can be obtained that

$$x_2 = (\dot{x}_1 - a_{11}x_1 - \Delta_1)/a_{12}$$  \hspace{1cm} (6)

Substituting (6) into (5) results to

$$\dddot{x}_1 = q_2\dddot{x}_1 + q_1\dot{x}_1 + q_0x_1 = bu + w$$  \hspace{1cm} (7)

where $q_2$, $q_1$, $q_0$, $b$ and $w$ are time-varying functions, and their expressions are presented as follows:

$$q_2 = a_{11}, q_1 = a_{12}, q_0 = -a_{22} - a_{32}$$

Substituting (4) into (3) results to

$$\dddot{x}_1 = a_{11}\dddot{x}_1 + a_{12}\ddot{x}_2 + a_{11}x_1 + a_{12}x_2 + \dddot{\Delta}_1 + a_{12}\{a_{22}\dddot{x}_1 + \dddot{\Delta}_1\}$$  \hspace{1cm} (10)

where $a_{11}$, $a_{12}$, $a_{22}$ and $a_{32}$ are the time-varying characteristic parameters. The main purpose of above deduction is to prove the existence of continuous characteristic model for its original model. Moreover, from the above theoretical deduction, it has been proven that the continuous characteristic model has the same output as that of its original model for the same input.

Remark 2. There are many choices for the intermediate parameters $\delta$ and $m$ appearing in (14), that is, the compress mapping function is not unique. It means that the original system has many continuous characteristic models with different characteristic parameters. The main purpose of above deduction is to prove the existence of continuous characteristic model, which is the most important of all for the IGC design conducted in this paper. Due to the characteristic parameters that are related to the coefficients and uncertainties of model (1), which are supposed to be unknown in assumption 1, all the characteristic parameters are also unknown.

In addition, for the stability of the closed-loop system, the characteristic parameters are supposed to
satisfy the following reasonable assumption in view of their physical backgrounds.

**Assumption 3.** The characteristic parameters \( a_1(t), a_2(t), b_1(t) \) and \( d(t) \) are bounded time functions having bounded derivatives.

### 2.2 Design Objective

The objective of IGC design can be elaborated as: Use the continuous characteristic model instead of its original model to design an IGC law in the presence of unknown model coefficients, mismatched uncertainties and states except the output of model, which should guarantee the missile intercepts the target with small miss distance and keep all the states of missile stable.

By accepting the concept that zeroing LOS angular rates will lead to a successful interception with zero miss distance, the specific guidelines for the IGC design are given as follows.

1. To sustain LOS angular rates \( \dot{\epsilon} \) and \( \dot{\eta} \) as zeros for making the miss distance small.
2. To maintain the roll angle \( \gamma \) in a small neighbourhood of zero.
3. To stabilize the states of the missile and be robust to the unknown coefficients and uncertainties existing in the missile dynamics.

### 2.3 IGC Law Design

As the characteristic model (16) is a fast time-varying linear system with a matched uncertainty, the adaptive dynamic surface control approach has been the most common method for such a kind of systems recently (Song et al., 2014). In this section, an adaptive dynamic surface control algorithm is developed to deal with the output control problem under Assumption 3.

The second-order continuous characteristic model (16) is written in the state-space form

\[
\begin{align*}
\dot{x}(t) &= A(t)x(t) + b(t)u(t) + d(t) \\
y(t) &= x(t)
\end{align*}
\]

Where \( u \in \mathbb{R} \) and \( y \in \mathbb{R} \) are the system input and output, \( x = [x_1, x_2]^T \in \mathbb{R}^2 \) is the system state vector, and
\[
A = \begin{bmatrix} 0 & 1 \\ -a_2(t) & -a_1(t) \end{bmatrix},
b(t) = \begin{bmatrix} 0 \\ b_1(t) \end{bmatrix},
d(t) = \begin{bmatrix} 0 \\ d(t) \end{bmatrix}.
\]

**Remark 3.** Since \( b_1(t) = b = a_2 b_1 \), meanwhile, \( a_{12} \leq \bar{a} < 0 \) and \( b_i \leq \bar{b} < 0 \) in Remark 1, it can be seen that \( b_1(t) > 0 \) and the sign of \( b_i(t) \) is known and constant throughout the engagement.

The robust output-feedback adaptive dynamic surface control design for system (18) is proceeded as follows:

**Step 1:** The first surface error is defined as
\[
S_1 = y
\]

By considering (18), the time derivative of first surface error is
\[
\dot{S}_1 = \dot{y} = -c_1S_1 + \left[ (x_2 - \alpha_1) + c_1 S_1 \right]
\]

where \( c_1 \) is a positive design parameter, \( \alpha_1 \) is the virtual control to be designed to stabilize (20) in the absence of \( (x_2 - \alpha_1) \).

Choosing the virtual control as
\[
\alpha_1 = -c_1S_1
\]

Let \( \alpha_1 \) as the input of the following first-order filter.
\[
\tau_2 \dot{x}_{2d} + x_{2d} = \alpha_1
\]

Where \( \tau_2 \) is the time constant, \( x_{2d} \) is the filter output.

**Step 2:** The second surface error is defined as
\[
S_2 = x_2 - x_{2d} = \dot{y} - x_{2d}
\]

Whose time derivative is
\[
\dot{S}_2 = -c_2S_2 + b_1(t)\left[ u + \theta^T \omega \right]
\]

Where \( c_2 \) is a positive design parameter, and
\[
\theta = \left[ \frac{1}{b_0}, \frac{1}{b_0} - a_0/b_0, \frac{1}{b_0} - a_1/b_0, \frac{1}{b_0} - d/b_0 \right]^T,
\]
\[
\omega = \left[ c_2S_2 - \dot{x}_{2d}, y, \dot{y}, 1 \right]^T.
\]

Define the estimate error as
\[
\hat{\theta} - \hat{\theta} = \hat{\theta}
\]

Where \( \hat{\theta} \) is the estimate of \( \theta \), and the adaptive law is defined as
\[
\dot{\hat{\theta}} = \Gamma S_2 \omega - \sigma \hat{\theta}
\]

Where \( \Gamma \) is a positive definite matrix, \( \sigma \) is a positive design parameter.

In final step, the actual system input is chosen as
\[
u = -\hat{\theta}^T \omega
\]

**Remark 4.** It is seen that the design procedure is quite simple compared with traditional adaptive controllers for linear time-varying (LTV) plants. The number of estimate parameters is small and constant, which is not
related to the model order and relative degree of its original system. In addition, the proposed scheme does not need a state observer any more. Thus, the algorithm does not introduce state observer errors, which may affect the rate of convergence and the output error to some extent.

2.4 Stability Analysis

The stability analysis is based on the Lyapunov stability theory. First, define the boundary layer error as

\[ y_i = x_{2i} - \alpha_i \]  

(29)

In view of (22) and (29), we have

\[ y_i \dot{y}_i = y_i \left[ \left( \alpha_i - x_{2i} \right)/\tau_z + \dot{\alpha}_i \right] \leq -y_i^2/\tau_z + \left| y_i \right| \eta_i \]  

(30)

Where \( \eta_i \) is a positive continuous function. Similarly to the linear time-invariant case (Wang & Lin, 2010), the existence of \( \eta_i \) can be checked with the Assumption 3.

Consider the following Lyapunov function candidate

\[ V = \left[ S_i^2 + S_2^2 + y_i^2 \right] + b_0(t) \dot{\theta}^T \Gamma^{-1} \dot{\theta} \]  

(31)

Where \( \dot{\theta} = \theta - \dot{\theta} \) is the estimate errors vector. As mentioned in Remark 3, there exists a constant \( b > 0 \) such that \( b_0(t) \geq b > 0 \), thus \( V \) is positive definite.

**Theorem 1:** On condition that assumption 1-3 are satisfied, the closed-loop system consisting of the equivalent continuous characteristic model (16) instead of its original LTV system (1) in the design process, the control law (28), the adaptive update law (27) and the first-order filters (22) is considered. If the positive constant \( R_0 \) given arbitrarily meets the following inequality

\[ V(0) \leq R_0 \]

there exist \( c_1, c_2, \tau_z, \Gamma \) and \( \sigma \) to guarantee that all closed-loop signals are uniformly bounded and the output error \( S_i \) is arbitrarily small with proper design parameters.

**Proof**

According to (27) - (30), the time derivative of (31) yields

\[ \dot{V} \leq -c_1 S_i^2 - c_2 S_2^2 + \sigma b_0(t) \dot{\theta}^T \Gamma^{-1} \dot{\theta} + S_i \left( S_i^2 + y_i^2 \right) + b_0(t) \dot{\theta}^T \Gamma^{-1} \dot{\theta} + \left( -y_i^2/\tau_z + \left| y_i \right| \eta_i \right) \]  

(32)

Using Young’s inequality, we have

\[ |y_i| \eta_i \leq \frac{y_i^2 \eta_i^2}{2\epsilon} + \frac{\epsilon}{2} \]

\[ \theta^T \Gamma^{-1} \dot{\theta} \leq -\frac{1}{2} \theta^T \Gamma^{-1} \theta + \frac{1}{2} \theta^T \Gamma^{-1} \theta. \]

Where \( \epsilon \) is a positive constant given arbitrarily. According to (33), the inequality (32) is written as

\[ \dot{V} \leq -c_1 S_i^2 - c_2 S_2^2 - \frac{1}{2} \sigma b_0(t) \dot{\theta}^T \Gamma^{-1} \dot{\theta} + \left( S_i^2 + \frac{y_i^2}{2} \right) + \frac{1}{2} \sigma b_0(t) \dot{\theta}^T \Gamma^{-1} \theta + \frac{1}{2} b_0(t) \dot{\theta}^T \Gamma^{-1} \dot{\theta} + b_0(t) \dot{\theta}^T \Gamma^{-1} \dot{\theta} + \left( -y_i^2/\tau_z + \frac{\eta_i^2}{2\epsilon} \right) y_i^2 + \frac{\epsilon}{2} \]  

(34)

Define the function

\[ \varphi = \sigma b_0(t) \dot{\theta}^T \Gamma^{-1} \theta/2 + b_0(t) \dot{\theta}^T \Gamma^{-1} \dot{\theta} + b_0(t) \dot{\theta}^T \Gamma^{-1} \dot{\theta} + \epsilon/2 \]  

(35)

Substituting (35) into (34) yields

\[ \dot{V} \leq -c_1 S_i^2 - c_2 S_2^2 - \frac{1}{2} \sigma b_0(t) \dot{\theta}^T \Gamma^{-1} \dot{\theta} + \left( -y_i^2/\tau_z + \frac{\eta_i^2}{2\epsilon} \right) y_i^2 + \left( S_i^2 + \frac{y_i^2}{2} \right) + \varphi \]  

(36)

Define the compact set

\[ \Omega_0 := \left\{ (S_i, S_2, y_i, \dot{\theta}) : V \leq R_0 \right\} \subset \mathbb{R}^7 \]  

(37)

Where \( R_0 \) is a given positive constant. Thus, from (37), the continuous functions \( \eta_i \) and \( \varphi \) have maximums on \( \Omega_0 \), say, \( M_i \) and \( M_\varphi \), respectively.

Choose

\[ c_i \geq 1 + \alpha_o, \ c_2 \geq 0.5 + \alpha_o, \ \frac{1}{\tau_z} \geq 0.5 + M_i^2/2\epsilon + \alpha_o, \ \sigma \geq 2\alpha_o. \]  

(38)

Where \( \alpha_o \) is a positive design parameter.

Substituting the inequality (38) into (36), we obtain

\[ \dot{V} \leq -2\alpha_o V + M_\varphi \]

(39)

Hence, if \( R_0 \geq V(0) \) and \( \alpha_o \) is chosen as

\[ \alpha_o \geq M_\varphi/2R_0 \]

(40)

then \( \dot{V} \leq 0 \) on \( V = R_0 \).

From the inequality (39), we can obtain that
\[0 \leq V \leq \left[ V(0) - \frac{M_e}{2\alpha_0} \right] e^{-2\alpha_0 t} + \frac{M_e}{2\alpha_0} \] (41)

Hence,

\[
\lim_{\tau \to 0} \left| S_i(t) \right| \leq \lim_{\tau \to 0} \sqrt{2V(t)} = \sqrt{-\frac{M_e}{\alpha_0}} (42)
\]

Therefore, the output error can be made arbitrarily small with a large value of \(\alpha_0\), which can be implemented by letting the values of \(c_1\), \(c_2\) and \(\sigma\) large enough and that of \(\tau_2\) small enough.

### 3 CONTROL DESIGN FOR THE YAW AND ROLL CHANNELS

SIMILARLY to the pitch channel, the yaw channel model for the integrated guidance and control loop is given by

\[
\begin{align*}
\dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + \Delta_1 \\
\dot{x}_2 &= a_{22}x_2 + x_3 + \Delta_2 \\
\dot{x}_3 &= a_{32}x_2 + b_3u + \Delta_3 \\
y &= x_1,
\end{align*}
\] (43)

Where \(x_1 = \eta\), \(x_2 = \beta\) and \(x_3 = \omega_x\) are system states; \(\eta\) is the azimuth angle of LOS; \(\beta\) is the sideslip angle; \(\omega_x\) is the body-axis yaw rate; \(u = \delta_y\) and \(y = \eta\) are system input and output; \(\delta_y\) is the rudder deflection; \(\Delta_1\), \(\Delta_2\) and \(\Delta_3\) are uncertainties; \(a_{11} = 2\tau_{/r}\), \(a_{12} = QSLm_y^\beta/J_y\), \(a_{22} = QSLm_x^\beta/mV\), \(a_{32} = QSLm_y^\beta/J_y\) and \(b_3 = QSLm_x^\beta/J_y\) are model coefficients; \(J_y\) is the yaw moment of inertia; \(c_d^\beta\) is the partial derivative of side force coefficient with respect to \(\beta\); \(m_x^\beta\) and \(m_y^\beta\) are the partial derivatives of yawing moment coefficient with respect to \(\beta\) and \(\delta_y\), respectively.

Obviously, the yaw channel model (43) takes the form of the pitch channel model. Thus, their adaptive control algorithms are similar and the design procedure of yaw channel is omitted.

Additionally, the roll channel model is given by

\[
\begin{align*}
\dot{x}_1 &= x_2 + \Delta_1 \\
\dot{x}_2 &= b_3u + \Delta_2 \\
y &= x_1,
\end{align*}
\] (44)

Where \(x_1 = \gamma\) and \(x_2 = \omega_x\) are the system states; \(\gamma\) is the roll angle; \(\omega_x\) is the body-axis roll rate; \(u = \delta_{\alpha}\) and \(y = \gamma\) are the system input and output; \(\delta_{\alpha}\) is the aileron deflection; \(\Delta_1\) and \(\Delta_2\) are systems uncertainties; \(b_2 = QSLm_y^\delta/J_y\) is the model coefficient; \(J_y\) is the roll moment of inertia; \(m_x^\delta\) is the partial derivative of rolling moment coefficient with respect to \(\delta_{\alpha}\).

Although the roll channel model (44) is much different from the pitch and yaw channels, the proposed scheme is not related to the model structure. So, the adaptive control algorithm can be utilized to stabilize system (44) is given as follows, the form of which is same as that of the pitch and yaw channels.

\[
\begin{align*}
S_1 &= y = \gamma \\
\alpha_1 &= -c_1S_1 \\
\tau_2\dot{x}_{2d} + x_{2d} &= \alpha_1 \\
S_2 &= \dot{y} - x_{2d} \\
\dot{\theta} &= \Gamma S_2 \omega - \sigma \hat{\theta} \\
u &= -\hat{\theta} \omega
\end{align*}
\] (45)

Where \(\hat{\theta}\) is the estimate of \(\theta\) defined in (25), \(\omega\) is defined in (25). The design parameters \(c_1\), \(c_2\), \(\tau_2\) and \(\sigma\) should satisfy the inequality (38).

### 4 SIMULATION RESULTS

In this section, the effectiveness of the newly proposed IGC scheme based on the continuous characteristic and the dynamic surface control technique is verified through the 6DOF nonlinear numerical simulation (Wang et al., 2016), where an air-to-surface missile is considered in its terminal homing to intercept a maneuvering target on ground.

As the forms of IGC laws for pitch, yaw and roll channels are the same as each other, in this section, we select the same design parameters for each channel. The design parameters of the novel IGC law are chosen as

\[
c_1 = 2, \quad c_2 = 4, \quad \tau_2 = 0.02, \quad \sigma = 1, \quad \Gamma = \text{diag}[2, 2, 2, 2].
\]

and the initial value of \(\hat{\theta}\) is set as \([0, 0, 0, 0]^T\).

For comparison, the traditional separate channel guidance and control scheme with proportional navigation guidance (PNG) plus proportional-integral-derivative (PID) control law is also carried out. For simplicity, the novel IGC law and traditional separate scheme are denoted as NIGC and TSGC, respectively.

In numerical simulations, the missile dynamics parameters are outlined in Table 1. \(c_{\alpha 0}\) is the zero-lift drag coefficient. \(c_{\alpha \alpha}\) and \(c_{\alpha \beta}\) are the partial derivatives of drag force coefficient with respect to \(\alpha\) and \(\beta\), respectively. \(c_{\gamma \beta}\) and \(c_{\delta_y}\) are partial derivatives of lift force coefficient with respect to \(\beta\) and \(\delta_y\), respectively. \(c_{\delta_{\alpha}}\) and \(c_{\delta_{\beta}}\) are partial derivatives of side force
coefficient with respect to $\alpha$ and $\delta$, respectively. $m^a_i$ and $m^b_i$ are partial derivatives of rolling moment coefficient with respect to $\alpha$ and $\beta$, respectively. The initial states of the missile are respectively set as $V = 600\text{m/s}$, $\theta = \psi = 0$, $\vartheta = 0.2\text{rad/s}$, $\omega = 0.1\text{rad/s}$.

In the inertial coordinate system, the initial position of the missile is set as $[0, 0, 0]^T$ m. The initial position, velocity, and acceleration of the target in the inertial coordinate system are set as $[8000, 0, 2000]^T$ m, $[-17, 0, 0]^T$ m/s and $[-1.5, 0, -10\cos(0.5\pi)]^T$ m/s$^2$, respectively.

Table 1. The Nominal Parameters of Missile Dynamics.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Name</th>
<th>Value</th>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>1200 kg</td>
<td>$m^a_i$</td>
<td>-0.38</td>
<td>$c_i$</td>
<td>0.4</td>
</tr>
<tr>
<td>$S$</td>
<td>0.42 m$^2$</td>
<td>$m^b_i$</td>
<td>1.06</td>
<td>$c_i^b$</td>
<td>57.2</td>
</tr>
<tr>
<td>$L$</td>
<td>0.68 m</td>
<td>$m^a_i$</td>
<td>-27.3</td>
<td>$c_i^a$</td>
<td>-0.1</td>
</tr>
<tr>
<td>$J_x$</td>
<td>100 kg-m$^2$</td>
<td>$m^b_i$</td>
<td>-13.3</td>
<td>$c_i^b$</td>
<td>5.7</td>
</tr>
<tr>
<td>$J_y$</td>
<td>5700 kg-m$^2$</td>
<td>$m^a_i$</td>
<td>-28.2</td>
<td>$c_i^a$</td>
<td>0.1</td>
</tr>
<tr>
<td>$J_z$</td>
<td>5600 kg-m$^2$</td>
<td>$m^b_i$</td>
<td>-14.0</td>
<td>$c_i^b$</td>
<td>-56.3</td>
</tr>
<tr>
<td>$\rho$</td>
<td>1.16 kg/m$^3$</td>
<td>$c_{\alpha}$</td>
<td>0.2</td>
<td>$c_i^\alpha$</td>
<td>0.1</td>
</tr>
<tr>
<td>$g$</td>
<td>9.8 m/s$^2$</td>
<td>$c_i^\alpha$</td>
<td>0.4</td>
<td>$m_c^\alpha$</td>
<td>0.45</td>
</tr>
</tbody>
</table>

The NIGC and TSGC approaches will be conducted in the case. The curves of locations, LOS angular rates, aerodynamic angles and deflection angles are presented in the Figure 1–Figure 3. The three dimensional (3D) trajectory of the missile with respect to NIGC and TSGC is shown in Figure 1. The curves of missile-target range and LOS angular rates are presented in Figure 2. The missile distance of the NIGC is 0.1537m, while, the missile distance of the TSGC is 8.3206m. The proposed IGC law has smaller missile distance than the classical separate channel scheme, and can guarantee the missile hits the target with higher accuracy. In addition, it is known that the NIGC scheme has better transient response performance in driving LOS angular rates to zero than that of TSGC approach from Figure 2(b). The angle of attack, sideslip angle and velocity bank angle are shown in Figure 3(a), and the deflection angles of the roll, yaw, and pitch control surface are presented in Figure 3(b).
According to Figure 2(b), for the TSGC scheme, LOS angular rates begin to diverge rather than converge to zero when the missile gets close to the target at the last moment. Meanwhile, it also makes the fin deflections increase largely and chatter, which affects the Euler angles and angular rates directly. But on the contrary, for the NIGC scheme, the LOS angular rates converge to a small neighbourhood of zero in finite time and do not diverge. Therefore, the missile distance of the NIGC scheme is much smaller than that of the TSGC scheme.

In order to verify the robustness of the IGC method, one hundred times of Monte Carlo simulations are conducted. Suppose that the coefficients of aerodynamics forces and moments all offset randomly by $-20\% \sim +20\%$ of their respective nominal values. The impact point deviation distributions of the Monte Carlo simulations are shown in Figure 4.

The mean missile distance of the NIGC is 0.2343m with a standard deviation 0.1449m. However, the mean missile distance of the TIGC is 1.0685m with standard deviation of 0.8861m. As a result, it can be obtained that the proposed NIGC scheme is more robust to uncertainties and has better performance during the interception than the classical TSGC scheme.

5 CONCLUSION

A novel adaptive IGC scheme for the homing missile is proposed based on the novel fully continuous characteristic model of the IGC model and the dynamic surface control technique. The major advantages of the proposed IGC scheme are drawn as follows:

(1) For the IGC model of each channel, the continuous characteristic model is established for the IGC law design. The continuous characteristic model is equivalent to its original model in output, that is, their outputs are identical when their inputs are the same.

(2) With the novel continuous characteristic model, all the model coefficients of the IGC dynamics model related to the missile-target range, relative velocity along the LOS and aerodynamic coefficients can be unknown. Moreover, the intermediate states such as attack angle, sideslip angle and angular rates of roll, yaw and pitch are also unavailable for the proposed IGC law which only utilizes the LOS angular rates.

(3) The dynamic surface control method is used for the continuous characteristic model to drive the LOS angular rates to an arbitrarily small neighborhood of zero and guarantee the stability of the closed-loop system.

(4) Nonlinear numerical simulations explicitly demonstrate the effectiveness of the proposed IGC scheme, which has better performance than the classical separate channel guidance and control scheme.

6 ACKNOWLEDGMENT

The authors would like to thank the reviewers for their constructive comments and suggestions.
7 DISCLOSURE STATEMENT

NO potential conflict of interest was reported by the authors.

8 REFERENCES


9 NOTES ON CONTRIBUTORS

Jun Zhou received a Ph.D. in Navigation, Guidance and Control from Northwestern Polytechnical University, Xi’an, China in 1993. He is currently a professor at the School of Astronautics, Northwestern Polytechnical University, Xi’an, China. He is also the director of Institute of Precision Guidance and Control. His research interests include guidance and control systems of missiles and space vehicles, flight control and simulation technique, fault tolerance, and software engineering.

Zhenzhen Ge is a Ph.D. candidate from Institute of Precision Guidance and Control, at the School of Astronautics, Northwestern Polytechnical University, Xi’an, China. He received a M.Sc. in Navigation, Guidance and Control from Northwestern Polytechnical University in 2015. His main research interests include guidance and control of hypersonic vehicles, adaptive control and artificial intelligence.
NARX Network based Driver Behavior Analysis and Prediction using Time-series modeling

Ling Wu¹, Haoxue Liu², Tong Zhu², Yueqi Hu³

¹ School of Vehicle Engineering, Xi'an Aeronautical University, Xi'an, Shaanxi, China
² Key Laboratory for Automotive Transportation Safety Enhancement Technology of the Ministry of Communication, Chang'an University, Xi'an, Shaanxi, China
³ School of Automobiles, Chang'an University, Xi'an, Shaanxi, China

KEY WORDS: NARX network, Driver, Extra-long tunnels, Heart rate, Traffic psychology

ABSTRACT
The objective of the current study was to examine how experienced and inexperienced driver behaviour changed (including heart rate and longitudinal speeds) when approaching and exiting highway tunnels. Simultaneously, the NARX neural network was used to predict real-time speed with the heart rate regarded as the input variable. The results indicated that familiarity with the experimental route did decrease drivers' mental stress but resulted in higher speed. The proposed NARX model could predict synchronous speed with high accuracy. These results of the present study concern how to establish the automated driver model in the simulation environment.

INTRODUCTION
ROAD tunnels, as part of a particular structure in road networks, provide convenient connections in mountainous cities and space-constrained areas (Yeung and Wong, 2014; Van der Hoeven, 2011). In China, the largest number of tunnels has been constructed with the most complicated geological conditions. Furthermore, the quickest development is also expected in the future (Yan et al., 2017). Although surface roads have more elevated crash rates than road tunnels (Yeung and Wong, 2013; Ma et al., 2009; Amundsen and Ranes, 2000; Lemke, 2000), the consequences of tunnel crashes are much more catastrophic with prominently higher fatality rates (Kircher and Ahlstrom, 2012). Regarding different tunnel segments, the entrance zone resulted in a higher crash risk (Amundsen and Engebretsen, 2009; Nussbaumer, 2007) and pathologic discomfort (Yeung and Wong, 2014) than further into the tunnel. Because tunnels are dirty, dark and monotonous (Noizet and Ricard, 2004), accompanied by the high contrast of inside and outside driving environments, which could trigger rare, but severe accidents (Laureshyn et al., 2010), it is of critical importance to investigate the effects from the drivers perspective in terms of micro-behavior and understand how the driver’s behavior changes when traveling through the tunnels.

The principal causes of traffic accidents have been extensively attributed to behavioral issues such as safe headway maintenance, speeding and lane keeping as well as the lack of vigilance, including fatigue, distraction, and inattention (Nussbaumer, 2007). Compared to open-road experiments, tunnel structures are found to cause increased attention, mental workload (Domenichini et al., 2017) and stress (Manseer and Riener, 2014), associated with lower driving speeds and differences in lateral deviations (distance to the tunnel wall) (Calvi et al., 2012). Other studies also examined the effects caused by different tunnel characteristics. Different tunnel lengths gave rise to differences in the proportion and distribution of fixation duration (He et al., 2010). A simulator study by Hirata et al. (2006) indicated that informing drivers of accidents ahead in a tunnel exerted a meaningful impact on safety. Previous literature has focused on the single aspects of potential workload (subjective evaluation), differences in attention distribution, speed or car-following behavior using either simulator tests or on-road tests. However, only a handful of field-operational studies provide thoughtful and comparative insight into both the variation of driver stress and speed when compared to open roads,
specifically in extra-long highway tunnels with more obvious characteristics.

A growing body of literature was devoted to automated driving and tried to simulate driver behavior under different driving conditions (Smith and Starkey, 1995; Zeeb et al., 2015). In this special driving environment, driver mental activity and corresponding speed may be influenced by the change of visual pattern and the tunnel wall effect (Törnros, 2000), which could be embodied in the detailed description of the simulation model. ECG (Electrocardiogram) activity, regarded as parameter skin conductance, have a high correlation with the perceived stress levels of the subject (Manseer and Riener, 2014; Healey and Picard, 2005). Hence, the objective of this study was to examine experienced and inexperienced driver behavior (including HR (heart rate) and speed) in extra-long highway tunnels (longer than 3 km) using on-road tests. To remove the effects of geometric differences and inhomogeneity of the traffic flow, two typical extra-long tunnels were selected, and vehicles were driven in both directions in the tunnel (north to south and south to north). The mean value was adopted to characterize the variations in driver stress and speed. The reason why both experienced drivers (staff working in this experimental route) and inexperienced drivers were selected was to determine whether the familiarity with the tunnels would decrease driver stress and lead to a safer speed range. The correlations between heart rate and speed were experimentally verified. In the final part of the paper, the NARX neural network was adopted to predict their longitudinal speeds based on the heart rate while driving through the extra-long highway tunnels.

The specific research questions were as follows: How does an extra-long highway tunnel affect driver stress and speed in the entrance zone, the exit zone, and entire tunnel area? Did all drivers reduce speed and exhibit increased heart rate in this driving environment? Did the familiarity with the route safely impact driver speed and stress? Could the NARX model simulate driver’s real-time speeds with higher reliability and validity?

2 MATERIALS AND METHODS

2.1 Experiment Route

EXTRA-LONG tunnels on the Lan-Shang (Lantian - Shangluo) highway, situated in the Shaanxi Province, were selected as an experiment route. The Lan-Shang highway is bi-directional with four lanes, consisting of 226 bridges and 35 tunnels (the accumulated length of tunnels in a single direction was 37.13 km). There are two extra-long tunnels located on this highway (namely, the Lijiahe Tunnel 3# with 4259 m and the Qinling Tunnel with 4748 m). Both the Lijiahe Tunnel 3# and Qinling Tunnel are extra-long unidirectional tunnels with two lanes. The speed limit within the tunnel is 60 km/h, and on an open expressway, it is 80 km/h. This experiment route gives an overall reflection of the influence caused by tunnel environment on the drivers' ECG signals, because it runs through the Qinling mountains. On-site observation of hourly traffic volumes indicated that the experiment route is mainly in a free flow status. In other words, drivers control the speed freely based on their driving style, vehicle conditions, and road conditions, without being affected by traffic.

2.2 Subjects

There were several criteria for recruiting subjects due to the potential danger triggered by the tunnel environment. Subjects were asked to have no less than three driving years of driving practice and have no serious accident experience (no more than 10000 RMB economic losses or someone dying in the crash) before. In total, 60 male subjects (no female subjects) participated in this study, including 30 experienced drivers (staff responsible for the management of two tunnels) and 30 inexperienced drivers (no driving experience with this experiment route before). Experienced drivers had an average age of 35.7 years old with an average of 14.2 driving years. Simultaneously, inexperienced drivers had an average age of 39.2 years old with an average of 13.6 driving years. Thus, these two types of drivers were almost balanced in their age and driving years. The detailed information of all participants is listed in Table 1.

2.3 Apparatus

The real-time speed was collected using the OBD (on-board diagnostic) system and saved in the SCM (single-chip microcomputer) software. Synchronized ECG data from the subjects during the experiment were recorded using the MP150 (16-channel multifunctional physiological recorder produced by BIOPAC, a United States company). The ECG module (ECG100C amplifier) was used, and the obtained ECG data can be analyzed either online or offline via an additional software AcqKnowledge.

<table>
<thead>
<tr>
<th>Driver attribute</th>
<th>Age (years old)</th>
<th>Actual driving experience (years)</th>
<th>Driving distance/year (ten thousand kilometers)</th>
<th>Familiar with the experiment route or not</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experienced</td>
<td>35.7±9.8</td>
<td>14.2±9.7</td>
<td>4.8±2.6</td>
<td>Familiar</td>
</tr>
<tr>
<td>Inexperienced</td>
<td>39.2±11.3</td>
<td>13.6±12.0</td>
<td>1.9±0.9</td>
<td>Not familiar</td>
</tr>
</tbody>
</table>

Table 1. Participants in this Study (Mean ± Standard Deviation).
2.4 Data Pre-processing

In accordance with the provisions in the Specification of Highway Route Design (JTG D20-2006) (Ministry of Transport of the People's Republic of China, 2006) on the road alignment, the lines outside the tunnel entrance should be well aligned with those inside. The plane lines outside the tunnel entrance within longer than 3s travel distance, and those inside the entrance within longer than 3s travel distance in design speed should be consistent, while the requirements for those vertical lines dictated that the traveling distance could be longer than 5s. Based on the fact that 80 km/h is the highest speed limit on an open expressway, the 3s travel distance and 5s travel distance were approximately 67 m and 111 m, respectively. In addition, the tunnel was divided into the entrance zone (within 250 m before and after the entrance), exit zone (within 250 m before and after the exit) and internal zone (the remaining area in the tunnel) when Yeung and Wong analyzed the road traffic accidents in Singapore expressway tunnels (Yeung and Wong, 2013).

Combined with the above mentioned specifications, the current study extended the scope to 300m outside the tunnel portal and divided each extra-long tunnel into entrance zone, internal zone, and exit zone. Specifically, the entrance zone consisted of 300m before the portal and 300m immediately after entering the tunnel. The exit zone was comprised of 300m before the tunnel exit and 300m after the exit of the tunnel. The remaining middle section was the internal zone.

To reflect the universality of the influence caused by the extra-long tunnel environment on the ECG signals, excluding the impact of certain factors such as the road shape and the safety facilities, both the selection of the typical tunnels and the data processing methods were adjusted. In detail, the entrance zone included both the Lijiahe Tunnel 3# and Qinling Tunnel in two directions (south direction and north direction). Similarly, the internal zone and exit zone in two directions of both the Lijiahe Tunnel 3# and Qinling tunnel were considered to generalize the ECG variation characteristics in the overall extra-long tunnel environment rather than a particular direction or condition. All subjects experienced a statistic before-experiment test to see if their heart rates were in a normal range for 60 to 100 beats per minute. The mean heart rate was presented since we adopted a within-subject measure (driving in both tunnels and open expressway).

We need to obtain the mean index to characterize the change at a specific position, because we obtained the data for both tunnels in both directions. Therefore, the method of cutting the traffic interval was proposed based on the above mentioned tunnel zone divisions. The entrance zone and the exit zone contain 12 traveling intervals with a range of 50m. In total, 20 consecutive points were selected to divide the internal zone into 19 traveling intervals with the same division rule adopted in an open expressway. Thus, in a way, the entire tunnel was composed of 43 intervals. Finally, the mean indicator for the same interval corresponding to both directions in both tunnels was calculated. Specifically, the mean value at an interval of 20m was obtained by Orthogonal Hermite Interpolation Method in the entrance zone and exit zone. The division diagram of tunnel zones and traveling intervals are presented in Figure 1.

3 RESULTS

3.1 Variation of Mean Heart Rate in the Entrance Zone

The mean HR of experienced drivers varied during the process of passing the tunnel entrance zone. As shown in Figure 2, the mean HR variation could be divided into four stages. Considering that dark adaptation took more time than light adaptation, the ending point of dark adaption could be represented by the second mean HR dropping point or the first mean HR stable point occurring after the first dropping point when drivers entered the tunnel portal. It could be observed that the first mean HR dropping point occurred at 20m after entering the tunnel portal. Hence, the impact of the entrance zone on the mean HR appeared after the first dropping point when drivers entered the tunnel portal. It could be observed that the first mean HR dropping point occurred at 20m after the tunnel entrance, followed by the first mean HR stable point at 180m. Thus, the position ranges of the entrance zone with an impact on experienced drivers appeared from 300m before and 180m after the portal. Simultaneously, within the range of 20m before and after the portal, the mean HR had a staggering growth increase of 0.47%.

The mean HR variation for inexperienced drivers when they passed the entrance zone could also be distributed into four stages, seen in Figure 3. The mean HR exhibited the second drop at 220m after entering the tunnel portal. Hence, the impact of the entrance zone on the mean HR of inexperienced
drivers mainly occurred from 300m before to 220m after the portal. Within the range of 120m before to 120m after the portal, the mean HR increased continuously by 1.76%.

### 3.2 Variation of Mean Heart Rate in the Exit Zone

The variation of the mean HR for experienced drivers in the exit zone could be divided into three stages, shown in Figure 4. Light adaptation is more rapid than dark adaptation. Therefore, the first mean HR dropping point occurred when drivers exited the tunnel, representing the end point of light adaptation. The impacting positions on the mean HR of experienced drivers appeared from 300m before the exit to 220m after the exit. Within the range of 80m before to 80m after the exit, the mean HR grew continuously by 1.94%.

The mean HR of inexperienced drivers when passing the exit zone could be distributed into four stages, seen in Figure 5. The mean HR of drivers fluctuated and increased due to the environment change exerted on their physiological index, although the mean HR slightly dropped between the 300m and 80m before the exit. Thus, the impacting range of the exit zone on the mean HR of inexperienced drivers apparently occurred from 300m before to 80m after the exit. Within the range of 80m before and after the exit, the mean HR went up continuously by 2.10%. 

---

**Figure 2.** Mean Heart Rate in the Tunnel Entrance Zone for Experienced Drivers.

**Figure 3.** Mean Heart Rate in the Tunnel Entrance Zone for Inexperienced Drivers.

**Figure 4.** Mean Heart Rate in the Tunnel Exit Zone for Experienced Drivers.

**Figure 5.** Mean Heart Rate in the Tunnel Exit Zone for Inexperienced Drivers.
3.3 Variation of Mean Heart Rate in Tunnels and Open Expressways

Figure 6 presents the mean HR variation for the two types of drivers when they passed through extra-long tunnels. Concerning the traveling direction, experienced drivers had a lower mean HR than that of inexperienced drivers at all intervals. In the process of entering and exiting the tunnel, the mean HR of both types of drivers began to increase farther from the entrance but closer to the exit. The curve of the mean HR for these two types of drivers presented a tendency of increase in the entrance zone, decreasing in the internal zone and again increasing in the exit zone. The mean HR had larger fluctuations in transition areas (entrance zone and exit zone) but mild fluctuations in the internal zone.

Due to their familiarity with the experimental route, experienced drivers had a lower mean HR than inexperienced drivers in all road sections, shown in Figure 7. Regarding tunnel sections, experienced drivers showed the highest mean HR in the entrance zone, while no apparent variation was found in the internal and exit zones. The mean HR of inexperienced drivers, however, who gradually adapted to the outside environment, was highest in the entrance zone and lowest in the exit zone. Both types of drivers were observed to have the highest mean HR in open expressways rather than in tunnels.

3.4 Variation of Mean Speed in the Entrance Zone

Figure 8 illustrates the mean speed variation at all entrance zone position points for experienced and inexperienced drivers. Little impact occurred in the range 100m from the entrance on the mean speed of experienced drivers, and the mean speed remained stable. From 100m to 20m before the portal, the mean speed increased by 6.20%; the mean speed remained steady from 20m before and 20m after the portal; and finally, the mean speed continued to increase. The mean speed variation for experienced drivers in the entrance zone showed a trend of increase-stable-increase.

Inexperienced drivers slowed down at a distance 300m to 80m before the portal, showing a drop of 5.61%; from 80m before to 120m after the portal, the mean speed remained stable; and then, the mean speed declined continuously by 4.10% until 220m after entering the tunnel. The mean speed variation for inexperienced drivers in the tunnel zone presented a trend of decrease-stable-decrease. The mean speed increased again once drivers were adapted to the tunnel environment, and this position point was defined as the ending point of dark adaptation. As a result, the impacting position of the entrance zone on the mean speed of inexperienced drivers occurred at the distance range from 300m before and 200m after the portal. Within the range of 20m before and after the portal, the mean speed dropped slightly. In the entrance zone, the variation of the mean speed of experienced drivers was twice than that of inexperienced drivers.
Figure 8. Mean Speed in the Tunnel Entrance Zone for Both Types of Drivers.

3.5 Variation of Mean Speed in the Exit Zone

As seen in Figure 9, during the process of approaching and leaving the tunnel exit, the mean speed of experienced drivers dropped slightly by 2.46%, and then, it dramatically and continuously declined by 16.00% at 20m from the exit until leaving the exit zone.

Inexperienced drivers within the range of 120m to the tunnel exit had a steady mean speed; from 120m before to 20m after the exit, the mean speed decreased steadily by 2.41%; and then, the mean speed plunged by 4.17% until 80m after the exit. Once drivers were adapted to the light, their mean speed began to increase markedly. The mean speed variation of inexperienced drivers in the exit zone showed a trend of decrease – plummet – increase. The ending point of the light adaptation process was represented by the position that witnessed a tendency of acceleration after the exit. As a result, the position ranges of the exit zone that impacted the mean speed of inexperienced drivers occurred at a distance from 120m before to 80m after the exit. Within the range of 20m before and after the exit, the mean speed of inexperienced drivers had a slight decline. In the entire exit zone, the mean speed variation for experienced drivers was twice of that of inexperienced drivers.

3.6 Variation of Mean Speed in Tunnels and Open Expressways

Figure 10 represents the mean speed of the two types of drivers in all tunnel intervals. With respect to the traveling direction, the mean speed of experienced drivers showed a trend of increase-stable-decrease. Inexperienced drivers, however, decelerated in the entrance zone, accelerated at a stable value in the internal zone, decelerated slightly in the exit zone, and accelerated again after exiting the tunnel. The mean speed of inexperienced drivers was lower than that of experienced drivers in each internal traveling zone interval.

Figure 9. Mean Speed in the Tunnel Exit Zone for Both Types of Drivers.

Figure 10. Mean Speed in the Entire Tunnel for Both Types of Drivers.
As seen in Figure 7, in the tunnel sections, all drivers had the minimum mean speed in the entrance zone, while experienced drivers showed the maximum mean speed in the internal zone, and inexperienced drivers showed the maximum mean speed in the exit zone. When compared to tunnels, open expressway drivers traveled at higher mean speeds.

### 3.7 Correlations between Heart Rate And Speed

There was no significant correlation between the driver’s heart rate and speed when the data of all participants driving in tunnels and on highways were considered. However, the heart rate correlated negatively with speed ($r = -0.109$, $p < 0.01$) when only the tunnel data was considered. Still, no significant correlation was found in the highway data including both experienced and inexperienced drivers.

When drivers and road sections were taken into consideration separately, the heart rate of the experienced drivers in tunnels suggested a significantly negative correlation with speed ($r = -0.115$, $p < 0.01$). However, the heart rate of experienced drivers on highways correlated positively with speed ($r = 0.275$, $p < 0.01$). Also, the heart rate of inexperienced drivers on highways manifested a negative correlation with speed ($r = -0.234$, $p < 0.01$), although no significant correlation between these two indicators in tunnels was observed.

### 3.8 Speed Prediction Using NARX Models

NARX neural network describes a discrete nonlinear system by using past input and output data (Narendra and Parthasarathy, 1990; Nissinen et al., 1999). As seen in Figure 11, the expression is shown as follows:

$$\hat{y}(t) = f[y(t-1), y(t-2), \ldots, y(t-n_y), x(t-1), x(t-2), \ldots, x(t-n_x)]$$  \hspace{1cm} (1),

Where $x(t)$ and $y(t)$ represent the input and output values. $1:2$ represents the time step of delay and $W$ represents the link weights. $b$ represents the threshold. That is to say, the next $y(t)$ value depends on the previous $y(t)$ and the previous $x(t)$.

Regarding the data pre-processing, the mean heart rate and speed were all mean values originating from the traveling intervals. Thus, it was necessary to convert the spatial data into time series data. Because the original data was based on the high sampling rate, the value of heart rate and speed value at 1m interval could be obtained by Hermite interpolation. The heart rate correlated with the speed in tunnels based on the previous analysis, so the NARX model could be used to predict the driver’s speed using the heart rate as the input variable.

The training effect of NARX neural network was shown in Figure 12. It can be seen from Figure 12(1) that the MSE (Mean Squared Error) started to increase after the 76 time-step, which proved that the training process could be ended and the error of the whole data set was 0.0031464. Figure 12(2) shows the time-series response for the target data and predicted data. The fitting coefficient of the training set, the validation set, the test set and the whole data set was $R = 0.99986$, $R = 0.99987$, $R = 0.99983$ and $R = 0.99986$ respectively. Therefore, it could be concluded that the NARX neural network model was effective in predicting the speed of the drivers when they drove through the tunnel section.

### 4 CONCLUSION

The objective of this study was to use field-operational tests to determine whether a driver’s behavior was affected when experienced drivers (familiar with the experimental route) and inexperienced drivers (unfamiliar with the experimental route) drove through extra-long highway tunnels and on open expressways. Two tunnels in both directions (north to south and south to north) were selected to conduct driving tests with real-time ECG signals, and driving speeds were synchronously collected. Finally, the mean indicator derived from traveling intervals was used to characterize the variation at specific positions.

Compared to experienced drivers, the impacting positions on inexperienced drivers were wider at the entrance zone but were narrower at the exit zone. In the process of entering and exiting the tunnel, the mean HR of both types of drivers began to increase at a distance far from the entrance but close to the exit, which demonstrated that the entrance and exit zones caused different impacting patterns on driver stress levels.
The mean speed variation for experienced drivers in the entrance zone showed a trend of increase-stable-increase, which implied that drivers did not consider the entrance zone as a high-risk area and thus decelerate to avert any potential accident when familiar with the experimental route. Experienced drivers selected in the current study were staff in the tunnel management organization, so it was expected that the mean speed would increase when the depressed driving environment transitioned to the open expressway. However, the mean speed did not show such a trend. One possible reason was that experienced drivers might know about the hidden speed measurement devices installed at the exit zone.

Inexperienced drivers who were unfamiliar with the experimental route slowed down in advance at a distance farther from the tunnel entrance but closer to the exit. This psychological tension compensates for predicted risks. Simultaneously, in both the entrance and exit zones, the mean speed variation of experienced drivers was much more than that of the inexperienced drivers. Within the range of 20m before and after the portal (either at the entrance or the exit), inexperienced drivers decelerated slightly. In comparison with the entrance zone, the impacting area of the exit zone on the mean speed of inexperienced drivers was narrower; representing that light adaptation was faster than dark adaptation.

Experienced drivers who were familiar with road conditions showed a lower mean HR and drove faster in the exit zone than in the entrance and internal zones. All drivers showed the highest mean HR and the maximum mean speed on the open expressway, which may explain why the open expressway was thought to be safer but produces higher stress and behavioral risks. The results of Person’s correlation also revealed that both experienced and inexperienced drivers had different behavioral patterns between heart rate and speed on highways and in tunnels, which indicated that different drivers tended to respond to the two driving environments differently.

The NARX nonlinear neural network provides an efficient and innovative method for predicting driver’s speeds based on the heart rate. These conclusions were beneficial to delineate individual differences in driver behavior model under specific driving environment, providing a basis for constructing automated driving model or the development and design for more and more advanced in-vehicle technologies.

4.1 Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 51108036; No. 51178054), the Special Fund for Basic Scientific Research of Central Colleges, Chang’an University (No. 310812161012; No. 310822161005), China Scholarship Council, and the Science and Technology Project of Shaanxi Department of Transportation (No. 15-12K).

5 REFERENCES

A. Calvi, M. R. De Blasis, and C. Guattari. (2012). An empirical study of the effects of road tunnel on...
A. Lareshy, Å. Svensson, and C. Hydén, (2010). Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. Accident Analysis & Prevention, 42(6), 1637-1646.
M. Manser & A. Riener, (2014, September). Evaluation of Driver Stress while Transiting Road Tunnels. In Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 1-6). ACM.

6 DISCLOSURE STATEMENT
No potential conflict of interest was reported by the authors.

7 NOTES OF CONTRIBUTORS
Ling Wu is currently working towards a Ph.D. degree in Transportation Engineering (Vehicle Operation Engineering) at the School of Automobiles, Chang’an University, Xi’an, Shaanxi, China. She mainly focuses on driver behavior analysis and modelling, like driver distraction, and car-following behavior.

Haoxue Liu works as a research professor in Transportation Engineering (Vehicle Operation Engineering) at the School of Automobiles, Chang’an University, Xi’an, Shaanxi, China. His research interest mainly focuses on traffic safety and driver behavior.
Tong Zhu is now working as an associate professor in Transportation Engineering (Vehicle Operation Engineering) at the School of Automobiles, Chang’an University, Xi’an, Shaanxi, China. His research interest mainly concentrates on driver psychology and driver behavior analysis.

Yueqi Hu is working towards a Ph.D. degree in Transportation Engineering (Vehicle Operation Engineering) at the School of Automobiles, Chang’an University, Xi’an, Shaanxi, China. Her research interest is driver distraction.
The Data Analyses of a Vertical Storage Tank using Finite Element SOFT Computing

Lin Gao, Mingzhen Wang

College of Architecture Engineering, Chongqing University of Arts and Sciences, Chongqing, China

ABSTRACT
With the rapid development of the petrochemical industry, the number of large-scale oil storage tanks has increased significantly, and many storage tanks are located in potential seismic regions. It is very necessary to analyze seismic response of oil storage tanks since their damage in an earthquake can lead to serious disasters and losses. In this paper, three models of vertical cylindrical oil storage tank in different sizes, which are commonly used in practical engineering are established. The dynamic characteristics, sloshing wave height and hydrodynamic pressure of the oil tank considering the liquid-structure coupling effect are analyzed by using ADINA finite element software, which are compared with the result of the standard method. The close numerical values of both results have verified the correctness and reliability of finite element model. The analytic results show that liquid sloshing wave height is basically in direct proportion to ground motion peak acceleration, the standard method of portion sloshing wave height calculation is not conservative. The hydrodynamic pressure generated by liquid sloshing caused by ground motion is not negligible compared with the hydrostatic pressure. The tank radius and oil height have a significant effect on the numerical value of hydrodynamic pressure. The ratio of the hydrodynamic pressure and hydrostatic pressure, which is named hydraulic pressure increase coefficients, is related to the height, which given by the GB 50341-2014 code in China have a high reliability. The seismic performances of tank wall near the bottom needs to be enhanced and improved in the seismic design of the oil tank.

KEY WORDS: Vertical storage tank, ADINA, Dynamic characteristic, Hydrodynamic pressure, Sloshing wave height

1 INTRODUCTION
An oil storage tank is a container to store oil. It is the main equipment in petrochemical engineering. Storage tanks can be divided into various forms by material, location, installation and shape. Among them, the vertical cylindrical oil storage tank is widely applied. All the previous earthquake disasters show that oil tanks are easily damaged in strong earthquakes. The breakage of oil tanks will not only lead to damages in its own structure direct economic losses, but also cause greater indirect economic losses and serious secondary disasters. Apart from its own inertia force and hydrostatic pressure, when the oil tank is subjected to seismic actions, the hydrodynamic pressure generated by the liquid sloshing in the tank is also the main damage reason of the tank. Therefore, it’s necessary to carry out seismic response analyses on the oil storage tank to study dynamic characteristics, oil sloshing wave height and hydrodynamic pressure distribution of the oil tank. At present, the acceleration, stress and displacement distributions of oil tanks under the actions of earthquakes are analyzed, but the study of sloshing wave height and hydrodynamic pressure are slightly less. In 1963, the hydrodynamic pressure of tanks assumed to be rigid under seismic excitation was analyzed by Housner (1963). The contribution of hydrodynamic pressure on a tank wall was divided into the impulsive pressure and convective pressure. Since the tanks were not absolutely rigid, Haroun and Housner (1981) showed that the hydrodynamic pressure was influenced significantly by the flexible behavior of tank walls. Veletsos (1974), Veletsos and Younan (1998) assumed that liquid storage tanks...
under horizontal ground motions were like a cantilever beam and the effects of the flexible tank wall was considered by applying a cantilever beam type mode to the equation of motion for the storage tank.

The seismic-isolating cylindrical liquid storage tanks were studied considering of fluid-structure interaction. Shekari, et. al. (2009) observed that seismic isolation was more effective in slender tanks in comparison with broad tanks and the liquid surface displacement increases due to seismic isolation especially in slender tanks. Ghaemmaghami and Kianoush (2010), Kianoush and Ghaemmaghami (2011), Hashemi, et. al. (2013) studied the dynamic response of the concrete rectangular liquid containing structure by using the theory and finite element method considering its three dimensional space structure and soil structure interaction. Moslemi and Kianoush (2012) carried out parametric studies on the dynamic behavior of cylindrical ground-supported tanks. Park, et. al. (2016) conducted experiments to study the dynamic behavior of a cylindrical liquid storage tank subjected to seismic excitation. The dynamic test results of the dynamic behavior characteristics including beam-type and oval-type vibration of a cylindrical liquid storage tank under horizontal earthquake excitation were obtained. Goudarzi and Danesh (2016) carried out a numerical investigation of a vertically baffled tank under seismic excitation. The results of reduction in sloshing wave height caused by the baffles were estimated for selected tanks subjected to the seismic excitations. A simple procedure to estimate the reduction in sloshing amplitude due to the presence of baffles was proposed and validated using the time history numerical results. At present, using software for computational analyses are the main research method, like the reference Tian and Jiao (2016). Manser, et. al. (2017) studied the maximum sloshing wave height in cylindrical metallic tanks by using ANSYS V11.0, and the results were compared with the Euro code 8.

In this paper, the ADINA finite element software is used to analyze the dynamic characteristics and seismic response of the vertical cylindrical storage tanks since ADINA software has powerful solvers and advantages in dealing with liquid-solid coupling problems, as described in the reference Gao, (2015). Three actual ground motion records on the Wenchuan earthquake are used to analyze the maximum sloshing wave height, the distributions of hydrodynamic pressure and the effective stress of the tank under earthquake action. And the finite element analysis results are compared with theoretical calculation results.

2 FLUID-SOLID COUPLING THEORY

THE dynamic effects caused by liquid shaking under earthquakes are applied to the tank body, and the tank body will generate a series changes simultaneously, including stress, displacement and shape. The existing analyses show that with the contacting time of the tank and the liquid in it increases, the interaction between the solid and liquid changed too, which refers to the mutual liquid-solid coupling problem. The liquid-solid coupling problems between tank and oil are analyzed in this paper.

3 TANK MODEL

THE large, medium and small-capacity oil tanks are selected, which are common in practical engineering. The capacity is 2000m³, 1000m³ and 500m³ respectively. The specific parameters of three oil tanks are listed in Table 1. The purpose of this setting is to compare the effects of different parameters on the seismic response of oil tanks. The abbreviations are set for three oil tanks for convenience. The capacity 500m³ is called tank A for example. The oil height is 90% of the total tank height. The tank bottom and wall are set as shell units. The oil in the tank is set as 3-D potential fluid unit considering its liquid sloshing characteristic. The unit assumes that the fluid is non-viscous, non-rotational, non-heat exchange and incompressible. To get more accurate results, linear potential-based fluid elements are used in the static and modal analyses, and subsonic potential-based fluid elements are used in the dynamic time history analyses. The subsonic potential-based fluid elements can better reflect the liquid movement under dynamic state.

After setting the unit types of oil tank, the material properties are determined. The properties of steel material are as follows: The elastic modulus is $2 \times 10^{11}$ N/m², the density is 7800 kg/m³ and the Poisson ratio is 0.3. The properties of liquid material are as follows: The bulk modulus is $1.767 \times 10^9$ N/m², the density is 812 kg/m³ and the damping ratio is 0.005.

The mesh of liquid part is divided into a copper type. Each side is divided into 18 copies, and the height direction is divided into one copy per 0.3 meters. The mesh at the intersection of the tank and liquid is the same. That is the common node. The upper tank wall on top of the liquid surface is divided into one copy per 0.3 meters along the height direction. Take tank B as an example, the mesh of tank and liquid are shown in Fig.1.

4 DYNAMIC CHARACTERISTIC ANALYSES

THE dynamic characteristics of oil tanks are analyzed previous to the seismic dynamic time history analyses. The basic period calculation formula is given in Code (2014) for Design of Vertical Cylindrical Welded Steel Oil Tanks (GB 50341-2014). The finite element and the standard calculation results are compared to verify the correctness of the established model. The basic period results of the three methods for three tanks are listed in Table 2.
Table 1. The Specific Parameters of Three Oil Tanks.

<table>
<thead>
<tr>
<th>Tank Abbreviation</th>
<th>Capacity (m$^3$)</th>
<th>Radius (m)</th>
<th>Total height (m)</th>
<th>Bottom thickness (mm)</th>
<th>Wall thickness (mm)</th>
<th>Storage oil height (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tank A</td>
<td>500</td>
<td>3.75</td>
<td>12.00</td>
<td>6</td>
<td>7</td>
<td>10.8</td>
</tr>
<tr>
<td>Tank B</td>
<td>1000</td>
<td>7.50</td>
<td>6.00</td>
<td>6</td>
<td>7</td>
<td>5.4</td>
</tr>
<tr>
<td>Tank C</td>
<td>2000</td>
<td>7.50</td>
<td>12.00</td>
<td>6</td>
<td>7</td>
<td>10.8</td>
</tr>
</tbody>
</table>

Table 2. The Basic Period Results of Two Methods for Three Tanks.

<table>
<thead>
<tr>
<th>Order</th>
<th>First order</th>
<th>Second order</th>
<th>Third order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard method</td>
<td>ADINA method</td>
<td></td>
</tr>
<tr>
<td>Tank A</td>
<td>2.867</td>
<td>2.863</td>
<td>1.675</td>
</tr>
<tr>
<td>Tank B</td>
<td>4.357</td>
<td>4.346</td>
<td>2.377</td>
</tr>
<tr>
<td>Tank C</td>
<td>4.076</td>
<td>4.071</td>
<td>2.376</td>
</tr>
</tbody>
</table>

The basic periods calculated results of two methods for three tanks are almost the same as can be seen from Table 2. The fore three orders sloshing periods of three tanks are also listed in Table 2. The fore three orders liquid vibratory types of tank B are displayed in Fig. 2.
The first order liquid sloshing period is the maximum among all sloshing periods for the same oil tank from Table 2. And with the order increase, the liquid sloshing periods reduce gradually. For the same modal order, when the storage oil heights are the same, the larger the tank radius, the longer the sloshing period. When the tank radiiues are the same, the oil height decreases, and the sloshing period increases. But when the order increases, the differences between the adjacent two periods are less significant. The above analyses show that the tank radius is the most important factor affecting the liquid sloshing period. As seen from Fig.2, liquid sloshing is violent, and deformation of the tank wall, the liquid is small compared to that of liquid. The low-frequency vibration of the oil tank is liquid sloshing. The liquid sloshing period is longish. The modal quality participation percentage of the first order is the maximum and has the greatest impact on the whole system. All the liquid sloshing vibration modes present the \( \cos \theta \) beam vibration type.

5 SEISMIC RESPONSE ANALYSES

The Rayleigh damping coefficients are calculated for seismic response analyses of the oil tank according to the dynamic characteristics analyses results. For the problems of nonlinear, long duration and large deformation, the Bathe time iteration method in ADINA software is used, because of its good accuracy and stability. Three groups of natural ground motion records named Heishuidiban, Baoji and Chencang in the Wenchuan earthquake are selected to simulate the dynamic response of oil tank under actual ground motion. The total duration of all ground motions are 30 seconds. The peak acceleration of all ground motions are adjusted to 100gal, 200gal and 400gal. The basic information of ground motions are shown in Table 3. The ground motions are both input to the X direction of models. And the results of each condition are extracted respectively after calculations.

5.1 Sloshing Wave Height

The wave height time history and stress nephogram graph of the maximum sloshing wave height in each working condition are extracted. Taking tank B as an example, the time history curve of the sloshing wave height under Heishuidiban ground motion at different acceleration peaks are shown in Fig.3. And the stress nephogram graph of the maximum wave height at 400gal peak acceleration is drawn in Fig. 4. The maximum shaking wave heights of the three tanks under unidirectional ground motion are listed in Table 4.

<table>
<thead>
<tr>
<th>Name of ground motion</th>
<th>Predominant period (s)</th>
<th>Site classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heishuidiban</td>
<td>0.095</td>
<td>II</td>
</tr>
<tr>
<td>Baoji</td>
<td>0.611</td>
<td>III</td>
</tr>
<tr>
<td>Chencang</td>
<td>1.138</td>
<td>III</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ground motion</th>
<th>Tank A</th>
<th>Tank B</th>
<th>Tank C</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSDB-EW-100gal</td>
<td>0.1048</td>
<td>0.0932</td>
<td>0.0870</td>
</tr>
<tr>
<td>HSDB-EW-200gal</td>
<td>0.2126</td>
<td>0.1879</td>
<td>0.1791</td>
</tr>
<tr>
<td>HSDB-EW-400gal</td>
<td>0.4282</td>
<td>0.3771</td>
<td>0.3641</td>
</tr>
<tr>
<td>BJ-EW-100gal</td>
<td>0.3495</td>
<td>0.3619</td>
<td>0.3694</td>
</tr>
<tr>
<td>BJ-EW-200gal</td>
<td>0.7020</td>
<td>0.7257</td>
<td>0.7440</td>
</tr>
<tr>
<td>BJ-EW-400gal</td>
<td>1.4046</td>
<td>1.4523</td>
<td>1.4935</td>
</tr>
<tr>
<td>CC-NS-100gal</td>
<td>0.7795</td>
<td>0.4478</td>
<td>0.4281</td>
</tr>
<tr>
<td>CC-NS-200gal</td>
<td>1.0410</td>
<td>0.8967</td>
<td>0.8609</td>
</tr>
<tr>
<td>CC-NS-400gal</td>
<td>2.0844</td>
<td>1.7933</td>
<td>1.7249</td>
</tr>
</tbody>
</table>
By comparative analysis, it is found that when the tank is not damaged, the ratio of sloshing wave heights for three tanks under 200gal peak acceleration and 100gal peak acceleration is close to two, and the ratio of sloshing wave heights for three tanks under 400gal peak acceleration and 100gal peak acceleration is close to four. That means the sloshing wave height is basically in direct proportion to the peak acceleration of ground motion. At the same time, the calculated results also show that the sloshing degrees of oil under different ground motions are different. The degree of surface sloshing under the long period ground motion is greater than that under the short period ground motion. The liquid sloshing under the ground motion actions could not be ignored since the maximum wave height is more than two meters in some conditions. The calculation formula of sloshing wave height in the oil tank under the horizontal earthquake action is given by note D.3.9 Specification for Design of Vertical Cylindrical Welded Steel Oil Tanks (GB 50341-2014), which is shown in Eq.1. The calculated damping ratio of oil is 0.005.

\[ h_x = 1.5\eta\omega R \]  

(1)

The sloshing wave height results of the finite element and the standard method for three tanks at 100gal peak acceleration are shown in Table 5.

<table>
<thead>
<tr>
<th>Tank name</th>
<th>Method</th>
<th>Heishui</th>
<th>Baoji</th>
<th>Chencang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tank A</td>
<td>Finite</td>
<td>0.1048</td>
<td>0.3495</td>
<td>0.7795</td>
</tr>
<tr>
<td></td>
<td>Standard</td>
<td>0.3246</td>
<td>0.3444</td>
<td>0.3444</td>
</tr>
<tr>
<td>Tank B</td>
<td>Finite</td>
<td>0.0932</td>
<td>0.3619</td>
<td>0.4478</td>
</tr>
<tr>
<td></td>
<td>Standard</td>
<td>0.5307</td>
<td>0.5705</td>
<td>0.5705</td>
</tr>
<tr>
<td>Tank C</td>
<td>Finite</td>
<td>0.0870</td>
<td>0.3694</td>
<td>0.4281</td>
</tr>
<tr>
<td></td>
<td>Standard</td>
<td>0.5526</td>
<td>0.5925</td>
<td>0.5925</td>
</tr>
</tbody>
</table>

The sloshing wave height results of the finite element method are smaller than that of the standard method for three tanks at 100gal peak acceleration.

5.2 Hydrodynamic Pressure

Structure designers don’t concern the real-time values of the structural responses under earthquake actions. Instead, they care about the maximum values of the structural responses to be used in structure design. To the analyzed vertical storage tank in this paper, the hydrodynamic pressures for the tank wall and bottom in the direction of the unidirectional seismic action are the largest. The maximum hydrodynamic pressure distribution of the tank wall in the height direction and the bottom radial direction at different peak acceleration ground motions are shown in Fig.5.

![Figure 5. Hydrodynamic Pressure Distribution at Different Peak Acceleration Ground Motions for Tank B.](image)

It can be seen from Fig.5 that the hydrodynamic pressure at the same position is increased with the increase of the peak acceleration under the same ground motion, and the increase multiple of hydrodynamic pressure is the same as the increase multiple of the peak acceleration. When the other conditions are the same, the hydrodynamic pressures under the Heishuidiban ground motion are larger than pressures under the Baoji and Chencang ground motion. The hydrodynamic pressures under the Baoji and Chencang ground motion are almost the same. Although the hydrodynamic pressure values of the tank wall and bottom under different ground motions
are different, the curve shapes, namely the distributions of hydrodynamic pressure, are basically the same. The hydrodynamic pressure distribution trend of the tank wall in the height direction for tank B shows a large bottom and a small top in Fig. 5a. Under some ground motions, the hydrodynamic pressure of the liquid surface position is the minimum, and the hydrodynamic pressure at one meter below the liquid surface is the minimum. The hydrodynamic pressures don’t change obviously at about 2m above the bottom, and the hydrodynamic pressure reaches the maximum at about one meter above the bottom of the tank. This is one of the main reasons for buckling failure of the bottom part of the tank wall under seismic actions. In Fig. 5b, no matter what kind of ground motion actions, the maximum hydrodynamic pressures along the radial direction are almost symmetrical about the center of the tank bottom. The hydrodynamic pressures at the center of the bottom are almost equal at different peak accelerations and in different ground motions. There is a positive correlation between the hydrodynamic pressure and the peak acceleration of ground motions under the different peak accelerations and in the same ground motion. Therefore, only the calculation results under the 100gal peak acceleration are employed at a later study.

In order to analyze the effect of tank radius to the hydrodynamic pressure distribution, the tank wall hydrodynamic pressure distributions under the same ground motions of tank A and tank C with the same heights are compared and plotted in Fig. 6. In order to analyze the effect of oil height to the hydrodynamic pressure distribution, the tank bottom pressure distributions under the same ground motions of tank B and tank C with the same radius are compared and plotted in Fig. 7.

The tank wall hydrodynamic pressures of tank A under different ground motions are larger than that of tank C can be seen in Fig. 6, and the hydrodynamic pressures at the same position even increase more than one time under long-period ground motions such as the Heishuidiban and Chencang ground motion. When the oil height remains the same, the smaller the tank radius, the larger the hydrodynamic pressure is. The regularity that the hydrodynamic pressure of the bottom part of tank wall is large and the hydrodynamic pressure near the liquid surface is small keeps unchanged. From Fig. 7, the tank bottom hydrodynamic pressures of tank C are larger than that of tank B under different ground motions, and the pressures at the same position even increase more than two times. When the tank radius remains the same, the larger the storage oil height, the larger the hydrodynamic pressure. The regularity remains unchanged that the hydrodynamic pressure near the center of tank bottom is the minimum, the pressure at the radius position is the maximum and the distribution is linear from the center to the radius.

In order to compare the relative value between hydrostatic and hydrodynamic pressures, the hydrostatic pressure, and the hydrodynamic pressure under different peak accelerations of Heishuidiban ground motion and the total hydraulic pressure of tank B are compared and plotted in Fig. 8. As the hydrodynamic pressure under Heishuidiban ground motion of tank B is larger than that under other ground motions, only the hydrodynamic distribution under Heishuidiban ground motion is shown here.
Figure 8. Distribution Conditions of Hydrostatic Pressure, Hydrodynamic Pressure and the Total Hydraulic Pressure.

It can be seen from Fig. 8 that the hydrostatic pressure and hydrodynamic pressure near free liquid surface are small. With the increase of oil depth, the hydrostatic pressure increases linearly, the incremental rate of hydrodynamic pressure decreases gradually, and the hydrodynamic pressure at about one meter to the bottom reaches the maximum. The distribution of total hydraulic pressure obtained by the superposition of hydrostatic and hydrodynamic at the same depth is approximately linear. The larger the peak acceleration of ground motion, the larger the hydrodynamic pressure at the same position, and the increase multiple is seismic coefficient k. The hydrodynamic pressure generated by liquid sloshing under ground motion is not negligible compared with the hydrostatic pressure. Therefore, the hydrodynamic pressure must be taken into account in the oil tank seismic design, and the value of hydrodynamic pressure should be reasonably estimated. The ratio of hydrodynamic pressure and hydrostatic pressure under 100gal peak acceleration ground motion of the three tanks are shown in Fig.9, which also includes the value given by the D.3.7 of Specification (GB 50341-2014). The ratio of hydrodynamic pressure and hydrostatic pressure is referred as hydraulic pressure increase coefficient. Since the hydrostatic pressure at the liquid surface is zero, the hydraulic pressure increase coefficient cannot be calculated. So the only the increase coefficients of hydraulic pressures at 0.3 meters lower than the surface are calculated in Fig. 9.

It can be seen from Fig. 9 that the hydraulic pressure increase coefficient increases gradually from tank bottom to liquid surface regardless of tank types. And the increase rate rises at it get closer to the surface. The hydraulic pressure increase coefficients listed in the code are exactly the same along the height of tank wall for calculation convenience and partial conservative. The results show that the hydraulic pressure increase coefficients listed in the code exceed most results by the finite element calculation. The finite element results under partial ground motions are larger than the code results in the range of one meter below liquid surface. Although the partial finite element value exceeds the code value, the hydrodynamic pressure near the liquid surface is smaller, which does not cause a significant change of the total hydraulic pressure. The oil tank is relatively safe near the liquid surface.

Moreover, if the standard values are all beyond the finite element values, it will lead to larger hydraulic pressure, and result in serious imbalance between the economy and function in structural design. The hydraulic pressure increase coefficient by standard calculation still has a high reliability based on the analyses results.

6 CONCLUSIONS

ADINA finite element software is used to analyze the dynamic characteristics and seismic responses of vertical storage tanks in this paper. The following four conclusions are obtained.

(1) The sloshing periods and vibration modes of oil in tanks are obtained through the dynamic
characteristics analyses. The fundamental period of liquid sloshing is calculated using the standard formula given by the Specification and compared with the finite element calculation results. The two values are close to each other, and the correctness and reliability of the finite element model are verified, which lays the foundation for seismic response analyses.

(2) Three natural ground motions are used to analyze the seismic response of oil storage tank, and the distribution characteristics of sloshing wave height are studied: the sloshing wave height is basically in direct proportion to the peak acceleration of the ground motion when the ground motion is the same. The surface sloshing under long period ground motions is fiercer than that under short period ground motions. The sloshing wave height calculated by the standard method is somewhat unsafe in some cases.

(3) The calculation results of hydrodynamic pressure show that the hydrodynamic pressure caused by liquid sloshing cannot be neglected under ground motions. The incremental rate of hydrodynamic pressure decreases gradually in the up-down liquid height direction. The hydrodynamic pressure changes not much at two meters above the tank bottom, and the hydrodynamic pressure is the maximum at about one meter from the tank bottom. The maximum hydrodynamic pressure in the radial direction of the tank bottom is almost symmetrical about the bottom center. The hydrodynamic pressure of bottom center is the smallest and the hydrodynamic pressure of bottom radius is the largest. When the oil height remains the same, the smaller the tank radius, the larger the hydrodynamic pressure is. When the tank radius keeps the same, the larger the storage oil height, the larger the hydrodynamic pressure is. The hydraulic pressure increase coefficients given by the Specification have a high reliability.

(4) The seismic performance of the bottom part of tank wall should be strengthened and improved in the tank seismic design.

7 ACKNOWLEDGMENT

THIS work was supported by Scientific Research Fund of Chongqing University of Arts and Sciences (R2015JJ06, 2017RJ32). This work was supported by Scientific Research Fund of Chongqing Municipal Education Commission (KJ1601132).

8 REFERENCES


NOTES ON CONTRIBUTORS

Lin Gao received Ph.D. degree in Structure Engineering at the Institute of Engineering Mechanics, China Earthquake Administration. She works at the College of Architecture Engineering, Chongqing University of Arts and Sciences in China. She is an associate professor. Her research interests are finite element calculation simulation analysis of engineering structure.

Mingzhen Wang received Ph.D. degree in Structure Engineering at the Institute of Engineering Mechanics, China Earthquake Administration. He works at the College of Architecture Engineering, Chongqing University of Arts and Sciences in China as a Lecturer. His research interests are finite element calculation simulation analysis of engineering structure.
Association Link Network based Concept Learning in Patent Corpus

Wei Qin, Xiangfeng Luo

School of Computer Engineering and Science, Shanghai University, China

ABSTRACT

Concept learning has attracted considerable attention as a means to tackle problems of representation and learning corpus knowledge. In this paper, we investigate a challenging problem to automatically construct a patent concept learning model. Our model consists of two main processes; which is the acquisition of the initial concept graph and refined process for the initial concept graph. The learning algorithm of a patent concept graph is designed based on the Association Link Network (ALN). A concept is usually described by multiple documents utilizing ALN here in concept learning. We propose a mixture-ALN, which add links between documents and the lexical level, compared with the ALN. Then, a heuristic algorithm is proposed to refine the concept graph, leading to a more concise and simpler knowledge for the concept. The heuristic algorithm consists of four phases; first, for simplifying bag of words for concept in patent corpus, we start to select a core node from the initial concept graph. Second, for learning the association rule for the concept, we searched important association rules around the core node in our rules collection. Third, to ensure coherent semantics of the concept, we selected corresponding documents based on the selected association rules and words. Finally, for enriching semantics of the refined concept, we iteratively selected core nodes based on the corresponding documents and restarted our heuristic algorithm. In the experiments, our model shows effectiveness and improvements in prediction accuracy in the retrieve task of the patent.

KEY WORDS: Association link network, Concept learning, Heuristic algorithm, Patent corpus, Retrieve task

1 INTRODUCTION

WITH the explosion increase of Internet documents, millions of new concepts appear on the Internet as documents and provides a good chance for web user to learn knowledge on it. Every day millions of people search for information through computer, mobile phone and etc. by search engines such as PatentList, Baidu, and Google. In the standard IR algorithm, the vector space mode (VSM) is used to represent user query and documents, and some predefined score function is designed for selecting relevant documents based on similarity between query and document, which is considered as a concept refined method based on user query. However, the standard IR algorithm, which is a precision-oriented task, is where a user finds an answer to their information needs that can typically be addressed by one or two relevant concepts. Such methods cannot satisfy a user’s requirement, when the user is typically ready to check possibly hundreds of relevant documents to summarize the concept. For example; the user wants to know “automobile engine” in the patent corpus. The standard IR algorithm may only provide one or two aspects about the automobile engine such as engine power or engine fuel consumption. In this paper, we search an algorithm to construct and refine concepts, which can learn the knowledge in the patent corpus, which can better represent a patent concept. The refined concept is represented as a smaller set of a document list and several core keywords and association rules.

Existing methods of the relevant concept of the learning method such as Hjorland (2009), and the concept theory and learning method by Hammer, et al. (2009), have been widely used to represent specific corpus or domain. However, they have encountered the following challenges: 1) How to construct a concept representation that could be understand both by machine and user, 2) How to construct a user readable concept graph automatically without assistance from domain experts, 3) A number of state-of-the-art
concept learning methods relying on specific domain and model which transfer another domain, 4) Learning concept approaches, which were adopted by various literature will take exponential growth of time with the increase number of documents.

These problems appear more serious in a patent corpus than a normal corpus such as a new article. A patent is hard to understand, because it exists of large numbers of terminology, which is not well understood both with machine and user. The technology used in a patent is broad, even some has cross-domain technology that requires a multi-domain expert to extract the concept from it.

In this paper, we address a concept learning task as a construct for the initial concept graph and refine the concept problem, which obtains sever documents and word document rules to represent the concept. In order to rely on the above four problems as mentioned, first, we use an extension form of ALN, which we call it MALN, that makes it more convenient for constructing a scoring function. Second, a heuristic algorithm is proposed based on the MALN that consist of four phases: 1) Select Core Nodes Phase 2) Select Association Rules Collection Phase 3) Select Document Phase and 4) Feedback to the MALN Phase. In the following section all the phase will be described in detail.

To address such challenges in the concept learning, we propose our model and make the following contributions:

1) A small set of documents, which incorporate keywords and association rules is used to represent a specific concept, which both increase readability of people and machine.
2) The process of building the ALN is fully automated without a domain expert assistant, moreover, manual participation is not required for the heuristic process.
3) Our model can transplant to any other domain with few changes to the model, and a training set is unnecessary for our model.

The remainder of this paper is organized as follows: We continue by covering the related work in Section 2. Then, we introduce the basic definition and construction method of the ALN and how we expand to the MALN in Section 3. Section 4 describes procedures of our heuristic algorithm in detail. Section 5 shows the experiment to our concept learning model. Finally, we conclude in Section 6.

2 RELATED WORK

2.1 Concept Represent Model

DIFFERENT methods on concept learning have been proposed in literature, and can be divided into three categories; expert based methods, statistics machine learning methods, and user memory based methods.

The expert based methods (Aizawa, 2003), (Guthrie, et al., 2006) required experts to be familiar with a specific domain that contains concepts the people want to know. (Angluin, 1988) attempted using queries to learn an unknown concept. Several types of queries are used and studied in the supervised learning framework. Jong, K. A. D. (1975) explored the use of genetic algorithms as a key process in the implementation of the concept learning system. He assumes that conceptual learning should be agregated, which implies that people always focus on several core concept nodes in the process of learning. Some people consider that concept learning should be based on ontology rules (Rouder & Ratcliff, 2006)), which is designed by an expert. In this method, the learning direction decisions are made simply and rely on simple relevant rules without considering rules that exist in a rules set. These concept learning methods need to be a manual participant or at least a semi manual designation of some feature, which are limited to the particular field. The result of learning is only the relationship among the features or the features and concepts, which greatly increases the people’s cognitive burden.

The statistics machine learning methods model concept is by one or more vectors. The vector space mode (VSM) (Salton, 1971)) has been successfully applied to the famous SMART document retrieval system. Contents of the document are reduced to vector space operations, and it uses spatial similarity to express semantic similarity, is intuitive and easy to understand. PLSA (Kushilevitz, et al., 1998) is an index retrieval method. The method and the traditional vector space model (VSM) used as a vector to represent the word (terms) and document(s), and the relationship between vectors (such as angle) to determine the relationship between words and documents. Exception is, LSA, which will be mapped to the document words and latent semantic space, which in addition to some of the "noise" of original in the vector space, improves the precision of information retrieval. Word2vec (Mikolov, et. al., 2013) is a distributed representation other than One-hot Representation in the traditional concept representation such as VSM mentioned above. It grants certain semantic meaning to words, so that we could easily calculate the similarity between two vectors. Although this type of approach could be easily handled by a machine, it lacks an intuitive understanding of the concept. Simultaneously, some of these methods only have an initial representation for the concept such as VSM; others need training for vectors and are hard to transfer to another domain such as Word2vec.

Some research represents a graph-based concept, which is a more interpretable concept of the representation that can be understood by both user and machine (Hussain, et al. 2014). ALN (Luo, et al., 2011) is one of a typical user memory based method for
concept representation, which is more perceptual intuition. ALN could not only be defined by formal but, also can convert into a concept graph naturally. It is efficient and effective for the construct concept, and the graph structure makes it easier to interfere in the following task. We consider ALN as our initial concept representation, and extend it to MALN for applying to the heuristic algorithm.

2.2 Evaluation of Concept Learning

At present, the evaluation of concept learning has not formed unified evaluation criteria. Researchers usually use artificial methods to evaluate the quality of construction concepts, or evaluate them through conceptual applications. In this paper, we focus on the IR method, which is one of most popular application based on construction concepts. IR’s query could be considered as part terminology, which is augmented by finding a relevant document. Some research already uses an image retrieval task to evaluate the result of an image concept domain (Huang, Hu, et al., 2016). It is common practice to use heuristic rules to construct a document ranking list in a search. Usually, rules are created based on the observation about the relationship between query and document. Most of a heuristic is designed for advanced similarity scoring function between query and document in a corpus. The most known framework is OKAPI (Robertson & Zaragoza, 2009), the Scoring function fused word frequency, and the document length is featured into one function. Subsequently, Query reformulation (Mahdabi & Crestani, 2014) method was proposed to avoid noise existence in the query of IR task, which successfully is applied to news articles and patent prior art search, another study, using query expansion via feature selection (Zhang, et al., 2016). Both the query formulation method and its expansion have shown better performance compared to using the query as a core concept. However, the refined method of IR cannot learn a complete concept representation when the user query is incomplete, which frequently occurs. And, when the only document selected by the IR method cannot represent the true nature of the concept. Therefore, we propose a new heuristic algorithm applied to select documents, keywords, and rules about the concept.

3 BUILDING MIXTURE ASSOCIATION SEMANTIC LINK NETWORK FOR THE CONTENT REPRESENTATION OF CONCEPT

3.1 Frame Work

In this paper, we propose a model for automatic concept learning that has good representation and quality from the patents data. We use ALN to construct and represent the concept that people desire. Then we generate a summarization from the given paper by using a network last phase. The two phases we will describe subsequently. The whole process of our model is shown in Figure 1:

![Figure 1. Concept Learning Model Framework.](image)

The pre-process is uses standard Natural Language Processing steps, which reduced the document noise in the patent corpus, and normalized the document, we will discuss it in our experiment. The Constructed concept is not only a concept representation, but also is the foundation for the subsequent use of the heuristic algorithm for learning concepts. This step will be illustrated later on. The selected document obtained by a heuristic algorithm is the core of our model, which selects a document, rules and keywords that could properly represent the concept. It is able to learn concepts effectively from the constructed concept graph. This step discussed in Section 4.

3.2 Building an Association Semantic Link Network

The traditional concept of construction and establishment of may requirement for artificial participation, and some even completely rely on relevant field experts. Therefore, although the accuracy of conventional methods are relatively high and easy to understand, it is difficult to handle growing number of documents in the Internet, especially new concepts.

The Association Semantic Link Network (ALN), (Luo, et al., 2011) is a resource organization model, which is used for extracting a core semantic and store correspondence knowledge. Given a document list comes from a query, ALN automatically learns the document list representation based on the co-occurrence information of a word and topological of the graph structure. The key principle of the ALN is the document list representation be converted into a word-based graph. The ALN is defined in (1):

$$ALN = (N, L)$$

where \( N = \{n_i | 1 \leq i \leq m \} \) is the corresponding vector represented document, and keywords are extracted from the document. The weight associated with the keyword is calculated by TF-IDF, where TF is the frequency of the keyword in the document and IDF is the inverse document frequency of the keyword.
length of $N$ indicates the number of keywords that exist in the document. $L = [L_{ij}]_{n \times n}$ is a rule matrix defined in (2):

$$L_{ij} = \frac{\#(i,j)}{\#(i) \#(j)}$$  

Here $\#(i,j)$ is the number of times word $j$ appears in the context of word $i$, $\#(i)$ is $i$ IDF frequency, $\#(j)$ is $j$ IDF frequency.

After making the connection using formula 1, ALN optimized itself the structure according to the small world (Collins, & Chow, 1998) and the scale free network theory (Yoo, & Hu, 2006). As ALN represents the semantic association and has the ability of extracting the core semantic from the document, which is consistent with the concept definition. We have sufficient reasons to use production of the ALN as an initial representation of the concept.

3.3 Building a Mixture Association Link Network

The simplest form of the ALN only represents graphs of lexical layers, however for a document list, it is necessary to represent both lexical layers and document layers. Therefore, we slightly extended the definition of the original ALN, so that it has stronger ability to express the relationship between the documents and the lexical, it is a Mixture Association Link Network (MALN), which is defined in (3):

$$MALN = (S, N, L, D)$$  

where the definition of $N$, $L$ is the same as formula 1. $S = \{s_i | 1 \leq i \leq s\}$ is the corresponding vector of the document that is related to the concept. We consider each document having same importance so that the vector weight set is 1. Given the document collection $S$ and association rules collection $L$, we have a dependent relationship from $S$ to $L$, which is defined as $D = [D_{ij}]_{s \times r}$; and $D_{ij} = 1$ indicates that document $i$ has rule $j$ (Goodman, et al. 2010).

According to the definition, MALN can be converted to a multi-level graph, which the upper layer represents $S$, the bottom layer represents $N$, $L$, and $D$ is the Links for the two layers. The graph is a foundation used for the heuristic algorithm. Figure 2 gives the graph illustration of the MALN.

4 HEURISTIC ALGORITHM FOR A CONCEPT SELECTOR

OUR algorithm first started from several core keywords as a starting point for MALN. Secondly, we find the best collection of association rules for the moment through the breadth traversal of graphs. Third, we select the most relevant document to become active according to the association rules collection. Fourth, the selected document gives feedback to the MALN and returns to phase 2. The algorithm stopped when the core keywords and association rules are selected.

In Fact, our heuristic algorithm for concept learning is simulated from the real user behavior describing the concept, and it is a clear oriented algorithm that is based on specific concepts, which a user wants to know. When a user wants a concept, he releases some obscure keywords, which always is the most relevant terminology about a concept. Then, he will search relevant documents based on these core keywords to know more about the knowledge concept, which expands the concept terminology and construct Link between the knowledge of the terminology. Finally, the user will have a good grasp of the concept by repeating the above mentioned two phase.

In this section, we describe the proposed heuristic algorithm. We start by describing the procedure of the four phases in our algorithm, and then we describe our approach procedure with an algorithm depicted.

4.1 Select Initial Core Nodes Phase

The meaning of the select core node is a common concept that can always be of one or more words to represent, where the core node comes from two main approaches. As our heuristic method is an iterative algorithm, second sources of the core nodes are only calculated when having an initial document list, which we will discuss in phase 4. At the beginning of heuristic algorithm, we accept two types of initial core nodes, which come from user specified keywords such as; query word or comes from a category description word; however we discover that the experiment received poor performance if we put all query words into the core nodes. Therefore, we designed a pre-defined formula (4) to remove noisy words existing in specified keywords:

$$CN = \{ w \in Q | DF < \alpha^1 \}$$  

Here, we calculate document frequency (DF) for every word in a specified keywords list, notice that we consider the query word type as same as a category.
description word. We removed words with higher average frequency and added all the rest of the words to the core node list. We show a graph depict of a selected core nodes phase in Figure 3:

![Figure 3. Select Initial Core Nodes.](image)

### 4.2 Select Association Rules Collection Phase

After selecting the core node from MALN, we will select rules around the core node, which we assume the user will search for relevant concepts given certain core nodes. The associate rules are stored in L, so our task is to select some maximizing weight sub-rules collection based on the core node, the rule weight is calculate by formula (5).

\[
RC_{ij} = \{ \text{rule } \in L_{ij} | \frac{\text{rule}(i, j)}{\sum_{(i,j) \in S} \text{rule}(i, j)} > \beta \} 
\]  

Where \( L_{ij} \) is representing of the definition of MALN and \( \beta \) is the threshold for selecting the Association Rules Collection. Consider the core nodes that are extracted from last phase as a starting point. We find the maximum weight association rules from surround that satisfy at least one of constraints add to association rules collection every time. The constraints include: 1) One end of rules must be an initial keyword node, because our rule collection phase aims to find the most relevant rules for the concept. 2) When the weight of the rule is the same, we prioritize select rules that where both ends nodes are in the core node, which have a great possibility of becoming an exclusive phrase in the concept. The constraint above mentioned is to ensure that the association rules could be a more complete representation for the concept. Figure 4 shows the result of a select association rules collection phase.

![Figure 4. Select Association Rules Collection.](image)

### 4.3 Select Document Phase

Different from the IR method, we select relevant documents based on both selected core nodes and a selected rule collection. The relation tuple D in MLAN, which is standard for which rules and keywords are contained in document D and are used to calculate the Score for each document in a corpus. We sum the weight of the selected core nodes and rule collection in the document, the highest score of the document is selected to our model: Scoring function is defined in formula (6):

\[
\text{Score}(d) = \frac{\sum_{\text{rule}(i, j) \in \text{RC}} \text{weight}_{\text{rule}} + \sum_{\text{word}(i) \in \text{core nodes}} \text{weight}_{\text{word}}}{\sum_{\text{rule}(i, j) \in \text{RC}} \text{weight}_{\text{rule}} + \sum_{\text{word}(i) \in \text{core nodes}} \text{weight}_{\text{word}}}
\]  

Where the \( \text{weight}_{\text{rule}} \) is the rule’s weight calculated in MALN and \( \text{weight}_{\text{word}} \) is the word’s weight in MALN. We normalize the document score by the sum of all rules weight and word score. Then several maximum score documents form Score(d) will be put into our model as one representation for our concept representation. Figure 5 shows the result of select document phase.

![Figure 5. Select Document.](image)
4.4 Feedback on MALN Phase

After we selected one or more documents into the concept representation, we will change the core nodes list and rules list and call it the feedback phase. In this phase, we focus on the problem of the change core nodes, in other words, which core node will be activated on the MALN. We calculate the words TFIDF value as a score function where a super threshold is selected into the core node. Then, the rules are reselected in the Select Rules Collection Phase. The words TFIDF function is defined as follows:

\[ CN_w = \{ w \in \bigcup_j D | \sum_{w \in j} freq(w) \cdot IDF(w) > \alpha^2 \} \]  

(7)

Here, \( w \) is one of the words that exist in a query concept words list or in a selected document. It is worth noting that we sum all frequencies of \( w \) in the selected document as TF, instead of only occurring in an individual document. In addition, the \( IDF(w) \) value is calculated in all the documents. Formula 6 indicates that we prefer selecting words that frequently occur in a document obtained from the previous phase, and prefer not to select a common sense word. Figure 6 shows the result of the feedback on the MALN phase.

We program our algorithm as follow:

**Algorithm 1. Heuristic algorithm to refine concept**

Input: MALN(S, N, L, D)
- User Keywords: UK
- Core Nodes: CN
- Rules Collection: RC
- Core Documents: CD

Output: Concept Represent = [core documents, keyword, document rule]

Initialize Concept Represent = \( \phi \)
CN=UK //Select Core Nodes  
While |SUMM|<LIMITED
  if iteration=1
    CN=UK
    for (rules around the \{CN\})
      If \( RC_w > \) threshold then // \( RC_w \)
        \( RC += RC_w \)
        CD+= argmax(Score(d))/
        for (nodes in \{CD\})
          If \( CN(w) > \) threshold then // \( CN(w) \) defined in formula 7
            CN += CN_w
      end for
  end while
return Concept Represent

5 Experiments and Analysis

5.1 Dataset and Pre-process

We conduct the experiment on the china patent document dataset from the State Patent Bureau. The China patent document consist of a title, abstract, claim, description, applicant, inventor, publication date and International Patent Classification codes (IPC). We utilize the IPC description as our initial core nodes in our model, and applied it to the classification task, which is the first upset order of the IPC document, then we redistribute the unordered the document to correspond to the IPC code. Our concept learning model is applied to the learn patent concept. We select four IPC categories from the third-level IPC, and the description of the corresponding IPC are: how to dry objects, deposit box, steam, toy car, own number of (356, 356, 357), 357 documents respectively. The goal of our model is make people learn these four concepts fast and effectively. We used a Stanford-segmenter to segment the Chinese content. Then the stop words that were provided by the Harbin University and BaiDu’s stop word list are removed from the document.

5.2 Evaluation Strategy of our Concept Learning Model

We evaluate the quality of our concept representation by performing a retrieval task. The former will evaluate a concept, which represents potential effectiveness in predicting if the other relevant documents belong to a specific concept. Specifically, the selected documents and keywords, and rules in the MALN from our leaning model will be considered as the retrieval centre of the concept. Once we have identified the retrieval centre, the rest of documents could be directly assigned to a corresponding Category based on some distance formula. The cos similarity, which is defined in formula (8) is used in our experiment as the distance formula:

\[ \text{If } RC_{ij} > \text{threshold then } \] // \( RC_{ij} \) defined in formula 5
\[ RC += RC_{ij} \]
end for
\[ CD += \arg\max(Score(d))/ \]
\[ \text{for (nodes in } \{CD\} \) }
\[ \text{If } CN(w_i) > \text{threshold then } \] // \( CN(w_i) \) defined in formula 7
\[ CN += CN_{w_i} \]
end for
end while
return Concept Represent
Here, $d_j$ document is a vector in $D$, which the weight is represented as TFIDF. $c_i$ is the centre of the concept, which is a vector consisting of a selected document, keyword, and rules extracted in our model. After all the documents in corpus are classified, we use a precision score criteria, which is commonly used in retrieving and clustering a task to evaluate our result.

Although we evaluate our concept learning model only using documents information, representation of the keywords and rules are more important for the concept. However, it is difficult to evaluate on just keywords and rules layer in the patent corpus. There is not an existing automatic method to directly evaluate two representations in a patent apart from human assistance as we know.

### 5.3 Results of the Concept Learning Model Evaluation

We take four classes extracted before as our test data sets. We need to learn the four independent concepts from the four classes. To address how we generate the selected concept representation from our model, we briefly describe the process of the learning concepts from our data sets. First, we put four categories of the corpus of documents together as our concept learning data without its category label. Second, based on the mixed corpus, MALN is constructed as an initial representation of the documents data. Third, we select a category description such as toy and car as our core node in the heuristic algorithm. Subsequently, the completed heuristic algorithm is applied for a select document, keywords and rules.

Through a great deal of the experiment, we found that of the hyper parameter, the $\beta$ defined in phase 2 in the heuristic algorithm will have greater impact on our model. We empirically seek to evaluate the threshold $\beta$ defined phase two in the heuristic algorithm yielding the best performance. Parameter $\beta$ varies from 0 to 1 with 0.2 intervals. In this experiment, we run 10 times our heuristic algorithm to select 10 documents and corresponding keywords, and rules as learned knowledge. When $\beta$ is 0, it represents our rule collection only having one rule, while results to a variety of documents will have the same score with a selected document. When $\beta$ is 1, the algorithm will degenerate to a global rule select function, which almost neglects a related constraint in formula (4). We chose $\beta$ as 0.2, 0.4, 0.6, 0.8, 1, 0 is excluded from our test data, because, the document selection phase becomes a random selection process. Figure 7 shows the precision score of our corpus from four categories retrieved as a qualitative is evaluated.

![Figure 7. Precision vs Threshold $\beta$ for Retrieving Task.](image)

Notably, there is a rather steep drop-off in performance when $\beta>0.4$, that is, too many irrelevant rules are included in the association rule collection. Therefore, the parameter of $\beta=0.4$ is set to the best threshold in our experiments.

Figure 8 reports the performances of the proposed methods and baseline BM25 in the IR system. We use the category description as a query to the BM25 and extract the most relevant document as a core concept representation. Then, a simple classifying method is proposed in evaluating the strategy and is applied to evaluate the precision score. In order to compare the two methods in a similar environment, the BM25 selected document list size is the same as our model.

![Figure 8. Models Performance vs. a Selected Document List Size for Two Selected Approaches.](image)

The results indicate that improvements of our model over the baseline BM25 are significant. For example; compared to the best result when the selected document list size is 20, our model improves 7%. The results also indicate that 30 documents are enough to represent the patent concept, which testifies the assumption of every concept that could be represented by a limited number of document, keywords and rules.

### 6 Conclusion

ACCURACY, conciseness and comprehensiveness are three important criteria for evaluating the constructed concept in a patent corpus. In this paper, we proposed an automatic concept learning model, a mixture ALN graph for running the heuristic algorithm.
for a patent concept learning, to improve accuracy, conciseness and comprehensiveness of patent concepts.

1) The mixture ALN, which is expanded from the ALN via the add relation rules between the document and lexical, are used to express the initial patent concept. This approach improves the comprehensiveness of the concept. And such knowledge representation of the concept can be expressed in graphs, which enables us to use a graph based heuristic algorithm.

2) The select core node phase aims to improve the concept accuracy, which focuses on the extract user designated concept keyword and concept terminology from a corresponding document. This approach removes a noise word, which improves concept accuracy.

3) The select association rule collection phase, which finds the most relevant association rules around the concept terminology. This approach will easily find additional terminology about concept so that improving both concept conciseness and comprehensiveness.

4) The select document phase and feedback phase in our heuristic algorithm, which can acquire relevant documents and reselect the core node and rules collection based on the document. This approach purifies the noise caused by the association rule expansion, while avoiding the overestimation of the irrelevant concept terminology.

7 ACKNOWLEDGMENT

The research work reported in this paper was supported in part by the National Science Foundation of China under grant nos. 61471232. This work was jointly supported by the Shanghai Science and Technology Cooperation Project under grant no.16550720400

8 REFERENCE


9 NOTES ON CONTRIBUTORS

Wei Qin received bachelor’s degree in 2013 from XiDian University, China. Currently, he is a student doctor at Shanghai University. His main research interests include concept learning and knowledge graph.

Xiangfeng Luo is a professor in the School of Computers, Shanghai University, China. Currently, he is a visiting professor at Purdue University. He received master's and Ph.D. degrees from Hefei University of Technology in 2000 and 2003, respectively. He was a postdoctoral researcher with the China Knowledge Grid Research Group, Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), from 2003 to 2005. His main research interests include web wisdom, cognitive informatics, and text understanding. He has authored or co-authored more than 50 publications and his publications have appeared in IEEE Trans. on Automation Science and Engineering, IEEE Trans. on Systems, Man, and Cybernetics-Part C, IEEE Trans. on Learning Technology, Concurrency and Computation: Practice and Experience, and New Generation Computing, etc. He has served as the Guest Editor of ACM Transactions on Intelligent Systems and Technology. Dr. Luo has also served on committees of a number of conferences/workshops, including Program Co-chair of ICWL 2010 (Shanghai), WISM 2012 (Chengdu), CTUW2011 (Sydney) and more than 40 PC members of conferences and workshops.
Multi-phase Oil Tank Recognition for High Resolution Remote Sensing Images

Changjiang Liu¹, Xuling Wu², Bing Mo¹ & Yi Zhang³

¹ School of Mathematics and Statistics, Key Lab of Enterprise Informationization and Internet of Things of Sichuan Province, Sichuan Province University Key Laboratory of Bridge Non-destruction Detecting and Engineering Computing, Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Zigong, China
² School of Foreign Languages, Sichuan University of Science and Engineering, Zigong, China
³ College of Computer Science, Sichuan University, Chengdu, China

ABSTRACT

With continuing commercialization of remote sensing satellites, the high resolution remote sensing image has been increasingly used in various fields of our life. However, processing technology of high resolution remote sensing images is still a tough problem. How to extract useful information from the massive information in high resolution remote sensing images is significant to the subsequent process. A multi-phase oil tank recognition of remote sensing images, namely coarse detection and artificial neural network (ANN) recognition, is proposed. The experimental results of algorithms presented in this paper show that the proposed processing technology is reliable and effective.

KEY WORDS: Artificial neural network, Hough circle transform, Multi-scale retinex with color restoration, Supervised classification

1 INTRODUCTION

REMOTE sensing is used in numerous fields, including geography, land surveying and earth science disciplines (Kracker, 1999; Meng, Wang & Cao, 2014; Obade & Lal, 2013; Schnebele & Cervone, 2013). It is also widely used in military, intelligence, commercial, economic, planning, and humanitarian fields (Irvine, Kimball, Regan, & Lepanto, 2014; Cao, Xu & Bian, 2011; Chen & Borken-Kleefeld). Nowadays, remote sensing generally refers to the use of satellite or aircraft-based sensor technologies to detect and classify objects. Hence, digital image processing and pattern recognition techniques play a great role in solving these problems.

At present, the processing technology of high resolution remote sensing image mainly is divided into two categories; frequency domain and temporal processing (Li, Zang, Zhang, Li & Wu, 2014; Wang, Huo & Feng, 2015). High resolution remote sensing image processing roughly includes five research perspectives; remote sensing image matching, fusion of multi source images and remote sensing image, edge detection of SAR images, remote sensing image classification and remote sensing image change detection (Zhu, Jiang, Zhou & Addison, 2016). It is critical to classify landmarks from satellite sensors for climate change analysis, urban area studies, forestry applications, risk and damage assessment, water quality assessment, crop monitoring (Jawak, Kulkarni, Luis, 2015; Bruzzone, Demir, 2014). Automatic classification for remote sensing images consists of supervised classification (Stumpf, Lachiche, Malet, Kerle & Puissant, 2014; Xia, Chanussot, Du & He, 2014; Wang, Liu, Xu, Dong and Yang, 2017) and unsupervised classification (Romero, Gatta & Camps-Valls, 2016).

Oil tank, as a high risk object, plays a vital role in the military, environmental protection, and commercial activities. This paper focuses on oil tank recognition from high resolution remote sensing images. Supervised classification is widely used in remote sensing recognition. However, it needs huge samples with a class label. It leads to its limitation in insufficient samples. Beyond that, handling large size remote sensing images is time intensive. Thus, we first analyze characteristics of the oil tank in remote sensing images and device a ROI based detection strategy in combination with shape detection and supervised classification. It is worth noting that target
oil tanks have three distinct properties; homogeneity in luminance, circular shape and texture feature of inside the oil tank. Therefore, we integrated traditional image processing with supervised classification. The homogeneity of luminance makes MSER effective in detecting blobs, which is a rough domain of target oil tank regions. In this paper, output of MSER is used as region of interest or ROI, which reduces the burden of calculation. Prior to circle detection, color and illumination calibration is realized by improved MSRCR. Subsequently, circular shaped regions are extracted by Hough circular transformation, which is thought of oil tank with high probability. We call the aforementioned procedures coarse detection. In fine detection, a high dimensional feature composed of color, texture and shape feature, altogether with class label is applied to ANN.

The remainder of the paper is organized as follows: Multi-phase oil tank recognition is presented in Section 2, including maximally stable extremal regions, multi-scale retinex with color restoration and Hough circular transformation. Experimental results are outlined in Section 3. Finally, we conclude and discuss some directions for future research in Section 4.

2 MULTI-PHASE OIL TANK RECOGNITION

In this paper, a two-phase oil tank recognition in remote sensing images is proposed. It is divided into two steps, in which highly probable oil tank targets are extracted in the first step and final recognition results are obtained by means of high dimensional feature and supervised neural network in the second step. In detail, ROI is determined by MSER algorithm. Then local retinex enhancement on ROI aforementioned is presented. To be followed, circular detection of enhanced local images is implemented. Therefore coarse detection of oil tanks is accomplished by circular detection on the basis of enhanced ROI. We call the above procedures coarse detection. In the second recognition phase, pattern recognition neural network is created on the analysis of color, texture and shape feature. So far, misidentification results are eliminated from coarse detection. Abridged general view of proposed oil tank recognition is shown in Figure 1.

2.1 Maximally Stable Extremal Regions (ROI setting)

Maximally Stable Extremal Regions, or MSERs are useful algorithm to find out scene elements of interest for gray-scale images, which are used in image matching and recognition (Matas, Chum, Urban & Pajdla, 2004; Chen, Tsai, Schroth, Chen, Grzeszczuk & Girod, 2011; Nistér & Stewénius, 2008). MSERs of the remote sensing image are shown in Figure 2, different blobs painted in various colors. The MSER detection uses a water shedding process to implement as described below.

![Remote Sensing Image and MSERs](image)

The gray-scale image is presented by function:

\[ I : \Omega \rightarrow S, \]

Where \( \Omega = [1...W] \times [1...H] \subset Z^2, \ S = [0,...,255] \).

Maximally Stable Extremal Regions are defined in mathematical form as follows:

1) Connecting relation \( A : A \subset \Omega \times \Omega \).

Considering a four-neighbourhood relation, for \( p, q \in \Omega \), if and only if \( \sum_{i=1}^{4} |p_i - q_i| \leq 1 \), point \( p \) and \( q \) is connected, and the relation can be denoted by \( p \in A(q) \), or \( q \in A(p) \).
2) Region D. For any \( p, q \in D \subset \Omega \), if there exists an ordered sequence \( p, c_1, c_2, \cdots, c_n, q \), such that \( p \in A(c_1),\ldots,c_i \in A(c_{i-1}),\ldots,c_n \in A(q) \), \( D \) will be called connected subset of \( \Omega \).

3) Boundary \( \partial D \). It can be presented as: \( \partial D = \{ q \in D \mid \exists \ p \in D \text{ s.t. } q \in A(p) \} \).

4) Extremal region D. If \( I(p) > I(q) \) (maximal extremal region) or \( I(p) < I(q) \) (maximal extremal region) always holds for all \( p \in D, q \in \partial D \).

5) Maximally Stable Extremal Regions. Assuming \( D_1, D_2, \ldots, D_n \) is a sequence of nested external region, namely \( D_i \subset D_{i+1} \), where \( D_i = \{ x \in \Omega \mid I(x) < i, i \in S \} \). Rate of region change is defined as:

\[
r(i) = \frac{|D_i - D_{i-1}|}{|D_{i-1}|}
\]

\( D_i \) is the MSER, if and only if \( r(i') \) is a local minimum. Here, \( | \cdot | \) represents the area of a region, \( \Delta \in S \) is a tiny change of gray level.

Suppose the element of region \( D_i \) assigned white pixel, visualization of \( D_i \) is shown in Figure 3.

![Figure 3. Visualization of the D_i with i = 50, 100, 150, 200.](image)

As we know, MSERs are shaped in ellipse, whose bounding box is parameterized by the top left corner \((X_{\min}, Y_{\min})\) and bottom right corner \((X_{\max}, Y_{\max})\).

Taking a bigger tolerance domain into account, the bounding box is expanded to the region with the two corners \((X_{\min} - d_i, Y_{\min} - d_i)\) and \((X_{\max} + d_i, Y_{\max} + d_i)\).

### 2.2 Multi-scale Retinex with Color Restoration

Land and McCann (1971) conceived the Retinex theory as a model of the lightness and color perception of human vision. Jobson et al. (1997a) developed a practical implementation of the Retinex theory and defined a single scale Retinex. In the following years, they proposed the multi-scale Retinex with color restoration (MSRCR) (Jobson, Rahman and Woodell, 1997b), which provided not only dynamic range compression with a small scale but also tonal rendition with a large scale.

Multi-scale Retinex (MSR) can be interpreted as:

\[
I_{MSR}^i(x, y) = \sum_{n=1}^{s} a_n \frac{\log(I(x, y) + 1.0)}{\log(I_{\sigma_n}(x, y) + 1.0)} * F_n(x, y)
\]

Where \( i = 1, 2, 3 \) represents the red, green and blue channel image of \( I \) respectively, \( a_n \) is the weight of the \( n \)-th scale of \( s \) scales, and "*" is convolution operator.

Convolution kernel function \( F_n(x, y) \) is defined as:

\[
F_n(x, y) = k_\sigma e^{-\frac{(x+y)^2}{2\sigma^2}}
\]

Where \( \sigma_n \) is the standard deviation of Gaussian function, corresponding to a scale in MSR, the \( F_n(x, y) \) is normalized by the \( k_\sigma \), such that \( \int F_n(x, y) dx dy = 1 \).

For the sake of consistency in color, the \( C_i(x, y) \) restoration function is introduced to MSR (Jiang, Woodell and Jobson, 2015):

\[
I_{MSRCR}^i(x, y) = C_i(x, y) I_{MSR}^i(x, y)
\]

Where \( C_i(x, y) \) is defined as:

\[
C_i(x, y) = \log \left( \frac{\eta I(x, y) + 1.0}{\sum_{i=1}^{s} (I_i(x, y) + 1.0)} \right)
\]

Where \( \eta \) is the control parameter.

Normally, \( s = 3 \), or \( \sigma_1 = 5, \sigma_2 = 20, \sigma_3 = 240 \) are used. In previous works (Liu, Cheng, Zhang & Basu, 2017), we proposed an explicit construction method for multiple scales \( s > 3 \) as to keep balance in high contrast and color restoration.
2.3 Hough Circle Transform

Hough circle transform works in a manner roughly analogous to the Hough line transform (Yuen, Prince, Illingworth & Kittler, 1990; Atherton & Kerbyson, 1999), in which it is parameterized by \( \theta \) (angle) and \( r \) (radius). The approach can be described as;

1) Read the remote sensing image \( I \).
2) Convert RGB image to gray scale by eliminating the hue and saturation information while retaining the luminance, denoted the output image by \( G \).
3) Calculate gradient of the \( G \) with the notation \( g_x, g_y \) and magnitude \( g = \sqrt{g_x^2 + g_y^2} \).
4) Find indices of the \( g \) whose value is greater than the threshold \( t_g \) and store the \( x \)- and \( y \)-coordinate in the variables \( E_x \) and \( E_y \).
5) Generate a series of equally spaced points between \( R_{\text{min}} \) and \( R_{\text{max}} \), where \( R_{\text{min}}, R_{\text{max}} \) is predefined minimal and maximal radius in application respectively. The sequence can be presented as:

\[
R = \{ R_{\text{min}}, R_{\text{min}} + d, \ldots, R_{\text{max}} + m \cdot d \}.
\]

Where \( m = \left\lceil \frac{R_{\text{max}} - R_{\text{min}}}{d} \right\rceil \), the \( d \) usually takes 0.5.
6) For any \( r \in R \), compute the accumulator array as follows:

\[
\begin{align*}
x_r &= E_x - R \left( \frac{g_x}{g} \right) \\
y_r &= E_y - R \left( \frac{g_y}{g} \right)
\end{align*}
\]

7) Accumulate the votes in the parameter plane, only retain the candidate \( R, x_r \) and \( y_r \), whose accumulator is greater than the threshold \( t_g \).

2.4 Artificial Neural Network Recognition (Fine Detection)

In remote sensing images, there are some circular objects, which are actually not oil tanks. The following context will discuss how to eliminate such cases. This phenomenon implies that shape is not the unique characteristics of oil tanks. Multiple trials conclude that high dimension feature vector composed of color, shape and texture can distinguish an oil tank.

2.4.1 Feature Eextraction

We consider four categories of feature: Color, texture, shape and HOG. They are defined as:

1) Color feature: Denote total number of pixels in the image by \( N \), the number of pixels with gray level \( l \) by \( n_l \) and the gray levels by \( L \). Thus the average of the \( k \)-th channel color \( u_k \), mean square deviation \( \sigma_k \), and the entropy \( e_k \) are calculated below:

\[
\begin{align*}
u_k &= \frac{1}{N} \sum_{i,j} I_{ik} \\
\sigma_k &= \sqrt{\frac{1}{N} \sum_{i,j} \left( I_{ik} - u_k \right)^2} \\
p_i &= \frac{n_{ik}}{N}, e_k = -\sum_{i=1}^{L} p_i \log_2 p_i \\
&(I = 1, 2, \ldots, L; k = 1, 2, 3)
\end{align*}
\]

2) Color histogram: A color histogram of an image is a representation of the distribution of colors. In this paper, HSV color space is chosen, in which the \( H \) component (hue) is divided into \( N_h \) bins, \( S \) component \( N_s \) bins. Henceforth, \( N_h, N_s \) types of colors are considered. When \( S \) component is less than a specific value, human regards as a gray scale, in that case we consider \( N_g \) gray levels. In all, color histogram is \( N_h \cdot N_s \cdot N_g \) dimension. In this paper, \( N_h = N_s = N_g = 10 \), therefore, color feature is 110 dimensional.

3) Histogram of oriented gradients (HOG descriptors): Hog descriptors are applied into computer vision and image processing, serving as feature descriptor in object detection. This technique calculates occurrences of gradient orientation in localized parts of an image. In this paper, we divide the entire range of gradient value into 90 bins.

4) Invariant moments: M. K. Hu (1962) proposed seven moments, which are translation, scale and rotation invariant. We chose the first 6 Hu moments in this paper.

5) Unified local binary pattern: Local binary pattern is a convincing texture descriptor which is widely used in many areas of image processing such as face recognition and defect detection, etc. (Guo, Zhang and Zhang, 2010). Ojala and Harwood (1996) proposed an unified local binary pattern, which simplified local binary pattern to 18 modes.

6) Gray level co-occurrence matrix (GLCM): Texture is an image attribute, which describes properties such as smoothness, coarseness, regularity, etc. It shows the organizational structure of the surface and its sequence. Gray level co-occurrence matrix (Haralick, 1973) is a common approach to describe texture by studying the related spatial features of a gray level. Based on GLCM, four indexes such as angular second moments, contrast, correlation and entropy, are also thought of image features.

7) Shape feature: The shape of an image is the most important visual feature in image features.
The descriptor of shape feature involves the description of contour boundary, or perimeter and the region within the boundary, or area. In this paper, we employ the width-height ratio and roundness of the minimum area bounding rectangle to describe the shape feature. Assuming the width and height of the bounding rectangle is $w$ and $h$ respectively. The width-height ratio is defined as:

$$\alpha = \frac{w}{h}.$$ 

The roundness is defined as

$$\gamma = \frac{C^2}{A},$$

Where $C, A$ is the perimeter and area of contouring region respectively.

### 2.4.2 Feed-forward Network to Solve Oil Tank Recognition

We design a two-layer feed-forward network with sigmoid hidden and softmax output neurons to train a neural network, which is able to get rid of the false oil tank objects from the first coarse detection. The network was trained with scaled conjugate gradient back propagation. A 239-150-2 neural network is defined, as shown in Figure 4.

![Figure 4. Structure of a Neural Network.](image)

### 3 EXPERIMENTAL RESULTS AND DISCUSSION

We analyzed dozens of high resolution remote sensing images, containing more than 1000 oil tanks. The algorithm was implemented using the Microsoft Visual Studio development platform supplemented by an Open CV (Computer Vision) library. Experimenting on remote sensing image with the size 2560*2560, the average execution time (CPU: Intel(R) Core(TM) i5-4210U (2.40 GHz), RAM: 8 GB and OS: Windows 7 (64 bit)) of the proposed algorithm is 1.29 s.

#### 3.1 Adaptive Radius $R_{\text{min}}$ and $R_{\text{max}}$

Ill-suited radius range in Section 2.3 will bring about imperfect detection results, as shown in Figure 5.

![Figure 5. Circle Detection with Various Radius Ranges.](image)

In general, the radius of oil tanks is in the range of 2.65 m-50 m. If spatial resolution is also know, it is easy to calculate the distance in pixel metric. According to this, the radius of experimental remote sensing images in this paper range in [1 21] pel.

#### 3.2 MSER

Based on Section 2.1, oil tank regions are extracted by the MSER algorithm, shown in Figure 6(a), in which the union of MSERs bounded by ellipse in green contains all oil tanks. Ellipse shaped MSERs are expanded to rectangular shaped ROIs, as shown in Figure 6(a). Figure 6(c) demonstrates that ROIs extracted only account for 14.7% area of original image. Therefore, it reduces 93% of processing time.

![Figure 6. MSERs and ROIs Extracted.](image)
3.3 Improved MSRCR based Enhancement

In comparison with the original image, standard MSRCR based enhancement algorithm offers a good contrast, as shown in Figure 7(b). However, it is not hard to find an enhanced image based on the standard MSRCR and cannot retain color consistency. Especially a region rich in texture is excessively smoothed in Figure 7(b). Whereas our method's enhanced image (see Figure 7(c)) performs well in contrast and tonal rendition.

3.4 Artificial Neural Network Recognition

Coarse detection will detect circular objects which is possibly oil tanks. In combination with artificial neural network recognition, namely fine detection, oil tank candidates are to be confirmed. Final detection results are seen in Figure 8 and Figure 9.

4 CONCLUSION

In this paper, a high-resolution remote sensing image is studied, and a multi-phase oil tank recognition strategy, namely from coarse to fine detection, is proposed. It helps to monitor the environment and prevent crises before they emerge. In this sense, our work significantly improves time performance and rate of identification. In future work, we will conduct experiments on more high resolution remote sensing images, and consider optimizing the algorithm to put it into airborne images. We will also investigate applying more samples to attain high recognition rate and consider accurate localization of oil tanks.

5 ACKNOWLEDGMENT

This work was supported in part by the Open Project of Key Lab of Enterprise Informationization and Internet of Things of Sichuan Province under Grant No. 2017WZY01, the open Project of the Artificial Intelligence Key Laboratory of Sichuan Province under Grant No. 2014RZY02 and 2017RZJ03, the Open Project of Sichuan Province University Key Laboratory of Bridge Non-destruction Detecting and Engineering Computing under Grant No. 2014QZY01, Natural Science Foundation of Sichuan University of Science and Engineering (SUSE) under Grant No.2015RC08, 2017RCL54 and 2017RCL23, and Educational reform project of SUSE under Grant No. JG-1707. The work was also supported in part by the National Natural Science Foundation of China (Grant No. 11705122) and the Training Programs of Innovation and Entrepreneurship of Sichuan Province for Undergraduates under Grant No.201710622088. The work was also supported in part by Guangxi Key Laboratory of Cryptography and Information Security under Grant No.GCIS201607.

6 REFERENCES


processing. Recent Patents on Computer Science, 09(999), 53-61.

7 DISCLOSURE STATEMENT
NO potential conflict of interest was reported by the authors.

8 NOTES OF CONTRIBUTORS
Changjiang Liu received Ph.D. degree from Sichuan University, Chengdu, in 2014. He worked as a visiting professor from August 2015 to August 2016 and January 2017 to April 2017 with the Department of Computing Science, University of Alberta, Edmonton, Canada. He is currently an associate professor with the Sichuan University of Science and Engineering, Zigong, China. His current research interests include image processing and pattern recognition.

Bing Mo received M.E. from Sichuan University of Science and Engineering, Zigong, in 2017. He is currently working as an engineer at Fujian Tietuo Machinery Co., Ltd. His current research interests focused on automation control.

Xuling Wu received M.S degree from the University of Electronic Science and Technology of China in 2012. She is now working as a lecturer at Sichuan University of Science and Engineering. Her research focused on foreign linguistics and applied linguistics.

Yi Zhang received Ph.D. degree from the National University of Singapore, Singapore, in 2010. He is currently an associate professor with the College of Computer Science, Sichuan University, Sichuan, China. His current research interests focused on image and vision computing.
Contents Continued

The SLAM Algorithm for Multiple Robots Based on Parameter Estimation  
   MengYuan Chen  
   593

The Lateral Conflict Risk Assessment for Low-Altitude Training Airspace Using Weakly Supervised Learning Method  
   Kaijun Xu, Xueting Chen, Yusheng Yao, and Shanshan Li  
   603

Simulation and Data Analysis of Energy Recovery Sensing on Parallel Hydraulic Hybrid Crane  
   Youquan Chen, Xinhui Liu, Xin Wang, and Jinshi Chen  
   613

Intelligent Control for Integrated Guidance and Control Based on Intelligent Characteristic Model  
   Jun Zhou, Zhenzhen Ge  
   623

NARX Network Based Driver Behavior Analysis and Prediction Using Time-Series Modeling  
   Ling Wu, Haoxue Liu, Tong Zhu, and Yueqi Hu  
   633

The Data Analyses of Vertical Storage Tank Using Finite Element SOFT Computing  
   Lin Gao, Mingzhen Wang  
   643

Association Link Network Based Concept Learning in Patent Corpus  
   Wei Qin, Xiangfeng Luo  
   653

Multi-Phase Oil Tank Recognition for High Resolution Remote Sensing Images  
   Changjiang Liu, Xuling Wu, Bing Mo, and Yi Zhang  
   663
Intelligent Automation & Soft Computing
An International Journal

Aims and Scope
Intelligent automation is achieved by implementing humanlike intelligence, machine learning, and soft computing (computational intelligence) tools and techniques, deep learning neural networks, genetic algorithm, genetic programming, clustering, statistical tools like Bayesian network, etc., recognizing that automation and intelligence are the focused issues of decades. Applications such as big data analytic, cloud-based networking and cyber-security are all considered.

Intelligent Automation and Soft Computing: An International Journal seeks to provide a common forum for the dissemination of accurate results about the world of intelligent automation, control, manufacturing, modeling and systems engineering. It is intended that the articles published in the journal will encompass both the short and the long term effects of soft computing and other related fields such as robotics, control, computer, vision, speech recognition, pattern recognition, data mining, "big data", data analytics, machine intelligence and machine translation. It further hopes it will address the existing and emerging relationships between automation, systems engineering, system of systems engineering and soft computing. The journal will publish original and survey papers on intelligent automation and systems engineering with an emphasis on current and potential applications of soft computing. It will have a broad interest in all engineering disciplines, computer science, and related technological fields such as medicine, biology operations research, technology management, and agriculture and information technology. Topics to be considered include but not limited to:

Intelligent Automation: Robotics, industrial inspection, design automation and rapid prototyping, scheduling, process control and automation, autonomous systems, computer integrated manufacturing, multi-agent systems, intelligent system of systems, MEMS, nanotechnology, stochastic learning automation, autonomous control, cyber-physical (system of systems) systems, applications in space- and earth-bound systems.

Soft Computing: Deep Neural networks and deep learning, neurocomputing, fuzzy logic, genetic algorithms, genetic programming, AI and expert systems, probabilistic reasoning, machine learning, learning control, fuzzy logic control, distributed intelligence, image processing and self-organizing systems, data mining, "big data" management, clustering, etc.

Paper Charge Policy: AutoSoft has a publication charge of US $730 for up to 12 pre-publication pages. Additional pages are at $100. This fee brings access to the entire history of the journal plus a hard copy of the journal in which their paper appeared.

Instructions for Contributors
The journal will consider two types of papers: articles and survey papers. Articles will be full-length papers (between 12-15 typeset pages in length) reporting original contributions or significant results. Papers emphasizing applications and survey (tutorials) are especially welcome. Prospective authors should upload their papers in PDF format on the site http://www.autosoftjournal.net of previously unpublished papers. To facilitate communication, the address, phone number, fax number, and e-mail address of the corresponding author should be given on the first page of the papers. Papers should be written in correct English and arranged in the following order: 1) title of the paper; 2) coordinates of all authors with affiliations and mailing address for each author; 3) an abstract (50-100 words) containing the purpose, scope, contribution of the paper; 4) up to 10 keywords with the first keyword being the most pertinent one; 5) introduction; 6) main body of the paper divided into sections and subsections where appropriate; 7) conclusion; 8) references; and 9) appendices (if any).

Use a Roman typeface for your manuscript whenever possible (i.e., Times or Times New Roman). All paragraphs in the manuscript should be double spaced and typed flush left. Do not indent paragraphs, but allow extra space between them. Italicize words to be printed in italics and bold words to be printed in bold.

Both equations and figures must be numbered sequentially using Arabic numerals. Captions for figures should be in sentence format with the first word capitalized and a period at the end. Equations should be numbered sequentially with the number enclosed in parenthesis and placed at the right margin.

Type the reference list in APA style. For any questions on the above, please refer to this site: www.apastyle.org/learn/quick-guide-on-references.aspx.

IMPORTANT PLEASE READ

SUBMISSION INSTRUCTIONS: Only electronic submissions are allowed. Please submit them in .pdf format on the website. MS word is an acceptable format. Figures should be embedded. Submit figures individually in a file as well in either JPG, .TIF, or .EPS format. Figures that have very thin lines sometimes do not print well so please be sure that lines on figures, graphs, and charts are 1 point. Each designated corresponding author will receive a proof to check the accuracy of the final typeset article from TSI Press Ms. Deb Rivard. Only corrections for typesetting errors are allowed. Extensive author alterations will incur an extra fee to the author. Correctly formatted articles with preset styles are available. Authors can access the template by going to: http://wacong.org/autosoft/auto/forms.html). We will be happy to answer any other questions you may have at tsiepress@gmail.com or Debra Rivard debrarivard@yahoo.com