



Laporan Akhir Projek Penyelidikan Jangka Pendek

**Development of Touch-Less Palm Print
Biometric Authentication System To
Smart Phone Using Android Operating
System**

**by
Dr. Dzati Athiar Ramli
Dr. Haidi Ibrahim**

2015



RU GRANT FINAL REPORT CHECKLIST

Please use this checklist to self-assess your report before submitting to RCMO.
Checklist should accompany the report.

NO.	ITEM	PLEASE CHECK (✓)		
		PI	JKPTJ	RCMO
1	Completed Final Report Form	✓		✓
2	Project Financial Account Statement (e-Statement)	✓		✓
3	Asset/Inventory Return Form (Borang Penyerahan Aset/Inventori)	✓		✓
4	A copy of the publications/proceedings listed in Section D(ii) (Research Output)	✓		✓
5	Comprehensive Technical Report	✓		✓
6	Other supporting documents, if any	-		-
7	Project Leader's Signature	✓		✓
8	Endorsement of PTJ's Evaluation Committee	✓	✓	✓
9	Endorsement of Dean/ Director of PTJ's	✓	✓	✓



RU GRANT FINAL REPORT FORM

Please email a softcopy of this report to rcmo@usm.my

A	PROJECT DETAILS
i	Title of Research: Development of Touch-Less Palm Print Biometric Authentication System to Smart Phone using Android Operating System
ii	Account Number: 1001/PELECT/814161
iii	Name of Research Leader: Dr. Dzati Athiar Ramli
iv	Name of Co-Researcher: 1. Dr. Haidi Ibrahim
v	Duration of this research: <p>a) Start Date : 15 July 2012</p> <p>b) Completion Date : 14 July 2015</p> <p>c) Duration : 3 years</p> <p>d) Revised Date (if any) : -</p>
B	ABSTRACT OF RESEARCH
	<p><i>(An abstract of between 100 and 200 words must be prepared in Bahasa Malaysia and in English. This abstract will be included in the Report of the Research and Innovation Section at a later date as a means of presenting the project findings of the researcher/s to the University and the community at large)</i></p> <p><i>The emerging of internet and wireless dimension has brought a new era in biometrics technology. Instead of operating the biometric system with static biometric device, mobile biometric system can be implemented and this approach leads to more efficient and reliable implementation. In this study mobile biometric system based on palm print modality based on Android operating system is developed. In order to execute mobile biometric system, efficient processing time and storage are some of the important factors that need to be considered. Algorithms involving palm print feature processing are evaluated so as to obtain optimum time and memory consumption. Several feature processing methods including Region of Interest (ROI), Principal Component Analysis (PCA), and Kernel Principal Component Analysis (KPCA) and a new approach in feature extraction called Reduced-Set Kernel Principal Component Analysis (RSKPCA) are investigated. In this project, it has been proven that the RSKPCA gives the best result for mobile biometric system based on palm print. Meanwhile, the android operating system is able to acquire the palm print image from the Android devices and send the user name and palm print image to the server via the internet. The server is able to receive the data from multiple clients at the same time, performs the verification</i></p>

Kemunculan baru dimensi internet dan teknologi tanpa wayar telah membawa era baru dalam teknologi biometrik. Selain sistem biometrik dengan peranti statik, sistem biometrik mudah alih boleh dilaksanakan dan pendekatan ini membawa kepada pelaksanaan yang lebih cekap dan efisien. Dalam kajian ini, sistem biometrik mudah alih berasaskan tapak tangan telah dibangunkan. Dalam kajian ini, sistem biometrik tapak tangan mudah alih berasaskan sistem pengoperasi android telah dibina. Untuk melaksanakan sistem biometrik mudah alih, masa pemprosesan dan penyimpanan yang cekap adalah faktor penting yang perlu dipertimbangkan. Algoritma-algoritma yang melibatkan pemrosesan ciri tapak tangan dinilai berdasarkan penggunaan masa dan memori yang optimum. Beberapa kaedah pemrosesan ciri termasuk Ruang Dikehendaki (ROI), Analisa Komponen Utama (PCA) dan Analisa Komponen Utama Kernel (KPCA) disiasat. Pendekatan baru dalam pengekstrakan ciri yang digelar Analisa Komponen Utama Kernel Set Dikurangi (RSKPCA) telah dikaji. Projek ini telah membuktikan bahawa pengekstrakan ciri menggunakan RSKPCA yang dicadangkan memberikan keputusan yang terbaik untuk sistem biometrik mudah alih berasaskan tapak tangan. Sementara, sistem pengoperasi android mampu memperoleh imej tapak tangan daripada peranti Android dan menghantar nama pengguna dan imej tapak tangan kepada pelayan melalui internet. Pelayan boleh menerima data dari pelbagai pelanggan pada masa yang sama, melaksanakan proses pengesahan dan kemudian menghantar hasil pengesahan kepada peranti Android.

C BUDGET & EXPENDITURE

i

Total Approved Budget : RM136,235.00

Yearly Budget Distributed

Year 1 : RM 51,545.00

Year 2 : RM 43,845.00

Year 3 : RM 40,809.53

Total Expenditure : RM136,235.00

Balance : RM 35.47

Percentage of Amount Spent (%) : 99.97%

Please attach final account statement (eStatement) to indicate the project expenditure

ii Equipment Purchased Under Vot 35000

No.	Name of Equipment	Amount (RM)	Location	Status
1.	Laptop ASUS	RM 4,985.00	Bilik 3.36 PPKEE USM	Baik
2.	Smart Phone Samsung S3	RM 2,070.00	Bilik 3.36 PPKEE USM	Baik
3.	Smart Phone HTC	RM 1,830.00	Bilik 3.36 PPKEE USM	Baik

Please attach the Asset/Inventory Return Form (Borang Penyerahan Aset/Inventori) – Appendix 1

D	RESEARCH ACHIEVEMENTS		
i	Project Objectives (as stated/approved in the project proposal)		
	No.	Project Objectives	Achievement
	1	Intelligent System Modeling Algorithm	July 2011
	2	Fusion System Modeling	April 2012
	3	Access Control System Application	Dec 2012
	4		
	5		
	6		

ii	Research Output		
	a) Publications in ISI Web of Science/Scopus		
	No.	Publication (authors,title,journal,year,volume,pages,etc.)	Status of Publication (published/accepted/ under review)
	1.	Haryati Jaafar, Salwani Ibrahim, Dzati Athiar Ramli. 2015. A Robust and Fast Computation Touchless Palm Print Recognition System Using LHEAT and the IFkNCN Classifier, <i>Computational Intelligence and Neuroscience</i> , pp. 1-17. ISSN 16875265. (ISI)	Published
	2.	Bakhtiar Affendi Rosdi, Haryati Jaafar, Dzati Athiar Ramli. 2015. Finger vein identification using fuzzy-based k-nearest centroid neighbor classifier. In: THE 2ND ISM INTERNATIONAL STATISTICAL CONFERENCE 2014 (ISM-II): Empowering the Applications of Statistical and <i>Mathematical Sciences</i> , 1643, 649-654. (ISI)	Published
	3.	Lydia Abdul Hamid, Dzati Athiar Ramli. 2013. Quality based Speaker Verification System using Fuzzy Inference Fusion Scheme. <i>Journal of Computer Science</i> . Volume 10, Issue 3. pp 530-540. ISSN 15493636. (SCOPUS)	Published
	4.	Noor Salwani Ibrahim, Haryati Jaafar, Dzati Athiar Ramli. 2014. Robust Palm Print Verification System Based On <i>Evolution Kernel Principal Component Analysis</i> , IEEE International Conference on Control System, Computing and Engineering 2014 (ICCSCE 2014), pp. 202-207. ISBN 978-1-4799-5685-2 (SCOPUS)	Published
	5.	Tan Wan Chien, Haryati Jaafar, Dzati Athiar Ramli, Bakhtiar Afendi Rosdi, Shahriza Shahrudin. 2014. Intelligent Frog Species Identificaion on Android Operating System, <i>International Journal of Circuits, Systems and Signal Processing (JCSSP)</i> , volume 8, pp. 137-148. ISSN 1998-4464. (SCOPUS)	Published
	6.	Haryati Jaafar, Dzati Athiar Ramli, Shahriza Shahrudin, 2013. MFCC based Frog Identification system in Noisy Environment. IEEE International Conference on Signal and .mage Processing Applications - IEEE ICSIPA 2013, pp 123-127. ISBN 978-1-4799-0267-5 (ISI).	Published

7.	23. Chia Chin Lip, Dzati Athiar Ramli. 2012. Comparative Study on Feature, Score and Decision Level Fusion Schemes for Robust Multibiometric Systems, In Sabo Sambath and Egui Zhu (Eds.), <i>Advances in Intelligent and Soft Computing</i> (Frontiers in Computer Education), Publisher Springer Berlin Heidelberg, volume 133, pp. 941-948. ISBN 978-3-642-275524. (SRINGER)	Published
8.	Aini Hafizah Mohd Saod, Dzati Athiar Ramli, 2011. Preliminary Study on Classification of Apraxia Speech using Support Vector Machine. <i>Australian Journal of Basic and Applied Sciences (AJBAS)</i> , vol 5(9):2060-2071, 2011. ISSN 1991-8178. (SCOPUS)	Published

b) Publications in Other Journals

No.	Publication (authors, title, journal, year, volume, pages, etc.)	Status of Publication (published/accepted/ under review)
9.	Lau Su Ching, Noor Salwani Ibrahim, Dzati Athiar Ramli. 2014. Development Of Multibiometric Verification System Based on Speech and Palmprint Information, <i>Australian Journal of Basic and Applied Science-AJBAS</i> , Vol 9, no 27, pp 223-228.	Published
10.	Mohd Azha Mohd Salleh, Noor Salwani Ibrahim, Dzati Athiar Ramli. 2014. Data Reduction on MFCC Features Based on Kernel PCA for Speaker Verification System. <i>WALIA Journal</i> , volume 30(S2), pp. 56-62. ISSN 1026-3861	Published
11.	Haryati Jaafar, Dzati Athiar Ramli. 2015. Robust Syllable Segmentation Of The Automatic Frog Calls Identification System. <i>International Conference on Environmental Forensics (IENFORCE2015)</i> .	Published
12.	Haryati Jaafar, Salwani Ibrahim, Dzati Athiar Ramli. 2014. The Local Histogram Equalization and Adaptive Thresholding for Hand-Based Biometric Systems, <i>Proceeding of the 1st International Conference on Mathematical Methods & Computational Techniques in Science & Engineering (MMCTSE 2014)</i> . Athens, Greece. November 28-30 2014, pp. 168-173. ISBN: 978-1-61804-256-9.	Published
13.	Salwani Ibrahim, Dzati Athiar Ramli. 2013. Evaluation on Palm-Print ROI Selection Techniques for Smart Phone based Touch-less Biometric System. <i>American Academic & Scholarly Research Journal</i> . Vol 5 No 5.	Published
14.	Haryati Jaafar, Dzati Athiar Ramli. 2013. A Review of Multibiometric System with Fusion Strategies and Weighting Factor. <i>International Journal on Computer Science and Engineering (IJCSE)</i> , Vol 2 No 4. pp. 258-165. ISSN 2319-7323.	Published

c) Other Publications

(book, chapters in book, monograph, magazine, etc.)

No.	Publication (authors, title, journal, year, volume, pages, etc.)	Status of Publication (published/accepted/ under review)
15.	Haryati Jaafar, Dzati Athiar Ramli, Bakhtiar Afendi Rosdi, Shahriza Shahrudin. 2014. Frog Identification System based on Local Means K-Nearest Neighbors with Fuzzy Distance Weighting. In H.A. Mat Sakim and M.T. Mustafa (Eds.) <i>Lecture Notes in Electrical Engineering (The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications – Innovation Excellence Toward Humanistic Technology)</i> , Publisher Springer Singapore, volume 291, pp. 153-160. ISBN 978-981-4585-41-5 (ISI, SPRINGER).	Published

d) Conference Proceeding			
No.	Conference (conference name,date,place)	Title of Abstract/Article	Level (International/National)
	3CA 2011, 1-2 December 2011,Macao, China.	Comparative Study on Feature, Score and Decision Level Fusion Schemes for Robust Multibiometric Systems	International Conference on Frontiers in Computer Education, (ICFCE)
# Please attach a full copy of the publication/proceeding listed above			
iii Other Research Ouput/Impact From This Project (patent, products, awards, copyright, external grant, networking, etc.)			
1. Networking with Vitrox Corporation and BorderPass. Development of palmprint biometric prototype during sabbatical leave (4 months)			

E HUMAN CAPITAL DEVELOPMENT			
a) Graduated Human Capital			
Student	Nationality (No.)		Name
	National	International	
PhD	1		1. Haryati Jaafar 2.
MSc	8	1	1. Lydia Abdul Hamid (Research) 2. Nor Salwani Ibrahim (Research) 3. Mohd Azhar bin Mohamad Salleh (ESDE) 4. Nurul Farhana Abd Razak (ESDE) 5. Abdullah Amir (ESDE) 6. Teh Choon Yan (ESDE) 7. Roziatul Nazirah Ishak (ESDE) 8. Aini Hafizah Mohd Saod (ESDE) 9. NurulHayati Che Rani (ESDE)
Undergraduate			1. Yeap Chin Meng 2. Chan Yeow Pang 3. Lau Wei Cheang 4. Lee Yong Chun 5. Tay Chui Hui

	11		6. Lau Su Ching 7. Clifford Loh Ting Yuan 8. Azman Aris 9. Muhammad Hazim 10. Tan Chee Kan 11. Chia Chin Lip
--	-----------	--	---

b) On-going Human Capital

Student	Nationality (No.)		Name
	National	International	
PhD	1		1. Nor Salwani Ibrahim
MSc		1	1. Amir Hajian
Undergraduate			1. 2.

c) Others Human Capital

Student	Nationality (No.)		Name
	National	International	
Post Doctoral Fellow			1. 2.
Research Officer	1		1. Lydia Abdul Hamid 2.
Research Assistant	1		1. Roslinda Roslan Azman Aris 2.
Others (Student)	1		1. Muhammad Yusof Abd Aziz 2.

F COMPREHENSIVE TECHNICAL REPORT

Applicants are required to prepare a comprehensive technical report explaining the project. The following format should be used (this report must be attached separately):

- Introduction
- Objectives
- Methods
- Results
- Discussion
- Conclusion and Suggestion
- Acknowledgements
- References

G PROBLEMS/CONSTRAINTS/CHALLENGES IF ANY

(Please provide issues arising from the project and how they were resolved)

-

H	RECOMMENDATION
	<p><i>(Please provide recommendations that can be used to improve the delivery of information, grant management, guidelines and policy, etc.)</i></p> <p>-</p>

Project Leader's Signature:



.....
Name : Dr. Dzati Athiar Khamli

Date : 27 Jun 2016

Dr. Dzati Athiar Khamli (PhD),
School of Electrical & Electronic Engineering,
JSM, Pulau Pinang.

I COMMENTS, IF ANY/ENDORSEMENT BY PTJ'S RESEARCH COMMITTEE

Very good research output
with 2 ISI journal.



Signature and Stamp of Chairperson of PTJ's Evaluation Committee

PROFESOR DR. MOHD. FADZIL BIN AIN

Name : Timbalan Dekan
(Penyelidikan, Siswazah dan Jaringan)
Date : Pusat Pengajian Kejuruteraan Elektrik & Elektronik
Kampus Kejuruteraan
Universiti Sains Malaysia



Signature and Stamp of Dean/ Director of PTJ

PROFESOR DR. MOHD RIZAL ARSHAD

Name : Dean
Date : School of Electrical & Electronic Engineering
Engineering Campus
Universiti Sains Malaysia

17.7.16

Purchase Requisition		Purchase Order		Suppliers		Maintenance		Financials		Coda Info		Reports		Admin	
UserCode: SHARIDA / USMKCLIVE / PELECT				Program Code: Votebook9100				Current Program : Votebook (Header)							
Current Date : 11/07/2016 10:29:55 AM				Version: 15.124, Last Updated at 01/07/2016				DB: 13.00, 09/18/2010 VB: 13.01, 03/14/2011				Switch Language : English /Malay			
Wildcard : eg. Like 100%, Like 10%1, Like %1															
Element 1: %		Element 2: %		Element 4: PELECT											
Element 5: 814161		Year: 2016													
Detail	Excel	Budget Rule	Budget Control	Account Description	Budget Account Code	Roll over	Budget	Cash Received	Advanced	Commit	Actual	Available	Percentage		
Detail	Excel	46	T	Projek Kumpulan Wang Uni Penyelidikan	1001.111.0.PELECT.814161	47,278.56	0.00	0.00	0.00	0.00	0.00	47,278.56	0.00%		
		46	T	SubTotal		47,278.56	0.00	0.00	0.00	0.00	0.00	47,278.56	0.00%		
Detail	Excel	47	T	Projek Kumpulan Wang Uni Penyelidikan	1001.221.0.PELECT.814161	-4,758.53	0.00	0.00	0.00	0.00	0.00	-4,758.53	0.00%		
Detail	Excel	47	T	Projek Kumpulan Wang Uni Penyelidikan	1001.223.0.PELECT.814161	500.00	0.00	0.00	0.00	0.00	0.00	500.00	0.00%		
Detail	Excel	47	T	Projek Kumpulan Wang Uni Penyelidikan	1001.227.0.PELECT.814161	-5,377.25	0.00	0.00	0.00	0.00	0.00	-5,377.25	0.00%		
Detail	Excel	47	T	Projek Kumpulan Wang Uni Penyelidikan	1001.228.0.PELECT.814161	1,000.00	0.00	0.00	0.00	0.00	0.00	1,000.00	0.00%		
Detail	Excel	47	T	Projek Kumpulan Wang Uni Penyelidikan	1001.229.0.PELECT.814161	-36,582.31	0.00	0.00	0.00	0.00	0.00	-36,582.31	0.00%		
		47	T	SubTotal		-45,218.09	0.00	0.00	0.00	0.00	0.00	-45,218.09	0.00%		
Detail	Excel	48	T	Projek Kumpulan Wang Uni Penyelidikan	1001.335.0.PELECT.814161	-2,025.00	0.00	0.00	0.00	0.00	0.00	-2,025.00	0.00%		
		48	T	SubTotal		-2,025.00	0.00	0.00	0.00	0.00	0.00	-2,025.00	0.00%		
		9999		GrandTotal		35.47	0.00	0.00	0.00	0.00	0.00	35.47	0.00%		



USM UNIVERSITI
SAINS
MALAYSIA

BORANG PENYERAHAN ASET / INVENTORI

A. BUTIR PENYELIDIK

1. NAMA PENYELIDIK : Dr. Dzat' Athiar Ramli
 2. NO STAF : AE 50205
 3. PTJ : PPKejuteraan Elektrik Elektronik
 4. KOD PROJEK : PELECT 1574161
 5. TARIKH TAMAT PENYELIDIKAN : 14 / 10 / 2015

B. MAKLUMAT ASET / INVENTORI

BIL	KETERANGAN ASET	NO HARTA	NO. SIRI	HARGA (RM)
1	RU ASUS 550 CM NB	AK00007126	D4N0CV4144 25176	4,985.00
2	SAMUNG GALAXY SIII	AK0309	RF1C7P35FDR	2,070.00
3	HTC ONE X	AK0309	SH257W111653	1,830.00

C. PERAKUAN PENYERAHAN

Saya dengan ini menyerahkan aset/ inventori seperti butiran B di atas kepada pihak Universiti:

(Dr. Dzat' Athiar Ramli) Tarikh: 25 Jun 2016

D. PERAKUAN PENERIMAAN

Saya telah memeriksa dan menyemak setiap alatan dan didapati :

- Lengkap
 Rosak
 Hilang : Nyatakan.....
 Lain-lain : Nyatakan masih digunakan oleh pelajar swasta

Diperakukan Oleh :

Tandatangan
Pegawai Aset PTJ

KHAIRUL ANUAR BIN AB. RAZAK
 Penolong Jurutera
 Pusat Pengajian Kejuruteraan Elektrik & Elektronik
 Universiti Sains Malaysia
 Kampus Kejuruteraan

Nama :
 Tarikh : 13/07/2016

*Nota : Sesalanan borang yang telah lengkap perlulah dikemukakan kepada Unit Pengurusan Harta, Jabatan Bendahari dan Pejabat RCMO untuk tujuan rekod.

Research Article

A Robust and Fast Computation Touchless Palm Print Recognition System Using LHEAT and the IFkNCN Classifier

Haryati Jaafar, Salwani Ibrahim, and Dzati Athiar Ramli

Intelligent Biometric Group, School of Electrical and Electronic Engineering, Universiti Sains Malaysia Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia

Correspondence should be addressed to Dzati Athiar Ramli; dzati@usm.my

Received 20 October 2014; Revised 25 April 2015; Accepted 29 April 2015

Academic Editor: Dominic Heger

Copyright © 2015 Haryati Jaafar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mobile implementation is a current trend in biometric design. This paper proposes a new approach to palm print recognition, in which smart phones are used to capture palm print images at a distance. A touchless system was developed because of public demand for privacy and sanitation. Robust hand tracking, image enhancement, and fast computation processing algorithms are required for effective touchless and mobile-based recognition. In this project, hand tracking and the region of interest (ROI) extraction method were discussed. A sliding neighborhood operation with local histogram equalization, followed by a local adaptive thresholding or LHEAT approach, was proposed in the image enhancement stage to manage low-quality palm print images. To accelerate the recognition process, a new classifier, improved fuzzy-based k nearest centroid neighbor (IFkNCN), was implemented. By removing outliers and reducing the amount of training data, this classifier exhibited faster computation. Our experimental results demonstrate that a touchless palm print system using LHEAT and IFkNCN achieves a promising recognition rate of 98.64%.

1. Introduction

Palm print recognition has been widely investigated for the last decade in the field of pattern recognition. Similar to fingerprint recognition, palm print technology is based on the aggregate of information presented in a friction ridge impression. Although the image quality of a fingerprint is robust because of multiple lines, wrinkles, and ridges, a palm print includes even more information. A palm print covers a wider area than a fingerprint and contains characteristics such as palmar creases and triradius that are useful for recognition [1]. More importantly, ridge structures remain unchanged throughout life, except for a change in size [2]. A palm print is distinctive and thick, enabling easy capture by low-resolution devices. Therefore, palm print detection systems have a low cost and require minimum user cooperation for extraction [3]. Most palm print biometrics utilizes scanners or charge-coupled device (CCD) cameras as the input sensor [4, 5]. Because users must touch the sensor to acquire their hand images, users are concerned about hygiene, particularly in public areas, such as hospitals, malls, and streets [6, 7]. Disease-causing organisms, such

as influenza virus, can be passed by indirect contact, and a susceptible individual can be infected from contact with a contaminated surface. The surface can become contaminated easily [6]; therefore, a touchless approach is required for palm print biometric technology.

The development of a touchless palm print recognition system is not straightforward. The hand position of the user during image acquisition is always changing. A touchless system does not require the user to touch or hold any platform or guidance peg. Users can open their hand, close their hand, or pose in a natural manner [6], and the hand can be deformed in other manners, including rotation, scale variability, and palm stretching, compared with touch-based systems [8]. Therefore, hand tracking and valley detection are challenging. As a result, hand tracking and region of interest (ROI) segmentation are difficult to implement. Complex backgrounds, poor ridge structures, and small image areas result in low-quality palm print images. The presence of noise/degradation (linear or nonlinear) and illumination changes [9] may reduce recognition accuracy. The computation times for the recognition process also must be considered. Because palm print systems consist of many major processes, such

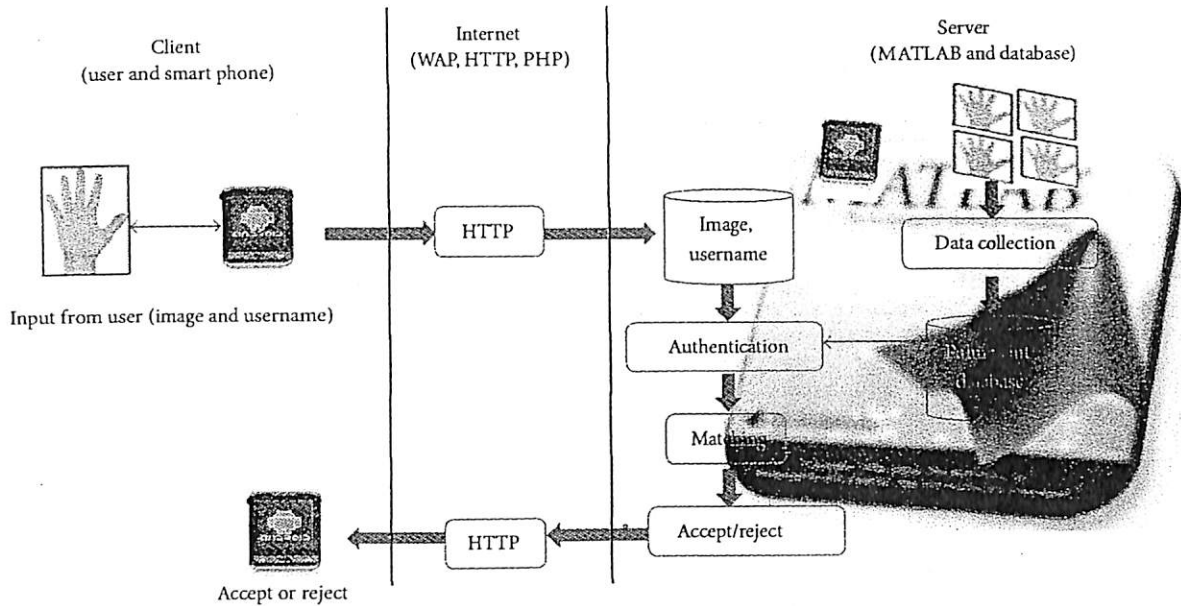


FIGURE 1: Overall research architecture.

as data acquisition, preprocessing, feature extraction, and classification, fast processing algorithms are crucial [10, 11].

This paper focuses on solutions for low-quality palm print images and computation times and includes a brief discussion of hand tracking and ROI segmentation. The overall research can be divided into three parts which are the client or smart phone side, internet side, and the server side which are illustrated as in Figure 1.

For the client side, the Android application for capturing biometric data is developed and it programs by using the latest few versions of Android OS, ranging from version 1.6 to version 4.1.2. Its programs support the mobile phone camera with the resolution up to 3.2 megapixels; hence only a few smart phones can be used for testing. Due to the existing camera application that varies for almost all smart phones and tablets, a customized camera application with the integration of enrolment and identification functions is developed for this research. The internet site is to connect the communication between smart phone and server and the connection is done via Wi-Fi and the PHP script is created to invoke the MATLAB program in the server.

The last part is server side where all the MATLAB programming including hand image identification, ROI extraction, palm print feature extraction, and pattern matching algorithms is written. The server software used in the project is free software where a personal computer serves as a server and has limited access from the client. Several palm print feature extraction algorithms which are based on subspace method are developed and evaluated for the fast and efficient mobile biometric system. Details of these operations can be found in Ibrahim and Ramli [12].

This study focused on the server side where two major contributions, that is, image enhancement and classification processes, have been developed to improve the quality of

touchless palm print recognition systems. We propose a local histogram equalization and adaptive thresholding (LHEAT) technique for image enhancement. This technique is an improved version of the local histogram equalization (LHE) and local adaptive thresholding (LAT) techniques. Unlike previous methods [13–16], we used the sliding neighborhood operation for faster computation [17]. To accelerate the recognition process, the improved fuzzy-based k nearest centroid neighbor (IFkNCN) was used as the classifier for the system. The sliding neighborhood operation in the LHEAT technique also reduces the processing time of the image enhancement stage compared with the baseline LHE and LAT techniques.

This paper is organized as follows. Section 2 presents related works and motivation. The proposed classifier for the palm print recognition system is described in Section 3. The experimental results are explained in Section 4, and Section 5 summarizes the work.

2. Related Works and Motivation

Many methods have been proposed to overcome the challenges associated with palm print recognition. Han and Lee [5] described two CMOS web cameras placed in parallel to segment the ROI of 1200 palm print images of identical size. The first camera captures the infrared image for hand detection, and the second camera is used to acquire the color image in normal lighting. The images are normalized using information on skin color and hand shape. The normalized images are then segmented to determine the ROI using the ordinal code approach and then classified with the Hamming distance classifier. Experimental results have shown that the equal error rate (EER) of the verification test is 0.54% and that the average acquisition time is 1.2 seconds. Feng et al. [18]

used the Viola-Jones method [19] to detect the hand position after capturing 2000 images. In this study, images were acquired in different positions with various lighting and cluster backgrounds. Subsequently, a coarse-to-fine strategy was used to detect the key points on the hand. The key hand points were then verified with the shape context descriptor before the images were segmented into the ROI. The boosting classifier cascade [20] has previously been applied, and the accuracy rate was 93.8%, with a 178 ms average processing time for one image. Michael et al. [2] described a touchless palm print recognition system that was designed using a low-resolution CMOS web camera to acquire real-time palm print images. A hand tracking algorithm, that is, skin color thresholding and hand valley detection algorithm, was developed to automatically track and detect the ROI of the palm print. The Laplacian isotropic derivative operator was used to enhance the contrast and sharpness of the palm print feature, and a Gaussian low-pass filter was applied to smooth the palm print image and bridge some small gaps in the line. The modified probabilistic neural network (PNN) was used to classify the palm print texture. The accuracy rate was greater than 90%. Similar to previous studies, Michael et al. [21] used local-ridge-enhancement (LRE) to enhance the contrast and sharpness of images of both the right and left hands. The LRE was used to determine which section of the image contains important lines and ridge patterns and then amplify only those areas. The support vector machine (SVM) was used, and the average accuracy rates for the left and right hands were 97% to 98%, respectively.

Although previous researchers have achieved greater than 90% accuracy, the palm print image was captured in a semi-closed environment in a boxlike setup with an illumination source on top. This setup results in clean images with prefixed illumination settings [22]. The high accuracy is not reflective of the real environment. In the present study, an Android smart phone was used to capture the images, allowing users to easily access their system every day. Because the images were captured in the real environment, they were exposed to different levels of noises and blurring because of variations in illumination, background, and focus. Noise can also be due to bit errors in transmission or introduced during the signal acquisition stage.

We propose a touchless palm print recognition system that can manage real environment variability. The two areas discussed are image enhancement and classification. In image enhancement, a LHEAT technique was used. The purpose of LHE is to ensure that the brightness levels are distributed equally [15, 23]. In the LHE, the image is divided into small blocks or local $N \times M$ neighborhood regions. Each block or inner window is surrounded by a larger block or outer window, which is used to calculate the mapping function lookup for the inner window. To remove the borders of the block, the mapping function is interpolated between neighborhood blocks [15]. The LHE is an excellent image enhancement method. However, in the palm print image, considerable background noise and variation in contrast and illumination exist. Occasionally, the LHE overenhances the image contrast and causes degradation of the image [13, 14, 16]. Then, the binarization technique, LAT, is applied. In LAT,

the threshold extracts the useful information from an image that has been enhanced by LHE and separates the foreground from the background with nonuniform illumination. Several methods, such as those described in Bersen, Niblack, Chow and Kaneko, and Sauvola [24], have been used to calculate the threshold values. Sauvola's method is most frequently used and was implemented here because of its promising results for degraded images.

In the pattern recognition system, there are two modes of recognition: verification and identification. This study focuses on the touchless palm print recognition system with identification mode. The identification mode is the time during which the system recognizes the user's identity by comparing the presented sample against the entire database to find a possible match [2]. Choosing the correct classification model becomes an important issue in palm print recognition to ensure that the system can identify a person in a short time. The k nearest neighbor (kNN) method is a nonparametric classifier widely used for pattern classification. This classifier is simple and easy to implement [25]. Nevertheless, there are some problems with this classifier; the performance of kNN often fails because of the lack of sample distribution information [26, 27] and not carefully assigning the class label before classification [28]. IFkNCN may resolve these limitations. This classifier incorporates centroid-based distance and fuzzy rule approaches with triangle inequality. The classifier removes the training samples that are far from the testing point or the query point by setting a threshold. The training samples that are located outside of the threshold are called outliers and defined as a noisy sample, which does not fit to the assumed class label for the query point. By removing the outliers, future processing focuses on the important training samples or candidate training samples, and this focus reduces the computational complexity in the searching stage. The query point is classified based on the centroid-distance and fuzzy rule system. The centroid-distance method is applied to ensure that the selected training samples are distributed sufficiently in the region of the neighborhood with the nearest neighbors located around the query point. Consequently, the fuzzy-based rule is used to solve the ambiguity of the weighting distance between the query point and its nearest neighbors.

3. Proposed Method

Figure 2 displays the overall procedure for a touchless palm print recognition system.

In this work, a new comprehensive collection of palm print database was developed. This database currently was containing 2400 color images corresponding to 40 users who were Asian race students where each user had 60 palm print images. This database will be released to the public as benchmark data and it can be downloaded from the website of Intelligent Biometric Group (IBG), Universiti Sains Malaysia (USM), for research and educational purposes. All the users who are taking part in the data collection are completely voluntary and each volunteer gave verbal consent before collecting the image. The age of the user ranged from

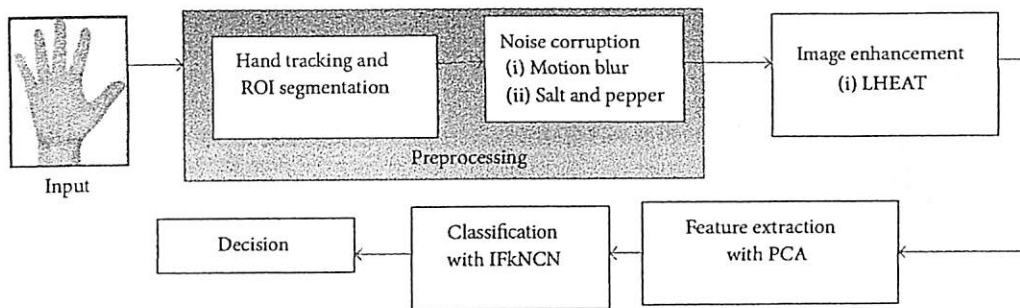


FIGURE 2: Block diagram of a touchless palm print recognition system.

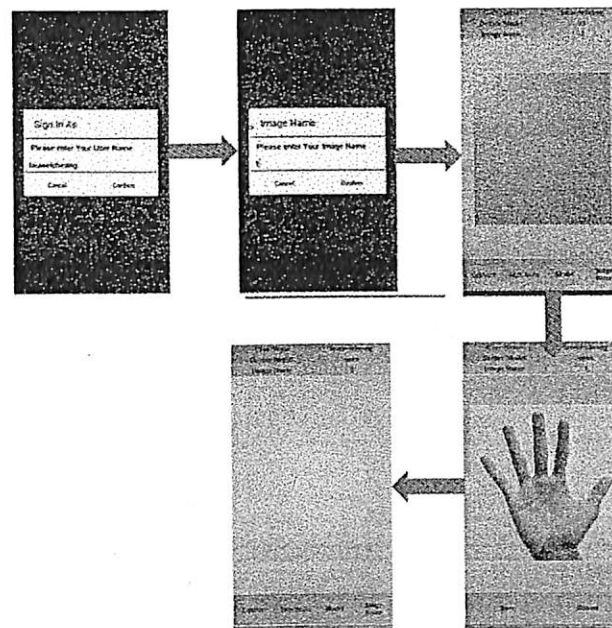


FIGURE 3: Data enrolment process.

19 to 23 years. An input image is acquired using a HTC One X Android mobile phone with 8 megapixels of image resolution and a stable background. The data collection is divided into 3 sessions; the first session is used for training purpose. The latter two sessions are used for testing purpose. The time interval for each session is in two weeks' time.

For enrolment process, a user needs to follow the instruction displayed on the smart phone screen as shown in Figure 3. Firstly, the user was required to sign in and key in the image name. Subsequently, the users were simply asked to put their palm print naturally in front of the acquisition device. A semitransparent pink color box acts as a constraint box to ensure the palm and fingers lie inside the box. The pixels that lie outside of the constraint will be cropped. So the distance between hand and device is set as constant. Once the image was captured, it was saved into the database and this process was repeated for new image and user.

As no peg or other tool is used in the system, the users may place their hands at different heights above the mobile phone camera. The palm image appears large and clear when

the palm is placed near the camera. Many line features and ridges are captured at near distance. However, if the hand is positioned too close to the mobile phone, the entire hand may not be captured in the image, and some parts of the palm print image may be occluded, as shown in Figure 4(a) [6]. When the hand is moved away from the camera, the focus fades, and some print information disappears (Figure 4(b)) [2]. The optimal distance between the hand and mobile phone is set according to the image preview in the enrolment process in Figure 3, enabling the whole hand image to be captured, as shown in Figure 4(c). Some examples of image of the whole palm print are shown in Figure 5.

The file were stored in JPEG format. Each folder was named as "S_x." "S_x" represents the identity of the user which ranges from 1 to 40. Each folder had 60 palm print images. During preprocessing, the image was segmented to determine the ROI. This process is called hand tracking and ROI segmentation. The image was then corrupted by adding noises, such as motion blur noise and salt and pepper noise. Subsequently, the LHEAT method was applied to enhance

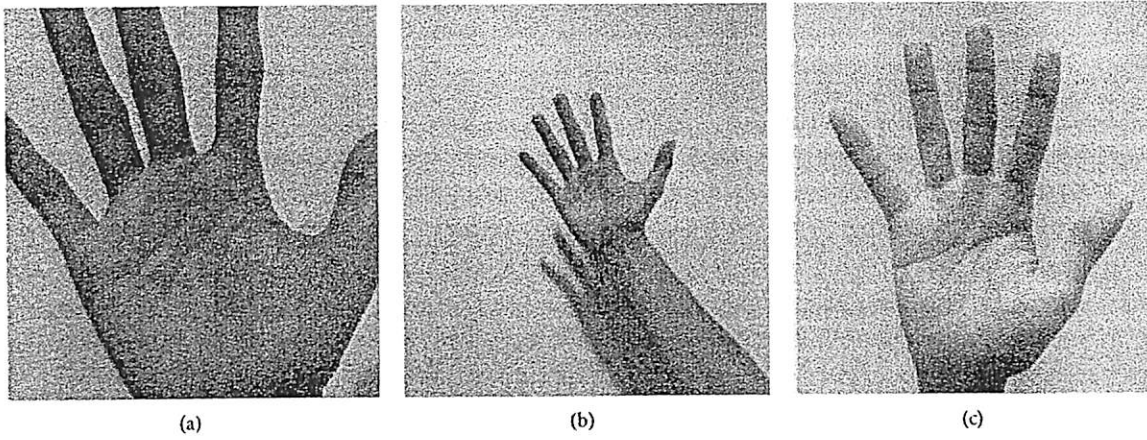


FIGURE 4: Hand image detection: (a) original RGB hand image; (b) binarized image. Hand position: (a) too close; (b) too far; and (c) suitable distance.

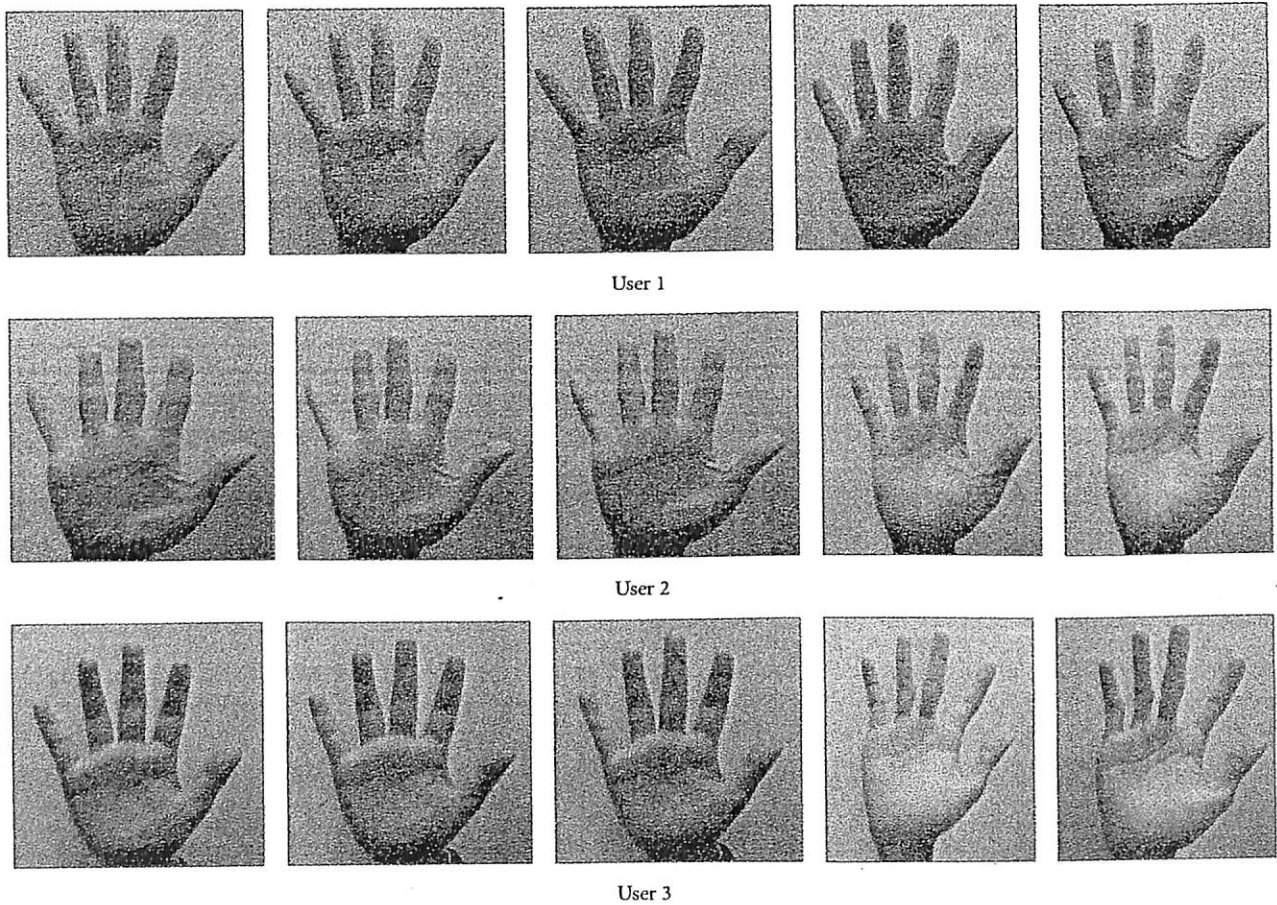


FIGURE 5: Original hand images captured by a smart phone camera for 5 different samples.

the image. Then, feature extraction was performed. Principle analysis component (PCA) was employed to extract the image data and reduce the dimensionality of the input data. Finally, the image was classified by the IFkNCN classifier.

3.1. *Preprocessing.* There are three major steps in the hand tracking and ROI segmentation stage: hand image identification, peak and valley detection, and ROI extraction [12]. In the hand image identification step, the RGB image is

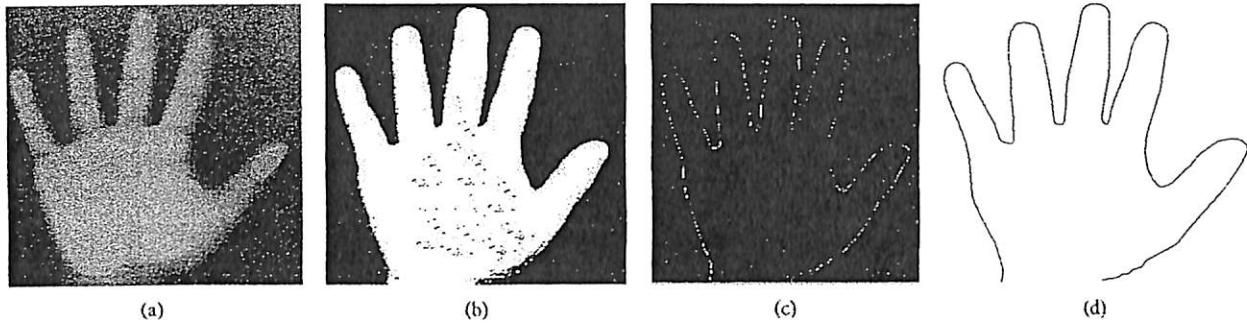


FIGURE 6: Hand image detection: (a) original RGB hand image; (b) binarized image; (c) hand contour with the Canny method; (d) perfect hand boundary plot.

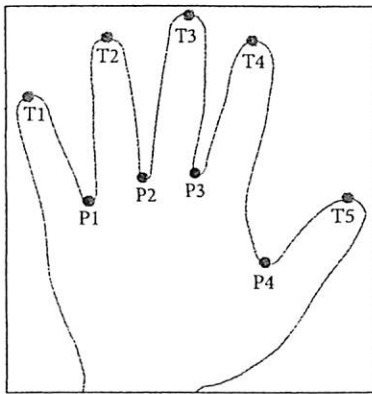


FIGURE 7: Five peaks and four valleys indicate the tips and roots of the fingers.

transformed into a grayscale image and then converted to a binary image. Because the lighting conditions in the camera setup are uncontrolled, straightforward hand identification is not possible. Noise results in many small holes. The noise and unsmooth regions are removed by filling the small holes in the hand region. Once the noise is removed, the edge of the image is detected using the Canny edge detection algorithm. The hand boundary of the image is traced before the perfect hand counter is acquired, as shown in Figure 6.

Because the image was captured without pegs or guiding bars, the palm print alignment varied in each collection. This variation caused the palm print image to be affected by rotation and may hamper accurate recognition. Therefore, the local minima and local maxima methods were used to detect peaks and valleys [29]. As shown in Figure 7, the peak and valley points in the hand boundary image were sorted and named before ROI segmentation.

The locations of three reference points, P1, P2, and P3, need to be detected in order to set up a coordinate system for palm print alignment. The size of ROI is dynamically determined by the distance between P1 and P3. It makes the ROI extraction scale invariant. To locate the ROI, a line is drawn between reference points; for example, P1 and P3 are shown in Figure 8(a) and labeled as "d." The image was then rotated using a command "imrotate" in MATLAB function

in order to ensure that the line was drawn horizontally as shown in Figure 8(b). The rotated image has the same size as the input image. A square shape was drawn, as shown in Figure 8(c), in which the length and width of the square were obtained as

$$a = d + \frac{d}{6.5}. \quad (1)$$

The ROI was segmented, and the region outside the square was discarded. Then, the ROI was converted from RGB to grayscale.

To investigate the performance of the proposed method in noisy environments, the ROI image was corrupted using motion blur noise and salt and pepper noise, as shown in Figure 9. The level of source noise (σ) was set to 0.13.

3.2. Image Enhancement. Image enhancement is an important process that improves the image quality. Similar to the LHE and LAT methods, in the LHEAT method, the input image is broken into small blocks or local window neighborhoods that contain a pixel. In the LHEAT, the LHE is firstly obtained to ensure an equal distribution of the brightness levels. The LAT is employed to extract the useful information of the image that had been enhanced by the LTE and separated the foreground from the nonuniform illumination background. An input image is broken into small blocks or local window neighborhoods containing a pixel. This is similar in the LHE, LAT, and LHEAT. Each block is surrounded by a larger block. The input image is defined as $X \in R^{H \times W}$, with dimensions of $H \times W$ pixels, and the enhanced image is defined as $Y \in R^{H \times W}$, with $H \times W$ pixels. The input image is then divided into the block $T_i = 1, \dots, n$ of window neighborhoods with the size $w \times w$, where $w < W$, $w < H$, and $n = \{(H \times W)/(w \times w)\}$.

Each pixel in the small block is calculated using a mapping function and threshold. The size of w should be sufficient to calculate the local illumination level, both objects, and the background [24]. However, this process results in a complex computation. To reduce the computation complexity and accelerate the computation, we used the sliding neighborhood operation [17]. Figure 10 shows an example of the sliding neighborhood operation. An image with a size of 6×5 pixels was divided into blocks of window

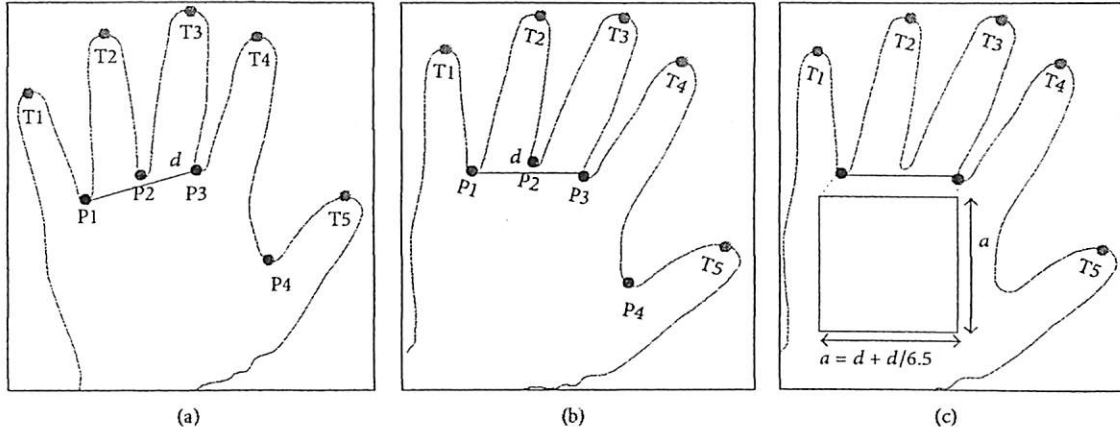


FIGURE 8: ROI segmentation process: (a) line drawn from P1 to P3; (b) rotated image; (c) ROI selection and detection.

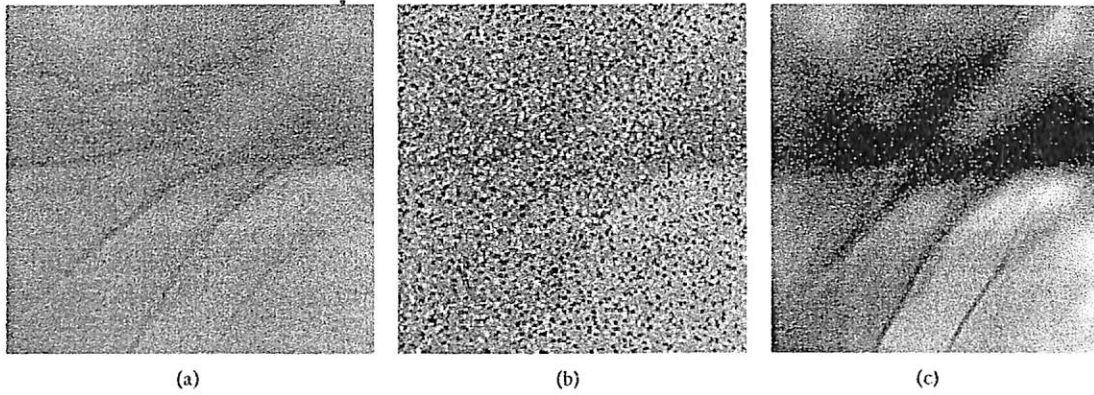


FIGURE 9: ROI image: (a) original; (b) degraded with salt and pepper noise; (c) degraded with motion blur noise.

neighborhoods with a size of 3×3 pixels. It is shown in Figure 10(a). The 6×5 image matrix was first rearranged into a 30-column ($6 \times 5 = 30$) temporary matrix, as shown in Figure 10(b). Each column contained the value of the pixels in its nine-row ($3 \times 3 = 9$) window. The temporary matrix was then reduced by using the local mean (M_i):

$$M_i = \frac{1}{N} \sum_{j=1}^n w_j, \quad (2)$$

where w was size of window neighborhoods, j was the number of pixels contained in each neighborhood, i was the number of columns in temporary matrix, and N was the total number of pixels in the block. After determining the local mean in (2), there was only one row left as shown in Figure 10(c). Subsequently, this row was rearranged into the original shape as shown in Figure 10(d).

There are three steps in the LHE technique: the probability density (PD), the cumulative distribution function (CDF), and the mapping function. The probability distribution of image PD for each block can be expressed as follows:

$$P(i) = \frac{n_i}{N} \quad \text{for } i = 0, 1, \dots, L-1, \quad (3)$$

where n_i is the input pixel number of level, i is the input luminance gray level, and L is gray level, which is 256.

Subsequently, the LHE uses an input-output mapping derived from CDF of the input histogram defined as follows:

$$C(i) = \sum_{i=0}^n P(i). \quad (4)$$

Finally, the mapping function is determined from the CDF as follows:

$$g(i) = M + [(x_i - M) \times C(i)], \quad (5)$$

where M is the mean value from (2).

Although the image has been enhanced, it remains mildly degraded because of the background noise and variation in contrast and illumination. The image was corrupted with two noises, motion blur noise and salt and pepper noise. The median filter, which has a 3×3 mask, was applied over the grayscale image. For an enhanced image, $g(i)$, $q(i)$ is the output median filter of length l , where l is the number of pixels over which median filtering takes place. When l is odd, the median filter is defined as follows:

$$q(i) = \text{median} \left\{ g(i-k : i+k), k = \frac{(l-1)}{2} \right\}. \quad (6)$$

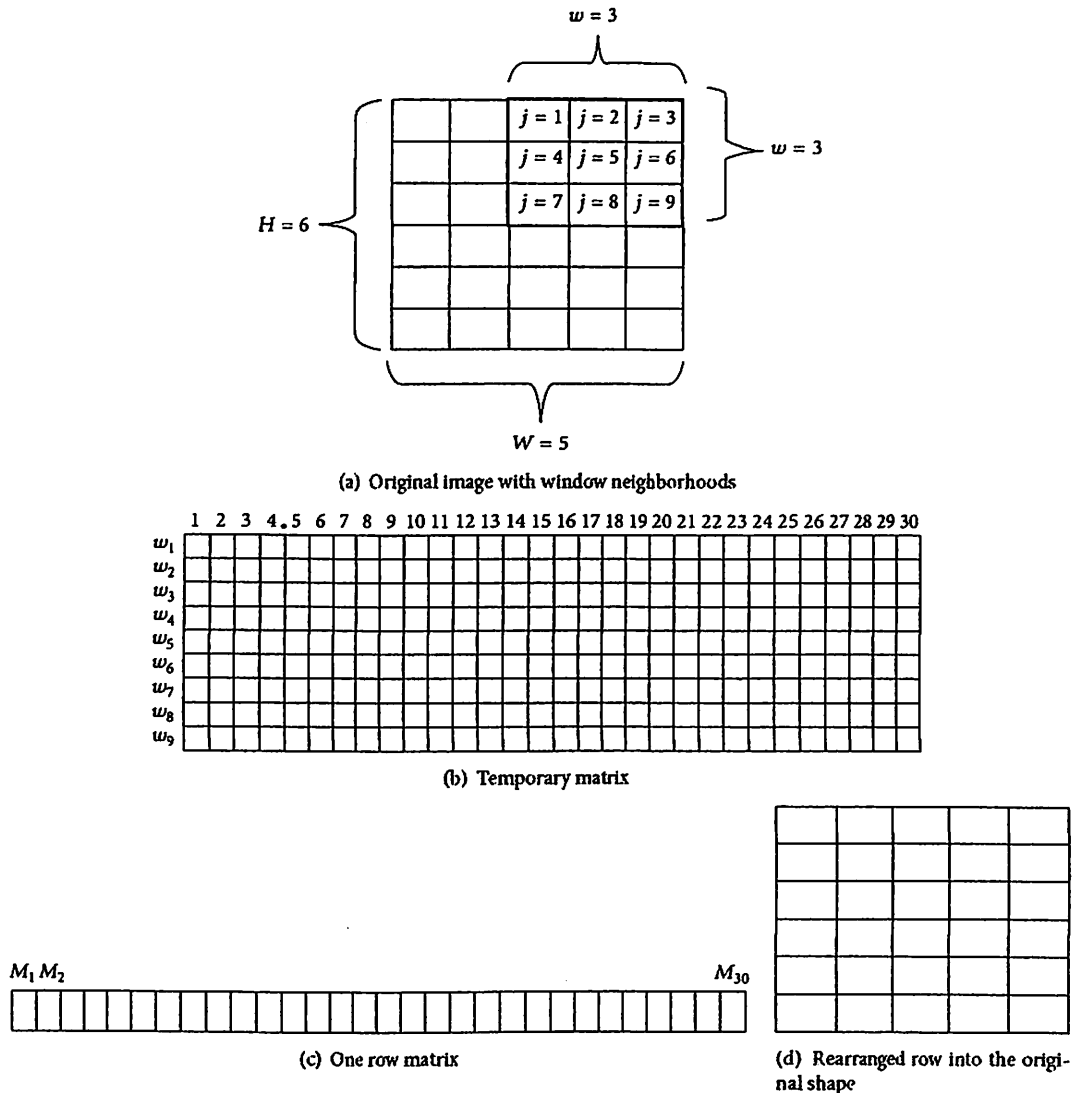


FIGURE 10: The sliding neighborhood operation.

When l is even, the mean of the two values at the center of the sorted sample list is used. The purpose of filtering is to reduce the effect of salt and pepper noise and the blur of the edge of the image.

Once the image has been filtered, the image is segmented using the LAT technique. The LAT separates the foreground from the background by converting the grayscale image into binary form. Sauvola's method was applied here, resulting in the following formula for the threshold:

$$T_h(i) = M \left[1 + k \left(\frac{Z}{R} - 1 \right) \right], \quad (7)$$

where T_h is the threshold, k is a positive value parameter with $k = 0.5$, R is the maximum value of the standard deviation,

which was set at 128 for grayscale image, and Z is the standard deviation which can be found as

$$Z = \sqrt{\frac{1}{N-1} \sum_{j=1}^n (w_j - M)^2}. \quad (8)$$

According to (8), the binarization results of Sauvola's method can be denoted as follows:

$$y(i) = \begin{cases} 1 & \text{if } q(i) > T_h(i) \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Figure 11 shows the comparison of output results after applying the LHE and LHEAT techniques. The detail in the enhanced image using LHEAT was sharper, and fine details, such as ridges, were more visible. Section 4.1 depicts

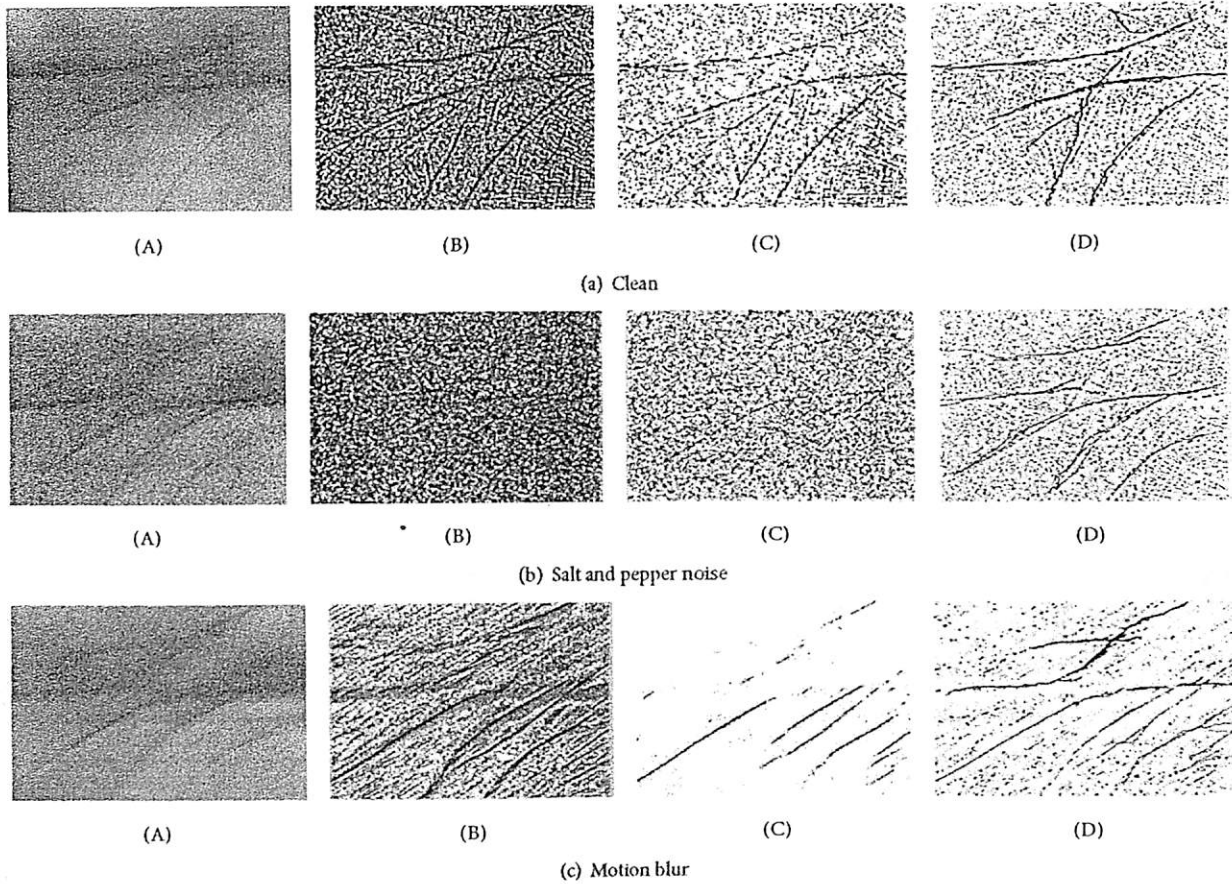


FIGURE 11: Comparison of image enhancement: (A) original image; (B) LHE; (C) LAT; and (D) LHEAT techniques.

the reduction in processing time and increased accuracy by applying the proposed image enhancement techniques.

3.3. Feature Extraction. Touchless palm print recognition must extract palm print features that can discriminate one individual from another. Occasionally, the captured images are difficult to extract because the line structures are discriminated individually. The creases and ridges of the palm cross and overlap one another, complicating the feature extraction task [30]. Recognition accuracy may decrease if the extraction is not performed properly.

In this paper, PCA was applied to create a set of compact features for effective recognition. This extraction technique has been widely used for dimensionality reduction in computer vision. This technique was selected because the features were more robust compared with other palm print recognition systems, such as eigenpalm [31], Gabor filters [32], Fourier transform [33], and wavelets [34].

The PCA transforms the original data from large space to a small subspace using a variance-covariance matrix structure. The first principle component shows the most variance while the last few principle components have less variance that is usually neglected since it has a noise effect.

Suppose a dataset $\{x_i\}$ where $i = 1, 2, \dots, N$ and x_i is rearranged in P^2 dimension. The PCA first computes the average vector of x_i and defined as

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (10)$$

whereas the deviations from x_i can be calculated by subtracting \bar{x} :

$$\Phi_i = x_i - \bar{x}. \quad (11)$$

This step obtains a new matrix:

$$A = [\Phi_1, \Phi_2, \dots, \Phi_n]. \quad (12)$$

That produces a dataset whose mean is zero. A is the $P^2 \times N$ dimensions.

Next, the covariance matrix is computed:

$$C = \sum_{i=1}^N \Phi_i \Phi_i^T = AA^T. \quad (13)$$

However, (13) will produce a very large covariance matrix which is $P^2 \times P^2$ dimensions. This causes the computation

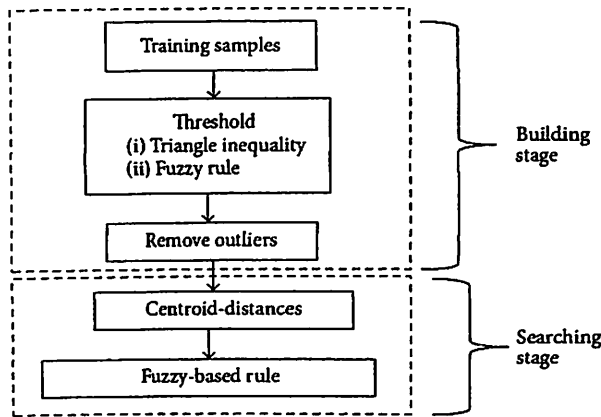


FIGURE 12: Architecture of the IFkNCN classifier.

required to be huge and the system may slow down terribly or run out of memory. As a solution, the dimensional reduction is employed where the covariance matrix is expressed as

$$C = A^T A. \quad (14)$$

Thus, the lower dimension of covariance matrix in $N \times N$ is obtained.

Next, the eigenvalues and eigenvectors of the C are computed. If the matrix $V = (V_1, V_2, \dots, V_p)$ contains the eigenvectors of a symmetric matrix C , then V is orthogonal, and C can be decomposed as

$$C = V D V^T, \quad (15)$$

where D is a diagonal matrix of the eigenvalues and V is a matrix of eigenvectors. Then, the eigenvalues and corresponding eigenvectors are sorted in the order to decrease the dimensions. Finally, the optimum eigenvectors are chosen based on the largest value of eigenvalues. The details of these procedures can be found in Connie et al. [30].

3.4. Image Classification. This section describes the methods used for the IFkNCN classifier. There were two stages for this classifier: the building stage and the searching stage (Figure 12). In the building stages, triangle inequality and fuzzy IF-THEN rules were used to separate the samples into outliers and train candidate samples. For the searching stage, the surrounding rule was based on centroid-distance, and the weighting fuzzy-based rule was applied. The query point was classified by the minimum distances of the k neighbors and sample placement, considering the assignment of fuzzy membership to the query point.

Building Stage. In this stage, the palm print images were divided into 15 training sets and 40 testing sets. The distance of testing samples or query point and training sets was calculated, and the Euclidean distance was used.

Given a query point y and training sets $T = \{x_j\}_{j=1}^N$, with $x_j = \{c_1, c_2, \dots, c_M\}$, N is the number of training sets, x_j is the sample from the training sample, M is the number class, and

c is the class label of M . The distance between the query point and training samples can be determined as follows:

$$d(y, x_j) = \sqrt{(y - x_j)^T (y - x_j)}, \quad (16)$$

where $d(y, x_j)$ is the Euclidean distance, N is the number of training samples, x_j is the training sample, and y is the query point.

The distances were sorted in *ascending* order to determine the minimum and maximum distance. The threshold was set such that the training samples fell within a selected threshold distance and were considered inliers. Otherwise, they were considered to be outliers. To determine the threshold, triangle inequality was applied. The triangle inequality method requires that the distance between two objects (reference point and training samples; reference point and query point) cannot be less than the difference between the distances to any other object (query point and the training samples) [35]. More specifically, the distance between the query point and training samples satisfies the triangle inequality condition as follows:

$$d(y, x_j) \leq d(x_j, z) + d(y, z), \quad (17)$$

where $d(y, z)$ is the distance from the query point to reference sample. In this study, the maximum distance obtained from (16) was assumed to be $d(y, z)$. For faster computation, the distance between training sample and reference sample $d(x_j, z)$ was discarded. To eliminate the computation of $d(x_j, z)$, (17) was rewritten as follows:

$$2d(y, x_j) \leq d(x_j, z) + d(y, z). \quad (18)$$

Because $d(y, x_j) \leq d(x_j, z)$, the value of $d(x_j, z)$ is not necessary, and (18) can be rearranged as follows:

$$d(y, x_j) \leq \frac{1}{2}d(y, z). \quad (19)$$

The choice of threshold values is important because a large threshold value requires more computation. A small threshold makes the triangle inequality computation useless. To tackle the problem, the candidate outlier detection can be expressed by the fuzzy IF-THEN rules. Each input set was modeled by two functions, as depicted in Figure 13.

The membership functions were formed by Gaussian functions or a combination of Gaussian functions given by the following equation:

$$f(x, \sigma, c) = e^{-(x-c)^2/2\sigma^2}, \quad (20)$$

where c indicates the center of the peak and σ controls the width of the distribution. The parameters for each of the membership functions were determined by taking the best performing values using the development set [21].

The output membership functions were provided as Outlierness = {High, Intermediate, Low} and were modeled as shown in Figure 14. They have distribution functions similar to the input sets (which are Gaussian functions).

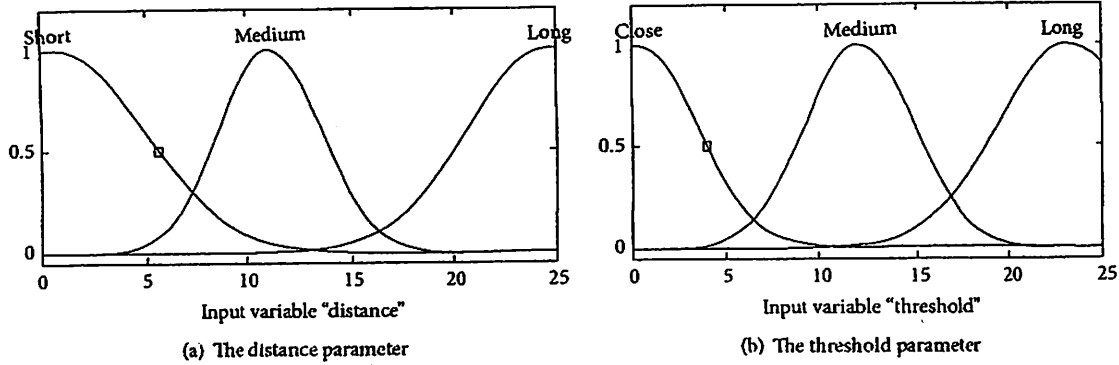


FIGURE 13: Input membership function.

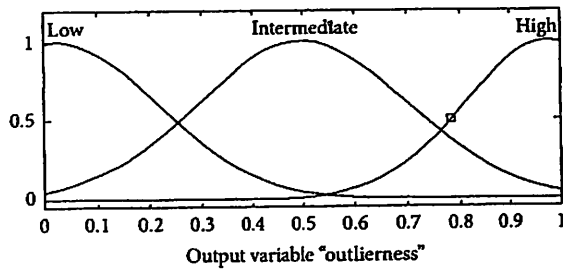


FIGURE 14: Output membership function.

The training sample was determined as an outlier if the distance of the training sample was long and the threshold was far and vice versa.

The Mamdani model was used to interpret the fuzzy set rules. This technique was used because it is intuitive and works well with human input. Nine rules were used to characterize the fuzzy rules. The main properties are as follows:

- (i) If the distance is short and threshold is small, then outlierness is low.
- (ii) If the distance is short and threshold is large, then outlierness is intermediate.
- (iii) If the distance is long and threshold is small, then outlierness is intermediate.
- (iv) If the distance is long and threshold is far, then outlierness is high.

The defuzzified output of the fuzzy procedure is influenced by the value of $d(y, x_j)$ and $d(y, z)$. The fuzzy performance with a training sample with $d(y, x_j) = 6.31$ and reference sample with $d(y, z) = 20$ is shown in Figure 15. The outlierness was 0.381, and the training sample was accepted as a candidate training sample. By removing the outlier, future processing only focuses on the candidate training samples.

Searching Stage. A surrounding fuzzy-based rule was proposed in which the rule is modified by the surrounding rule and the applied fuzzy rule. The main objective of this stage

was to optimize the performance results while considering the surrounding fuzzy-based rules which are as follows:

- (i) The k centroid nearest neighbors should be as close to the query point as possible and located symmetrically around the query point.
- (ii) The query point is classified by considering the fuzzy membership values.

Given a query point y , a set of candidate training samples $T = \{x_j \in R^m\}_{j=1}^N$, with $x_j = \{c_1, c_2, \dots, c_M\}$, where N is the number of training samples, x_j is the training sample, M is the number of classes, and c is the class label of M , the procedures of the IFkNCN in building stage can be defined as follows:

- (i) Select the candidate training sample as the first nearest centroid neighbor by sorting the distance of the query point and candidate training sample. Let the first nearest centroid neighbor be x_1^{NCN} .
- (ii) For $k = 2$, find the first centroid of x_1^{NCN} and the other candidate training samples are given as follows:

$$x_2^C = \frac{x_1^{NCN} + x_j}{2} \quad (21)$$

- (iii) Then, determine the second nearest centroid neighbors by finding the nearest distance of the first centroid and query point.
- (iv) For $k > 2$, repeat the second step to find the other nearest centroid neighbors by determining the centroid between the training samples and previous nearest neighbors:

$$x_k^c = \frac{1}{k} \sum_{j=1}^k x_j^{NCN} + x_j \quad (22)$$

- (v) Let the set of k nearest centroid neighbors $T_{jk}^{NCN}(y) = \{x_{jk}^{NCN} \in R^m\}_{j=1}^k$ and assign the fuzzy membership of the query point in every k nearest

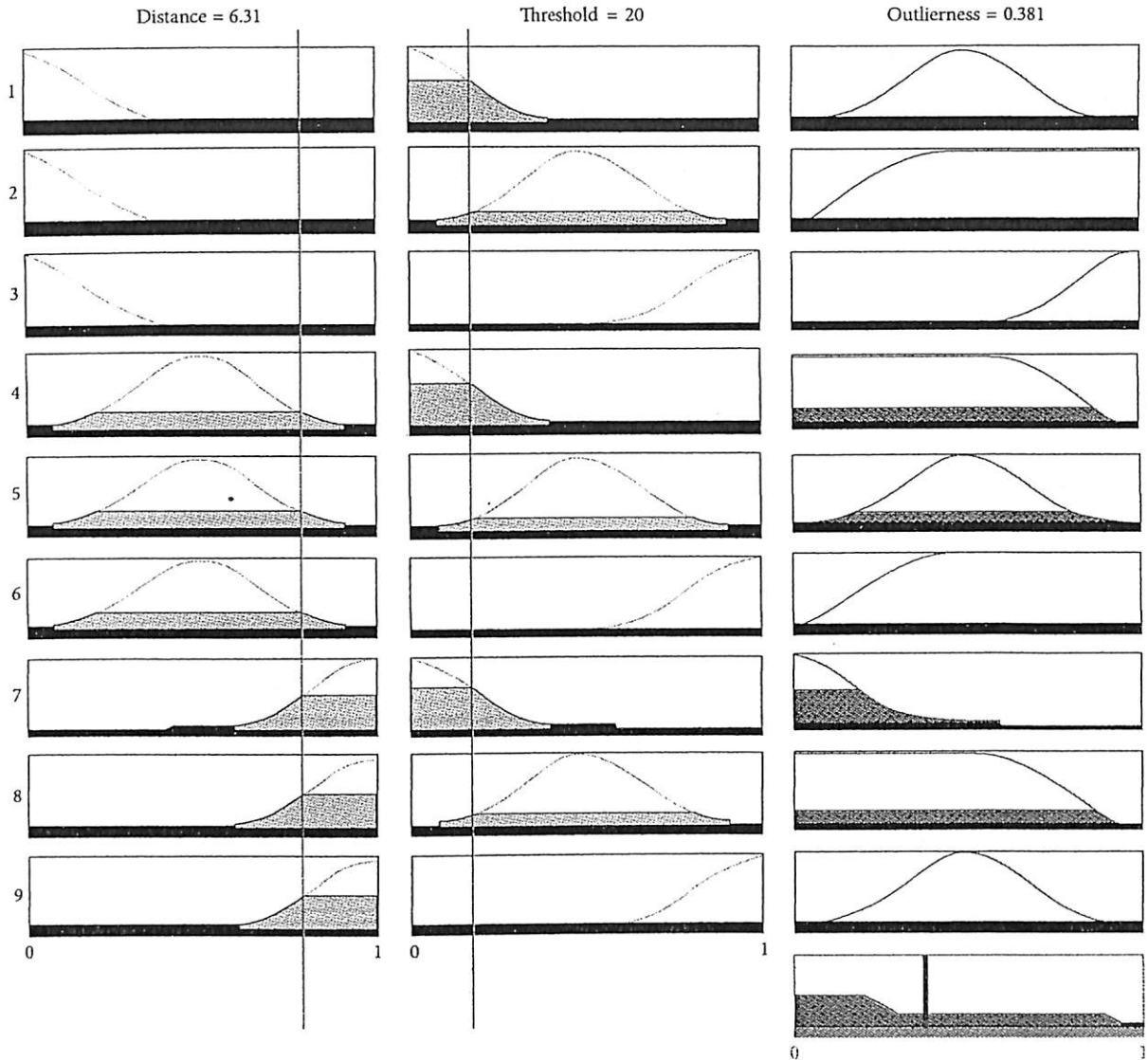


FIGURE 15: Example of the fuzzy IF-THEN rules.

centroid neighbor. The fuzzy membership is as follows:


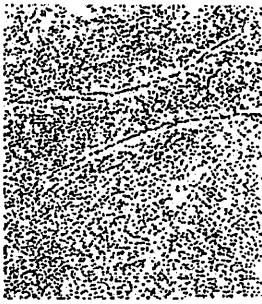
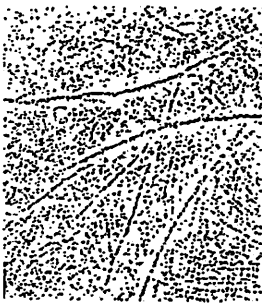

$$u_i^{\text{NCN}}(y) = \frac{\sum_{j=1}^k u_{ij} \left(1 / \|y - x_{jk}^{\text{NCN}}\|^{2/(m-1)} \right)}{\sum_{j=1}^k 1 / \|y - x_{jk}^{\text{NCN}}\|^{2/(m-1)}}, \quad (23)$$

where $i = 1, 2, \dots, c$ is the number of classes, u_{ij} is the membership degree of training sample, x_{jk} selected as the nearest neighbor, $\|y - x_{jk}^{\text{NCN}}\|$ is the L -norm distance between the query point x and its nearest neighbor, and m is a fuzzy strength parameter, which is used to determine how heavily the distance

is weighted when calculating each neighbor's contribution to the fuzzy membership values.

- (vi) For the value of the fuzzy strength parameter, the value of m is set to 2. If m is 2, the fuzzy membership values are proportional to the inverse of the square of the distance, providing the optimal result in the classification process.
- (vii) There are two methods to define u_{ij} . One definition uses the crisp membership, in which the training samples assign all of the memberships to their known class and nonmemberships to other classes. The other definition uses the constraint of fuzzy membership;

TABLE I: Performance with different sizes of the window neighborhood.

w	3	11	15	19
Image				
Time (s)	0.07	0.84	1.09	2.30

that is, when the k nearest neighbors of each training sample are found (say x_k), the membership of x_k in each class can be assigned as follows:

$$u_{ij}(x_k) = \begin{cases} 0.51 + 0.49 \left(\frac{n_j}{k} \right) & j = i \\ 0.49 \left(\frac{n_j}{k} \right) & j \neq i, \end{cases} \quad (24)$$

where n_j denotes the number of neighbors of the j th training samples.

The membership degree u_{ij} was defined using the constraint of fuzzy membership. The fuzzy membership constraint ensures that higher weight is assigned to the training samples in their own class and that lower weight is assigned to the other classes.

- (ix) The query point to the class label can be classified by obtaining the highest fuzzy membership value:

$$C(y) = \arg \max (u_i^{\text{NCN}}(y)). \quad (25)$$

- (x) Repeat steps (i) to (vii) for a new query point.

4. Experimental Results

As mentioned in Section 3, this study was conducted based on 2400 palm print images from 40 users. For each user, 15 images from the first session were randomly selected for training samples and the remaining 40 images from the second and third session were used as testing samples. Therefore, a total of 600 (15×40) and 1600 (40×40) images were used in the experiment. In order to gain an unbiased estimate of the generalization accuracy, the experiment was then run 10 times. The advantage of this method is that all of the test sets are independent and the reliability of the results can be improved.

Two major experiments, image enhancement and image classification, were conducted to evaluate the proposed touchless palm print recognition system. In the image enhancement experiment, three experiments were performed. The first experiment determined the optimal size of

the window neighborhood for the LHEAT technique. The second experiment validated the usefulness of the image enhancement technique by comparing the results with and without applying the image enhancement technique. The third experiment compared the proposed LHEAT technique with the LHE [23] and LAT [24] techniques. In the image classification, the first experiment determined the optimal value of k and size of feature dimensions for the IFkNCN classifier and compared the performance of the IFkNCN with kNN [25], k nearest centroid neighborhood (kNCN) [27], and fuzzy kNN (FkNN) [28] classifiers.

The performance for both image enhancement and image classification experiments was evaluated based on processing time and classification accuracy (C_A), where the C_A is defined as follows:

$$C_A = \frac{N_C}{N_T} \times 100\%, \quad (26)$$

where N_C is the number of query points, which is classified correctly, and N_T is the total number of the query points.

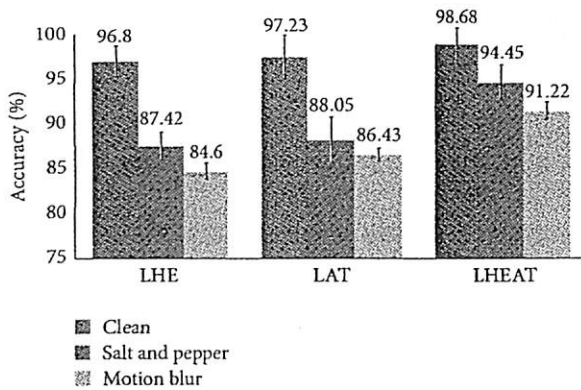
All experiments were performed in MATLAB R2007 (b) and tested on Intel Core i7, 2.1 GHz CPU, 6 G RAM, and Windows 8 operating system.

4.1. Image Enhancement. To determine the optimal size of window neighborhood, w for the proposed method, a clean image was obtained, and the values of w were set to 3, 9, 15, and 19. The performance result was based on image quality and processing time. The results are shown in Table 1. The window neighborhood of $w = 15$ provided the best image quality. Although the image quality for $w = 19$ was similar to $w = 15$, the processing time was longer. Therefore, to size the window neighborhood, $w = 15$ was used in the subsequent experiments.

This section also validates the utility of the image enhancement techniques discussed in Section 3.2. In this experiment, the palm print features were extracted using PCA with a feature dimension size fixed at 80. Then, the IFkNCN classifier was obtained, in which the value of k was set to 5. Table 2 shows the performance results with and without applying the image enhancement techniques. An improvement gain of approximately 3.61% in the C_A was

TABLE 2: Comparison of the image enhancement techniques.

Method	C_A (%)		
	Clean	Salt and pepper noise	Motion blur noise
Without image enhancement	96.40 \pm 1.14	86.40 \pm 2.07	88.80 \pm 1.48
With LHEAT technique	98.42 \pm 0.55	90.40 \pm 0.89	93.60 \pm 0.89

FIGURE 16: Performance of the LHE, LAT, and LHEAT methods for C_A .

achieved when the proposed image enhancement method was applied. Although the performance decreased because of degradation in image quality in the corrupted image, the image enhancement technique was able to recover more than 90% of the image compared with results without the image enhancement technique.

The next experiment investigated how the proposed LHEAT technique compared with previous techniques, such as LHE and LAT. The settings used in this experiment were the same as in the previous experiments. The result of the three experiments is shown in Figure 16. LHEAT performed better than LHE and LAT, yielding a C_A of more than 90% for the clean and corrupted images. LHE enhanced brightness levels by distributing the brightness equally and recovered original images that were over- and underexposed. When LAT was applied, the threshold changed dynamically across the image. LAT can remove background noise and variations in contrast and illumination. LHE and LAT in LHEAT complement one another and yield promising results.

LHEAT gives another advantage over other methods in terms of its simplicity in computation. Normally, LHE and LAT require a time complexity of $O(w^2 \times r^2)$ with an image of size $(n \times n)$ with a size of window neighborhood $(w \times w)$. However, in the proposed LHEAT technique, the time complexity is $O(n^2)$ because the sliding neighborhood is only used to obtain local mean (M) and local standard deviation (Z). Hence, the time required for LHEAT is much closer to global techniques. Figure 17 shows a comparison of computation times during the image enhancement process. The LHEAT technique outperformed the LHE and LAT techniques.

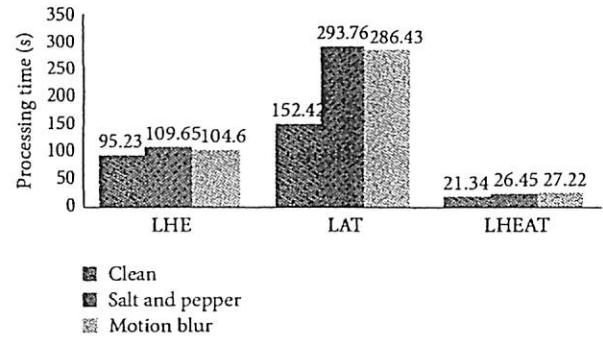


FIGURE 17: Performance of LHE, LAT, and LHEAT in processing time.

4.2. Image Classification. Following the image enhancement experiments, the efficiency and robustness of the proposed IFkNCN classifier were evaluated. The first experiment in this section determined the optimal k value for the IFkNCN classifier. To avoid situations in which the classifier "ties" (identical number of votes for two different classes), an odd number for k , such as 1, 3, 5, 7, 9, 11, 13, 15, and 17, was used, and the size of feature dimension was fixed to 80. Comparison results are summarized in Table 3. IFkNCN achieved the highest C_A results when k was 5 and 7. The best C_A values were $98.54 \pm 0.84\%$ ($k = 5$), $94.02 \pm 0.54\%$ ($k = 5$), and $91.20 \pm 1.10\%$ ($k = 7$) for clean, salt and pepper noise, and motion blur images, respectively. Because there was only a 0.12% difference between $k = 7$ and $k = 5$ for IFkNCN in motion blur images, the value of k is set to 5 to ease the calculation in the subsequent experiments. The results also showed that increasing the value of k further lowers the C_A . When k increases, the number of nearest neighbors of the query point also increases. In this situation, some training samples from different classes, which have similar characteristics, were selected as the nearest neighbor, and these training samples were defined as overlapping samples. Misclassification often occurs near class boundaries in which an overlap occurs.

The second experiment determined the optimal feature dimension size for the IFkNCN classifier. The k value was set to 5, and the size of the feature dimension was set to 20, 60, 80, 100, and 120. The results are shown in Table 4. As expected, the palm print recognition achieved optimal results when the size of the feature dimension was set to 120. However, the value also had the highest processing time. When the feature dimension was set to 100, the processing time was reduced twofold lower than the feature dimension of 120. The difference in C_A between the 100 and 120 feature dimensions was relatively small (approximately 0.10%). Therefore, a feature dimension of 100 was selected as the optimal value for IFkNCN, and this size was used for the next experiment.

The subsequent experiment evaluated the proposed classifier. A comparison of IFkNCN with other previous nearest neighbor classifiers, such as kNN, kNCN, and FkNN, was performed. The optimal parameter values, that is, $k = 5$ and

TABLE 3: Comparison of the CA results for different k values (results are in %).

Image	$k = 1$	$k = 3$	$k = 5$	$k = 7$	$k = 9$	$k = 11$	$k = 13$	$k = 15$	$k = 17$
Clean	96.02 ± 1.14	96.35 ± 0.95	98.54 ± 0.84	98.12 ± 0.98	97.67 ± 1.16	96.58 ± 1.24	96.34 ± 0.64	96.82 ± 1.14	96.34 ± 1.02
Salt and pepper	91.12 ± 0.82	93.54 ± 1.26	94.02 ± 0.54	93.84 ± 0.96	93.21 ± 1.12	93.15 ± 1.45	93.02 ± 0.98	92.34 ± 1.26	91.89 ± 0.66
Motion blur	88.02 ± 1.34	89.72 ± 1.22	91.08 ± 0.98	91.20 ± 1.10	90.33 ± 0.88	89.78 ± 0.45	89.54 ± 0.66	88.96 ± 1.82	89.02 ± 1.82

TABLE 4: Comparison of IFkNCN of different feature dimension values.

Dim.	Clean		Salt and pepper		Motion blur	
	Time (s)	C_A (%)	Time (s)	C_A (%)	Time (s)	C_A (%)
20	0.65	93.32 ± 1.22	0.74	91.50 ± 2.01	0.99	89.62 ± 1.52
40	0.86	93.56 ± 1.00	0.83	92.06 ± 1.88	1.03	90.12 ± 0.94
60	1.17	95.34 ± 0.94	1.15	92.95 ± 1.05	1.64	90.95 ± 1.32
80	1.54	98.64 ± 1.26	1.44	93.67 ± 1.22	1.71	91.02 ± 0.98
100	1.32	98.96 ± 0.55	1.46	94.11 ± 1.14	1.92	92.45 ± 1.14
120	5.43	99.02 ± 1.25	4.98	94.21 ± 1.35	5.24	92.49 ± 1.32

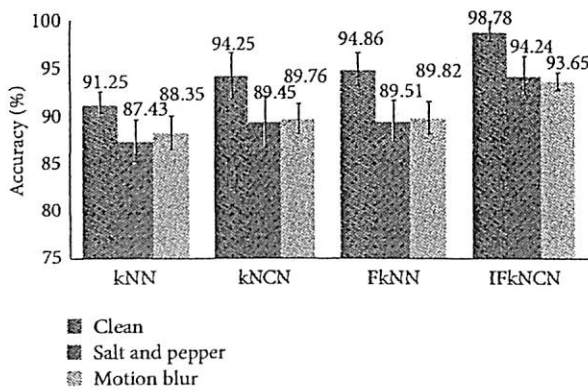


FIGURE 18: Comparison of IFkNCN with kNN, kNCN, and FkNN classifiers based on C_A .

a feature dimension of 100, were used. The overall performance results based on C_A are described in Figure 18. By utilizing the strength of the centroid neighborhood while solving the ambiguity of the weighting distance between the query point and its nearest neighbors, the IFkNCN classifier outperformed the kNN, kNCN, and FkNN classifiers. The C_A of the IFkNCN increased approximately 7.53%, 6.81%, and 5.3% in the clean, salt and pepper, and motion blur images, respectively, compared with kNN, kNCN, and FkNN.

In addition to better accuracy, the proposed IFkNCN classifier also had better processing times in all conditions, as shown in Figure 19. By using the triangle inequality and fuzzy IF-THEN rules, the training samples that were not relevant to additional processing were removed. Accuracy did not decrease, but the processing time was 2.39 s, whereas the processing times for kNN, kNCN, and FkNN were 7.82 s, 109.17 s, and 9.59 s, respectively.

The time required to execute each process, that is, image preprocessing, image enhancement, feature extraction, and

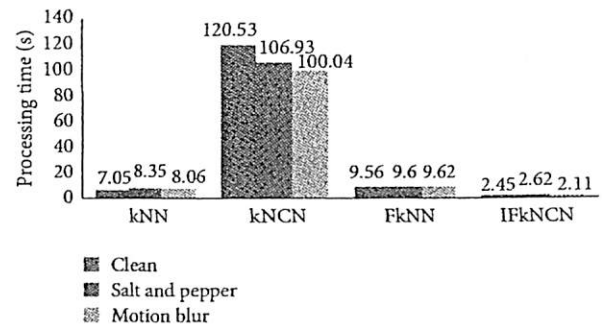


FIGURE 19: Comparison of IFkNCN with the kNN, kNCN, and FkNN classifiers based on processing time.

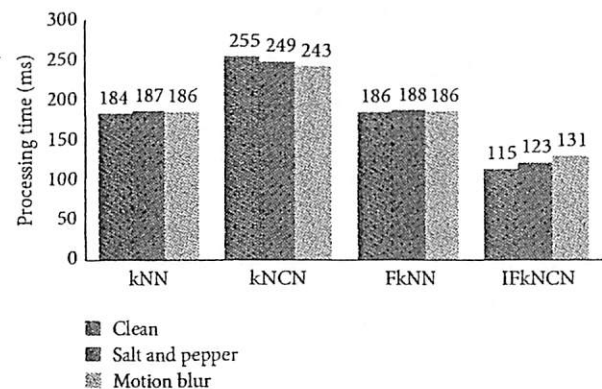


FIGURE 20: Processing speed of a touchless palm print system.

image classification, in the touchless palm print recognition is shown in Figure 20. The reported time is the average time required to process an input image from a user. The total time to identify a user was less than 130 ms.

The speed demonstrated by the proposed system demonstrates that it has the potential for implementation in real-world applications.

5. Conclusions and Future Works

This paper presents a touchless palm print recognition method using an Android smart phone. The proposed system is accessible and practical. In addition, the device is cost-effective and does not require expensive hardware. This paper focused on image enhancement and image classification. To enhance the quality of the acquired images, we propose the LHEAT technique. Because the sliding neighborhood operation is applied in the LHEAT technique, the computation was much faster compared with previous techniques, such as LHE and LAT. The proposed technique was also able to reduce noise and increase the dominant line edges in the palm print image. Moreover, this method works well in noisy environments. This paper also presents a new type of classifier, called IFkNCN, that has advantages compared with the kNN classifier. The major advantage of the IFkNCN classifier is that it can remove the outliers and that its computation is efficient. Extensive experiments were performed to evaluate the performance of the system in terms of image enhancement and image classification. The proposed system exhibits promising results. Specifically, the C_A with the LHEAT technique was more than 90%, and the processing time was threefold lower than with the LHE and LAT methods. In addition, the C_A achieved by the IFkNCN method was improved to more than 90% for clean and corrupted images, and the processing time was less than 120 ms, which was substantially less compared with the other tested classifiers. The proposed touchless palm print system is convenient and able to manage real-time recognition challenges, such as environmental noise and lighting changes.

Although the purpose of this research has been achieved, there are some aspects that need to be taken into consideration for future work. Firstly, in order to ensure the development of touchless palm print system is more applicable in real application, experiment in various types of noises needs to be extracted before the ROI extraction. So the filtered process can be improved before the subsequent process is applied. Secondly, additional algorithms in the image enhancement can be added to improve the LHEAT performance, especially when the image is captured in various types of illumination, background, and focus. However, addition of other algorithms may slow down the speed of this technique. Thus, this problem should be considered if the online or real-time processing algorithm is required. For the classification process, the code optimization could be conducted to increase the computational efficiency of the IFkNCN classifier during the searching stage. Since the complexity of each training sample in searching stage is high, the code optimization process will be beneficial in offering better solution to overcome this complexity problem.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to express their gratitude for the financial support provided by Universiti Sains Malaysia Research University Grant 814161 and Research University Postgraduate Grant Scheme 8046019 for this project.

References

- [1] Y. Zhou, Y. Zeng, and W. Hu, "Application and development of palm print research," *Technology and Health Care*, vol. 10, no. 5, pp. 383–390, 2002.
- [2] G. K. O. Michael, T. Connie, and A. B. J. Teoh, "Touch-less palm print biometrics: novel design and implementation," *Image and Vision Computing*, vol. 26, no. 12, pp. 1551–1560, 2008.
- [3] P. Somvanshi and M. Rane, "Survey of palmprint recognition," *International Journal of Scientific & Engineering Research*, vol. 3, no. 2, p. 1, 2012.
- [4] H. Imtiaz and S. A. Fattah, "A wavelet-based dominant feature extraction algorithm for palm-print recognition," *Digital Signal Processing*, vol. 23, no. 1, pp. 244–258, 2013.
- [5] W.-Y. Han and J.-C. Lee, "Palm vein recognition using adaptive Gabor filter," *Expert Systems with Applications*, vol. 39, no. 18, pp. 13225–13234, 2012.
- [6] G. K. O. Michael, C. Tee, and A. T. Jin, "Touch-less palm print biometric system," in *Proceedings of the International Conference on Computer Vision Theory and Applications*, pp. 423–430, 2005.
- [7] H. Sang, Y. Ma, and J. Huang, "Robust palmprint recognition base on touch-less color palmprint images acquired," *Journal of Signal and Information Processing*, vol. 4, no. 2, pp. 134–139, 2013.
- [8] X. Wu, Q. Zhao, and W. Bu, "A SIFT-based contactless palmprint verification approach using iterative RANSAC and local palmprint descriptors," *Pattern Recognition*, vol. 47, pp. 3314–3326, 2014.
- [9] A. K. Jain and J. Feng, "Latent palmprint matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 1032–1047, 2009.
- [10] L. Fang, M. K. H. Leung, T. Shikhare, V. Chan, and K. F. Choon, "Palmprint classification," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 2965–2969, October 2006.
- [11] H. Imtiaz and S. A. Fattah, "A spectral domain dominant feature extraction algorithm for palm-print recognition," *International Journal of Image Processing*, vol. 5, pp. 130–144, 2011.
- [12] S. Ibrahim and D. A. Ramli, "Evaluation on palm-print ROI selection techniques for smart phone based touch-less biometric system," *American Academic & Scholarly Research Journal*, vol. 5, no. 5, pp. 205–211, 2013.
- [13] T. Celik, "Two-dimensional histogram equalization and contrast enhancement," *Pattern Recognition*, vol. 45, no. 10, pp. 3810–3824, 2012.
- [14] M. Eramian and D. Mould, "Histogram equalization using neighborhood metrics," in *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision*, pp. 397–404, May 2005.

- [15] B. Kang, C. Jeon, D. K. Han, and H. Ko, "Adaptive height-modified histogram equalization and chroma correction in YCbCr color space for fast backlight image compensation," *Image and Vision Computing*, vol. 29, no. 8, pp. 557–568, 2011.
- [16] T. R. Singh, S. Roy, O. I. Singh, and K. Singh, "A new local adaptive thresholding technique in binarization," *International Journal of Computer Science Issues*, vol. 8, no. 6, p. 271, 2012.
- [17] J. L. Semmlow, *Biosignal and Medical Image Processing*, CRC Press, 2011.
- [18] Y. Feng, J. Li, L. Huang, and C. Liu, "Real-time ROI acquisition for unsupervised and touch-less palmprint," *World Academy of Science, Engineering and Technology*, vol. 78, pp. 823–827, 2011.
- [19] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, pp. 1511–1518, December 2001.
- [20] N. Vasconcelos and M. J. Saberian, "Boosting classifier cascades," in *Advances in Neural Information Processing Systems*, pp. 2047–2055, 2010.
- [21] G. K. O. Michael, T. Connie, and A. B. J. Teoh, "A contactless biometric system using multiple hand features," *Journal of Visual Communication and Image Representation*, vol. 23, no. 7, pp. 1068–1084, 2012.
- [22] C. Methani, *Camera based palmprint recognition [Doctoral Dissertation]*, International Institute of Information Technology, Hyderabad, India, 2010.
- [23] H. Zhu, F. H. Y. Chan, and F. K. Lam, "Image contrast enhancement by constrained local histogram equalization," *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 281–290, 1999.
- [24] Y.-T. Pai, Y.-F. Chang, and S.-J. Ruan, "Adaptive thresholding algorithm: efficient computation technique based on intelligent block detection for degraded document images," *Pattern Recognition*, vol. 43, no. 9, pp. 3177–3187, 2010.
- [25] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [26] X. Wu, V. Kumar, J. Ross Quinlan et al., "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2008.
- [27] B. B. Chaudhuri, "A new definition of neighborhood of a point in multi-dimensional space," *Pattern Recognition Letters*, vol. 17, no. 1, pp. 11–17, 1996.
- [28] J. Wang, P. Neskovic, and L. N. Cooper, "Improving nearest neighbor rule with a simple adaptive distance measure," *Pattern Recognition Letters*, vol. 28, no. 2, pp. 207–213, 2007.
- [29] L. Q. Zhu and S. Y. Zhang, "Multimodal biometric identification system based on finger geometry, knuckle print and palm print," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1641–1649, 2010.
- [30] T. Connie, A. Teoh, M. Goh, and D. Ngo, "Palmprint recognition with PCA and ICA," in *Proceedings of the Image and Vision Computing*, Palmerston North, New Zealand, 2003.
- [31] G. Lu, D. Zhang, and K. Wang, "Palmprint recognition using eigenpalms features," *Pattern Recognition Letters*, vol. 24, no. 9–10, pp. 1463–1467, 2003.
- [32] W. K. Kong, D. Zhang, and W. Li, "Palmprint feature extraction using 2-D Gabor filters," *Pattern Recognition*, vol. 36, no. 10, pp. 2339–2347, 2003.
- [33] W. Li, D. Zhang, and Z. Xu, "Palmprint identification by Fourier transform," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 4, pp. 417–432, 2002.
- [34] A. Kumar and H. C. Shen, "Recognition of palmprints using wavelet-based features," in *Proceedings of the IEEE International Conference on Systematic, Cybernetics and Informatics (SCI '02)*, Orlando, Fla, USA, 2002.
- [35] A. Berman and L. G. Shapiro, "Selecting good keys for triangle-inequality-based pruning algorithms," in *Proceedings of the IEEE International Workshop on Content-Based Access of Image and Video Database*, pp. 12–19, Bombay, India, 1998.

Finger Vein Identification using Fuzzy-based k-Nearest Centroid Neighbor Classifier

Bakhtiar Affendi Rosdi^a, Haryati Jaafar^b and Dzati Athiar Ramli^c

^aIntelligent Biometric Group, School of Electrical and Electronic, USM Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia

^aeebakhtiar@usm.my, ^bharyati.jaafar@yahoo.com, ^cdzati@usm.my

Abstract.

In this paper, a new approach for personal identification using finger vein image is presented. Finger vein is an emerging type of biometrics that attracts attention of researchers in biometrics area. As compared to other biometric traits such as face, fingerprint and iris, finger vein is more secured and hard to counterfeit since the features are inside the human body. So far, most of the researchers focus on how to extract robust features from the captured vein images. Not much research was conducted on the classification of the extracted features. In this paper, a new classifier called fuzzy-based k-nearest centroid neighbor (FkNCN) is applied to classify the finger vein image. The proposed FkNCN employs a surrounding rule to obtain the k-nearest centroid neighbors based on the spatial distributions of the training images and their distance to the test image. Then, the fuzzy membership function is utilized to assign the test image to the class which is frequently represented by the k-nearest centroid neighbors. Experimental evaluation using our own database which was collected from 492 fingers shows that the proposed FkNCN has better performance than the k-nearest neighbor, k-nearest-centroid neighbor and fuzzy-based-k-nearest neighbor classifiers. This shows that the proposed classifier is able to identify the finger vein image effectively.

Keywords: Finger vein, Biometrics, Nearest neighbor, Fuzzy, Classifier

PACS :07.05.Mh

INTRODUCTION

In the modern world, there is a high demand to authenticate and identify individuals automatically. Consequently, the technology such as personal identification number (PIN), smart card or passwords have been introduced [1]. However, these technologies are inadequate since they can be duplicated, misplaced, stolen and easy to be accessed by an imposter. For this reason, biometric has been introduced in the late 90s to recognize a person based on the physiological or biological characteristics [10]. Compared to the classical user authentication system, biometric technology provides a level of assurance that simply cannot be faked, stolen or lost. Due to the specific physiological or behavioral characteristic possessed by the user, this technology is more secure and reliable to be implemented in various fields such as door access controls, criminal investigations, logical access points and surveillance applications [11].

There are various kinds of modalities that can be utilized in the biometric systems such as fingerprint, iris, face, hand geometry, palm print, gait, voice and signature [1]. Among the available biometrics, face, iris and fingerprint are the most widely used modalities. However, there are some disadvantages that come along with these biometric modalities. For example, in the face biometrics, the users' faces will be changed over time. Moreover, in order to recognize faces accurately, the image must be acquired at a good pose. This is not always possible and can be very difficult to do in some environments [12]. As for the fingerprint, if the finger gets damaged and/or has one or more marks on it, identification becomes increasingly hard. Furthermore, the system requires the users' finger surface to have a point of minutiae or pattern in order to be matched with the registered data. This will be a limitation factor for the security of the algorithm [13]. On the other hands, the disadvantages of the iris biometric system are some individuals are difficult to capture and the iris can be easily obscured by eyelashes, eyelids, lens and reflections from the cornea. There is also a lack of existing data, which deters the ability to use for background or watch list checks [14].

To overcome the limitations of aforementioned biometric systems, a new technology based on finger vein pattern has been developed [2]. Recently, this biometric system has received a lot of researcher's attention due to their high user acceptance and exhibit some excellent advantages in this application. As compared to conventional biometrics such as fingerprint, face and iris, the features of the finger vein are inside the skin surface, which makes it difficult to be duplicated. Thus, it is more secure compared to other modalities and leads to the high recognition accuracy. In addition, as the veins are located inside the body; it is less likely to be influenced by changes in the

The 2nd ISM International Statistical Conference 2014 (ISM-II)
AIP Conf. Proc. 1643, 645-654 (2015); doi: 10.1064/1.4907507
2015 AIP Publishing LLC 978-0-7354-1281-1/\$30.00

weather or physical condition of the individual. Moreover, the rushes, cracked and rough skin does not affect the result of recognition [2].

To date, a number of methods have been studied to improve the accuracy of finger vein recognition. For example, a finger vein extraction method using repeated line tracking [15], local binary pattern (LBP) [16, 17], principal component analysis (PCA) [18], Gabor Wavelets and Circular Gabor Filter [19] were proposed. However, it was found most of the current available approaches for finger vein recognition are mainly focused on the feature extraction process. Apart from extracting the finger vein features, the classification is also a crucial factor that needs to be considered. To the best of our knowledge, only a few researchers pay attention to the classification process. For example, Yang et al. [20] propose to use Support Vector Machine (SVM) to classify the finger vein images. However, the biggest difficulty in the SVM is the model for classification is generated from the training stage using the sampling data. If the parameter values are not set properly, then the classification outcomes will be less than optimal. Another study on the classification of the finger vein can be found in [21]. In [21], a new type of classifier called Local Mean based K-nearest centroid neighbor (LMkNCN) is proposed to classify finger vein patterns. In the LMkNCN, a local mean vector of k nearest centroid neighbors from each class for a training sample or query point is well positioned to sufficiently capture the class distribution information. Nevertheless, the weighting issues in assigning the class label before classification is not studied carefully [6]. As in the kNN, the LMkNCN has an identical weight for making classification decisions, regardless of their distances to the query point is inappropriate or not.

In order to enhance the classification process of finger vein, an extensive improvement classifier of kNN [3] called a fuzzy-based k-nearest centroid neighbor (FkNCN) classifier is proposed in this paper. This classifier is obtained based on the centroid-distance and fuzzy rule system which have been applied in kNCN [7] and FkNN [6]. The centroid-distance is applied to ensure that the selected training samples are distributed sufficiently in the region of the neighborhood with the nearest neighbors located around the query point. Consequently, the weighting fuzzy-based rule is employed to solve the ambiguity of the weighting distance between the query point and its nearest neighbors. Compared to the kNN, kNCN and FkNN, in the FkNCN, the query point is classified not only depending on the minimum distances of the k neighbors and how the samples are placed around it, but also taking into account the assigning fuzzy membership to the query point. In addition, FkNCN is applicable to problems with a limited number of training samples since the region of the centroid neighborhood is bigger than other neighborhoods employed in the FkNN and kNN. This is an advantage of the FkNCN since a restricted number of training samples is often encountered in the classification process.

The rest of this paper is organized as follows: The proposed classifier for the finger vein identification system is described in Section 2. The experimental results are explained in Section 3, and this paper is concluded in Section 4.

THE PROPOSED FUZZY BASED K-NEAREST CENTROID NEIGHBOR CLASSIFIER

This section aims to provide a description of the proposed fuzzy based k-nearest centroid neighbor (FkNCN) classifier. The main objective of this classifier is to optimize the performance while considering the surrounding-fuzzy based rules, which are (i) the k centroid nearest neighbors should be close to the query point as possible and located symmetrically around the query point and (ii) the query point is classified by taking the fuzzy membership values into account.

Given a query point y , a set of training samples, $T = \{x_j \in R^m\}_{j=1}^N$, with $x_j = \{c_1, c_2, \dots, c_M\}$ where N is the number of training samples, x_j is the training sample, M is the number of class, and c is the class label of M . The procedures of the FkNCN are as follows;

- i) Select the training sample as the first nearest centroid neighbor by sorting the distance of the query point and training sample using the Euclidean distance given as;

$$d(y, x_j) = \sqrt{(y - x_j)^T (y - x_j)} \quad (1)$$

- ii) Let the first nearest centroid neighbour be x_1^{NCN} . For $k = 2$, find the first centroid of x_1^{NCN} and the other training samples given as;

$$x_2^c = \frac{x_1^{NCN} + x_j}{2} \quad (2)$$

- iii) Then, determine the second nearest centroid neighbor by finding the nearest distance of the first centroid and query point.

- iv) For $k > 2$, repeat the second step to find the other nearest centroid neighbors by determining the centroid between the training samples and previous nearest neighbors;

$$x_k^c = \frac{1}{k} \sum_{l=1}^k x_j^{NCN} + x_j \quad (3)$$

- v) Let the set of k nearest centroid neighbors $T_{jk}^{NCN}(y) = \{x_{jk}^{NCN} \in R^m\}_{j=1}^k$, assign the fuzzy membership of the query point in every k nearest centroid neighbor. The fuzzy membership is given by;

$$u_i^{NCN}(y) = \frac{\sum_{j=1}^k u_{ij} \left(\frac{1}{\|y - x_{jk}^{NCN}\|^{2/(m-1)}} \right)}{\sum_{j=1}^k \left(\frac{1}{\|y - x_{jk}^{NCN}\|^{2/(m-1)}} \right)} \quad (4)$$

where $i = 1, 2, \dots, c$, c is the number of classes, u_{ij} is the membership degree of training sample, x_{jk} selected as the nearest neighbor, $\|y - x_{jk}^{NCN}\|$ is the L-norm distance between the query point x and its nearest neighbor and m is a fuzzy strength parameter which is used to determine how heavily the distance is weighted when calculating each neighbor's contribution to the fuzzy membership values.

For the value of the fuzzy strength parameter, m the value of m is set to 2. This is because, if m is 2, the fuzzy membership values are proportional to the inverse of the square of the distance and this gives the optimal result in the classification process [8].

There are two ways to define u_{ij} [8]. One is using the crisp membership where the training samples assign all of the memberships to their known class and non-memberships to other classes. The other way is using the constraint fuzzy membership, i.e. when the k nearest neighbors of each training sample are found (say x_k), the membership of x_k in each class is assigned as;

$$u_{ij}(x_k) = \begin{cases} 0.51 + 0.49(n_j/k) & j = i \\ 0.49(n_j/k) & j \neq i \end{cases} \quad (5)$$

where n_j denotes the number of neighbors of j th training samples.

In this paper, the membership degree u_{ij} is defined by using the constraint fuzzy membership. The reason why the constraint fuzzy membership is used, is to ensure higher weight is assigned to the training samples in its own class while lower weight is assigned to the other classes [9].

- vi) Classify the query point to the class label by obtaining the highest fuzzy membership value;

$$y = \operatorname{argmax} (u_i^{NCN}(y)) \quad (6)$$

- vii) Repeat the steps (i) to (vi) for new query point.

In order to show that the proposed classifier outperforms better than the other classifiers, a comparison between the FkNCN and the other three classifiers i.e. kNN, kNCN and FkNN is demonstrated using a problem with a limited number of training samples. Figure 1 shows the two-dimensional decision area of two classes with 15 training samples of class 'o' and 10 training samples of class 'o'. Figure 1(a) shows all the training samples that are able to classify in class 'o', while kNN produces the erroneous decision boundary as three of the training samples from class 'o' are misclassified. In another case, Figure 1(b) shows the decision area produced by kNCN. Although there is a slight improvement in the kNCN, where only two of the training samples are misclassified in class 'o', there is a training sample misclassified in class 'o'. Meanwhile, Figure 1(c) shows the decision area produced by FkNN. In this figure, all the training samples are classified correctly. However, there are two disjoint areas. That is, the FkNN will assume that these disjoint areas contain outliers. On the other hand, Figure 1(d) illustrates the decision area of FkNCN. In this figure, all the training samples are classified successfully to their class label. It can be seen that the boundary line is more flexible than the other three classifiers, and this causes the FkNCN to be able to produce the best decision area and to make it more optimal in the limited training samples. As the result, the FkNCN can be more suitable for making classification decision, and this simple example shows that the FkNCN can handle the problem of the distribution of training samples and the limited training samples better than the other classifiers.

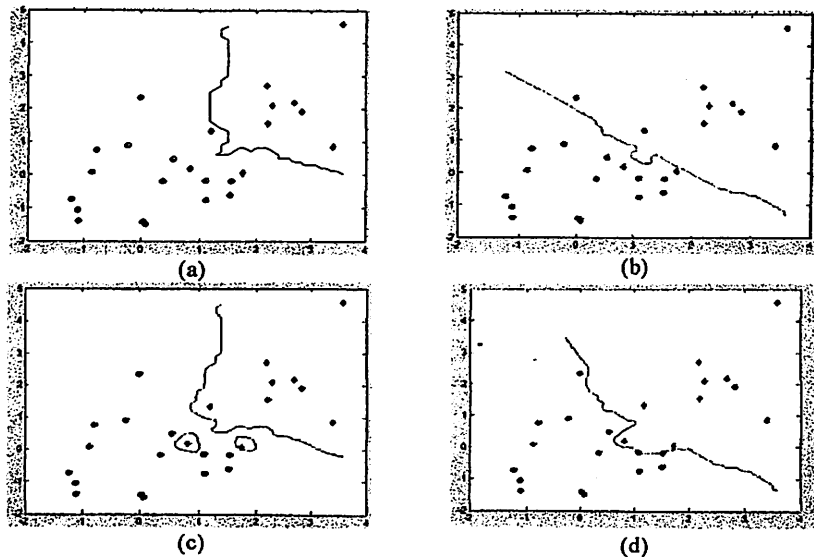


FIGURE 1. Example of decision area for (a) kNN (b) kNCN (c) FkNN (d) FkNCN

EXPERIMENTAL RESULTS

In this section, a comparative study on the performance of FkNCN on finger vein dataset has been investigated and compared with kNN, kNCN and FkNN. The performance of the classifier is determined based on the classification accuracy (CA) where the CA is defined as;

$$C_A = \frac{N_C}{N_T} \times 100\% \quad (6)$$

where N_C is the number of query point which is classified correctly, and N_T is the total number of the query point. The classifiers were implemented in Matlab R2007 (b), and the experiments were conducted on Intel Core i7, 2.1GHz CPU, 6G RAM and Windows 8 operating system.

Finger Vein Image Database

To evaluate the performance of the proposed classifier, we use our own finger vein image database [2] that can be downloaded from the following website : <http://blog.eng.usm.my/fendi/> . The database was obtained from 123 volunteers who were staffs and students (83 males and 40 females). The age of the subjects ranged from 20 to 52 years old. Each subject provided four fingers i.e. left index, left middle, right index and right middle resulting 492 finger classes employed. The images were acquired in two sessions, separated by more than two weeks' time. Each finger was captured six times in every session. Thus, a total of 5904 ($123 \times 4 \times 6 \times 2$) images were obtained from two sessions. The images from the first session were used as training data while the images in the second session were used as test data. The spatial and depth resolutions of the images were 640×480 pixels and 256 grey levels, respectively. As the focus of this paper is on the classification of finger vein image, the extracted region of interest (ROI) of the images were used in the experiments. Few examples of the extracted ROI of the finger vein images are shown in Figure 2.

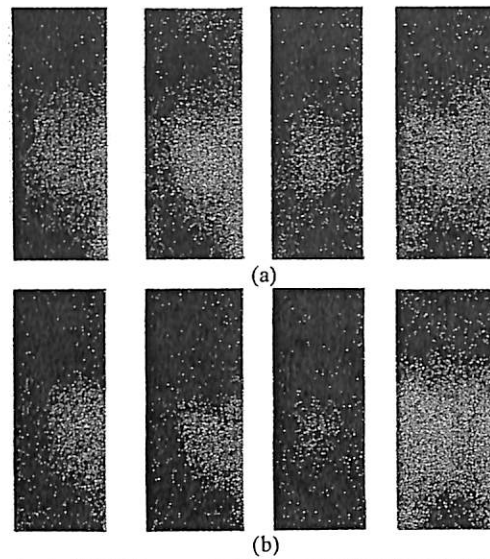


FIGURE 2. Example of extracted ROI finger vein image collection from (a) the first session and (b) the second session

Finger vein classification

In order to show the effectiveness of the FkNCN classifier in the finger vein identification, the ROI images have been classified without going through any image enhancement and feature extraction stages. The images are resized using the ratio from 0.1 to 1.0, and the obtained grayscale values are normalized to the unit norm.

Table 1 shows the performance of different classifiers, i.e. kNN, FkNN, kNCN and FkNCN where the value of k in each classifier is set to 3. It clearly shows supremacy of FkNCN over other classifiers. It was observed that the FkNCN significantly performed better than the kNN, FkNN and kNCN in each resize ratio. The best classification result is when the image was resized to 0.7 with 81.13%. The experimental results also show that when the size of the image is too small, the recognition accuracy is low. This is because some of the important features are lost when the resize ratio is reduced.

TABLE 1. Results of performance evaluation on classification				
Resize Ratio	kNN	FkNN	kNCN	FkNCN
0.1	76.32	76.82	78.42	80.79
0.2	76.86	76.86	78.52	80.79
0.3	76.90	77.30	78.56	80.76
0.4	76.96	77.34	78.59	80.93
0.5	77.03	77.19	78.61	81.03
0.6	76.10	76.99	78.64	81.09
0.7	76.96	77.37	78.61	81.13
0.8	76.96	77.35	78.59	81.10
0.9	76.68	77.39	78.56	81.09
1.0	76.49	77.41	78.53	81.09

CONCLUSION

In this paper, an empirical work of the centroid neighborhood and fuzzy-rule based system called a fuzzy-based k -nearest centroid neighbor (FkNCN) has been proposed and successfully implemented for finger vein database. The proposed classifier aims at exploiting the strength of the centroid neighborhood while solving the ambiguity of the weighting distance between the query point and its nearest neighbors. The FkNCN computes the k nearest centroid neighborhood for each class separately. Then, the fuzzy membership is constructed to assign the query point to the right class label. In this paper, the proposed FkNCN has been compared with the kNN, kNCN and FkNN. A series of experiments, based on the different ratio of image size from 0.1 to 1.0 have been performed to determine the

competence of the proposed classifier. The optimum classification accuracies were up to 77.03%, 77.41%, 78.64% and 81.13% for kNN, kNCN, FkNN and FkNCN, respectively. Results indicate that the FkNCN provides the best recognition accuracy among the classifiers.

ACKNOWLEDGMENTS

This work is supported by Universiti Sains Malaysia Research University Grant No. 1001/PELECT/814116 and Post Graduate Incentive Grant No. 1001/PELECT/8023013.

REFERENCES

1. H. Jaafar and D. A. Ramli, "A Review of Multibiometric System with Fusion Strategies and Weighting Factor" *International Journal of Computer Science Engineering (IJCSE)*, 2(4), 2013, pp. 158-165.
2. M. S. M. Asaari, S. A. Suandi and B. A. Rosdi, "Fusion of Band Limited Phase Only Correlation and Width Centroid Contour Distance for finger based biometrics" *Expert Systems with Applications*, 41, 2014, pp. 3367-3382.
3. T. Cover and P. Hart, "Nearest Neighbour Pattern Classification" *IEEE Transactions on Information Theory*, 13(1), 1967, pp. 21-27.
4. S. B. Imandoust and M. Bolandraftar, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background" *Int. Journal of Engineering Research and Application*, 3(5), 2013, pp. 605-610.
5. B. B. Chaudhuri, "A new definition of neighbourhood of a point in multi-dimensional space" *Pattern Recognition Letters*, 17(1), 1996, pp. 11-17.
6. J. M. Keller, M. R. Gray and J. A. Givens, "A Fuzzy K-Nearest Neighbor Algorithm" *IEEE Trans. Syst., Man, Cybern. SMC*, 15(4), 1985, pp. 580-585.
7. J. S. Sanchez, F. Pla and F. J. Ferri, "Improving The k-NCN Classification Rule Through Heuristic Modifications" *Pattern Recognition Letters*, 19, 1998, pp. 1165-1170.
8. H. L. Chen, B. Yang, G. Wang, J. Liu, X. Xin, S. J. Wang and D. Y. Liu, "A Novel Bankruptcy Prediction Model Based On An Adaptive Fuzzy K-Nearest Neighbor Method" *Knowledge-Based Systems*, 24(8), 2011, pp. 1348-1359.
9. T. W. Chua, and W. W. Tan, "A New Fuzzy Rule-Based Initialization Method For K-Nearest Neighbor Classifier" in *IEEE International Conference on Fuzzy Systems*, Aug 2009. pp. 415-420.
10. J. P. Campbell, D. A. Reynolds and R. B. Dunn, "Fusing High-And Low-Level Features for Speaker Recognition" in *Proceedings of Eurospeech*, 2003, pp. 2665 - 2668.
11. A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition" *IEEE Transactions on Circuit and Systems For Video Technology*, 14(1), 2004, pp. 4-20.
12. L. Akarun, B. Gökberk, and A. A. Salah, "3D Face Recognition for Biometric Applications" *13th European Signal Processing Conference (EUSIPCO)*, 2005.
13. C. Le, and R. Jain, "A Survey of Biometrics Security Systems" Washington University in St. Louis. (2009).
14. R. P. Wildes, "Iris recognition: an emerging biometric technology" *Proceedings of the IEEE*, 85(9), 1997, pp. 1348-1363.
15. N. Miura, A. Nagasaka and T. Miyatake, "Feature Extraction of Finger-Vein Patterns Based on Repeated Line Tracking and its Application to Personal Identification" *Mach. Vision Appl.* 15, 2004, pp. 194-203.
16. B. Rosdi, C. Shing, and S. Suandi, "Finger Vein Recognition Using Local Line Binary Pattern" *Sensors*, 11, 2011, pp. 11357-11371.
17. E. C. Lee, H. C. Lee and K. R. Park, "Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction" *International Journal of Imaging Systems and Technology*, 19(3), 2009, pp. 179-186.
18. G., Yang, X., Xi, and Y. Yin, "Finger Vein Recognition Based on (2D) 2 PCA and Metric Learning" *BioMed Research International*, 2012. pp. 1-9.
19. J. Yang, J. Yang and Y. Shi, "Combination of Gabor Wavelets and Circular Gabor Filter for Finger-Vein Extraction" *Emerging Intelligent Computing Technology and Applications*, Springer Berlin Heidelberg, 2009, pp. 346-354.
20. J. D. Wu and C. T. Liu, "Finger-vein pattern identification using SVM and neural network technique" *Expert Systems with Applications*, 38 (11), 2011. pp. 14284-14289.
21. A. K., Mobarakeh, S. M., Rizi, S., Nazari, J. P., Gou, and B. A. Rosdi, "Finger Vein Recognition Using Local Mean Based K-Nearest Centroid Neighbor Classifier" *Advanced Materials Research*, 628, 2013. pp. 427-432.

Search

Alerts

My list

My Scopus

Back to results | < Previous **3 of 27** Next >
 LinkSource |  ScienceDirect | View at Publisher | Export | Download | More...

Journal of Computer Science

Volume 10, Issue 3, 2014, Pages 530-543

Quality based speaker verification systems using fuzzy inference fusion scheme (Article)

Hamid, L.A., Ramli, D.A.

Intelligent Biometric Research Group (IBG), School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia

Abstract

View references (21)

Performances of single biometric speaker verification systems are outstanding in clean condition but drop significantly in noisy condition. Implementation of multibiometric systems is one of the solutions to this limitation. However, in order to ensure the performances of multibiometric systems are sustained, the optimum weight for the fusion system must be determined correctly according to the quality of current data. This study proposes the use of Fuzzy Inference System for weight inference. Two traits i.e., speech and lip are used while Support Vector Machine (SVM) is employed as the classifier in this study. The speech features are extracted using the Mel Frequency Cepstrum Coefficient (MFCC) method and the lip features are extracted using Region of Interest (ROI) method. The performances of single modal system (i.e., speech and lip) and multibiometric systems with sugeno and mamdani approaches are compared at different quality conditions in this study. Experimental results prove that the use of Fuzzy Inference System as weight inference is a very promising approach. For 15 dB SNR speech signal and 0.2 lip quality density, the GAR performances at FAR equals 0.1% for Mamdani-type, Sugeno-type, lip and speech systems are observed as 94, 95, 86 and 7%, respectively. In short, the proposed fusion scheme based on Fuzzy logic is able to maintain the performance of fusion system especially when one of the biometric sources is in noisy condition due to its capability to infer the correct fusion weight according to current data quality. © 2014 Science Publications.

Author keywords

Biometrics; Fuzzy logic fusion scheme; Mamdani-type; Multibiometric system; Single biometric system; Sugeno-type

ISSN: 15493636 Source Type: Journal Original language: English

DOI: 10.3844/jcssp.2014.530.543 Document Type: Article

References (21)

View in search results format

 Page Export | Print | E-mail | Create bibliography

- Becchetti, C., Ricotti, L.P.**
1 (1999) *Speech Recognition: Theory and C++ Implementation*, p. 407. Cited 115 times.
1st Edn., Wiley, New York, ISBN-10: 0471977306, pp
- Ben-Yacoub, S., Abdeljaoued, Y., Mayoraz, E.**
2 **Fusion of face and speech data for person identity verification**
(1999) *IEEE Transactions on Neural Networks*, 10 (5), pp. 1065-1074. Cited 207 times.
doi: 10.1109/72.788647

About Scopus
What is Scopus
Content coverage
Scopus Blog
Scopus API

Language
日本語に切り替える
切换到简体中文
切换到繁體中文

Customer Service
Help and Contact
Live Chat

About
Elsevier
Terms and Conditions
Privacy Policy



Copyright © 2015 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.
Cookies are set by this site. To decline them or learn more, visit our Cookies page.

Cited by 0 documents

Inform me when this document is cited in Scopus:

 Set citation alert | Set citation feed

Related documents

Comparative study on feature, score and decision level fusion schemes for robust multibiometric systems

Lip, C.C., Ramli, D.A.
(2012) *Advances in Intelligent and Soft Computing*

Performances of speech signal biometric systems based on signal to noise ratio degradation

Ramli, D.A., Samad, S.A., Hussain, A.
(2010) *Advances in Intelligent and Soft Computing*

Performances of qualitative fusion scheme for multibiometric speaker verification systems in noisy condition

Hamid, L.A., Ramli, D.A.
(2012) *Journal of Applied Sciences*

View all related documents based on references

Find more related documents in Scopus based on:

 Authors | Keywords

Metrics

3 Mendeley Readers 55TH PERCENTILE

View all metrics

QUALITY BASED SPEAKER VERIFICATION SYSTEMS USING FUZZY INFERENCE FUSION SCHEME

Lydia Abdul Hamid and Dzati Athiar Ramli

Intelligent Biometric Research Group (IBG), School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia

Received 2013-01-01; Revised 2013-02-07; Accepted 2013-12-03

ABSTRACT

Performances of single biometric speaker verification systems are outstanding in clean condition but drop significantly in noisy condition. Implementation of multibiometric systems is one of the solutions to this limitation. However, in order to ensure the performances of multibiometric systems are sustained, the optimum weight for the fusion system must be determined correctly according to the quality of current data. This study proposes the use of Fuzzy Inference System for weight inference. Two traits i.e., speech and lip are used while Support Vector Machine (SVM) is employed as the classifier in this study. The speech features are extracted using the Mel Frequency Cepstrum Coefficient (MFCC) method and the lip features are extracted using Region of Interest (ROI) method. The performances of single modal system (i.e., speech and lip) and multibiometric systems with sugeno and mamdani approaches are compared at different quality conditions in this study. Experimental results prove that the use of Fuzzy Inference System as weight inference is a very promising approach. For 15 dB SNR speech signal and 0.2 lip quality density, the GAR performances at FAR equals 0.1% for Mamdani-type, Sugeno-type, lip and speech systems are observed as 94, 95, 86 and 7%, respectively. In short, the proposed fusion scheme based on Fuzzy logic is able to maintain the performance of fusion system especially when one of the biometric sources is in noisy condition due to its capability to infer the correct fusion weight according to current data quality.

Keywords: Biometrics, Single Biometric System, Multibiometric System, Fuzzy Logic Fusion Scheme, Sugeno-type, Mamdani-type

1. INTRODUCTION

Previously, the traditional verification uses passwords, keys or smart cards which are less secure since few problems may occur due to forgotten password, duplicated keys or stolen smart cards. Nowadays, biometric data for verification systems are commercially used in data security, internet access, ATMs, network logins, credit cards and government records. More studies on biometric system have been done by researchers due to the increase of requirement of automatic information processing in many industrial fields (Chia and Ramli, 2011). Biometrics is defined as the development of statistical and mathematical methods applicable to data analysis problems in the biological sciences. Biometrics is also a

technology, which uses various individual attributes of a person to verify his or her identity. Biometric characteristics can be divided into two main classes i.e., physiological and behavioral characteristics. Physiological characteristics refers to the human body such as face, fingerprints, palm print, iris, DNA, hand geometry and finger vein structure while behavioral characteristics are related to the actions of a person such as voice, keystroke dynamics, gait, typing rhythm and signature (Jain *et al.*, 2004). This study implements biometric system for speaker verification systems. Speaker verification system is used to verify a person's claim from the enrollment database by using speech signal as the input data.

Single biometric systems have to face few limitations such as non-universality, noisy sensor data, large intra-

Corresponding Author: Lydia Abdul Hamid, Intelligent Biometric Research Group (IBG), School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia

user variations and susceptibility to spoof attacks. For example, a single biometric system uses voice patterns to identify the individuals may fail to operate because of a noisy data signal captured by the system. Limitations faced by single biometric system can be overcome by applying the multibiometric system. Multibiometric system enhanced the matching accuracy of a biometric system in noisy condition as well as increases the population coverage with multiple traits (i.e., lip, iris, voice and face). Studies on multibiometrics are further discussed in Ben-Yacoub *et al.* (1999) and Pan *et al.* (2000). Besides that, multibiometric system may continuously operate even though a certain trait is unreliable due to user manipulation, sensor or software malfunctions. However, this is only true when fusion scheme is done at the decision level where hard decision fusion for example or operator is executed. For the score level decision fusion, the multibiometric systems are at its best performance only when all traits operate in clean condition. In noisy condition, the unreliable speech signal tends to cause the system to obtain false scores for genuine and imposter signal. This problem does not occur in clean condition since both speech and lip signal gives reliable scores for genuine and imposter signal.

This study proposes the use of quality based score fusion approach to improve the performances of multibiometric systems. The quality based fusion depends on the input current condition. This method is very useful to ensure the speaker verification system is at its best performance especially in noisy condition. The quality based fusion implements the quality measure identification system to identify the quality of sample data. Researches on quality measure identification system have been discussed in Fierrez-Aguilar *et al.* (2005) and Nandakumar *et al.* (2008). In order to take full advantage of the quality based fusion approaches, this study implements the fusion mechanism for different biometric information. For this purpose, Fuzzy Inference System is developed so as to infer the optimum weight for robust and reliable multimodal biometric based security systems. The use of fuzzy logic as the fusion scheme for quality based fusion approach improves the system performances.

According to Vasuhi *et al.* (2010), the fuzzy logic decision-making is approximately the same with the human decision-making. Fuzzy design can accommodate the ambiguities of human languages and logics. It provides both an intuitive method for describing systems in human terms and automates the conversion of those system specifications into effective models. Fuzzy logic

has the ability to add human-like subjective reasoning capabilities to machine intelligences as described in Prade and Dubois (1996). General block of fuzzy logic with Mamdani-type and Sugeno-type is shown in Fig. 1. Fuzzification is the process where each input is assigned to a linguistic variable. Degree of membership can be obtained from the linguistic variable. The degrees of membership are combined using fuzzy rules which may be expressed in terms such as "if x is A, then y is B". The process of converting the fuzzy output based on the strength of membership is called defuzzification. Defuzzification is used in fuzzy modeling and in fuzzy logic control to convert the fuzzy outputs from the systems to crisp values.

There are two types of Fuzzy Inference System (FIS) i.e., mamdani and sugeno. A Mamdani-type FIS has fuzzy inputs and a fuzzy output. For Mamdani-type, the input is transformed into a set of linguistic variable during the fuzzification process. The Fuzzy Inference System (FIS) uses the input variables and fuzzy rule to derive a set of conclusion which will be used during the defuzzification process. A crisp number is the output of the defuzzification process (Jassbi *et al.*, 2007). Mamdani-type FIS is widely accepted for capturing expert knowledge. It allows us to describe the expertise in more intuitive and human-like manner. The advantages of the Mamdani-type FIS are it have widespread acceptance, intuitive and well-suited to human inputs. However, Mamdani-type FIS entails a substantial burden.

In short, both Mamdani-type and Sugeno-type are similar in term of the fuzzification and rule evaluation process. The main different between Mamdani-type and Sugeno-type is the output of Sugeno-type is linear or constant. Besides that, Mamdani-type uses defuzzification method to extract the output while Sugeno-type uses weighted average method to extract the output. Sugeno-type FIS is computationally effective and works well with optimization and adaptive techniques, which makes it is very attractive in control problems, particularly for dynamic nonlinear systems. So that it works well with linear technique and well-suited to mathematical analysis FLT, 2010.

The first objective of this study is to analyze the performances of single modal system i.e., speech and lip at different quality conditions. Consequently, the Fuzzy Inference System is designed for weight inference. Finally, the performances of the fusion systems with weight inferred from FIS are compared to the performances of the single systems.

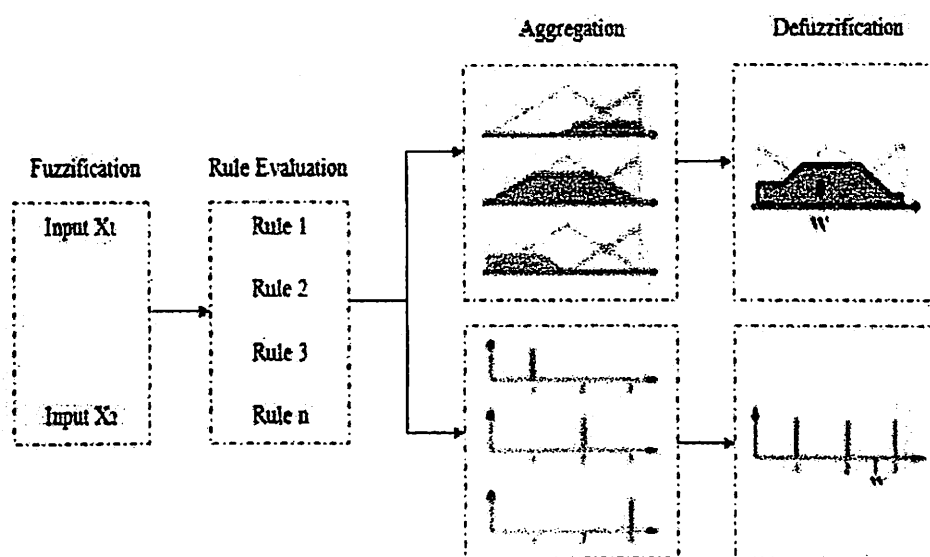


Fig. 1. Fuzzy logic with Mamdani-type and Sugeno-type

2. MATERIALS AND METHODS

Data Acquisition: In data acquisition, voice which is continuous electrical signal is converted to digital signal using a sampler and Analog-to-Digital (A/D) converter. The digitization process consists of sampling, quantization and coding. Sampling process is discussed extensively in (Rabiner and Schafer, 1978). After sampling process, the sampled signal is discrete in the time domain but still continuous in the amplitude domain. The quantization process divides the continuous amplitude range into finite subrange (Furui, 2000). Finally, the coding process is done by assigning these finite values into a sequence of codes for binary number representation.

In this study, the audio and visual data are obtained from Audio-Visual digit database (Sanderson and Paliwal, 2001). The database consists of 20 repetition of number zero from 37 different subjects. Mel Frequency Cepstrum Coefficient (MFCC) is used to obtain the features for speech modality. This study uses 12 MFCC features to form the feature vector. The data is collected in 32 kHz, 16-bit mono format. For the lip verification, the Region of Interest (ROI) of lip images are cropped and stored as JPEG files with resolution of 512×384 pixels. The ROI method to extract the lip features in this study as discussed in (Potamianos *et al.*, 2000; Iyengar *et al.*, 2001).

The database is divided to two sessions which are training and testing. During the enrolment process, 2220 audio data are developed for all 37 subjects. For training purposes, 740 data are used to train the system. Each subject is treated as the claimant and the other subjects as

the imposters during the verification process. Therefore, the database has 40 testing data from the authentic speaker and 1440 from the imposter speaker. The visual data consists of 60 sequences of images (20 for training and 40 for testing) where each sequence consists of 10 images. In total, 22200 data are developed for all 37 subjects. Similar to speaker verification, each subject is treated as the claimant and the other subjects as the imposters during the verification process. Hence, the database has 400 testing data from the authentic lip image and 14400 from the imposter lip image.

2.1. Feature Extraction

A preemphasis of high frequencies is required to compress the signal dynamic range by flattening the spectral tilt in order to raise the SNR. The first order FIR filter is used to filtering the speech signal. The use of window function is important to minimize the signal discontinuities at the beginning and end of each frame by zeroing out the signal outside the region of interest. This study implements the Mel Frequency Cepstrum Coefficient (MFCC) processing to extract the audio features. There are few steps involved in MFCC process. First, all frames of the signal are computed using discrete Fourier transform. Next, the filter bank processing formed the spectral features at defined frequency at its exit. After that, log energy computation which consists of computing the logarithm of the square magnitude of the filter bank is performed. Finally, the mel frequency cepstrum is computed (Becchetti and Ricotti, 1999).

2.2. Classification

This study implements the Support Vector Machine (SVM) as classifier. A SVM performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. SVM mode is a supervised learning method that generates input-output mapping functions from a set of labeled training data. The foundation of Support Vector Machines (SVM) has been developed as discussed in (Vapnik, 1995) and becomes popular and accepted nowadays due to many attractive features and promising empirical performance. Theory regarding SVM is further explained in (Gunn, 1998). In brief, decision boundary in support vector machine can be explained as presented in Fig. 2.

The SVM identifies the data points that are found to lie at the edge of an area in space which is a boundary from one class to another. The space between regions containing data points in different classes as being the margin between those classes. SVM is used to identify a hyperplane that separates the classes. The maximum margin between the different classes is found. An advantage of this method is that the modeling only deals with these support vectors, rather than the whole training dataset.

2.3. Fusion Scheme

A fuzzy fusion mechanism for robust and reliable multimodal biometric based security systems is developed. The use of fuzzy logic system as the fusion scheme improves the system performances. For this experiment, the fuzzy logic system consists of two inputs (speech and lip) and one output (weight). The parallel nature of the rules is one of the most important aspects in fuzzy logic (Hellmann, 2001). Initially, the input verification scores (speech and lip) are scaled to some range of score by using the min-max normalization equation as in Equation (1):

$$\hat{S}_i = \frac{s_i \min_{i=1}^K s_i}{\max_{i=1}^K s_i - \min_{i=1}^K s_i} \tag{1}$$

where denote the *i*th match score output and *K* is the number of the match scores available in the set (Jain *et al.*, 2005).

The fuzzy logic system procedures are proposed as below (Zadeh, 1965; 1984).

Step 1: Fuzzification

In this study, there are two fuzzy models for Mamdani-type and Sugeno-type, respectively. Each model has two inputs, speech and lip and one output which is weight. Figure 3 shows the fuzzy inference

system using Mamdani-type and Sugeno-type method in Matlab Fuzzy Toolbox.

Next, the inputs are identified and the degree of each input is determined according to appropriate fuzzy sets via membership function. The membership functions are Gaussian shapes because it can covers several values in one membership. The inputs are always a crisp numerical value. For input 1 (speech), the interval is varied between [0, 40] SNR and for input 2 (lip), the interval is varied between [0, 1] quality density. The output (weight) is varied between [0, 1].

Then, the speech fuzzy set is modeled for three mfs: speech (Qlow), speech (Qmed) and speech (Qhigh) and three mfs are also modelled for the lip fuzzy set: lip (Qlow), lip (Qmed) and lip (Qhigh) as shown in Fig. 4. For the output fuzzy set, three mfs: weight (Qlow), weight (Qmed) and weight (Qhigh) are used. Output for Mamdani-type and Sugeno-type are as illustrated in Fig. 5.

Step 2: Rule Evaluation

For this study, there are nine rules for the system. From the experiment, lip performs better than speech. Therefore, this study relies more on lip since uncertainty inputs condition are involved during the process. For example, when both speech (Qhigh) and lip (Qlow) are determined, the weight output is mapped to weight (Wmed). Rule editor is used to define the rules for each model. The rule editor for each model is shown in Fig. 6:

```

IF speech (Qlow) IF speech (Qmed) IF speech (Qhigh)
AND lip (Qhigh) AND lip (Qhigh) AND lip (Qhigh)
THEN (Wlow) THEN (Wlow) THEN (Wmed)
IF speech (Qlow) IF speech (Qmed) IF speech (Qhigh)
AND lip (Qmed) AND lip (Qmed) AND lip (Qmed)
THEN (Wlow) THEN (Wlow) THEN (Whigh)
IF speech (Qlow) IF speech (Qmed) IF speech (Qhigh)
AND lip (Qlow) AND lip (Qlow) AND lip (Qlow)
THEN (Wmed) THEN (Wmed) THEN (Whigh)
    
```

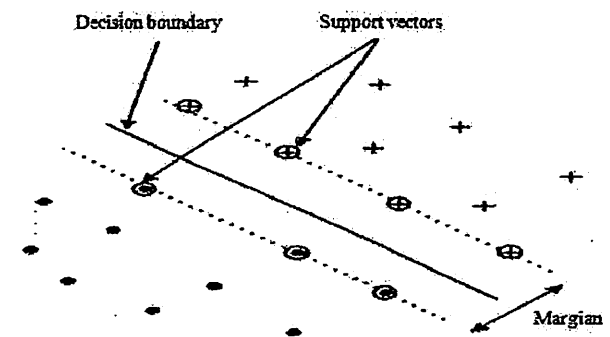


Fig. 2. Decision boundary in support vector machine

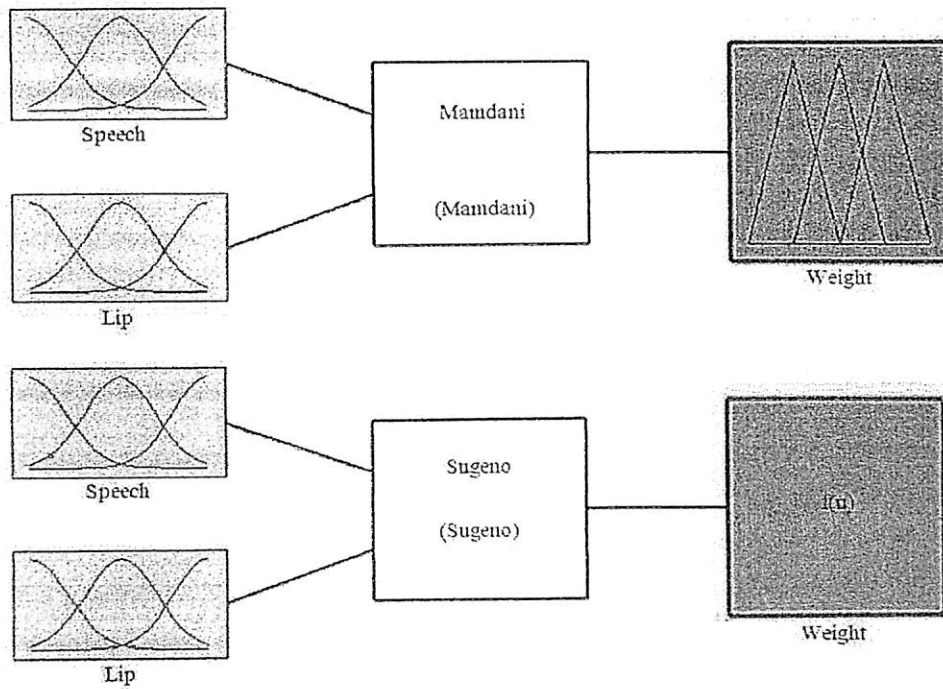


Fig. 3. Fuzzy Inference in Fuzzy Matlab Toolbox for Mamdani-type (top) and Sugeno-type (bottom)

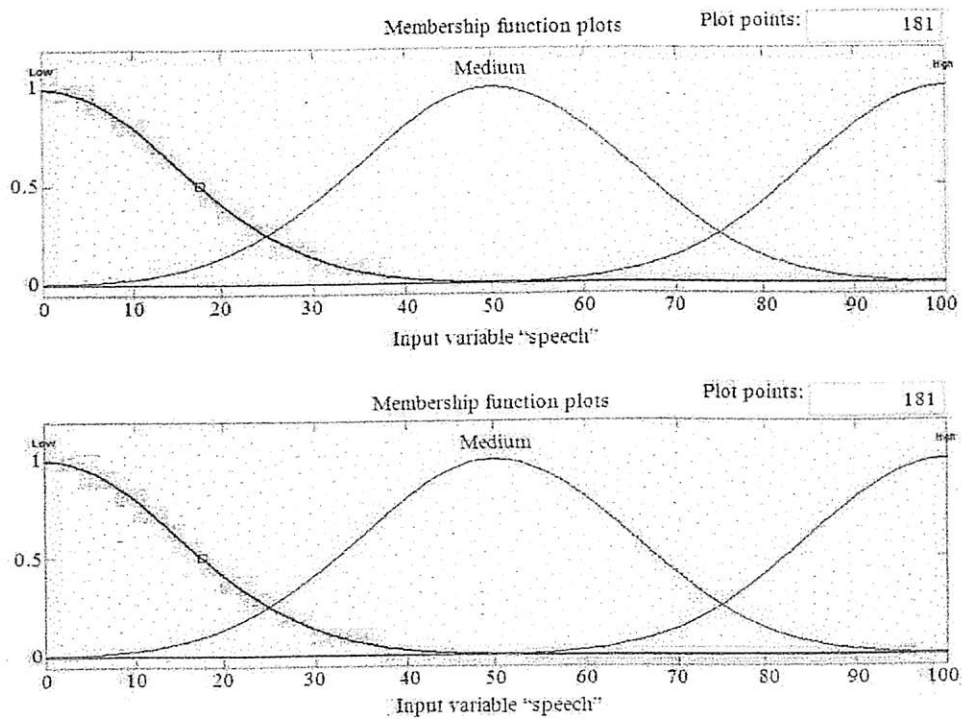


Fig. 4. Input Speech (top) and Input Lip (bottom) for Mamdani-type and Sugeno-type

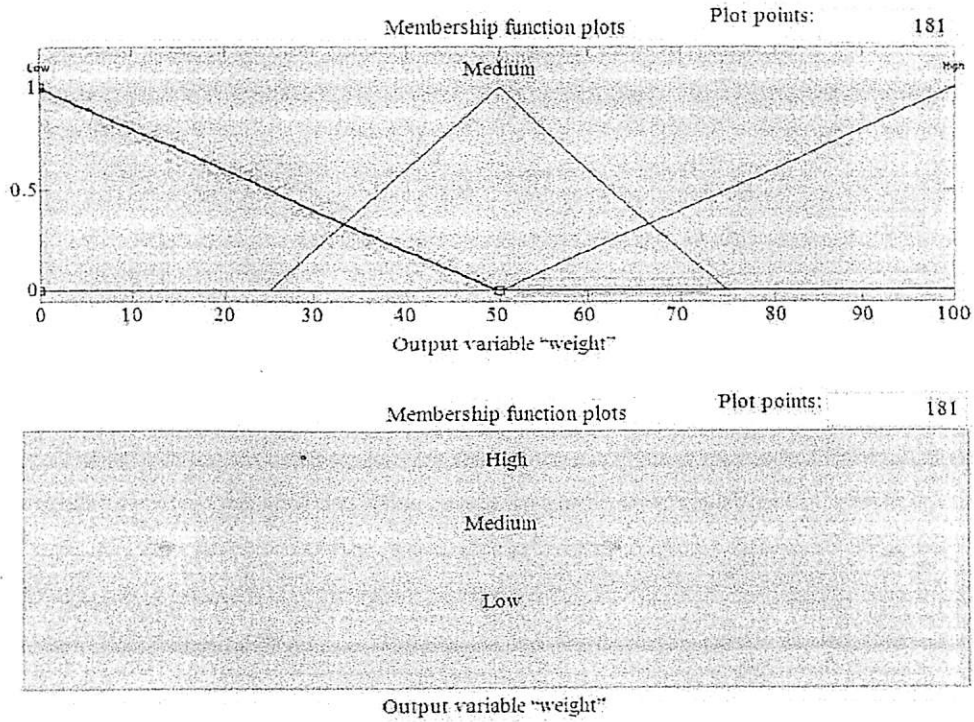


Fig. 5. Output for Mamdani-type (top) and Sugeno (bottom)

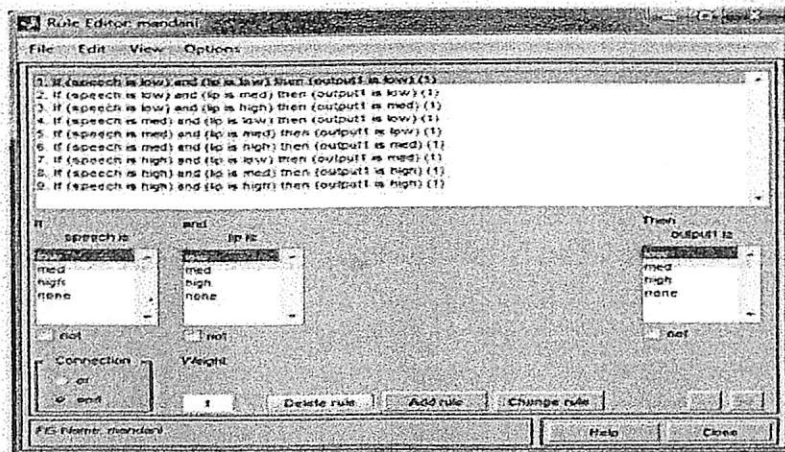


Fig. 6. Rule editor in fuzzy inference

Step 3: Aggregation

Aggregation is the process of unification of the outputs of all rules. The membership functions for all rules are scaled and combined into a single fuzzy set. The aggregation's inputs are the list of scaled

membership functions and the output is one fuzzy set for each output variable. The Mamdani-type method and Sugeno-type method for aggregating the fuzzy rules and computing the output are shown in Fig. 7 and 8, respectively. All the rules must be combined and tested in order to make a decision.

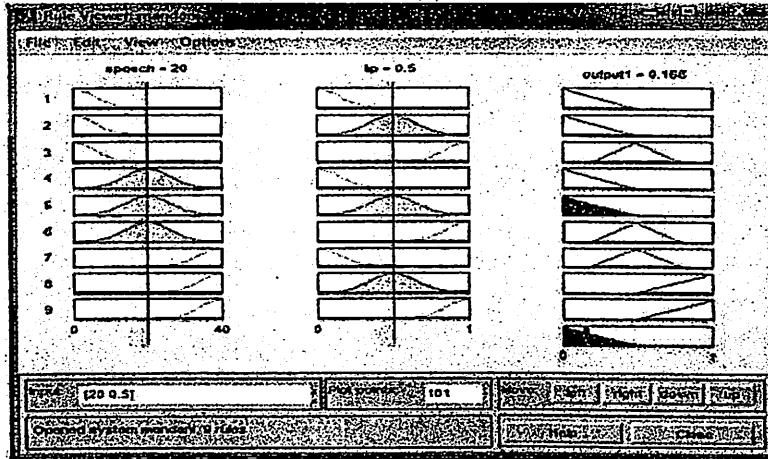


Fig. 7. Aggregation and defuzzification methods for Mamdani-type

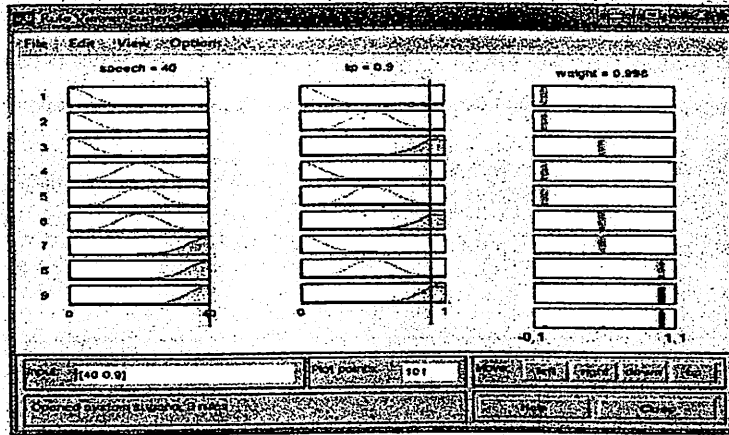


Fig. 8. Aggregation and defuzzification methods for Sugeno-type

Step 4: Defuzzification

The output of aggregation will be used as input for the defuzzification process and the output is a single number (weight). For defuzzification process, the Mamdani-type applied the centroid calculation method in order to obtain the centre of area under the curve while the Sugeno-type used the weighted average of few data points' method. The output (w) obtained from fuzzy logic system is implemented as in Equation (2) in order to calculate the fusion scores:

$$Y = wX_{speech} + (1-w)X_{lip} \tag{2}$$

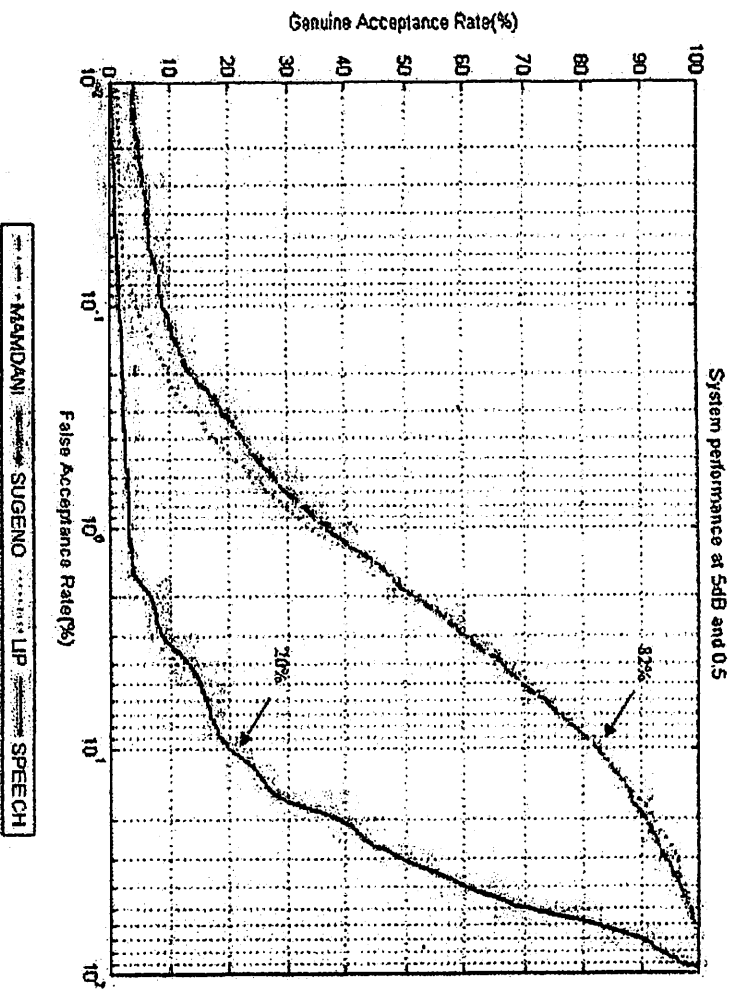
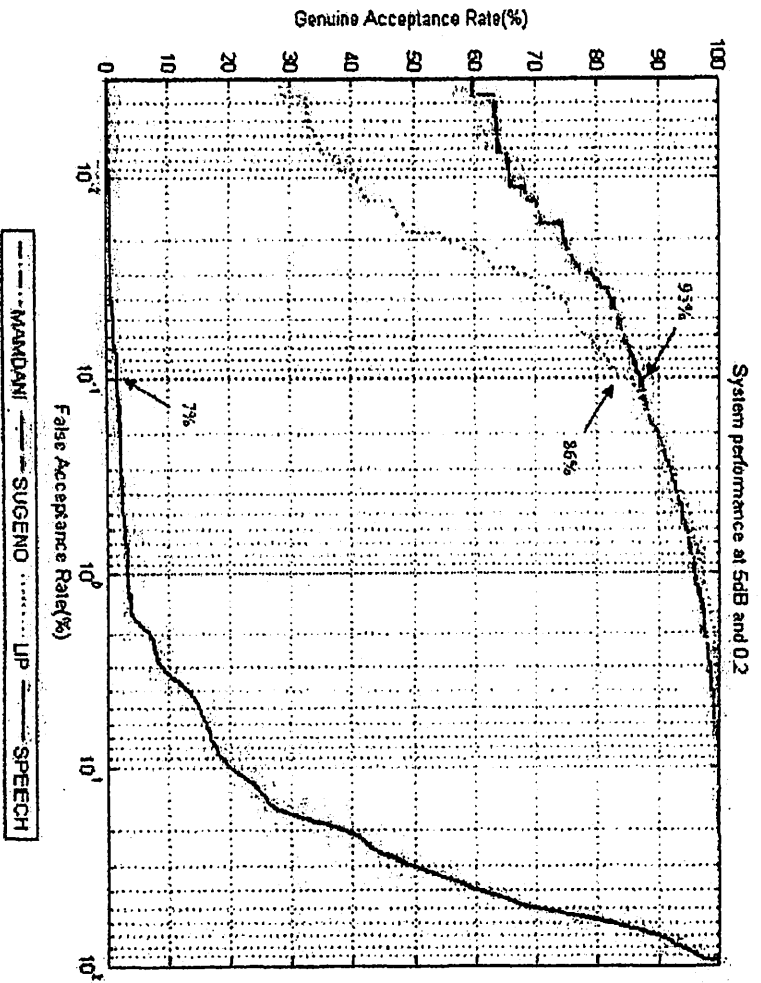
where, Y is the score and W is the weight applied to speaker's modality input data which are and respectively.

3. RESULTS

System performances for fuzzy logic fusion using Mamdani-type and Sugeno-type based on equal error rate (EER) at different levels of SNR are shown in Table 1 and 2, respectively. System performances based on receiver operation characteristic (ROC) showing the tradeoff between GAR and FAR percentages are then presented in Fig. 9-11.

Some results obtained by the single biometric and multibiometric system using Mamdani-type and Sugeno-type fusion method are also compared in terms of GAR and FAR at certain condition of speech and lip quality as illustrated in Fig. 9-11.

Figure 9 shows the performances of fusion systems compared to single systems at 5dB SNR with 0.2, 0.5 and 0.8 quality densities.



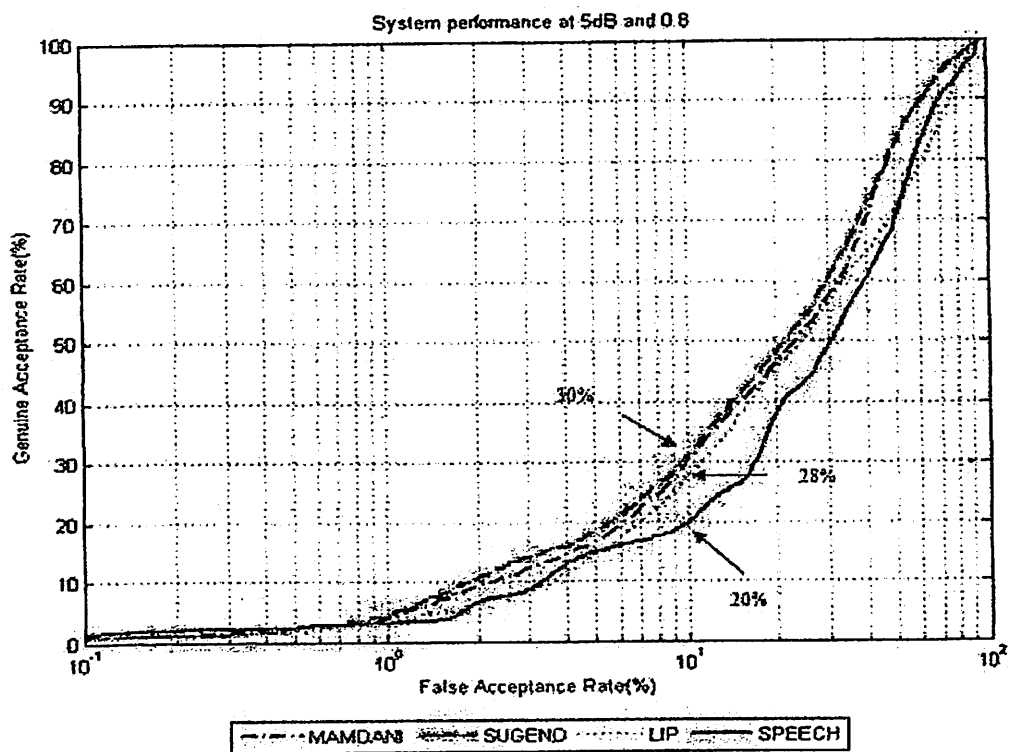
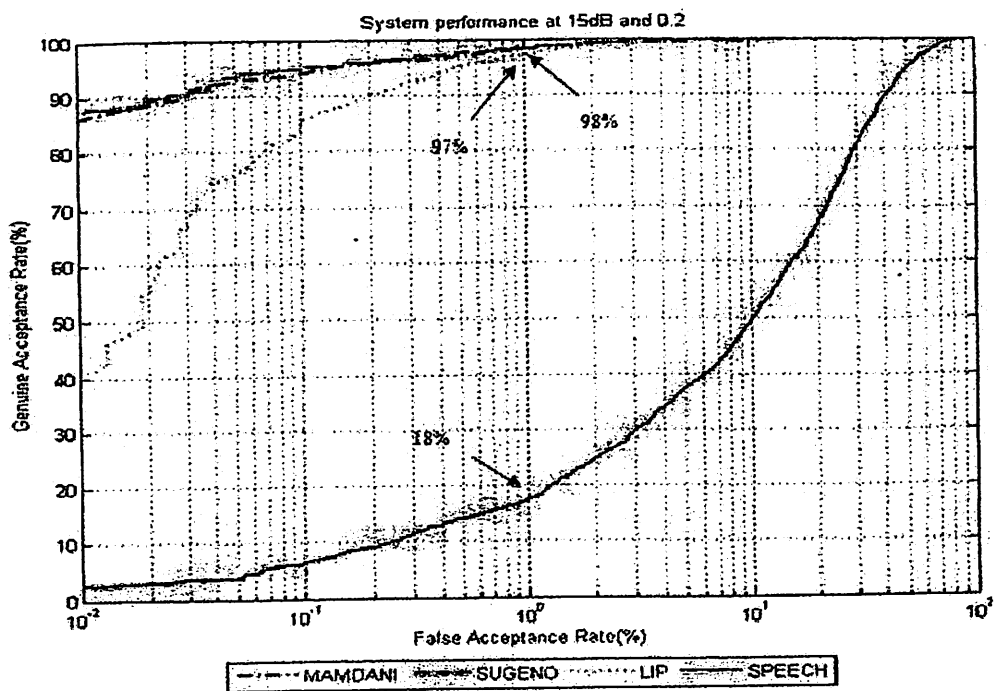


Fig. 9. The performances of fusion systems compared to single systems at 5dB SNR with 0.2, 0.5 and 0.8 quality densities



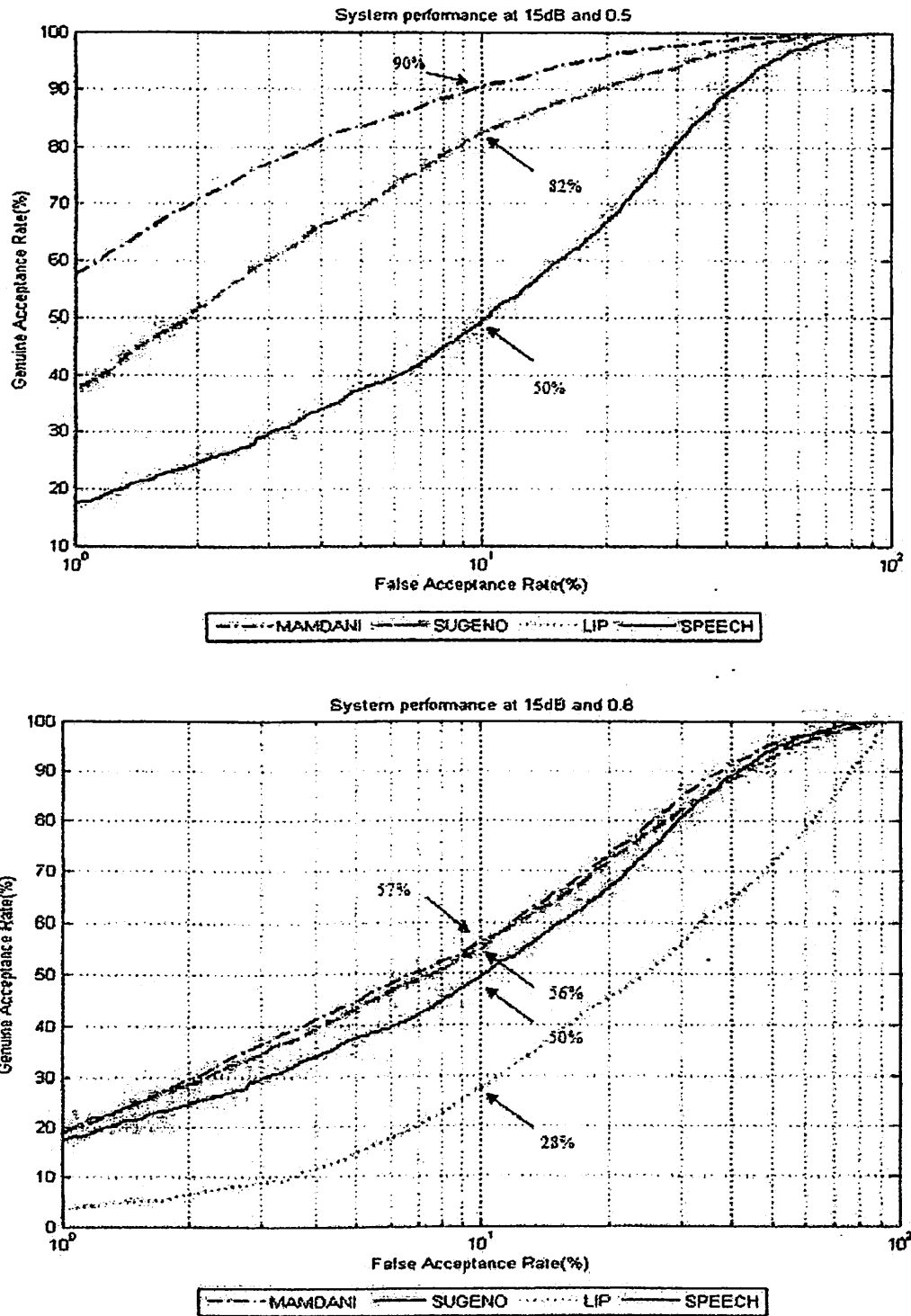
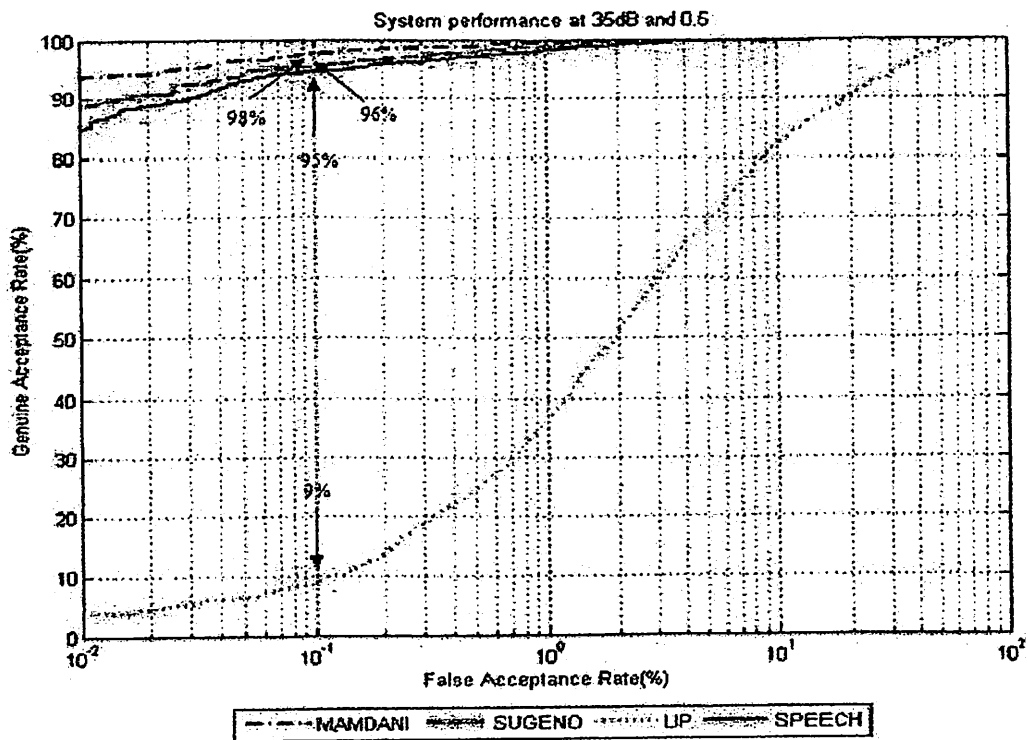
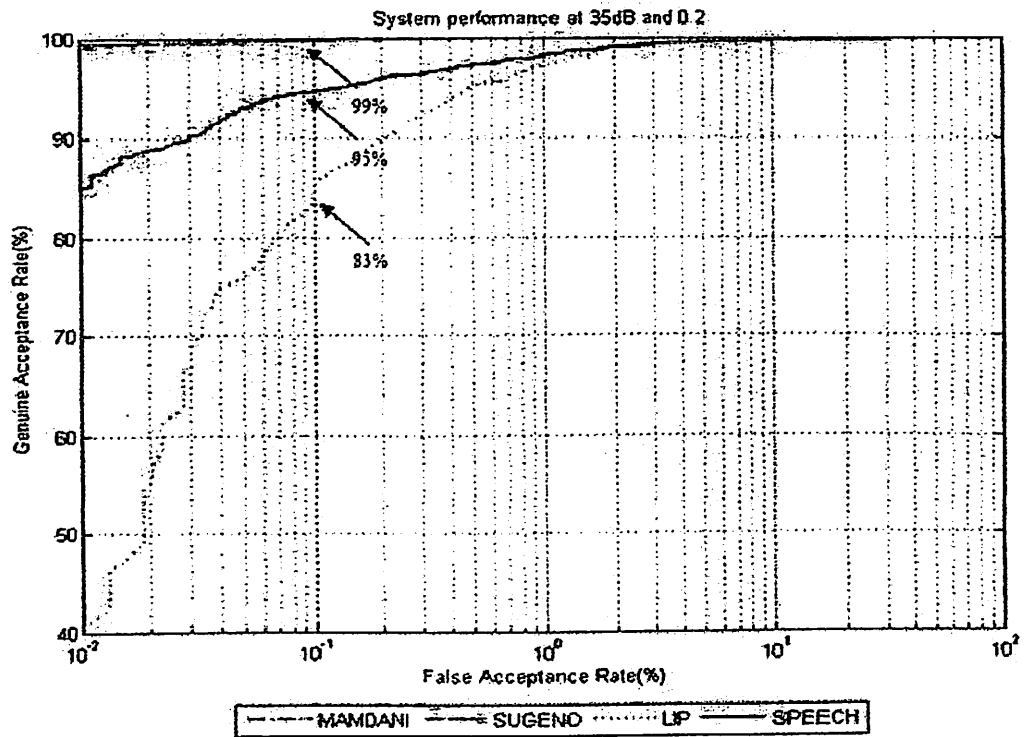


Fig. 10. The performances of fusion systems compared to single systems at 15 dB SNR with 0.2, 0.5 and 0.8 quality densities



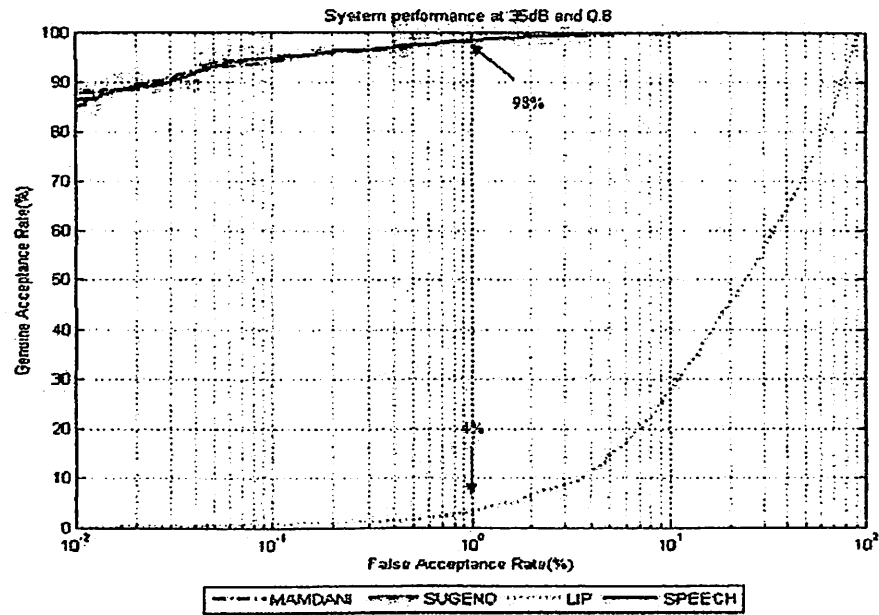


Fig. 11. The performances of fusion systems compared to single systems at 35 dB SNR with 0.2, 0.5 and 0.8 quality densities

Table 1. EER performances for fuzzy logic fusion using Mamdani-type

Audio										
Visual	clean	40dB	35dB	30dB	25dB	20dB	15dB	10dB	5dB	-5dB
Clean	0.0428	0.0493	0.0529	0.0566	0.1036	0.2993	0.4774	0.8443	1.5429	2.1105
0.1	0.0492	0.0648	0.0601	0.0591	0.2018	0.3069	0.5818	1.1421	1.9454	2.3003
0.2	0.0511	0.0882	0.1104	0.0779	0.3904	0.7104	1.1997	2.4062	3.6421	5.7645
0.3	0.1384	0.3388	0.3010	0.3463	1.0126	1.8816	2.5723	4.0465	5.9056	10.0475
0.4	0.2056	0.6278	0.6072	0.9552	1.4251	3.8081	5.7508	7.6079	9.0465	15.0956
0.5	0.2964	0.7066	0.7423	1.4011	3.9054	5.9223	9.7147	11.9257	13.5839	20.7664
0.6	0.3119	0.7873	0.8399	3.0261	5.1242	9.1122	15.2843	17.6605	20.5227	25.1253
0.7	0.3805	0.7883	1.1562	4.6678	6.8975	10.4255	18.1961	23.7960	28.0265	29.9903
0.8	0.4377	0.7883	1.2106	5.0221	9.5918	16.8290	23.1231	28.9611	35.1328	39.5665
0.9	0.5622	0.7742	1.4884	5.6034	13.4722	19.4998	25.0901	32.2325	39.0888	43.2836

Table 2. EER performances for fuzzy logic fusion using Sugeno-type

Audio										
Visual	clean	40dB	35dB	30dB	25dB	20dB	15dB	10dB	5dB	-5dB
Clean	0.0339	0.0489	0.0593	0.0627	0.2855	0.7642	0.9362	1.0072	1.1032	2.2117
0.1	0.0477	0.0666	0.0703	0.0976	0.8643	1.0811	1.0745	1.1924	2.0057	2.7555
0.2	0.0593	0.1342	0.1389	0.1952	1.2284	1.5907	1.6216	2.9034	3.7993	5.9015
0.3	0.3928	0.6607	0.6747	0.3987	2.8913	3.7172	3.8082	4.1225	5.1523	11.6776
0.4	0.5692	1.0801	0.9619	0.9196	6.5869	8.2226	8.3333	8.3343	8.3352	15.9945
0.5	0.6943	1.1421	1.1684	1.4310	9.4002	9.5126	10.6730	13.6806	13.6890	21.1034
0.6	0.6943	1.1355	1.1983	2.3020	10.2787	18.9921	21.5531	21.6282	21.8300	25.6724
0.7	0.8033	1.1233	1.2509	4.9278	12.6997	21.2828	23.6693	25.9741	27.6971	31.6770
0.8	0.8223	1.1515	1.2678	5.5572	13.1742	23.3183	24.8433	29.5069	36.4613	39.9001
0.9	0.8749	1.1780	1.2744	5.8708	14.9231	23.3183	26.1684	32.3931	39.1047	44.0005

When system at 5dB SNR and 0.2 quality density, GAR performances for Mamdani-type, Sugeno-type, lip and speech are evaluated as 88, 88, 83 and 2%, respectively, at 0.1% FAR. Meanwhile, at 5dB SNR and 0.5 quality density, GAR performances are observed as 82, 82, 81 and 20% for Mamdani-type, Sugeno-type, lip and speech, respectively at 10% FAR. Consequently, at 5dB SNR and 0.8 quality density, GAR performances for Mamdani-type, Sugeno-type, lip and speech equals to 30, 30, 28 and 20%, respectively at 10% FAR.

Subsequently, the performances of fusion systems compared to single systems at 15dB SNR with 0.2, 0.5 and 0.8 quality densities are illustrated in Fig. 10. When system at 15dB SNR and 0.2 quality density, GAR performances are observed as 94, 95, 86 and 7% for Mamdani-type, Sugeno-type, lip and speech respectively, at 0.1% FAR. Meanwhile, at 5dB SNR and 0.5 quality density, GAR performances are observed as 90, 82, 82 and 50% for Mamdani-type, Sugeno-type, lip and speech, respectively at 10% FAR. At the same FAR, i.e., 10%, when system at 5dB SNR and 0.8 quality density, GAR performances for Mamdani-type, Sugeno-type, lip and speech equals to 57, 56, 28 and 50%, respectively.

Finally, the performances of fusion systems compared to single systems at 35 dB SNR with 0.2, 0.5 and 0.8 quality densities are illustrated in Fig. 11 below. The GAR performances for Mamdani-type, Sugeno-type, speech and lip are observed as 99%, 99%, 95% and 83%, respectively at 0.1% FAR when system at 35dB SNR and 0.2 quality density. While system at 35dB SNR and 0.5 quality density, the GAR performances for Mamdani-type, Sugeno-type, speech and lip are defined as 97, 96, 95 and 10%, respectively at 0.1% FAR. GAR performances of 96, 96, 96 and 2% are then observed for Mamdani-type, Sugeno-type, speech and lip, respectively at 0.1% FAR when system at 35dB SNR and 0.8 quality density.

4. DISCUSSION

From the experimental results illustrated in Fig. 9-11, it is observed that fusion systems based on Mamdani-type FIS and Sugeno-type FIS are able to increase the performances of single systems i.e., speech and lip when one of the traits is in clean condition or under minor quality degradation. Fusion systems based on Sugeno-type FIS and Mamdani-type FIS are observed as the most outstanding systems compared to the other fusion schemes.

Consequently, when both of the traits are severely corrupted by noise, the performances of single system

tend to decrease. However, by implementing Sugeno-type FIS and Mamdani-type FIS fusion schemes, the systems are able to maintain its performances.

5. CONCLUSION

This study concludes a multibiometric verification system that combines both speaker and lip verification using fuzzy logic with Mamdani-type and Sugeno-type. Experimental results show that Mamdani-type and Sugeno-type are quite similar in accuracy performance and much better compared to the performances of single biometric systems. As a conclusion, the limitation faced by score level fusion in multibiometric system can be overcome using the fuzzy logic system due to its capability to infer the optimum weight according to the quality of verification data.

6. ACKNOWLEDGEMENT

This research is supported by the following research grants: Research University (RU) Grant, Universiti Sains Malaysia, 100/PELECT/814098 & 100/PELECT/814161 and Short Term Grant 304/PELECT/60311048, Universiti Sains Malaysia.

7. REFERENCES

- Becchetti, C. and L.P. Ricotti, 1999. *Speech Recognition: Theory and C++ Implementation*. 1st Edn., Wiley, New York, ISBN-10: 0471977306, pp: 407.
- Ben-Yacoub, S., Y. Abdeljaoued and E. Mayora, 1999. Fusion of face and speech data for person identity verification. *IEEE Trans. Neural Netw.*, 10: 1065-1074. DOI: 10.1109/72.788647
- Chia, C.L. and D.A. Ramli, 2012. Comparative study on feature, score and decision level fusion schemes for robust multibiometric systems. *Frontier Comput. Educ.*, 133: 941-948. DOI: 10.1007/978-3-642-27552-4_123
- Fierrez-Aguilar, J., J. Ortega-Garcia, J. Gonzalez-Rodriguez and J. Bigun, 2005. Discriminative multimodal biometric authentication based on quality measures. *Patt. Recogn.*, 38: 777-779. DOI: 10.1016/j.patcog.2004.11.012
- Furui, S., 2000. *Digital Speech Processing: Synthesis and Recognition*. 2nd Edn., Dekker, New York, ISBN-10: 0824704525, pp: 476.
- Gunn, S.R., 1998. *Support vector machines for classification and regression*. University of Southampton.

- Hellmann, M., 2001. Fuzzy logic introduction," epsilon nought radar remote sensing tutorials. The Pennsylvania State University.
- Iyengar, G., G. Potamianos, C. Neti, T. Faruque and A. Verm, 2001. Robust detection of visual ROI for automatic speechreading. Proceedings of the IEEE 4th Workshop on Multimedia Signal Processing, Oct. 03-05, IEEE Xplore Press, Cannes, pp: 79-84. DOI: 10.1109/MMSP.2001.962715
- Jain, A., K. Nandakumar and A. Ross, 2005. Score normalization in multimodal biometric systems. *Patt. Recogn.*, 38: 2270-2285. DOI: 10.1016/j.patcog.2005.01.012
- Jain, A.K., A. Ross and S. Prabhakar, 2004. An introduction to biometric recognition. *IEEE Trans. Circuits Syst. Video Technol.*, 14: 4-20. DOI: 10.1109/TCSVT.2003.818349
- Jassbi, J., S.H. Alavi, P.J.A. Serra and R.A. Ribeiro, 2007. Transformation of a mamdani FIS to first order sugeno FIS. Proceedings of the IEEE International Fuzzy Systems Conference, Jul. 23-26, IEEE Xplore Press, London, pp: 1-6. DOI: 10.1109/FUZZY.2007.4295331
- Nandakumar, K., Y. Chen, S.C. Dass and A.K. Jain, 2008. Likelihood ratio-based biometric score fusion. *IEEE Trans. Patt. Anal. Mach. Intell.*, 30: 342-347. DOI: 10.1109/TPAMI.2007.70796
- Pan, H., Z.P. Liang and Z.P. Huang, 2000. Fusing audio and visual features of speech. Proceedings of the International Conference on Image Processing, Sep. 10-13, IEEE Xplore Press, Vancouver, BC., pp: 214-217. DOI: 10.1109/ICIP.2000.899333
- Potamianos, G., A. Verma, C. Neti, G. Iyengar and S. Basu, 2000. A cascade image transform for speaker independent automatic speechreading. Proceedings of the IEEE International Conference on Multimedia Expo, Jul. 30-Aug. 02, IEEE Xplore Press, New York, pp: 1097-1100. DOI: 10.1109/ICME.2000.871552
- Prade, H. and D. Dubois, 1996. What are fuzzy rules and how to use them. *Fuzzy Sets Syst.*, 84: 169-185. DOI: 10.1016/0165-0114(96)00066-8
- Rabiner, L.R. and R.W. Schafer, 1978. *Digital Processing of Speech Signals*. 1st Edn., Prentice Hall, Englewood Cliffs, ISBN-10: 0132136031, pp: 512.
- Sanderson, C. and K.K. Paliwal, 2001. Noise compensation in a multi-modal verification system. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 07-11, IEEE Xplore Press, Salt Lake City, UT., pp: 157-160. DOI: 10.1109/ICASSP.2001.940791
- Vapnik, N., 1995. *The nature of Statistical Learning Theory*. 2nd Edn., Springer-Verlag GmbH, New York, ISBN-10: 0387945598, pp: 188.
- Vasuhi, S., V. Vaidehi, N.T.N. Babu and T.M. Treesa, 2010. An efficient multi-modal biometric person authentication system using fuzzy logic. Proceedings of the 2nd International Conference on Advanced Computing, Dec. 14-16, IEEE Xplore Press, Chennai, pp: 74-81. DOI:10.1109/ICOAC.2010.5725365
- Zadeh, L.A., 1965. Fuzzy sets. *Inform. Control*, 8: 338-353. DOI: 10.1016/S0019-9958(65)90241-X
- Zadeh, L.A., 1984. Making computers think like people. *IEEE Spectrum*, 8: 26-32. DOI: 10.1109/MSPEC.1984.6370431

Search

Alerts

My list

My Scopus

Back to results | < Previous 5 of 27 Next >

[LinkSource](#) | [SCIENCE@DIRECT®](#) | [View at Publisher](#) | [Export](#) | [Download](#) | [More...](#)

Proceedings - 4th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2014

30 March 2015, Article number 7072715, Pages 202-207

4th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2014; PARKROYAL Penang Resort Batu Ferringhi, Penang; Malaysia; 28 November 2014 through 30 November 2014; Category number CFP1414R-ART; Code 111756

Robust palm print verification system based on evolution of kernel principal component analysis (Conference Paper)

Ibrahim, S. , Jaafar, H. , Ramli, D.A.

Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia Engineering Campus, Nibong Tebal, Pulau Pinang, Malaysia

Abstract

View references (14)

Palm print is an emerging type of biometric that attracts researchers in biometrics area. As compared to the other biometric traits such as face, fingerprint and iris, the image quality of a fingerprint is robust with more information can be employed even though it is in low resolution. A new approach in feature extraction called evolution of kernel principal component analysis (Evo-KPCA) was proposed to speed up the processing time in the extraction stage. It used a reduced set density estimate (RSDE) to define a weighted gram matrix. As a result, the Evo-KPCA only extracted the most relevant and important information from a dataset. A total of 2400 palm print images was collected from three types of android mobiles. An experimental evaluation showed that the Evo-KPCA performed well in term of processing and accuracy compared to the region of interest (ROI), principle component analysis (PCA) and kernel principal component (KPCA) with the Genuine Acceptance Rates (GAR) of more than 98% and shorter processing time of less than 0.5s. © 2014 IEEE.

Author keywords

Evo-KPCA; palm print; RSDE; weighted gram matrix

Indexed keywords

Engineering controlled terms: Biometrics; Extraction; Feature extraction; Image segmentation

Evo-KPCA; Gram matrices; Kernel principal component; Kernel principal component analyses (KPCA); Palmprints; Principle component analysis; RSDE; The region of interest (ROI)

Engineering main heading: Principal component analysis

ISBN: 978-147995688-9 Source Type: Conference Proceeding Original language: English

DOI: 10.1109/ICCSCE.2014.7072715 Document Type: Conference Paper

Sponsors: Publisher: Institute of Electrical and Electronics Engineers Inc.

References (14)

View in search results format

[Page](#) | [Export](#) | [Print](#) | [E-mail](#) | [Create bibliography](#)

1 Jaafar, H., Ramli, D.A.

1 (2013) *International Journal of Computer Science Engineering (IJCSE)*, 2 (4), pp. 158-165.

About Scopus
What is Scopus
Content coverage
Scopus Blog
Scopus API

Language
日本語に切り替える
切换到简体中文
切换到繁體中文

Customer Service
Help and Contact
Live Chat

About Elsevier
Terms and Conditions
Privacy Policy



Copyright © 2015 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.
Cookies are set by this site. To decline them or learn more, visit our Cookies page.

Cited by 0 documents

Inform me when this document is cited in Scopus:

[Set citation alert](#) | [Set citation feed](#)

Related documents

A robust and fast computation touchless palm print recognition system using LHEAT and the IFkNCN classifier

Jaafar, H. , Ibrahim, S. , Ramli, D.A.
(2015) *Computational Intelligence and Neuroscience*

Realizing hand-based biometrics based on visible and infrared imagery

Michael, G.K.O. , Connie, T. , Chin, T.C.
(2010) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*

Palmprint recognition based on modified dct feature and RBF neural network

Yu, P.-F. , Xu, D.
(2008) *Proceedings of the 7th International Conference on Machine Learning and Cybernetics, ICMLC*

View all related documents based on references

Find more related documents in Scopus based on:

[Authors](#) | [Keywords](#)

Robust Palm Print Verification System Based on Evolution of Kernel Principal Component Analysis

Salwani Ibrahim

Intelligent Biometric Group,
School of Electrical and Electronic,
Universiti Sains Malaysia Engineering
Campus, 14300, Nibong Tebal,
Pulau Pinang, Malaysia
salwani.ibrahim@gmail.com

Haryati Jaafar

Intelligent Biometric Group,
School of Electrical and Electronic,
Universiti Sains Malaysia Engineering
Campus, 14300, Nibong Tebal,
Pulau Pinang, Malaysia
haryati.jaafar@yahoo.com

Dzati Athiar Ramli

Intelligent Biometric Group,
School of Electrical and Electronic,
Universiti Sains Malaysia Engineering
Campus, 14300, Nibong Tebal,
Pulau Pinang, Malaysia
dzati@usm.my

Abstract— Palm print is an emerging type of biometric that attracts researchers in biometrics area. As compared to the other biometric traits such as face, fingerprint and iris, the image quality of a fingerprint is robust with more information can be employed even though it is in low resolution. A new approach in feature extraction called evolution of kernel principal component analysis (Evo-KPCA) was proposed to speed up the processing time in the extraction stage. It used a reduced set density estimate (RSDE) to define a weighted gram matrix. As a result, the Evo-KPCA only extracted the most relevant and important information from a dataset. A total of 2400 palm print images was collected from three types of android mobiles. An experimental evaluation showed that the Evo-KPCA performed well in term of processing and accuracy compared to the region of interest (ROI), principle component analysis (PCA) and kernel principal component (KPCA) with the Genuine Acceptance Rates (GAR) of more than 98% and shorter processing time of less than 0.5s.

Index Terms— Evo-KPCA, RSDE, weighted gram matrix, palm print

INTRODUCTION

Today's complex demands for reliable authentication and identification methods are increasing rapidly. Initially, the traditional technologies such as personal identification number (PIN), smart cards and passwords were introduced [1]. However, they had a number of inherent disadvantages such as duplication, misplacing and hacking. Therefore, biometrics were introduced in the late 90s to recognize a person based on the physiological or biological characteristics [2]. The biometric technology is inherently more reliable. It is capable to provide a level of assurance for the preventions of duplication, stealing and hacking. Due to the specific physiological or behavioral characteristics that are possessed by the users, this technology is able to be implemented in various fields such as door access controls, criminal investigations, logical access points and surveillance applications [3].

Currently, there are various kinds of modalities of the biometric systems that are either widely used or developed such as the fingerprint, iris, face, hand geometry, palm print, gait, voice and signature [1]. The most widely used biometric modalities are the face, iris and voice. However, there are some drawbacks that come along with these biometric modalities. In the face biometrics, the users' faces change over the time. In order to recognize faces accurately, the image must be captured at an appropriate fixed position. This is not always possible and can be very difficult to be done in some environments [4]. On the other hand, the disadvantage of the iris biometric system is that some of the individuals' irises are difficult to be captured as they can be easily obscured by the eyelashes, eyelids, lens and reflections from the cornea. It is also lacked of existing data, thus deterring it to be used for background or watch checklist [5]. The main drawback of the voice biometric is that it cannot work effectively in the presence of noise [2].

Other than the aforementioned biometric system, the hand-based biometric systems such as the fingerprint and palm print recognitions have also received a lot of attention from researchers due to their high user acceptance and excellent advantages in their application [6]. Thus, the fingerprint is the most mature biometric technology while the palm print recognition has been widely investigated since the last decade in the pattern recognition field. The palm print recognition is similar to fingerprints, such that it is based on the aggregate of information presented in a friction ridge impression. The image quality of a fingerprint is robust because of its multiple lines, wrinkles and ridges but the palm print covers even more information and the ridge structures remain unchanged throughout the life, except for a change in size [7]. A palm print is distinctive and thick, therefore enabling an easy capture by using the low-resolution devices. As the results, the cost of detection system and the number of users needed to extract the features can be decreased [8].

To date, a number of palm print extraction methods have been studied to improve the accuracy of palm print recognition, for example the PCA [9], eigenpalm [10], Gabor filters [11], Fourier Transform [12] and wavelets [13]. The PCA has been widely used compared to the other methods because its features are more robust than the other palm print recognition systems. It is basically suitable to be applied for dimensionality reduction in a computer vision followed by the introduction of KPCA to map some points to the higher dimension space [14, 15]. This is done by obtaining the kernel manifold learning algorithm, where the eigen decomposition of $n \times n$ kernel matrix or gram matrix k is formed. The KPCA approach provides more advantages than the PCA as the nonlinear features give out a more compact information. However, in some applications, the computation of eigen decomposition can slow down the process in KPCA. Therefore, an Evo-KPCA was proposed to devise a considerable faster eigen decomposition of kernel k .

A common starting point for the data reduction process in terms of probability $f(x)$ due to the density can be described by the distribution of the data. The kernel density estimate (KDE) is known as one of the nonparametric ways to evaluate the probability density function of a random variable in which the RSDE was the major influence in the propounded Evo-KPCA. Instead of using $n \times n$ gram matrix, the RSDE defined a weighted $m \times m$ gram matrix where $m \ll n$. This new gram matrix significantly reduced the computational cost by avoiding the computation of the full kernel matrix. As a result, the processing time can be decreased and a better performance of classification or recognition can be achieved.

The proposed Evo-KPCA for the palm print verification system is described in Section II. The experimental results are explained in Section III and the conclusion is being made in Section IV.

THE PROPOSED EVOLUTION OF KERNEL PRINCIPAL COMPONENT ANALYSIS (EVO-KPCA)

The Evo-KPCA was performed by employing the knowledge of kernel smoothing and learning with integral operator as shown below. Intuitively, the integral equation of KPCA is defined as:

$$(Kf)(x) = \int_{y_1}^{y_2} k(x,y)f(y)p(y)dy \tag{1}$$

The Equation (1) implies a kernel smoothing of the density. For example an operator k is applied to $f(x)$ where $(Kf)(x)$ is an output, $f(y)$ is an input, $kx, (y)$ is the chosen kernel function, x is an output variable and y is an input variable.

Assuming that a set of points (samples) of $N, X: x_i, i, \dots, n$ is drawn from the density $f(x)$, the empirical estimate of probability density $f(x)$ using X is given as:

$$f(x) \approx \frac{1}{n} \sum_{i=1}^n \delta(x_i, x) \tag{2}$$

Then, the smoothed approximation of Equation (1) is obtained as:

$$\hat{f}(x) = (Kf)(x) \approx \frac{1}{n} \sum_{i=1}^n k(x_i, x) \tag{3}$$

where n is a bandwidth or smoothing parameter.

Equation (3) is known as KDE and it can converge to $f(x)$ under certain conditions. The utilization of RSDE, $\hat{p}(x)$ was implemented and resulted in Equation (4) to avoid the usage of an expensive computation to compute the smoothed approximation, $\hat{p}(x)$. The equation (4) is

$$\hat{p}(x) = \frac{1}{m} \sum_{i=1}^m w_i k(c_i, x) \tag{4}$$

where $W = \{w_1, w_2, \dots, w_m\}$, $C = \{c_1, c_2, \dots, c_m\}$ and $m \ll n$. When it was compared to equation (2), the empirical density in generating $\hat{p}(x)$ under the kernel smoother K was given by:

$$p(x) \approx \frac{1}{m} \sum_{i=1}^m w_i \delta(c_i, x) \tag{5}$$

It then led to the eigendecomposition problem with the reduced gram matrix of :

$$\tilde{K} \tilde{Q}_i = \tilde{\lambda}_i \tilde{Q}_i, \quad \tilde{K}_{ij} = \sqrt{w_i} k(c_i, c_j) \sqrt{w_j} \text{ for } c_i, c_j \in C \tag{6}$$

The proposed Evo-KPCA replaced the gram matrix k empirically as shown below:

$$K \alpha_i = \lambda_i Q_i, \quad K_{ij} = k(x_i, x_j) \tag{7}$$

The gram matrix K in the empirical eigen problem was surrogated by a density weighted as:

$$\tilde{K} = WK^c W^T \tag{8}$$

where $K^c_{ij} := k(c_i, c_j)$ and $W = \text{diag}(\sqrt{w_1}, \dots, \sqrt{w_m})$ was the weight matrix and K^c was an $m \times m$ matrix.

The nodes C were chosen by a sampling that was obtained from the distribution of $f(x)$. The weight of a node was given by a fraction of points that contributed to the node where $\sum_{i=1}^m w_j = 1$. Once C was selected and the weight was computed using a RSDE, the original data was discarded. The following algorithm summarizes the proposed Evo-KPCA framework.

Summary for the proposed Evo-KPCA:

Input: Set of samples $X = \{x_1, x_2, \dots, x_m\}$

1. A reduced set density estimator was applied to X for a computation where $W = \{w_1, w_2, \dots, w_m\}$, $C = \{c_1, c_2, \dots, c_m\}$, and $n \ll m$.
2. The empirical measure generating \hat{p} under the kernel smoother K was given by:

$$p(x) \approx \frac{1}{m} \sum_{i=1}^m w_i \delta(c_i, x)$$

3. The Gram matrix K in the empirical eigen problem was replaced by a density weighted surrogate of :

$$\tilde{K} = WK^c W^T$$

where $K^c_{ij} := k(c_i, c_j)$, $W = \text{diag}(\sqrt{w_1}, \dots, \sqrt{w_m})$ was the weight matrix and K^c was an $m \times m$ matrix.

4. The eigenvector decomposition was performed as:

$$K \alpha_i = \lambda_i Q_i$$

5. The eigenvectors were reweighted as:

$$Q_i = w^{1/2} \alpha_i$$

Output: The eigen functions was computed as:

$$\tilde{Q}_i(x) = \sum_{i=1}^n Q_i k(c_i, x)$$

EXPERIMENTAL RESULTS

In this section, a comparative study on the performance of Evo-KPCA on the palm print database had been investigated and compared with the ROI, PCA and KPCA. The support vector machine (SVM) classifier was employed to calculate the score of the pattern matching between training and testing data.

Generally, the evaluation of the performance of a palm print verification system is based on the calculation values of False Acceptance Rates (FARs), GARs, False Rejection Rates (FRRs) and Equal Error Rates (EERs). FAR refers to the proportions of unauthorized individuals that are granted with an access by the system and it is defined as:

$$FAR = \left(\frac{\text{number of imposter} > \text{threshold}}{\text{number of imposter}} \right) \times 100\% \quad (9)$$

FRR is the percentage of wrongly rejected individuals over the number of genuine accesses:

$$FRR = \left(\frac{\text{number of genuine} < \text{threshold}}{\text{number of genuine}} \right) \times 100\% \quad (10)$$

GAR is the percentage of the number of correctly accepted individuals over the number of genuine accesses.

$$GAR = 1 - FRR = \left(\frac{\text{number of genuine} \geq \text{threshold}}{\text{number of genuine}} \right) \times 100\% \quad (11)$$

EER measures the effectiveness of the system. It can be found when FAR=FRR.

The results discovered by the research are shown in the Receiver Operating Characteristic (ROC) curves.

The experiments were implemented in Matlab R2007 (b) and conducted using Intel Core i7, 2.1GHz CPU, 6G RAM and Windows 8 operating system.

Palm print Database

The database was obtained from an Intelligent Biometric Group (IBG) for research and educational purposes. It contained 2400 color images captured from 40 young Asian volunteered users, who were students. The age range of the users was from 19 to 23 years old and each of them had provided a verbal consent before a photography session. An input image was captured using an Android application which can be run on most of the Android devices. The devices used were HTC One X, Samsung Galaxy S3 and Samsung Galaxy Tablet 2 and they were positioned at a fixed background. The taken image was later converted to grey image with a depth scale resolution of 256. Fig. 1 shows the flow chart of the development of the Android application.

In this application, the users were simply asked to place their palm facing the acquisition device. There was no peg or other tool was used in the system thus enabling the users to place their hands at different heights facing the mobile phone camera. The palm image appeared as a large and clear image when it was placed near to the camera as there were many line features and ridges can be captured within a shorter distance. Fig. 2 shows an example of graphic user interface (GUI) for data collection using the application in three different types of Android mobile. The files were saved in

JPEG format. Some of hand image samples in the database are shown in Fig. 3.

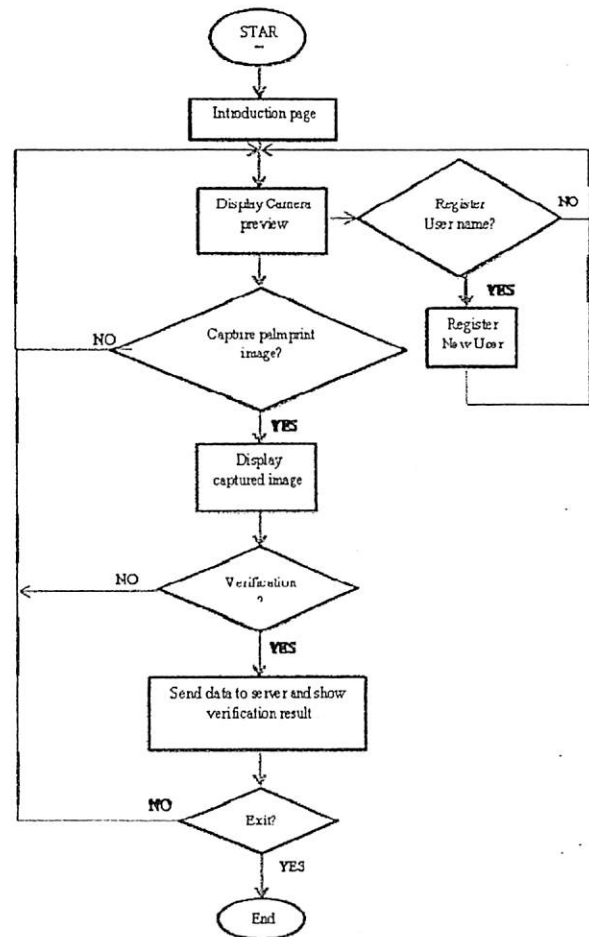


Fig. 1. The flow of Enrolment and Verification in an Android Application Development

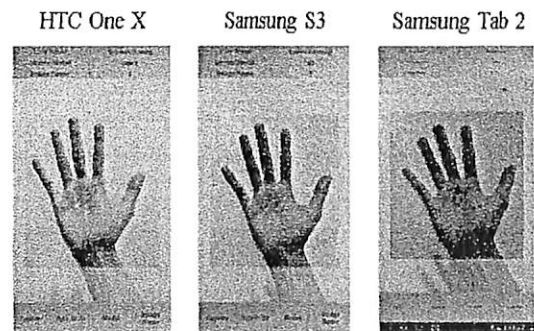


Fig. 2. The GUI for Data Collection Using Android Application

The captured image was then proceeded by using the Matlab application in the pre-processing stage. There were three major steps employed which were hand image detection, peak and valley detections and ROI extraction. In the hand

image detection, the tracking of hand boundary from a background was obtained followed by the peak and valley detections to trace the finger tips and valley respectively. Finally, the ROI was calculated and extracted based on the peak and valley data.

There were 60 palm print images from each users were captured, in which 20 images were used as a training set and the rest were used as the testing one. Thus, a total of 800 images and 1600 images were obtained for training and testing set respectively. As the focus of this paper was on the feature extraction of the palm print image, the extracted ROIs of the images were used in the experiments and they are shown in Fig. 4.

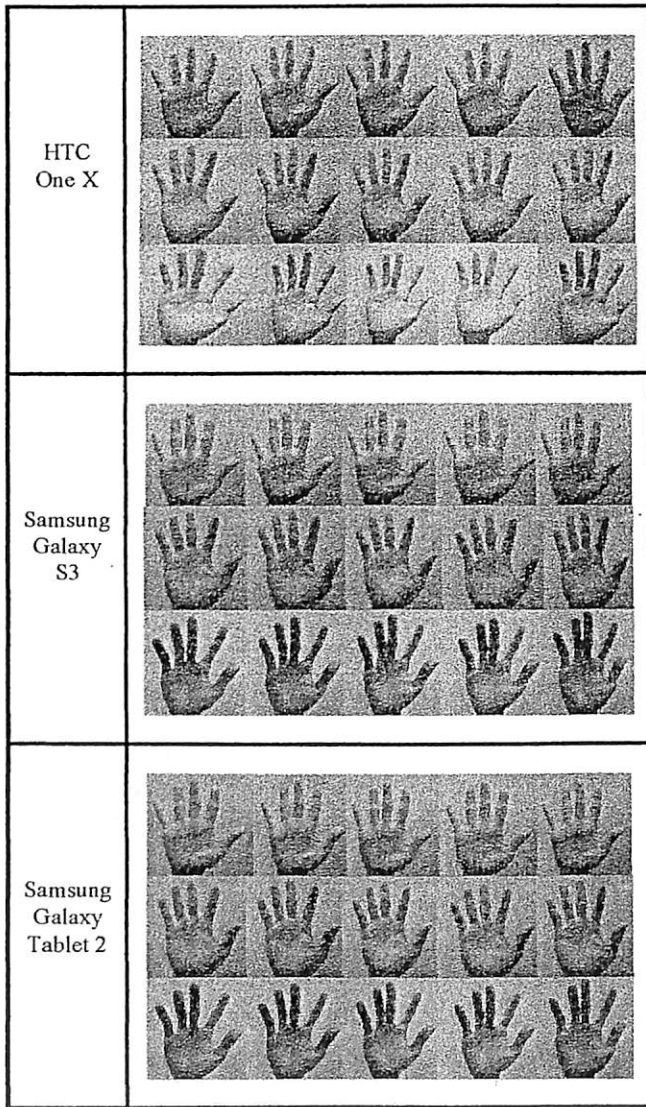


Fig. 3. The samples of hand images captured by using Android Application stored in Database

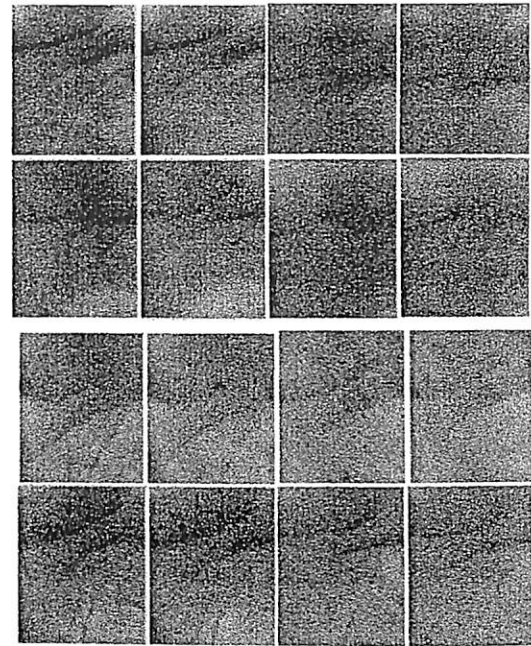


Fig. 4. The examples of extracted ROI palm print image collection

Performance Evaluation

The palm print features were extracted by using a constant feature dimension of size 64×64 . Based on the aforementioned Evo-KPCA algorithm, the calculation depended on the value of nodes C in computing the weights. Hence, it was compulsory to determine the value of C in Evo-KPCA where C was set at 30 in this experiment.

The performances of the palm print verification system by employing four approaches are illustrated in Fig. 5. The GARs at FAR of 1% of the ROI, PCA, KPCA and Evo-KPCA were observed at 98.6%, 97.5%, 99.4 and 99.1% respectively. Their percentages at FAR=10% increased to approximately 99.8% for ROI, KPCA and Evo-KPCA, and 98.1% for PCA.

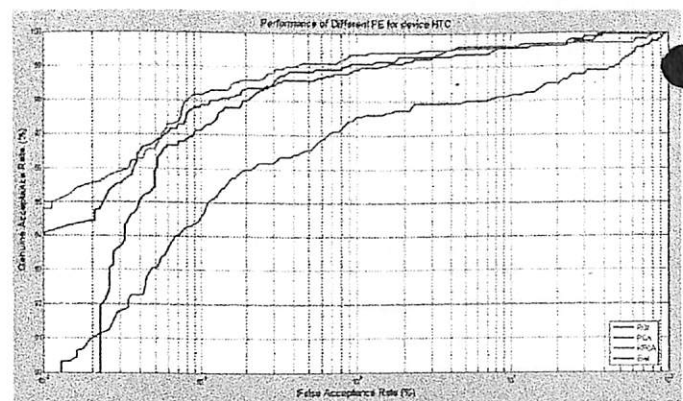


Fig. 5. The comparison of ROC curves of different feature extractions for HTC One X device

Fig. 6 shows the comparison of ROC curves for four feature extraction approaches experimented using Samsung Galaxy S3 device. The result shows that when FAR=0.1%, the KPCA performed the best with GAR value of 98.5%, while the Evo-KPCA attained the worst result which was lower than 97%. However, at FAR=1%, the GAR value for Evo-KPCA increased to be more than 99%, meaning that its performance was the most outstanding one compared to the other features.

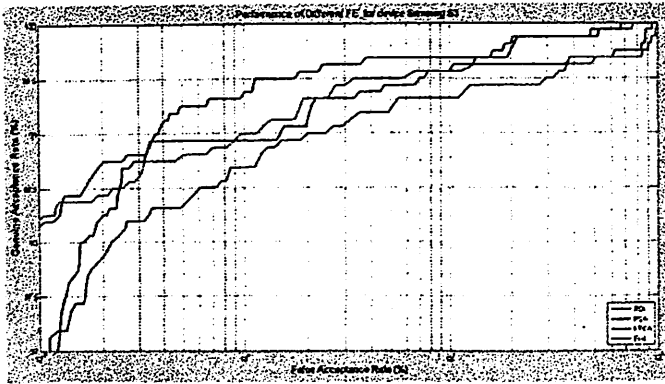


Fig. 6. The comparison of ROC curves of different feature extractions for Samsung Galaxy S3 device

The performance results of ROI, PCA, KPCA and Evo-KPCA using Samsung Galaxy Tablet 2 device are illustrated in the ROC curves in Fig. 7. The GAR values at FAR=1% of ROI and KPCA features were 99.5%. The values were 99% for the PCA and Evo-KPCA. The GAR percentages at FAR=10% increased to 99.5% for ROI, PCA, KPCA and Evo-KPCA. It was observed that when FAR was lower than 0.01%, Evo-KPCA was gradually performed better than the other approaches.

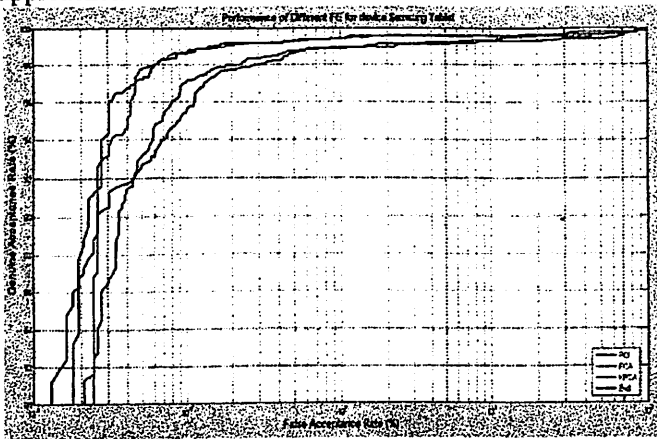


Fig. 7. The comparison of ROC curves of different feature extractions for Samsung Galaxy Tablet 2 device

Table 1 indicates the EER performances and time consuming for matching process for the ROI, PCA, KPCA and Evo-KPCA approaches. The KPCA features from HTC One X and Samsung Galaxy Tablet 2 had the best EER percentage compared to the other features. However, Evo-

KPCA consumed the shortest processing time with no much difference than the KPCA in EER percentage. In addition, Samsung Galaxy S3 proved that the Evo-KPCA had attained the best EER percentage and the lowest time consumed in the matching process.

HTC One X	EER (%)	Time (Sec)
ROI	1.0369	26.0318
PCA	2.1931	50.8717
KPCA	0.7973	46.3044
Evo-KPCA	0.9159	0.2760
Samsung Galaxy S3	EER (%)	Time (Sec)
ROI	1.0673	25.8105
PCA	1.1987	61.5136
KPCA	0.9824	52.2588
Evo-KPCA	0.6859	0.4841
Samsung Galaxy Tab 2	EER (%)	Time (Sec)
ROI	0.6306	26.2614
PCA	1.0649	47.2986
KPCA	0.6290	53.1162
Evo-KPCA	0.8865	0.4434

EER PERCENTAGES OF BIOMETRIC SYSTEMS WITH DIFFERENT FEATURE EXTRACTIONS BY USING THREE DIFFERENT DEVICES

CONCLUSION

An empirical work of the Evo-KPCA was proposed and successfully implemented in the palm print database. The propounded approach aimed to exploit the strength of Evo-KPCA while minimizing the computation in the gram matrix such that the processing time can be reduced. Therefore, an alternative formulation of the RSDE was proposed to devise a considerable faster eigen decomposition of kernel k by allowing the extraction of the most relevant and important information from a large dataset. In order to acquired an effective Evo-KPCA, it was compared with the ROI, PCA and KPCA. A series of experiments based on different android mobile phones were set to determine the competency of the proposed feature. It can be concluded that the Evo-KPCA provided the GAR value for more than 98% and it took the shortest processing time which was less than 0.5s.

ACKNOWLEDGMENT

This work was financially supported by Research University Grant 814161 and Research University-Post Graduate Grant Scheme 8046019.

REFERENCES

[1] H. Jaafar and D. A. Ramli, "A Review of Multibiometric System with Fusion Strategies and Weighting Factor" in International Journal of Computer Science Engineering (IJCSSE), vol.2, no.4, pp. 158-165, 2013.
 J.P. Campbell, D.A. Reynolds and R.B. Dunn, "Fusing High-And Low-Level Features for Speaker Recognition" in INTERSPEECH, 2003.

- A.K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition. Circuits and Systems for Video Technology" IEEE Transactions, vol. 14. No. 1, pp. 4-20, 2004.
- L. Akarun, B. Gökberk, and A.A. Salah, "3D Face Recognition for Biometric Applications" 13th European Signal Processing Conference (EUSIPCO), 2005.
- R.P. Wildes, "Iris recognition: an emerging biometric technology" Proceedings of the IEEE, vol. 85. No. 9, pp.1348-1363, 1997.
- L. Zhang, D. Zhang, and H. Zhu, "Online finger-knuckle-print verification for personal authentication" Pattern Recognition, vol. 43, pp. 2560-2571, 2010.
- G. K. O. Michael, C. Tee, A. T. Jin, "Touch-less palm print biometrics: Novel design and implementation," Image Vis Comput. Vol. 26, pp. 1551-1560, 2008.
- P. Somvanshi, M. Rane, "Survey of palmprint recognition," International Journal of Scientific & Engineering Research, vol. 3, No. 2, pp. 1, 2012.
- T. Connie, A. Teoh, M. Goh, and D. Ngo, "Palmprint Recognition with PCA and ICA," In Proc. Image and Vision Computing, New Zealand, 2003.
- G. Lu, D. Zhang, K. Wang, "Palmprint recognition using eigenpalms features," Pattern Recognit Lett, vol. 24, pp. 1463-1467, 2003.
- W. K. Kong, D. Zhang, W. Li, "Palmprint feature extraction using 2-D Gabor filters," Pattern Recognition, vol. 36. pp. 2339-2347, 2003.
- W. Li, D. Zhang, Z. Xu, "Palmprint identification by Fourier transform," International Journal of Pattern Recognition and Artificial Intelligence, Vol. 16, No. 04, pp. 417-432, 2002.
- A. Kumar, H. C. Shen, "Recognition of palmprints using wavelet-based features," Proc. Intl. Conf. Sys., Cybern., 2002.
- M. Welling, "Kernel Principal Components Analysis. Advances in neural information processing systems," Vol. 15, pp. 70-72, 2003.
- Q. Wang, "Kernel principal component analysis and its applications in face recognition and active shape models", arXiv preprint arXiv:1207.3538, 2012.

Search

Alerts

My list

My Scopus

[LinkSource](#) | [Export](#) | [Download](#) | [More...](#)

International Journal of Circuits, Systems and Signal Processing

Volume 8, 2014, Pages 137-148

Intelligent frog species identification on android operating system

(Article)

Tan, W.C.^a, Jaafar, H.^a, Ramli, D.A.^a, Rosdi, B.A.^a, Shahrudin, S.^b^a Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia^b School of Pharmaceutical Sciences, USM, 11800 Pulau Pinang, Malaysia

Abstract

In this paper an Intelligent Frog Species Identification System (IFSIS) which works as a sensor is developed. It is designed to assist the nonexperts to recognize frog species according to frog bioacoustics signals for environmental monitoring. IFSIS consists of Android devices and a server. Android device is used to record frog call signal and to display the details of the detected frog species once the identification is processed by the server. Meanwhile, feature extraction and identification process of the frog call signal are done on Intel atom board which works as server. The Mel Frequency Cepstrum Coefficient (MFCC) is used as feature extraction technique while the classifier employed is Support Vector Machine (SVM). Experimental results show that the performances of 95.33% has been achieved which proves that IFSIS can be a viable automated tool for recognizing frog species.

Author keywords

Android device; Frog call; Mel Frequency Cepstrum Coefficient; Support Vector Machine

Indexed keywords

Engineering controlled terms: Feature extraction; Speech recognition; Support vector machines; Display devices; Extraction; Feature extraction; Speech recognition; Support vector machines

Android device; Android operating systems; Environmental Monitoring; Feature extraction techniques; Frog calls; Identification process; Mel frequency cepstrum coefficients; Species identification; Automated tools

Engineering main heading: Android (operating system); Android (operating system)

ISSN: 19984464 Source Type: Journal Original language: English

Document Type: Article

Publisher: North Atlantic University Union

Ramli, D. A.; Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia; email:dzati@usm.my
© Copyright 2014 Elsevier B.V., All rights reserved.

Cited by 2 documents

Potassium carbonate-treated palm kernel shell adsorbent for congo red removal from water
Zhi, L.L., Zaini, M.A.A.
(2015) Jurnal Teknologi

Investigation on the possibility of using entropy approach for classification and identification of frog species
Ng, C.H., Dayou, J., Ho, C.M.
(2015) Jurnal Teknologi

View all 2 citing documents

Inform me when this document is cited in Scopus:

[Set citation alert](#) | [Set citation feed](#)

Related documents

Find more related documents in Scopus based on:

[Authors](#) | [Keywords](#)

Metrics

2 Citations

2.03 Field-Weighted Citation Impact

2 Mendeley Readers

66TH PERCENTILE

View all metrics

Top of page

About Scopus
What is Scopus
Content coverage
Scopus Blog
Scopus API

Language
日本語に切り替える
切换到简体中文
切换到繁體中文

Customer Service
Help and Contact
Live Chat

About Elsevier
Terms and Conditions
Privacy Policy



Copyright © 2015 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.
Cookies are set by this site. To decline them or learn more, visit our Cookies page.

Intelligent frog species identification on android operating system

W. C. Tan, H. Jaafar, D. A. Ramli, B. A. Rosdi and S. Shahrudin

Abstract— In this paper an Intelligent Frog Species Identification System (IFSIS) which works as a sensor is developed. It is designed to assist the non-experts to recognize frog species according to frog bioacoustics signals for environmental monitoring. IFSIS consists of Android devices and a server. Android device is used to record frog call signal and to display the details of the detected frog species once the identification is processed by the server. Meanwhile, feature extraction and identification process of the frog call signal are done on Intel atom board which works as server. The Mel Frequency Cepstrum Coefficient (MFCC) is used as feature extraction technique while the classifier employed is Support Vector Machine (SVM). Experimental results show that the performances of 95.33% has been achieved which proves that IFSIS can be a viable automated tool for recognizing frog species.

Keywords— Android device, Mel Frequency Cepstrum Coefficient, Support Vector Machine, Frog call.

I. INTRODUCTION

FROGS are the most common group of amphibians. Many ecologists suggest that amphibians, such as frogs, are good biological indicators because of the health of frogs is signifying the health of the whole ecosystem [1,2,3]. This is due to three reasons. Firstly, frogs require suitable habitat for both the terrestrial and aquatic environment. Secondly, frogs

This work was supported in part by Universiti Sains Malaysia Short Term Grant 60311048, Research University Grant 814161 and Research University-Post Graduate Grant Scheme 8046019.

W.C. Tan, is with Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia (e-mail: twc0317@gmail.com).

H. Jaafar, is with Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia (e-mail: haryati_jaafar@yahoo.com).

D.A. Ramli is with Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia (corresponding author to provide phone: +604-5996028; e-mail: dzati@usm.my).

B.A. Rosdi is with the Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia (e-mail: cebakhtian@usm.my).

S. Shahrudin is with the School of Pharmaceutical Sciences, USM, 11800 Pulau Pinang, Malaysia (email: shahriza18@usm.my)

are in the intermediate positions of the food chains. The third reason is the skin of frogs is permeable which can easily absorb toxic and pollutants [3].

These phenomena which can be observed in our surroundings can become signs to the environmental disturbances. As reported in [4,5]. Frogs have survived for the past 250 million years in countless ice ages, asteroid crashes and other environmental disturbances but yet, one-third of these amphibian species are on the verge of extinction nowadays. So, this should be served as an alarm call to humans that if drastically wrong in our environment. Hence, smart environment monitoring is needed so as to preserve the world from frog species elimination.

Apart from for environmental monitoring, another important factor which encourages this research is owing to the discovery of the secreted peptide on frog's skin. Since, numerous bacteria are now able to develop resistance against formerly drug or antibiotic which can cause a serious threat to public health, the scientists have rekindled their interest in other alternatives for new antimicrobial agents. This new peptides have evolved a chemical resources for body protection and oxidant scavenging activities[6].

A group of Russian researchers have discovered that over 76 different antimicrobial peptides on the skin of the European Common Brown Frog (*Rana Temporaria*) and these peptides have potential in preventing both pathogenic and antibiotic resistance [7]. The Alkaloid Epibatidine was also found on the skin of an Ecuadorian poison frog and it is proved to work as a powerful painkiller [8,9,10]. Finally, as frog eggs and oocytes are also involved in the cloning and embryology research so this activity will also be benefited by this proposed sensor.

Since frogs bring many advantages to ecosystem and some certain species are important for medical researches, an automatic recognition of frog species is needed. Currently, detecting and localizing certain frog species is commonly done manually by experts who is capable in recognizing the morphological characteristic of the frog. In this process, frogs or portion of the frogs need to be localized and then captured. This requires only experts with sufficient experiences and intuition to conduct the procedure. Nevertheless, the numbers of the qualified expert in this field is very limited. Furthermore, this procedure also involves intensive field sampling which is troublesome to be done manually using human visual sense [11, 12].

Due to almost all animals generate sounds either for communication or as a by-product of their living activities, so in this study, an automatic detection of frog species based on frog call is investigated. Besides, we often hear animal sound or vocalization rather than see the animal in the forest. As animals generate sounds to communicate with members of the same species, thus their vocalizations have evolved to be species-specific. Hence, recognizing animal species based on their vocalization is more effective.

Current developed frog species identification system is based on the architecture of audio biometric identification system [13,14]. Typically, this system architecture is divided into five modules which are frog call signal acquisition, signal pre-processing, feature extraction and pattern matching using classifier as illustrated in Fig. 1. This current developed frog species identification system is only implemented in computer or laptop. Due to frog species identification are always taken place in outdoor such as forest, river-side, rural area, and wetlands [15] then, the system which built with laptop are not really feasible and user-friendly. As a result, it is imperative to improve the current frog identification system to become portable and more practical. Thus, more samples can be collected and the field work can be done efficiently, conveniently, and more cost-effective. In this study, an automated system based on Intel Atom board and hand-held device android smartphone is proposed. This system is a client-server based system, where the Android smartphone acts as client and Intel Atom board acts as server. Another similar approach involved system monitoring and detection has also been researched; but for outdoor plant detection which can be found in [16]. Then, the use of smartphone platform for human behavior cognition and sensing context was reported in [17].

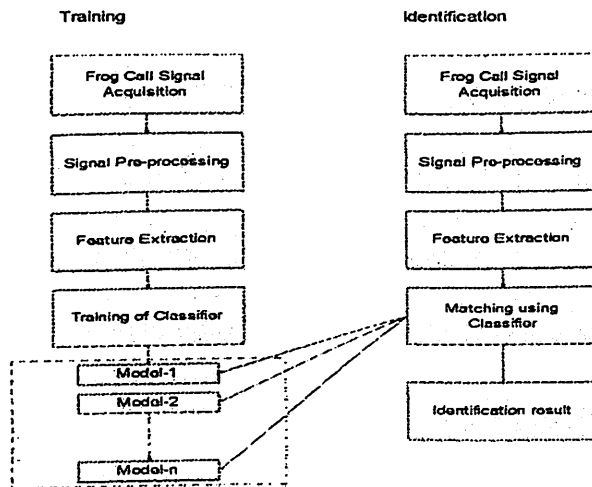


Fig. 1. Frog species identification system.

The first objective of this study is to develop data acquisition module on an Android device. Subsequently, the second objective is to develop feature extraction and classification algorithm based on Mel-Frequency Cepstrum Coefficient (MFCC) technique and Support Vector Machine

(SVM) on the Intel atom board. Finally, the last objective is to set up client-server communication between Android device and Intel Atom hence to integrate the whole system as real time frog call identification which can work as frog species identification sensor.

II. METHODOLOGY

Overview of the proposed implementation of the Intelligent Frog Species Identification System (IFSIS) is shown as in Fig. 2. The overall system is described as follows:

2. The overall system is described as follows:

- Input stage - Audio file of frog call is obtained either recorded by using microphone of the Android device or loaded offline from its memory cell. This audio file is then sent to the server which is an atom processor via Hypertext Transfer Protocol (HTTP).
- Processing stage - A Hypertext Preprocessor (PHP) script on the server invokes the server-side application to initiate the identification process. The recorded frog call audio are processed so as to extract the features. Subsequently, the extracted features are fed to the intelligent classifier for the identification of the types of species.
- Output stage - The identification result, which is the frog species that has been identified based on the input sound wave will be generated by the server. This result is then sent to the Android device for displaying purpose.

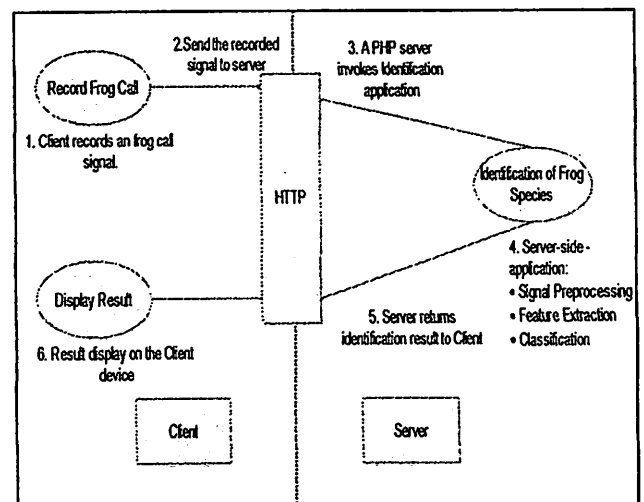


Fig. 2. Overview of IFSIS

A. Data acquisitions







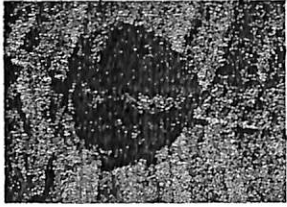


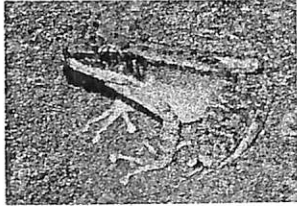





Two sources of frog call samples for data acquisition are database collected by IBG Research Group, PPKEE, USM and recorded samples using Android-powered device (Samsung Galaxy S3).

The frog call samples obtained from IBG Research Group were recorded at Sungai Sedim, in Kulim, Kedah, Malaysia from 8.00 pm to 12.00 pm using Sony Stereo IC Recorder ICD-AX412F supported with Sony electret condenser microphone. The sounds samples were recorded in wav files at a sampling frequency of 44.1 kHz and are then converted

to 16-bit mono. The recording dataset include samples of 15 species where the scientific name, common name and images are tabulated as in Table 1 [18,19]. The recordings were later analyzed by Praat software with the following parameters i.e

call durations, average calls and standard deviations of frog calls. It was observe, depends on the species, the number of calls varies from as low as 61 and high as 148 where the average of their calls are 0.25 to 1.2s as shows in Fig. 3.

Table 1. List of frog call samples used in the project

Image, scientific name and common name		
<p><i>Hylarana glandulosa</i> Rough sided frog</p> 	<p><i>Polypedates leucomystax</i> Common tree frog</p> 	<p><i>Microhyla heymonsi</i> Taiwan rice frog</p> 
<p><i>Phrynooides aspera</i> River toad</p> 	<p><i>Kaloula baleata</i> Flower pot toad</p> 	<p><i>Fejervarya limnocharis</i> Grass frog</p> 
<p><i>Kaloula pulchra</i> Asian painted bullfrog</p> 	<p><i>Philautus njobergi</i> Bubble-nest frog</p> 	<p><i>Hylarana labialis</i> White-lipped frog</p> 
<p><i>Odorrana hosii</i> Poisonous rock frog</p> 	<p><i>Duttaphrynus melanostictus</i> Black-spectacled toad</p> 	<p><i>Genus ansonia</i> Stream toad</p> 
<p><i>Philautus petersi</i> Kerangas bush frog</p> 	<p><i>Microhyla butleri</i> Painted chorus frog</p> 	<p><i>Rhacophorus appendiculatus</i> FILLED tree frog</p> 

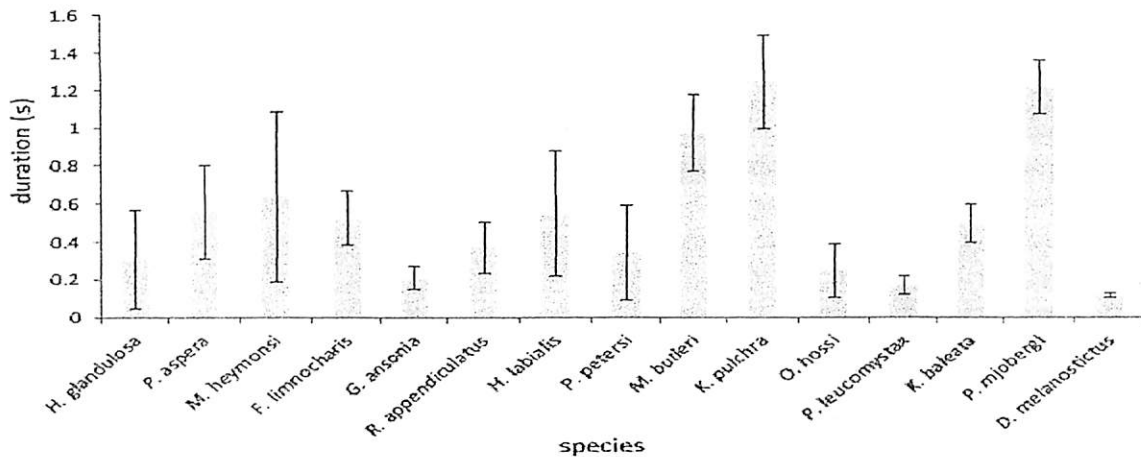


Fig. 3. Average and standard deviations for call duration of frog calls

Fig. 4 shows an example of calls waveform and spectrogram from *Microhyla heymonsi* and *Microhyla butleri*. From the figure, the waveforms from each call looked similar. However, each species has different calls based on how the individual frog permanently changes its calls. The changing of calls is occur in a wide range of frequencies and some are long, lasting several seconds, while others last only in fraction of a second.

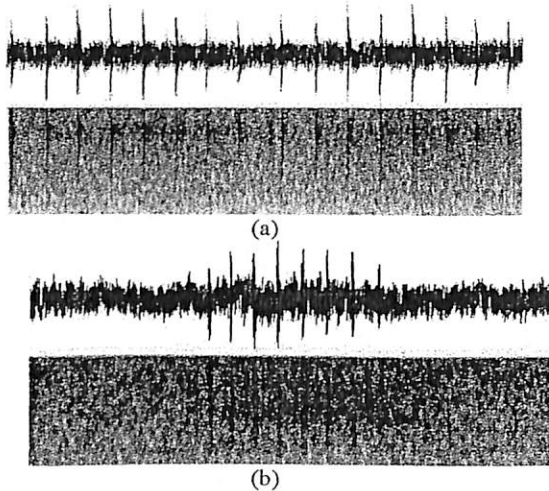


Fig. 4. Waveform and spectrogram for (a) *Microhyla heymonsi* (b) *Microhyla butleri*.

The frog call samples obtained from IBG Research Group is then played using speaker. The sound from the speaker is recorded using Samsung Galaxy S3. These recorded sounds are used for testing purpose for system evaluation.

B. Development of Client System in Android Device

The softwares required to develop the Android application are Eclipse IDE, Android SDK and Unified Modeling Language (UML). Eclipse IDE and Android SDK are the Android Developer Tool (ADT) which is used to develop the IFSIS Android Application for the Android client device. Unified

Modeling Language (UML) is used to design the Android application for the system. A very user-friendly graphical user interface (GUI) is then designed for the application. The use case diagram which is used to determine the requirement of the specifications of the Android application is illustrated in Fig. 5.

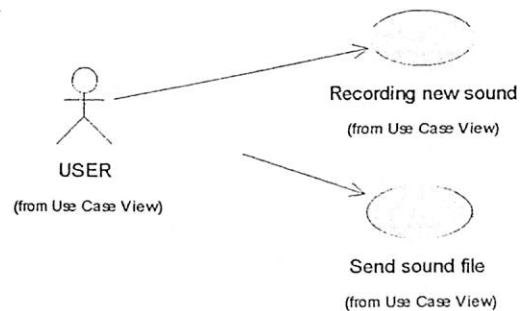


Fig. 5. Use case diagram of Android application.

IFSIS Android application is a graphical user interface application which can be run on most of the Android device. This application is used to record frog calls signal, save the signal into audio file in WAV format, upload the audio to the server, and download result from the server. The overall flowchart of this application is shown in Fig. 6. This application consists of two major parts i.e. the recording and uploading of the audio file. The Android multimedia framework includes support for capturing and encoding a variety of common audio formats. In order to perform the audio capturing or recording, some variables need to be declared and initialized. Subsequently, *Android MediaRecorder* instance is created in the next step. Next, the system needs to set the audio source, output file format, output file name, and audio encoder according to our requirements. After that, users start recording and stop recording by calling functions. Before the process ends, the system releases the *MediaRecorder* instance in order to clear

the memory. The recorded sound signal is saved in WAV format on the Android device memory.

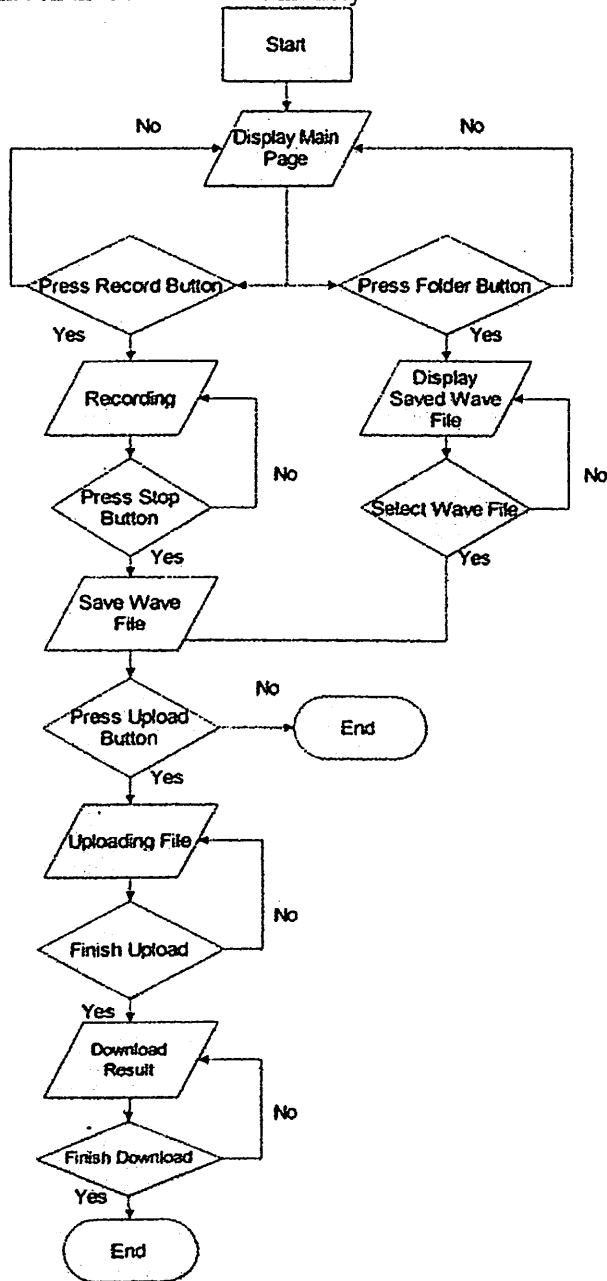


Fig. 6. Overall flowchart of IFSIS Android application

The flowchart for recording part is shown in Figure 7a.

After the frog call is recorded, it can be uploaded to the server for identification process. The application will use Hypertext Transfer Protocol (HTTP) to send and receive data. Android includes two HTTP clients: HttpURLConnection and Apache HttpClient; both of them support HTTP configurable timeouts, IPv6, and connection pooling. For IFSIS android application, HttpURLConnection is used. Next, the system creates a buffer of maximum size which is enough for the

audio file to be uploaded. After that, the system will read the file and write it into form. If necessary, multipart form data is sent, and close the file input streams. While the file is being sent to the server, the application will continuously read the response from the server. The upload process is complete after the server received the file. The flowchart of uploading file is shown in Fig. 7b.

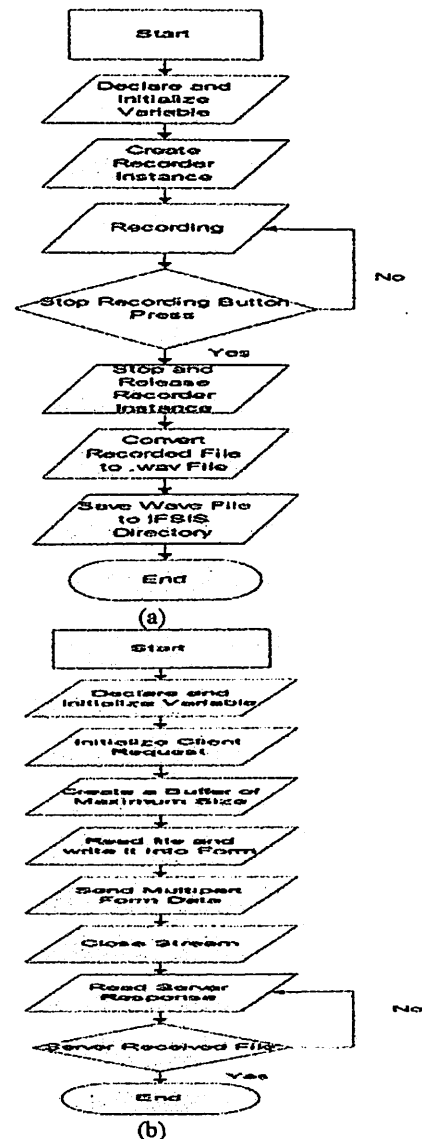


Fig. 7. Flowchart of (a) recording function and (b) uploading function developed for Android application

C. Development of Server System in Intel Atom Board

i) Server system requirement

The main hardware of the server is Intel Atom Innovation Kit 3 with some connected peripherals, such as hard disk, liquid crystal display (LCD), mouse, keyboard and wireless modem. The hard disk is used to store the operating system, files and software needed in the project. LCD monitor screen,

mouse and keyboard are used to set up the identification system in the server. A wireless modem router is connected to the Gigabit Ethernet port of the server using LAN cable. Once the connection between modem and server is established, the server is ready to receive files from client devices via wireless network and send identification result back to the client. Fig. 8 shows the hardware architecture of the server part in this project. The software required is XAMPP 1.8.1 with PHP 5.4.7, Scilab 5.4.1 and Matlab 2011b. XAMPP is a free and open source cross-platform web server package. It is used to receive requests from clients and return the requested content to them. PHP is the languages used for the web server development. In this project, a PHP script is written to facilitate the communication between client and server. This script is used to receive audio files from clients, invokes Scilab script, and sends results back to the clients. Scilab is an open source, cross-platform numerical computational package with a high-level programming language. The processing of sound signal data and species identification is implemented using Scilab and Matlab.

ii) Frog call signal processing and species identification Signal Pre-Processing

Once the audio file is received by IFSIS server, the server system will read the sound signal and execute signal pre-processing process on the signal. Signal pre-processing process in this system includes noise reduction, syllable segmentation, signal pre-emphasis, framing and windowing. These steps are employed in order to reduce computing time and increase the accuracy of the identification system.

Noise Reduction Using Band-Pass Filter.

Although it is impossible to remove all noises from the recorded sound signal, it can be minimized to certain acceptable level. The recorded frog call signals which are normally corrupted by various types of noise can reduce the accuracy of identification. In reality, noise is not merely from the environment but it can be due to the residual electronic noise signal. This electronic noise gives rise to acoustic noise heard as 'hiss' and it is high in frequency. Therefore, low-pass filter is used in this project so as to reduce the noise in the recorded sound signal. The cut-off frequency of the low pass filter is set lower than the noise frequencies so that the interested bandwidth can be preserved while the 'hiss' noise is filtered. In this project, Scilab predefined function 'filter' and 'zpbutt' are used to perform low pass filter on the recorded sound signal. Besides, Matlab predefined function 'filter' and 'butter' are used for the same purpose. The steps to develop low-pass filter are described as follows:

1. Obtain zeros and poles by using 'zpbutt' function in Scilab or 'butter' function in Matlab. This function computes the poles of a Butterworth filter of order n and cutoff frequency, f_c .
2. Obtain low-passed signal by using 'filter' function. The filter is set up using the zeros and poles computed in step 1.

Syllable segmentation

A syllable is a sound that a frog produces with a single blow of air from its lungs. Compared to human, frog syllables seem to be slightly less complex than human due to no-vowel-consonant and less intricate grammar [20]. In the past work, it has been indicated that zero-crossing rate (ZCR) and short-time energy (STE) are the two most important time domain and low level features which play major role in end point detection and syllable segmentation of speech [21,22]. In this project, ZCR is used together with STE for syllable segmentation. The steps for syllable segmentation using STE and ZCR are as follows:

1. The low-pass filtered signal is blocked into small frames of 20 milliseconds. The filtered signal waveform consists of a long sequence of sampled values. Thus, it is useful to break the long sequence into small frames which are quasi-stationary. The more sample points in a frame it has the less stationary it is. Therefore, a 20 millisecond frame size is chosen to compromise between sufficient sample points for accurate analysis and the quasi-stationary assumption. For a short-term sound signal (the n^{th} frame sound after framing and windowing) is as shown in (1)

$$x_n(m) = x(m)w(n-m), \quad n-N+1 \leq m \leq n \quad (1)$$

Where $w()$ is window function and n is the sample that the analysis window is centered on, and N is the window size.

2. The STE of each frames is computed using (2)

$$E_n = \sum_{m=n-N+1}^n [x(m)w(n-m)]^2 \quad (2)$$

Where $x(m)$, $m = 1, \dots, N$ is the audio samples of the n^{th} frame. This simple feature can be used for detecting silent part in audio signals.

3. ZCR is a simple measure of the frequency content of a signal, especially true for narrowband signals such as sinusoids. The ZCR of each frames is computed using (3):

$$Z_n = \frac{1}{2} \sum_{m=0}^{N-1} |x(m) - x(m+1)| \quad (3)$$

Z_n is especially helpful for detecting speech from noisy background or begin and end point detection,

4. Mean and standard deviation of E_n and Z_n for the first 100 millisecond of signals are computed. It is assumed that no voiced part in this interval.
5. The maximum value of E_n from all of the frames is determined.
6. E_n thresholds are computed based on results of steps 4 and 5. This thresholds are upper threshold (ETU) and lower threshold (ETL) computed by taking some percentage of the peaks over the entire interval. Threshold for zero crossings (ZCT) based on zero crossing distribution for unvoiced speech is computed.

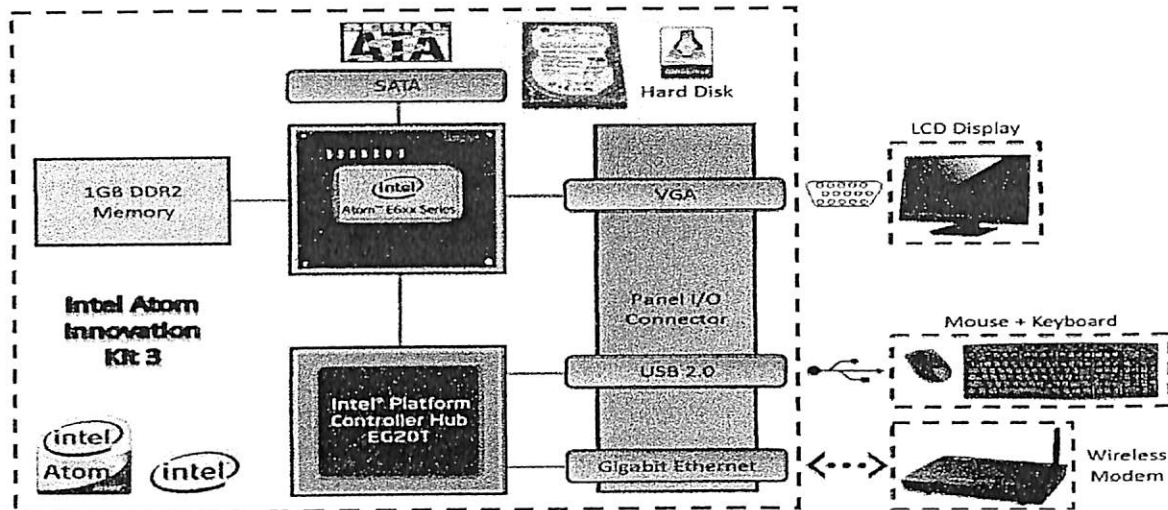


Fig. 8. Overall diagram of the server-side devices.

7. E_n thresholds are computed based on results of steps 4 and 5. These thresholds are upper threshold (ETU) and lower threshold (ETL) computed by taking some percentage of the peaks over the entire interval. Threshold for zero crossings (ZCT) based on zero crossing distribution for unvoiced speech is computed.
8. The frame with E_n exceeds the ETU threshold is determined. Find a putative starting point (N1) where E_n crosses ETL from below and a putative ending point (N2) where E_n crosses ETL from above.
9. Move backwards from N1 by comparing Z_n to ZCT, and find the first point where Z_n smaller than ZCT; similarly move forward from N2 by comparing Z_n to ZCT and finding last point where Z_n smaller than ZCT.
10. The first point and last point in step 8 are the starting point and ending point of a single syllable of voiced part signal.

Signal Pre-emphasis

The segmented syllable is first pre-emphasized to compensate the high-frequency part that was suppressed during the call production mechanism of frogs. It can also amplify the high-frequency syllables of frog call to obtain similar amplitude for all syllables. This is important because high-syllables have smaller amplitude relative to low-frequency syllables. Signal pre-emphasis is also applied to prevent numerical instability. Pre-emphasis of the segmented syllable frog call signal is implemented by filtering it with a first order FIR filter. The transfer function of this filter is in z-domain as follow:

$$H(z) = 1 - \alpha z^{-1} \quad 0 \leq \alpha \leq 1 \quad (4)$$

α being the pre-emphasis parameter. Essentially, pre-emphasis filter is a first order high-pass filter in time domain.

The relationship between pre-emphasized signal and input signal is shown as follow:

$$x'(n) = x(n) - \alpha x(n-1) \quad (5)$$

A typical value for α is 0.95. This value of α gives rise to a more than 20 dB amplification of the high frequency spectrum.

Framing and Windowing

A frame-based analysis is essential for speech signals. This short-term processing is performed by framing and windowing methods. The pre-emphasized signal $x'(n)$ is framed and windowed into succession windowed sequences $x_t(n)$, $t = 1, 2, \dots, T$, known as frames. This frame can be processed individually as:

$$x'_t(n) \equiv x'(n - t \cdot Q), \quad 0 \leq t \leq T, 0 \leq n \leq N \quad (6)$$

$$x_t(n) \equiv w(n) \cdot x'_t(n) \quad (7)$$

N is the number of samples in a frame and $w(n)$ is the impulse response of the window. Each frame is shifted by a temporal length Q given Q is smaller than N . The number of samples overlapped of one frame to the previous frame is equal to $N - Q$. This means that a total of $N - Q$ samples at the beginning of a particular frame $x_{t+1}(n)$ are duplicated from the end of the previous frame $x_t(n)$.

In this project, Q and N are set in order to overlap the frames in 50%. These frames must be in quasi-stationary so that the digitized sound signal can be represented by frames. After framing, the processing step is followed by windowing each individual frame to minimize the signal discontinuities at the beginning and end of each frame. The concept of using this step is to use the window to taper the signal to 0 at the beginning and end of each frame. This is very important because discontinuity at the begin point and last point of a frame will introduce undesirable effects in the frequency response. Frequency response of the signal is computed in feature extraction methods. Effectively, the signal is cross-multiplied by a window function as follows:

$$x'_t(n) = x_t(n)w(n), \quad 0 \leq n \leq N - 1 \quad (8)$$

There are several types of window functions such as rectangle, Hanning, Hamming, Blackman and Kaiser. Hamming window is a typical window applied most frequently. This window has the form,

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1 \quad (9)$$

Hamming window exhibits lower side lobes and wider main lobe than other windows. This characteristic of Hamming window reduces the leakage effect and resolution of sound signal. Thus, Hamming window is a good choice in speech recognition, because leakage will give negative effect on the signal and a high resolution is not required [23].

Feature Extraction Using MFCC

The extraction of important parametric representation of frog call signals is a crucial task in order to achieve better identification and recognition performance. Mel Frequency Cepstral Coefficients (MFCC) is one of the most commonly used feature extraction method in speech recognition. MFCC takes human hearing perception sensitivity with respect to frequencies into consideration.

After the frog call sound signal is pre-emphasized, framed, and windowed, MFCC is used to extract meaningful parameter in the frog call sound signal. The steps to implement MFCC in this project are as follows:

1. Discrete Fourier Transform (DFT) of each frame is computed for each frame. Each frame of N samples is converted from time domain into frequency domain in this step. The Fourier Transform is to convert the convolution of the glottal pulse and the vocal tract impulse response in the time domain. The DFT of all frames of the pre-processed frog call signal is:

$$X_t(\omega) = \sum_{n=1}^N x'_t(n) e^{-j\omega n}, \quad 1 \leq k \leq K \quad (10)$$

Where n is the number of samples in a frame, and k is the domain index of the DFT.

2. The signal spectrum is then processed by mel filter bank processing. The frequencies range in signal spectrum is very wide and voice signal does not follow the linear scale. Therefore, the magnitude of frequency is multiplied by Mel filter bank. This is to obtain the log energy of each triangular band-pass filter in the filter bank. The filter bank used in this project is consists of 24 triangular band-pass filter that is emphasize on processing the spectrum which frequency is below 1 kHz.

The positions of these filters are equally spaced along the Mel frequency scale and related by following equation:

$$f_{mat} = 2595 \log_{10} \left(1 + \frac{f}{700}\right) \quad (11)$$

Where f_{mat} is the subjective pitch in Mels corresponding to a frequency in Hz. Psychophysical studies have shown that human perception of the sound frequency contents for speech signals does not

follow a linear scale. Therefore, Mel scale is used to measure the subjective pitch of each tone with an actual frequency, f , measured in Hz.

3. Normally, log energy is obtained by computing the logarithm of square magnitude of the coefficients $Y_t(m)$. $Y_t(m)$ is the m^{th} filter bank output. In this project, the log energy is obtained by computing logarithm of the magnitude of the coefficients. This is done for reducing the complexity of computing.
4. Inverse DFT is computed on the logarithm of the magnitude of the filter bank output as shown following:

$$y_t^{(m)}(k) = \sum_{m=1}^M \log\{|Y_t(m)|\} \cdot \cos\left(\frac{k\left(m-\frac{1}{2}\right)\pi}{m}\right), \quad k=0, \dots, L \quad (12)$$

Where M is the number of triangular filters in the Mel filter bank, and L is the number for mel-scale cepstral coefficients. The obtained features are referred to as mel-scale cepstral coefficients, or MFCC. In this project, the value of N is 20 and L is 12. The energy within a frame is added to the 12 number of mel-scale cepstral coefficients.

5. Delta cepstrum—the first and second order time derivatives of 13 number of features which are the frame energy and mel-scale cepstral coefficients is computed.

$$y_t = \left\{ y_t^{(m)}(k), a_t, \Delta\{y_t^{(m)}(k)\}, \Delta\{a_t\}, \Delta^2\{y_t^{(m)}(k)\}, \Delta^2\{a_t\} \right\} \quad (13)$$

According to (13), the results from first and second derivatives are added as new features. Hence, a 39-dimensional MFCC features per frames is extracted from the digitized frog call sound signal. Each feature set consists of 12 mel cepstrum coefficient, one log energy and 13 first delta cepstrum and 13 second delta cepstrum.

Identification Using SVM Classifier

A user inputs frog call signal into server system using Android client during identification process. This frog call signal is pre-processed and its meaningful parameters are extracted. These parameters or features are used to generate a testing frog call template. Then, the testing template is categorized into one species or others based on the score of the testing template with the trained model. The process of computing similarity of two different features is known as feature matching. In this project, feature matching process is carried out by Support Vector Machine (SVM) [24].

Similar to other classifiers, SVM requires training data to build model which will be used as reference to predict a new set of data into one category or the others. In this project, there are 15 species of frogs need to be identified and recognized. Thus, multi-class SVM method is used to accomplish this task. There are several ways to construct SVM classifiers for more than two classes such as one-against-all, one-against-one, and DAGSVM methods. The

SVM multi-class classification implemented in the project is the one-against-all method. In this method, 15 SVM models is constructed as 15 species of frog are going to be identified. The i^{th} SVM is trained with all of the samples in the i^{th} class with positive labels, and negative labels for the rest of the samples.

SVM assign or predict the class of x by using the following decision function:

$$\text{class of } x \equiv \text{argmax}_{i=1, \dots, k} ((\omega^i)^T \phi(x) + b^i) \quad (14)$$

The largest value of the i^{th} decision function indicates that x is in i^{th} class.

In this project, 'libsvm_svmtrain' and 'libsvm_svmpredict' function in Scilab are used for SVM model training and identification, respectively. On the other hand, Matlab predefined function 'svmtrain' and 'svmpredict' are used for the same purpose.

The steps to build SVM model are listed as follows:

1. 20 samples of training data per each of the frog species is collected. This total up to 300 samples of frog call signal is used for training purpose.
2. Feature extraction process is then executed on the training samples.
3. The extracted features are then resized to 4096 feature point. This is done to ease the model training processes.
4. The i^{th} SVM model is built or trained with all of the samples in the i^{th} class with positive labels, and negative labels for samples in all other classes. The function 'libsvm_svmtrain' with polynomial kernel is used to train the SVM model.

The steps to predict or identify frog species using SVM are listed as follows:

1. A new sample of frog call data is sampled.
2. This sample is the testing data which will then proceed with feature extraction.
3. The extracted features are then resized to 4096 feature point. This is done to ease the identification processes.
4. 'libsvm_svmpredict' function is used to determine the matching rate for the testing data based on the trained SVM model.

D. Client-Server Communication using PHP

A Hypertext Preprocessor (PHP) script is written to facilitate client-server communication between Android device and Intel Atom board. This script allows a client to upload a recorded audio and returns the corresponding result from identification process. After the server is set up, the script will stand by and wait for receiving file from Android client. Once the file is successfully received, the script will invoke a Scilab script hence the identification process starts. The result from the identification process will send back to the client by the script. The flowchart of the PHP script is shown in Figure 7.

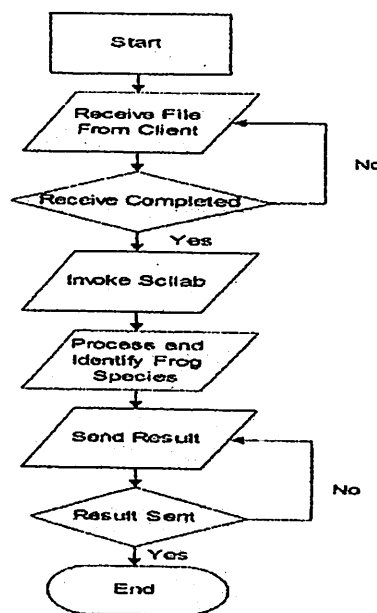


Fig. 9. Flowchart of PHP script which facilitates the client-server communication.

The steps of web server configuration are described as follows:

1. To start the server, the server machine i.e. Intel Atom Innovation Kit 3 is booted up and connected to a wireless router using a LAN cable.
2. In this project, XAMPP is used as web server application. It is one of the most robust, and offering cross platform. As the operating system Intel Atom board 3 is Fedora Linux, the Linux version of XAMPP is downloaded and installed.
3. After installation, SELinux in Fedora Linux operating system need to be deactivated. Only super user of the Linux Machine able to deactivate the SELinux. To log in as super user, the following command is used in the terminal:

→ su

Next, the system would request for super-user's password in the terminal. After successfully log in as super user, SELinux in Fedora Linux operating system is deactivated by using the following command line in a terminal:

→ setenforce 0

4. The web server only can be started once the SELinux is deactivated, by calling the following command line in the terminal.
→ /opt/lamp/lamp start
5. The written PHP script named 'v1.php' is then stored in the server folder path as follows:
→ /opt/lampp/htdocs/vup1
6. After step 5 is completed, the server is now ready to communicate with the client.

Fig. 10 illustrates the overall structure of the developed IFSIS.

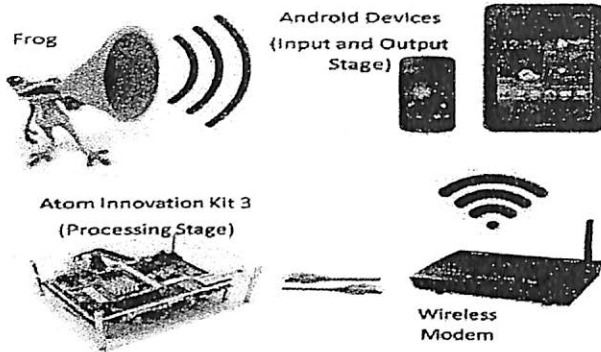


Fig. 10. Overall Structure of IFSIS

III. RESULTS AND DISCUSSION

In this experiment, 30 samples of frog call syllables are evaluated for each species. Training data which consists of 10 samples from each species are used to build the SVM model, while 20 samples from each species are randomly selected to test the system. Based on the results in Table 2, Scilab-IFSIS achieves 95.33% of accuracy which is considered high and this reaches the expected accuracy which has been set to 90%. The accuracy of Matlab-IFSIS is 95.67% which is slightly higher than the accuracy of Scilab-IFSIS.

Confusion matrix in Table 2 and 3 shows the true positives and false positives of the Scilab-IFSIS system on frog calls samples. It can be observed from the confusion

matrix that eight samples are tabulated under 'unknown' column. These samples of frog call are unidentified by the system. Besides that, there are six false positives are made by the system which are three from *Philautus Mjobergi*, and each from *Hylarana Labialis*, Genus *Ansonia*, and *Microhyla Butleri*. Out of 300 samples of frog call, a total number of 286 samples are correctly identified by the system based on the number of true positives.

A. Performances of IFSIS in term of processing time

The processing time of the identification processes are also recorded to evaluate the efficiency of IFSIS. In order to calculate the processing time, 15 frog call samples (1 sample for each species) with duration of 15 seconds were taken as the testing samples to record the processing time. The processing time is defined as the time taken to upload the frog call audio file samples by Android client to server until the result of the identification is displayed on the client. The difference of the processing time for each identification process is caused by the length of frog call syllables. Each species of frog exhibits unique syllable trend which is also different in length. The longer syllable of the frog call, the longer time is taken for the identification process. The results are tabulated in Table 4.

As observed from the above table, the processing time using Matlab-IFSIS is 24.00 sec. This performance is better compared to Scilab-IFSIS which is 27.17 sec in average. This system is considered efficient if the processing time is short and this achieves the expectation of the project. The expected processing time was set to 60 seconds or less.

Table 2. True positive and false positive of Matlab-IFSIS

Actual Class	sp	Predicted class															ACC (%)				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		Unknown			
1	20																			100	
2		18																		2	85
3			20																		100
4				20																	100
5					20																100
6						18														2	90
7							3	15												2	75
8									20												100
9										20											100
10											19									1	95
11												19									95
12													19								95
13														20							100
14															19						95
15																				20	100
		Mean Recognition Accuracy																	95.33		
Sp1: HylaranaGlandulosa		Sp6: RhacophorusAppendiculatus					Sp11: OdorranaHosii														
Sp2: PhrynoidisAspera		Sp7: HylaranaLabialis					Sp12: PolypedatesLeucomystax														
Sp3: MicrohylaHeymonsi		Sp8: PhilautusPetersi					Sp13: KaloulaBaleata														
Sp4: FejervaryaLimnocharis		Sp9: MicrohylaButleri					Sp14: PhilautusMjobergi														
Sp5: Genus Ansonia		Sp10: KaloulaPulchra					Sp15: DuttaphrynusMelanostictus														

Table 3. True positive and false positive of Scilab-IFSIS

Actual Class	sp	Predicted class															ACC (%)	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		Unknown
1	20																	100
2		17															3	85
3			20															100
4				20														100
5					20													100
6						18											2	90
7						3	15										2	75
8								20										100
9									20									100
10										19							1	95
11											19			1				95
12												19				1		95
13													20					100
14														19	1			95
15																20		100
		Mean Recognition Accuracy															95.33	
Sp1: HylaranaGlandulosa		Sp6: RhacophorusAppendiculatus					Sp11: OdorranaHosii											
Sp2: PhrynoidisAspera		Sp7: HylaranaLabialis					Sp12: PolypedatesLeucomystax											
Sp3: MicrohylaHeymonsi		Sp8: PhilautusPetersi					Sp13: KaloulaBaleata											
Sp4: FejervaryaLimnocharis		Sp9: MicrohylaButleri					Sp14: PhilautusMjobergi											
Sp5: Genus Ansonia		Sp10: KaloulaPulchra					Sp15: DuttaphrynusMelanostictus											

- Displaying the result on the android device.

Table 4. Processing time using Scilab-IFSIS and Matlab-IFSIS

Scientific name	Processing time (second)	
	Matlab-IFSIS	Scilab-IFSIS
HylaranaGlandulosa	20.34	23.02
PhrynoidisAspera	26.07	30.80
MicrohylaHeymonsi	21.33	24.12
FejervaryaLimnocharis	22.77	25.26
Genus Ansonia	22.97	25.19
RhacophorusAppendiculatus	24.21	27.85
HylaranaLabialis	25.83	29.66
PhilautusPetersi	22.99	25.19
MicrohylaButleri	23.05	26.33
KaloulaPulchra	27.80	31.58
OdorranaHosii	25.12	28.83
PolypedatesLeucomystax	25.61	29.13
KaloulaBaleata	23.04	25.71
PhilautusMjobergi	28.82	32.48
DuttaphrynusMelanostictus	20.06	22.44
Average time (second)	24.00	27.17

As a conclusion, in term of accuracy and processing time of the IFSIS, Matlab is slightly powerful than Scilab. However, Scilab-IFSIS is still a better option for this project in order to minimize the development cost for the whole system.

The android application GUI is shown as shown in Fig. 11. It consists of three main layout which facilitates user to perform the following procedures:

- Recording the frog call signal.
- Uploading the frog call audio file to Atom board (server).

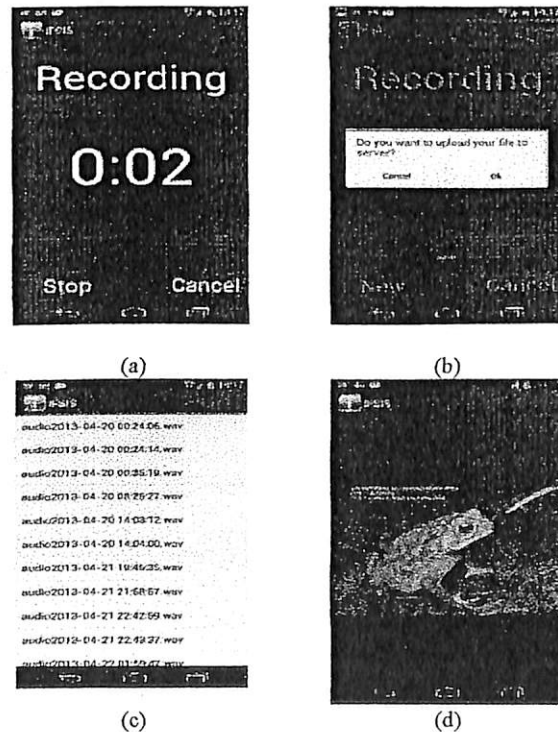


Fig. 11. User interface of the Android application for (a) recording sound signal, (b) alert dialog after users press “Stop” button, (c) audio file directory, and (d) result

IV. CONCLUSION

Intelligent Frog Species Identification system (IFSIS) has achieved high accuracy in identifying frog species correctly and efficiently. The identification of frog species which can be done using smartphone interface takes advantages of the android-device's graphic capabilities to make the system easy to use and portable. The system GUI assists user to record frog call signal, upload to-be-recognized frog call audio file to server, and read the identification result while the processing can be done via server on atom processor. From the experimental results, it shows that the proposed system is promising and can work as remote sensing for frog species identification. Thus, with the innovative, practical, reliable and affordable IFSIS, we may now suggest that "Now, everybody can be a physiological research expert in identifying frog species."

ACKNOWLEDGMENT

The authors would like to thank the financial support provided by Universiti Sains Malaysia Short Term Grant 60311048, Research University Grant 814161 and Research University-Post Graduate Grant Scheme 8046019 for this project.

REFERENCES

- [1] P. Kathryn, "Where have all the frogs and toads gone," *Bioscience*, vol. 40, no. 6, pp. 2-4, 1996.
- [2] T.J.C. Beebe, R.A. Griffiths, "The amphibian decline crisis: a watershed for conservation biology?," *Biological Conservation*, vol. 125, no.3, pp. 271-285, 2005.
- [3] C. Carey, "Kathryn Phillips: 1995, Tracking the Vanishing Frogs, Penguin Books," *Climatic Change*, vol. 37, no. 3, pp. 565-567, 1997.
- [4] N.H. Shubin, F. A. Jenkins, "An Early Jurassic jumping frog," *Nature* vol. 377, no. 6544, pp.49-52, 1995.
- [5] S.N. Stuart, J.S. Chanson, N.A. Cox, B.E. Young, A.S.L. Rodrigues, D.L. Fischman, R.W. Waller, "Status and Trends of Amphibian Declines and Extinctions Worldwide," *Science*, vol. 306, no. 5702, pp. 1783-1786, 2004.
- [6] X. Liu, R. Liu, L. Wei, H. Yang, K. Zhang, J. Liu, and R. Lai, "Two novel antimicrobial peptides from skin secretions of the frog, *Rana nigrovittata*," *Journal of Peptide Science*, vol. 17, pp. 68-73, 2010.
- [7] T.Y. Samgina, E.A. Vorontsov, V.A. Gorshkov, E. Hakalehto, O. Hanninen, R.A. Zubarev, A.T. Lebedev, "Composition and Antimicrobial Activity of the Skin Peptidome of Russian Brown Frog *Rana temporaria*," *Journal of Proteome Research*, vol. 11, no. 12, pp. 6213-6222, 2012.
- [8] A. Gomes, B. Giri, A. Saha, R. Mishra, S.C. Dasgupta, A. Debnath, A. Gomes, "Bioactive molecules from amphibian skin: their biological activities with reference to therapeutic potentials for possible drug development," *Indian journal of experimental biology*, vol. 45. No. 7, pp. 579-593, 2007.
- [9] C. Qian, T. Li, T.Y. Shen, L. Libertine-Garahan, J. Eckman, T. Biftu, S. Ip, "Epibatidine is a nicotinic analgesic," *European journal of pharmacology*, vol. 250, no. 3, pp.R13-R14, 1993.
- [10] S.C. Clayton, A.C. Regan, "A total synthesis of (\pm)-epibatidine," *Tetrahedron letters*, vol. 34, no. 46, pp. 7493-7496, 1993.
- [11] G.G. Yen, Q. Fu, "Automatic frog calls monitoring system: a machine learning approach," *International Journal of Computational Intelligence and Applications*, vol. 1, no. 02, pp.165-186, 2001.
- [12] A. Taylor, G. Watson, G. Gordon, H.M. Callum, "Monitoring frog communities: an application of machine learning," *Proceedings of Eighth Innovative Applications of Artificial Intelligence Conference*, Portland Oregon, 1996.
- [13] C.J. Huang, Y.J. Yang, D.X. Yang, Y.J. Chen, "Frog classification using machine learning techniques," *Expert Systems with Applications* vol. 36, no. 2, pp. 3737-3743, 2009.
- [14] H. Jaafar, and D. A. Ramli, "Automatic syllables segmentation for frog identification system," presented at the 2013 9th IEEE Int. Colloquium on Signal Processing and its App, 90-95.
- [15] A. Jansen, M. Healey, "Frog communities and wetland condition: relationships with grazing by domestic livestock along an Australian floodplain river," *Biological Conservation*, vol. 109, no. 2, pp. 207-219, 2003.
- [16] V. Dworak, J. Selbeck, K.H. Dammer, M. Hoffmann, A. Zarezadeh, C. Bobda, "Strategy for the Development of a Smart NDVI Camera System for Outdoor Plant Detection and Agricultural Embedded Systems," *Sensors*, vol. 13, no. 2, pp. 1523-1538, 2013.
- [17] L. Pei, R. Guinness, R. Chen, J. Liu, H. Kuusniemi, Y. Chen, L. Chen, J. Kaistinen, "Human Behavior Cognition Using Smartphone Sensors," *Sensors*, vol. 13, no. 2, pp. 1402-1424, 2013.
- [18] S. Shahrudin, J. Ibrahim, M.S. Anuar, "The Amphibian Fauna of Lata Bukit Hijau, Kedah, Malaysia," *Russian Journal of Herpetology*, vol. 18, no. 3, pp. 221-227, 2011.
- [19] D.R. Frost, "Amphibian Species of the world: an Online Reference," Version 5.5 (31 January 2011). American Museum of Natural History, New York, 2011.
- [20] D. Stowell, M.D. Plumbley, "Birdsong and C4DM: A survey of UK birdsong and machine recognition for music researchers," Tech. Rep. C4DM-TR-09-12, Queen Mary University of London, 2010.
- [21] C. Panagiotakis, G. Tziritas, "A speech/music discriminator based on RMS and zero-crossings," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 155-166, 2005.
- [22] A. Ghosal, R. Chakraborty, S. Haty, B.C. Dhara, S.K. Saha, "Speech/music classification using occurrence pattern of zer and ste," *Third International Symposium on Intelligent Information Technology Application*, 2009.
- [23] J.W. Picone, "Signal modeling techniques in speech recognition," *Proceedings of the IEEE*, vol. 81, no. 9, pp. 1215-1247, 1993.
- [24] V. Vapnik, *The nature of statistical learning theory*, Springer-Verlag, New-York, 1995.

MFCC Based Frog Identification System In Noisy Environment

Haryati Jaafar^{#1}, Dzati Athiar Ramli^{#2}, Shahriza Shahrudin^{#3}

^{#1,2} *Intelligent Biometric Group, School of Electrical and Electronic*

*Universiti Sains Malaysia, Engineering Campus
14300 Nibong Tebal, Pulau Pinang, Malaysia*

¹ haryati_jaafar@yahoo.com

² dzati@eng.usm.my

^{#3} *School of Pharmaceutical Sciences*

Universiti Sains Malaysia

11800 Minden, Pulau Pinang, Malaysia

³ shahriza18@usm.my

Abstract—Identification of frog sound is useful tool and competent in biological research and environmental monitoring. In contrast with traditional methods that not practical due to the time consuming, expensive or detrimental to the animal's welfare, this study proposes an automatic frog call identification system. 750 data species that recorded from Malaysia forest is used as data signals and have been corrupted by 10dB and 20dB noise to determine the performance of accuracy in noisy environment. MFCC parameter is employed as feature extraction. An analysis of signals for different number of MFCCs (8, 12, 15, 20 and 25) is presented and the results are provided using MFCC, Delta Coefficients (Δ MFCC) and Delta Delta Coefficients ($\Delta\Delta$ MFCC). Subsequently, kNN classifier is applied to evaluate the performance in the frog identification system. The results show the accuracy range from 84.67% to 85.78%, 61.33% to 68.89% and 59.33% to 67.33% in clean environment, 10dB and 20dB, respectively.

Index Terms—Frog identification system, mel frequency cepstrum coefficients, signal to noise ratio, kth nearest neighbour

I. INTRODUCTION

Frogs have been playing important roles to human society as bio-indicators of whole ecosystem due to their biphasic life and they can servers as a warning when something contaminating the environment that could be harmful to human society. Moreover, frogs can also secrete substances in medical field. The chemical compound in their skin include amines, alkaloid and peptides can play as poison, antibiotics and pain reliever that have significant potential application for human health [1]. Frogs produce a variety of sound to show their presence, mating ritual and defend their territory. Their sound can receive over varying distance that allows an obstructive detection of their existence [2].

Different techniques that relates feature extractions and classifiers have been studied and proposed to identify the frog species based on their sound automatically [3-6]. The most common features used is mel frequency cepstral coefficients (MFCC). This feature tends to uncorrelated, computationally efficient, has been resilience to noise and able to perform

results in higher accuracy [7]. For example, Lee *et al* [3] studied averaged MFCC as feature extraction and Linear discriminant analysis (LDA) was used as a classifier to identify 30 kinds of frog calls and 19 kinds of cricket calls. Another related study, Shih *et al* [4] used MFCC to transform 27 kinds of frog calls into feature vectors and combined two or three samples with different types of frogs to create mixed class samples before identifying the frog calls using support vector machine (SVM) classifier.

However, most of the earlier studies take directly from human speaker recognition tasks and used default value that commonly more suitable for human. For example, 8-12 MFCCs are commonly used in human speaker recognition since human can hear up to 20kHz and typically talk in 2 to 3kHz since 8-12 MFCC is suitable values due to the higher order MFCCs are less useful for human speaker. Nonetheless, the number and type of MFCCs to be used for improved classification for frog species has yet to be determine. In the frog world, most of the them can hear sounds up to 38kHz depend on their species [8]. So, in order to get higher identification accuracy, higher order MFCCs with information on syllables content can be revealed.

Another potential problem of most frog identification system is unsatisfactory performance in noisy environments. In real condition, the recordings may contain interference background noise for examples sound of running water, and the variable nature of weather i.e, sound of the wind or other animal calls in the background. Furthermore, their calls also depend on the corresponding changes in temperature and rainfall where this also contributes to the limitations in acquiring clean data for the experiment [9]. Therefore, effort is necessary to find the impact noise and other distortions in order to give the great performances in the frog identification system.

This paper determines the suitable methods of feature extraction which gives the highest performance accuracy for frog identification system. In addition, the experiment is conducted based on clean, 20dB and 10dB Signal to Noise Ratio (SNR) of data signal. The feature extraction based on 8,

12, 15, 20 and 25 MFCC is executed in this study. Also, delta (Δ MFCC) and delta delta coefficients ($\Delta\Delta$ MFCC) were calculated and compared in this study. Consequently, the k th nearest neighbour (kNN) is used as a classifier in the pattern matching process. The paper outline as follows. The methodology includes data acquisition, pre-processing, feature extraction and classifier presented in Section II. Section III describes the experimental results and discussion. Finally, conclusions are summarized in Section IV.

II. METHODOLOGY

A. Data Acquisition

The frog sound were recorded from locations around Baling and Kulim, Kedah, Malaysia using Sony Stereo IC Recorder ICD-AX412F supported with Sony electret condenser microphone in 32-bit wav files at a sampling frequency of 48kHz. The sounds were recorded next to a running stream from 8.00 pm to 12.00pm. The database consist 750 audio data which obtained from 15 species as listed in Table I and has been simulated with Additive White Gaussian Noise (AGWN).

TABLE I
FROG SPECIES DATABASE

Scientific name	Family	Common name
<i>Hylarana glandulosa</i>	Microhylidae	Rough sided frog
<i>Phrynoedis aspera</i>	Bufoidea	River toad
<i>Kaloula pulchra</i>	Microhylidae	Asian painted bullfrog
<i>Odorrana hossi</i>	Ranidae	Poisonous rock frog
<i>Polypedates leucomystax</i>	Rhacophoridae	Common tree frog
<i>Kaloula baleata</i>	Microhylidae	Flower pot toad
<i>Philautus mjobergi</i>	Rhacophoridae	Bubble-nest frog
<i>Microhyla heymonsii</i>	Microhylidae	Taiwan rice frog
<i>Hylarana labialis</i>	Ranidae	White-lipped Frog
<i>Philautus petersi</i>	Rhacophoridae	Kerangas bush frog
<i>Microhyla butleri</i>	Microhylidae	Painted chorus frog
<i>Rhacophorus appendiculatus</i>	Rhacophoridae	Friiled tree frog
<i>Fejervarya limnocharis</i>	Dicroglossidae	Grass frog
<i>Genus ansonia</i>	Bufoidea	Stream toad
<i>Duttaphrynus melanostictus</i>	Bufoidea	Black-spectacled toad

B. Syllables Segmentation

A syllable is basically a sound that frog produces with a side blow of air from the lungs. Compared to the human, frog syllables seem to be slightly less complex than human due to no-vowel-consonant and less intricate grammar [10]. The segmentation techniques described here is based on STE and STAZCR where the principle of the techniques is to determine the endpoint of syllable boundaries accurately where the endpoint is used to detect the syllable signal that has been segmented [9]. In the STE technique, energy of a call is another parameter for classifying voiced/unvoiced parts. The voiced part has high energy than unvoiced part due to the

periodicity. The STE function applied for the recorded signal is defined by the following expression;

$$E_n = \frac{1}{N} \sum_{m=1}^N [x(m)w(n-m)]^2 \quad (1)$$

Where E_n is the energy of the sample n of the signal, $x(m)$ is the discrete-time signal and $w[m]$ is a hamming window of size N .

On the other hands, STAZCR is often used as a part of the front-end processing in automatic speech recognition system. During the frog signal processing, the amplitude of the unvoiced part normally have much higher values and vice-versa. The ZCR is the rate at which signal changes from positive to negative and back and defined as;

$$Z_n = \frac{1}{2N} \sum_{m=1}^N |\text{sgn } x(m) - \text{sgn}[x(m-1)]| w(n-m) \quad (2)$$

where

$$\text{sgn}[x(m)] = \begin{cases} 1, & x(m) \geq 0 \\ -1, & x(m) < 0 \end{cases} \quad (3)$$

Fig. 1 shows an example of the syllable segmentation for a *Kaloula Baleata*'s call. The red lines in Fig. 1(a) shows the detected syllable is only happened when the frog call energy has been analysed in the high energy as shown in Fig. 1(b). In opposition, during the syllable concealed, the STAZCR in Fig. 1(c) is relatively low. Hence, the analysis of the STAZCR was further applied to enhance the accuracy of the syllable segmentation.

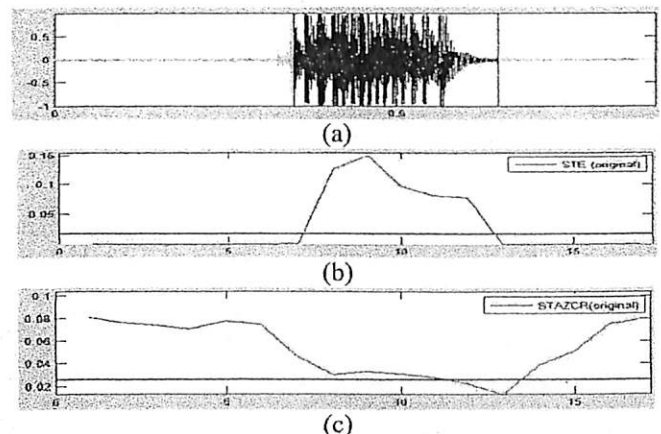


Fig. 1. An example of syllable segmentation

C. Pre-processing

The syllables characteristic of audio at higher frequency has smaller amplitude relative to low frequency syllables. Thus, a pre-emphasis of high frequencies needed to obtain similar amplitude for all syllables. Pre-emphasis is implemented by filtering the speech signal with a first order FIR filter whose transfer function in the z -domain is;

$$H(z) = 1 - \alpha z^{-1} \quad 0 \leq \alpha \leq 1 \quad (4)$$

In essence, a pre-emphasis filter in time domain is a first order high pass-filter;

$$X'(n) = x(n) - \alpha x(n-1) \quad (5)$$

where α is the pre-emphasis parameter. A typical value for α is specifying as 0.95. This has rise to a more than 20 dB amplification of the high frequency spectrum as stated by Ricotti [11] and Ramli *et al* [12].

In automatic speech recognition, Hamming window most frequently applied since high resolution is not required, whose impulse response is a raised cosine impulse and defined as;

$$W(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N} - 1\right) & n = 0, \dots, N-1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

D. Feature Extraction

MFCC is selected due to the features are robust to noise which is suitable to be implemented in outdoor environment that contain interference of background noises such as sound of wind, running water and other animal calls. The operation of this system is based on two types of filter which are linearly and logarithmically spaced and processes on the Fourier transform of $x_t(n)$: $X_t(e^{j\omega})$. The $X_t(e^{j\omega})$ is evaluated only for discrete number of ω values [13].

There have several steps in MFCC processing. The first step is computation of the Discrete Fourier Transform (DFT) of all frames of the signal. By considering $\omega = \frac{2\pi k}{N}$, the DFT of all frames of the signal, $x_t(k)$ is obtained as:

$$x_t(k) : X_t(e^{j\frac{2\pi k}{N}}), k = 0, \dots, N-1 \quad (8)$$

The computational complexity can also be reduced if the number of samples N is a power of 2. The result obtained after this step is called as signal's spectrum.

A filter bank processing is the second step in MFCC processing. Filter banks properly integrate a spectrum at defined frequency and spectral features are obtained after this process. The outputs of the filter bank are denoted as $Y_t(m)$, $1 \leq m \leq M$ where M is number of band-pass filters. In general, a set of 24 band-pass filter is used. Subsequently, computation of the log energy is the third step which computes the logarithm of the square magnitude of the filter banks outputs, $y_t(m)$. The final step for MFCC processing is mel frequency cepstrum computation that performs the inverse DFT on the logarithm of the magnitude of the filter bank output:

$$y_t^m(k) = \sum_{m=1}^M \log\{Y_t(m)\} \cdot \cos\left(k\left(m - \frac{1}{2}\right)\frac{\pi}{M}\right) \quad k = 0, \dots, L \quad (9)$$

In this study, the database of MFCC features consists of 750 set of MFCC features from 15 species with 50 syllables signal data in each species. The syllable is divided into overlapping frames of 256 samples with 50% overlap. Five different numbers of MFCCs i.e., 8, 12, 15, 20 and 25 with one log energy coefficient are applied. In addition, Δ MFCC

and $\Delta\Delta$ MFCC were calculated to measure temporal change in parameters and delta parameters [14]. The overall process of the MFCC is shown in Fig. 2.

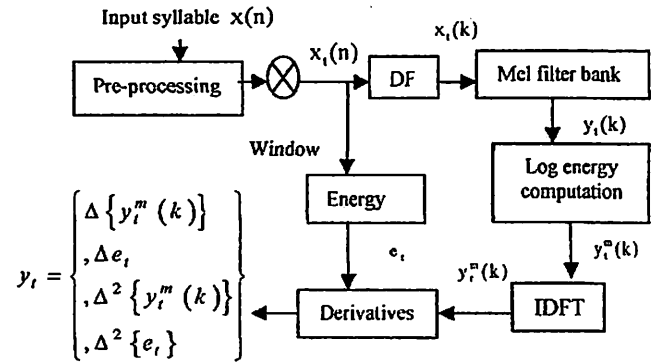


Fig 2. MFCC block diagram

E. Classification

The classifier obtained in this study is kNN where this classifier is a nonparametric classifier that employs lazy learning. This classifier is one of the most fundamental and simple classification techniques and has been practiced in various sound analysis [15]. Given a set of parameters, kNN classifier will find the nearest neighbour among training data by determining the minimum distance between query instance (testing set) and each training set. In the absence of prior knowledge, most kNN classifier use Euclidean distance with $d_E(x, y)$ to measure the distance or similarity between query instance and training set. The Euclidean distance is defined as;

$$d_E(x, y) = \sum_{i=1}^N \sqrt{(x_i^2 - y_i^2)} \quad (10)$$

where x, y are training and testing samples composed of feature N , respectively.

Each query instance will be compared with training set. A classification combination method that combines the selected training set and the query instance is then applied. The simplest classification method is the voting method where the class label of the query instance is determined based on the majority voting among the k nearest training samples category.

The k values represent an important role in kNN classification. Generally, k -values of kNN should be determined in advance and the best choice of k -values depends on the data. Normally, larger values of k will reduce the effect of noise on the classification but cause boundaries between classes less distinct [16].

III. EXPERIMENTAL RESULTS

The experiments are implemented using Matlab R2010(b) and have been test in Intel Core i5, 2.1GHz CPU, 2G RAM and Window 7 operating system. In this experiment, the data of 50 syllables have been extracted by MFCC. 20 syllables are

used for training and 30 for testing. The value of $k=3$ is used with Euclidean distance has been employed as classifier. The classification accuracy (C_A) is defined as;

$$C_A = \frac{N_C}{N_T} \times 100\% \quad (11)$$

where N_C is the number of syllables which are recognized correctly and N_T is the total number of test syllables.

The first experiments were based on different number of MFCCs and the results by MFCC, Δ MFCC and $\Delta\Delta$ MFCC is shown in Table II.

TABLE III
FROG IDENTIFICATION RESULTS BASED ON DIFFERENT NUMBER OF MFCC

SNR(dB)	Num.	MFCC (%)	Δ MFCC (%)	$\Delta\Delta$ MFCC (%)
Clean	8	82.44	82.44	82.44
	12	84.67	84.44	84.44
	15	85.78	85.33	85.33
	20	85.33	85.33	85.33
	25	85.33	85.33	79.56
20	8	68.89	68.82	68
	12	68.44	68.22	68.22
	15	68.44	68.44	68.44
	20	68	67.33	68.67
	25	67.56	67.33	61.33
10	8	65.76	65.76	65.76
	12	65.56	65.78	66
	15	66.67	65.78	65.55
	20	66.67	66	67.33
	25	62.67	65.56	59.33

Results in Table II show the accuracy tends to increase in clean recording with increasing number of MFCC and gives the greatest accuracy for 15 MFCC with 85.78%. However, the accuracy slightly decreases for 20 MFCC and tend to support to 85.33%.

The results show when using values for feature extraction of human speaker recognition which are 8-12 MFCC, the accuracies of frog identification system are slightly decrease. However, the accuracy is expected to be higher when the data signal is clean due to higher level MFCC are more likely robust to low-level noise [17].

For the second experiment, the results of MFCC, Δ MFCC and $\Delta\Delta$ MFCC at different levels of SNR value is shown in Fig. 3.

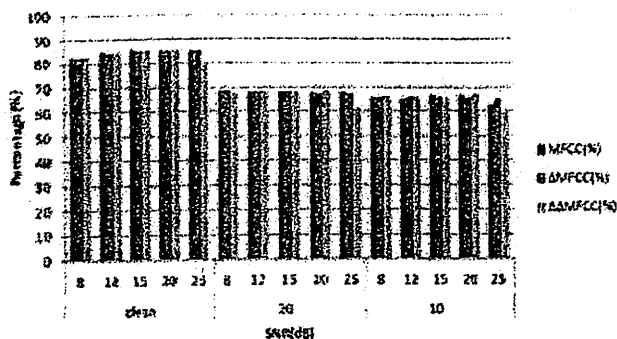


Fig. 3. Frog identification results based on different level of SNR

Fig. 3 shows the accuracy dropped less than 60% when level of SNR values increased. It was also observed that different level of SNR had different effects on the identification accuracy. For example, in 10dB (20 MFCC), the accuracy value of Δ MFCC gave much poorer identification accuracy but enhance with little different in identification frequency in $\Delta\Delta$ MFCC.

Based on the experiments, the best identification accuracy is obtain due to the considering if using 15 MFCC feature in clean recording condition. Table III lists the analytical results of the 15 MFCC in clean recording.

TABLE IIIII
FROG IDENTIFICATION RESULTS IN 15 MFCC

Scientific name	Syllables detected	Percentage (%)
<i>Hylarana glandulosa</i>	30	100
<i>Kaloula pulchra</i>	30	100
<i>Odorrana hossi</i>	20	66.67
<i>Polypedates leucomystax</i>	29	96.67
<i>Kaloula baleata</i>	30	100
<i>Philautus mjobergi</i>	29	96.67
<i>Duttaphrynus melanostictus</i>	30	100
<i>Phrynoidis aspera</i>	24	80
<i>Microhyla heymonsi</i>	17	70.83
<i>Fejervarya limnocharis</i>	25	83.33
<i>Genus ansonia</i>	29	96.67
<i>Rhacophorus appendiculatus</i>	30	100
<i>Hylarana labialis</i>	30	100
<i>Philautus petersi</i>	6	20
<i>Microhyla butleri</i>	27	90
Total	386	85.78

The results show that six species have successfully been identified with 100% accuracy. Nevertheless, a species such as *Philautus petersi* is clearly misclassified with only 20% of accuracy. It should be noted that some species is difficult to classify due to the recorded location and their syllables are keep changing when recorded.

IV. CONCLUSION

Frogs play a central role in many ecosystems since this species used to control the insect population, and as food source for many larger animals. Physiological research reported that certain frog species contain antimicrobial substances which is potentially and beneficial in overcoming certain health problem. As a result, there is an imperative need for an automated frog species identification to help people in physiological research in detecting and localizing certain frog species lived in Malaysia forest.

In this paper, different number of MFCC and level of SNR values were employed to automatic identification frog species based on their sound. The data were tested based on human speaker recognition tasks and higher order MFCC. Results suggest that the frog identification system can give a better result in higher level of identification accuracy, particularly when the feature extraction methods and classifiers are

modified to better suit frog sound. Accuracies of 59.33% to 85.78% were achieved from different number of MFCC and SNR levels indicating the excellent potential of MFCC and kNN classifier in the frog identification system.

[17] E.J.S. Fox, "Call-independent identification in birds," PhD. thesis, University of Western Australia, 2008.

ACKNOWLEDGMENT

The authors would like to thank the financial support provided by Universiti Sains Malaysia Short Term Grant, 304/PELECT/60311048, Research University Grant 814161 and Research University Grant 814098 for this project.

REFERENCES

- [1] C.R. Bevier, A. Sonnevend, J. Kolodziejek, N. Nowotny, P.F. Nielsen, and J.M. Conlon, "Purification and characterization of antimicrobial peptides from the skin secretions of the mink frog *Rana septentrionalis*," *Comp Biochem Physiol*, vol. 139, no. 1-3, pp. 31-8, 2004.
- [2] M.K. Obrist, G. Pavan, J. Sueur, K. Riede, D. Llusia, and R. Márquez, *Bioacoustic approaches in biodiversity inventories*, In: Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories, Abc Taxa, 2010, vol. 8, pp. 68-99.
- [3] C. H. Lee, C. H. Chou, C. C. Han, R. Z. Huang, "Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis," *Pattern Recognition Letters*, vol. 27, no. 2, pp. 93-101, 2006.
- [4] P.Y. Shih, M.T. Lin, J.F. Wang, and S.F. Lei, "Mixed-class identification for Taiwan frog vocalizations using support vector machines," National Cheng Kung University Department of Electrical Engineering, 2009.
- [5] C.J. Huang, Y.J. Yang, D.X. Yang, and Y.J. Chen, "Frog classification using machine learning techniques," *Expert Systems with Applications*, vol. 36, pp. 3737-3743, 2009.
- [6] N.C. Han, S.V. Muniandy, and J. Dayou, "Acoustic classification of Australian anurans based on hybrid spectral-entropy approach" *Journal of Applied Acoustic*, vol. 72, pp. 639-645, 2011.
- [7] P.J. Clemins, M. T. Johnson, K. M. Leong, and A. Savage, "Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations," *Journal of the Acoustical Society of America*, vol. 117, no. 2, pp. 956-963, 2005.
- [8] D. Roy, "Courtship in frogs: role of acoustic communication in amphibian courtship behavior," *Resonance*, 1996.
- [9] H. Jaafar, and D.A. Ramli, "Automatic syllables segmentation for frog identification system," in *2013 IEEE Int. Colloquium on Signal Processing and its App*, vol. 9, pp. 224-228, 2013.
- [10] D. Stowell, and M.D. Plumbley, "Birdsong and C4DM: A survey of UK birdsong and machine recognition for music researchers," Queen Mary University of London, C4DM-TR-09-12 Tech. Rep., 2010.
- [11] C. Bechetti, and L.R. Ricotti, *Speech recognition: Theory and C++ implementation*. Illustrated, reprint 1 Edn., Wiley: England, 1999, pp. 407.
- [12] D.A. Ramli, S.A. Samad and A. Hussain, "A multibiometric speaker authentication system with SVM audio reliability indicator," *IAENG International Journal of Computer Science (IJCS-Special Issues)*, vol. 36, no. 4, pp. 313-321, 2009.
- [13] M.H. Hasan, H. Jaafar, and D.A. Ramli. "Evaluation on score reliability for biometric speaker authentication system," *Journal of Computer Sciences*, vol. 8, no.9, pp. 1554-1563, 2012.
- [14] P. Somervuo, A. Harma, and S. Fagerlund, "Parametric representations of bird sounds for automatic species recognition," *IEEE Transactions Audio, Speech, and Language Processing*, vol. 14, pp. 2252-2263, 2006.
- [15] P. Parveen, "Face Recognition Using Multiple Classifiers," *IEEE International Conference on Tools with Artificial Intelligence*, 2006, pp. 179-186.
- [16] O.C. Ai, M. Hariharan, S. Yaacob and L.S. Chee, "Classification of speech dysfluencies with MFCC and LPCC features," *Journal Expert Systems with Applications: An International Journal*, vol. 39, no. 2, pp. 2157-2165, 2012.

Comparative Study on Feature, Score and Decision Level Fusion Schemes for Robust Multibiometric Systems

- Chia Chin Lip
- , Dzati Athiar Ramli

Abstract

Multibiometric system employs two or more behavioral or physical information from a person's traits for the verification and identification processes. Many researches have proved that multibiometric system can improve the performances of single biometric system. In this study, three types of fusion levels i.e feature level fusion, score level fusion and decision level fusion have been tested. Feature level fusion involves feature concatenation of the features from two modalities before the pattern matching process while score level fusion is executed by calculating the mean score from both biometrics scores produced after the pattern matching. Finally, for the decision level fusion, the logic AND and OR are performed on the final decision of the two modalities. In this study, speech signal is used as a biometric trait to the biometric verification system while lipreading image is used as a second modality to assist the performance of the single modal system. For speech signal, Mel Frequency Cepstral Coefficient (MFCC) is used as speech features while region of interest (ROI) of lipreading is used as visual features. Consequently, support vector machine (SVM) is executed as classifier. Performances of the systems for each fusion level is compared based on accuracy percentage and Receiver Operation Characteristic (ROC) curve by plotting Genuine Acceptance Rate (GAR) versus False Acceptance Rate (FAR). Experimental results show that score level fusion performance is the most outstanding method followed by feature level fusion and finally the decision level fusion. The accuracy percentages using 20 training data are observed as 99.9488%, 99.7534% and 99.6639% for the score level fusion, feature level fusion and decision level fusion, respectively.

Keywords

Multi-modal speech signal fusion level verification biometrics

References

1. Jain, A.K., Ross, A., Prabhakar, S.: *An Introduction to Biometric Recognition*. *IEEE Transactions on Circuits and Systems for Video Technology* 14(1), 4–20 (2004)[CrossRef](http://dx.doi.org/10.1109/TCSVT.2003.818349) (<http://dx.doi.org/10.1109/TCSVT.2003.818349>)
2. Reynolds, D.A.: *An Overview of Automatic Speaker Recognition Technology*. *IEEE Transactions on Acoustic, Speech and Signal Processing* 4, 4072–4075 (2002)
3. Ramli, D.A., Samad, S.A., Hussain, A.: *A Multibiometric Speaker Authentication System with SVM Audio Reliability Indicator*. *IAENG International Journal of Computer Science* 36(4), 313–321 (2008)
4. Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A.: *Recent Advances in Visual and Infrared Face Recognition—A Review*. *Computer Vision and Image Understanding* 97(1), 103–135 (2005)[CrossRef](http://dx.doi.org/10.1016/j.cviu.2004.04.001) (<http://dx.doi.org/10.1016/j.cviu.2004.04.001>)
5. Marcialis, G.L., Roli, F.: *Fingerprint Verification by Fusion of Optical and Capacitive Sensors*. *Pattern Recognition Letters* 25, 1315–1322 (2004)[CrossRef](http://dx.doi.org/10.1016/j.patrec.2004.05.011) (<http://dx.doi.org/10.1016/j.patrec.2004.05.011>)
6. Yong, F.A., Xiao, Y.J., Hau, S.W.: *Face and Palmprint Feature Level Fusion for Single Sample Biometrics Recognition*. *Neurocomputing* 70, 1582–1586 (2007)[CrossRef](http://dx.doi.org/10.1016/j.neucom.2006.08.009) (<http://dx.doi.org/10.1016/j.neucom.2006.08.009>)
7. Zhou, X., Bharu, B.: *Feature Fusion of Side Face and Gait for Video based Human Identification*. *Pattern Recognition* 41, 778–795 (2008)[MATH](http://www.emis.de/MATH-item?1132.68671) (<http://www.emis.de/MATH-item?1132.68671>) [CrossRef](http://dx.doi.org/10.1016/j.patcog.2007.06.019) (<http://dx.doi.org/10.1016/j.patcog.2007.06.019>)
8. Jain, A., Nandakumar, K., Ross, A.: *Score Normalization in Multimodal Biometric Systems*. *Pattern Recognition* 38, 2270–2285 (2005)[CrossRef](http://dx.doi.org/10.1016/j.patcog.2005.01.012) (<http://dx.doi.org/10.1016/j.patcog.2005.01.012>)
9. Cetingul, H.E., Erzän, E., Yemez, Y., Tekalp, A.M.: *Multimodal Speaker/Speech Recognition using Lip Motion, Lip Texture and Audio*. *Signal Processing* 86, 3549–3558 (2006)[CrossRef](http://dx.doi.org/10.1016/j.sigpro.2006.02.045) (<http://dx.doi.org/10.1016/j.sigpro.2006.02.045>)

Preliminary Study on Classification of Apraxia Speech using Support Vector Machine

Aini Hafizah Mohd Saod, Dzati Athiar Ramli

Intelligent Biometric Research Group (IBG),
 School of Electrical & Electronic Engineering, USM Engineering Campus,
 Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia.

Abstract: Apraxia is one of speech disorder problems which lead to the difficulty of the movement of muscle for speech production that encountered among children. Research shows that Apraxia children are unable to compete among their peers and this cause to the interpersonal problems. Early detection of Apraxia for speech therapy is important and becomes one of the solutions to this problem. In this study, an Automatic Speech Recognition (ASR) system for early detection of Apraxia of speech is suggested. A database of normal and Apraxia speech samples is collected for the modeling and testing process. Mel Frequency Cepstral Coefficient (MFCC) and Linear Prediction Coding (LPC) are extracted as speech features while for the classification of normal and Apraxia signals; a Support Vector Machine (SVM) is used as classifier. Four types of systems for validation are developed namely data dependent-speaker dependent, data independent-speaker dependent, data dependent-speaker independent and data independent-speaker independent systems. Performance evaluations are measured using percentage of accuracy and Mean Squared Error (MSE). Experimental results prove that the ASR system can be a viable system for early detection of Apraxia. Accuracy percentages for data dependent-speaker independent system and data independent-speaker independent system are observed as 99.86% and 81.88% using the MFCC method while 99.31% and 78.54% for the LPC method.

Key words: Apraxia, speech disorder, speech recognition system, speech therapy.

INTRODUCTION

Speech disorders refer to the communication problems which correspond to the areas of oral motor function. These disorders are related to difficulties in producing speech sounds which lead to deterioration of the quality of the produced speech. It ranges from making simple sound deviation to the lack of ability in using oral motor mechanism for functional speech (Tian-Swee *et al.*, 2007). According to Shriberg *et al.*, (2006), there are four main categories of speech disorders, which are omission, addition, distortion and substitution. Omission occurs when the children drop out sounds or syllables whereas addition occurs when the children add an extra sound or syllable to a word. Subsequently, distortion is a condition when the children pronounce a word correctly, but one of the sounds involved is not correct. Lastly, substitution is the case when the children consistently substitute one sound for another. Researches from previous works have indicated that children with speech disorder problems are found less intelligible than normal children of the same age.

Apraxia of speech is one of speech disorder problems focused on speech pathology research. According to Ogar *et al.*, (2005), Apraxia is a neurogenic speech disorder resulting from disability of the sensor motor commands to program the movement of muscles for speech production. It can be divided into three main categories, which are Apraxia among children, stroke-associated Apraxia and stress-induced Apraxia. However, Apraxia among children is different from Apraxia in adults which due to strokes, head injuries or stress as the children were born with the disorder. It is the least common form of Apraxia, which constitutes 15% of all reported cases, but it is the most difficult to treat. From Bowen (2009), there are several characteristics of Apraxia speech. For instance, the utterances are not clearly spoken, although there may be some exceptions such as for a very clear word like 'no'. The speech errors affect vowels and sometimes the sound that are used in different words are the same sound. Apart from that, unusual intonation, pausing and stress patterns and pronunciation of the same word in several different ways are also observed. Examples of the Apraxia words regarding the characteristics in Apraxia speech is tabulated in Table 1 as reported by Bowen (2009).

Table 1: Characteristics in Apraxia speech.

Characteristics	Examples
Words not clearly spoken	'ball' → 'or', 'bor' 'knee' → 'dee'
Speech errors affect vowels	'milk' → 'mih', 'muh', 'meh'
Inconsistency in pronunciations	'me' → 'ee', 'dee', 'bee', 'mee'
Different words produce the same sound	'happy', 'puppy' → 'hah hee' 'people', 'purple' → 'pur pur'
Unusual intonation pattern	'play' → 'psay'
Unusual pausing pattern	'top' → 'to..pp' 'arm' → 'ar..mm'
Unusual stress pattern	'pie' → 'pa..ee'

Corresponding Author: Intelligent Biometric Research Group (IBG), School of Electrical and Electronic Engineering, USM Engineering Campus, University Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia.

E-mail: dzati@eng.usm.my

Speech therapy session is one of the treatments that can be given to the Apraxia children. It is a clinical area concerned with disorder of human communication (Cleuren, 2003). Children with Apraxia must be taught with the skills required to program and sequence the movements for speech and they are required to practice the skills deliberately (Kumin, 2006). Early detection of Apraxia for speech therapy is imperative in order to avoid children from struggling and using incorrect oral communication during their learning years. Besides, it can shorten the duration of speech therapy treatment. According to Flipsen (2006), a general formula used by speech therapist to calculate the expected conversational intelligibility levels of preschoolers talking to unfamiliar people is given as equation (1),

$$\text{Age in years} / 4 \times 100 \% = \% \text{ understood by stranger} \quad (1)$$

From this formula, it can be calculated that children of aged 1, 2, 3 and 4 are 25%, 50%, 75% and 100% intelligible to strangers, respectively. According to Gordon-Brannan *et al.* (2000), children of four years with speech intelligibility score less than 66% should be considered as having speech disorders and become the candidates for speech therapy and treatment. Thus, for normal children, unfamiliar listeners should be able to understand at least 66% of the spoken by four year old children.

Normally, any speech impairments can only be traced by experts or therapists. However, as an alternative technique, Automatic Speech Recognition (ASR) system can be used for early detection of Apraxia. Various techniques of ASR for diagnosing speech disorders have been introduced over the years. One example of the techniques is the implementation of adaptive signal processing to compare the speech signal of children with speech disorders and normal children using feedback mechanism technique (Gudi *et al.*, 2010). The resultant curve fitted is used to tune the constant values of speech disordered children by calculating the deviation values using correlation technique. In another study, implementation of ASR utilizing Hidden Markov Model technique has been done by Tian-Swee *et al.*, (2007). The speech pattern of normal and speech disordered children are used to train the model for classifying the problem of speech disorder. The developed system also provides text-to-speech system to guide the user during diagnosis process.

Another technique is the use of data mining technique in order to obtain the best results of speech disorders diagnosis in condition of maximum efficiency. Data mining involves analysis of mass data to produce a particular enumeration of data patterns. The proposed system gives information that enables the implementation of personalized therapy programs according to the characteristics of the speech disordered children (Danubianu, 2009). In a research reported by Georgopoulos & Malandraki (2005), Fuzzy Cognitive Map (FCM) model is used to develop a diagnosis system for Apraxia. The implementation of Artificial Intelligence (AI) technique such as FCM model can integrate with decision making process in order to differentiate different types of Apraxia speech. Meanwhile, another diagnostic assessment by using ASR methods for Apraxia speech has also been described in Hosom *et al.*, (2007). Training of Artificial Neural Network (ANN) is performed in order to evaluate the speech production in this study. The main objective of this study is to evaluate the feasibility of classifying normal and apraxia words using SVM classifier. Experiments are also conducted so as to seek out the suitable apraxia words and feature extraction methods for the ASR system application.

MATERIALS AND METHODS

In this research, the apraxia and normal speech signal data is obtained by recording and saving the audio samples into WAV format files using digital audio editor. The digitized audio signals are 16 bit, monophonic and the sampling rate is 44.1 kHz, 705 kbps. Figure 1 shows the platform of Goldwave digital audio editor for recording and editing the speech signal samples. Apparently, the figure shows a recorded of one sentence with 20 repetitions of speech signals. For data collection, the recording process is carried out in three different sessions. For each session, the speaker performed 20 repetitions of 20 different normal and Apraxia sentences. A total of 3600 audio samples from all session are collected for the system development.

Feature Extraction and Classification:

Feature extraction is a process of transforming input data into reduced representation set of informative features which involves the speech signal processing method (Lim Sin *et al.*, 2009). For the purpose of speech recognition, speech samples are divided into frames and features are extracted from each frame. During feature extraction, speech features are extracted into a sequence of feature vectors in order to be classified in classification stage (Chulhee *et al.*, 2003). Before the feature extraction process taking place, 3 basic steps of pre-processing i.e. pre-emphasis, framing and windowing are executed as shown in Figure 2.

The first feature extraction used in this research is Mel Frequency Cepstral Coefficient (MFCC). All the parameters and steps involved for this method is illustrated as in Figure 3. This procedure produces 12 mel cepstrum coefficients, one log energy coefficient, their delta and delta-delta coefficients for each frame as features (Ramli *et al.*, 2011; Ramli *et al.*, 2010). Linear Prediction Coding (LPC) is the second method for feature extrac-

tion that has been employed in this study. It can be described as a linear combination of speech samples where unique set of predictor is determined by minimizing the sum of the squared differences between actual speech samples and predicted speech samples (Kondo, 2000; Kesarkar, 2003). The implementation of LPC processing is illustrated in Figure 4. The parameters for the sampling frequency, pre-emphasis, frame length and window length used at each stage of the processing are also indicated in this figure. For this method, each feature set consists of 14 cepstrum coefficients per frame.

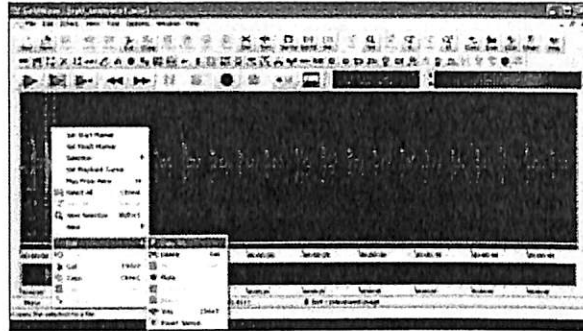


Fig. 1: Platform of digital audio editor.

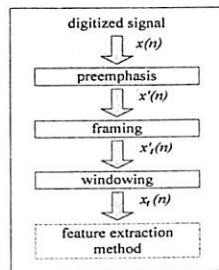


Fig. 2: Speech signal pre-processing process.

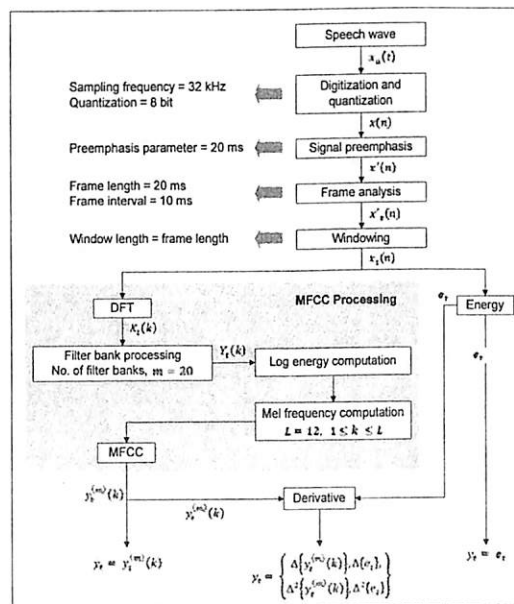


Fig. 3: Implementation of MFCC processing.

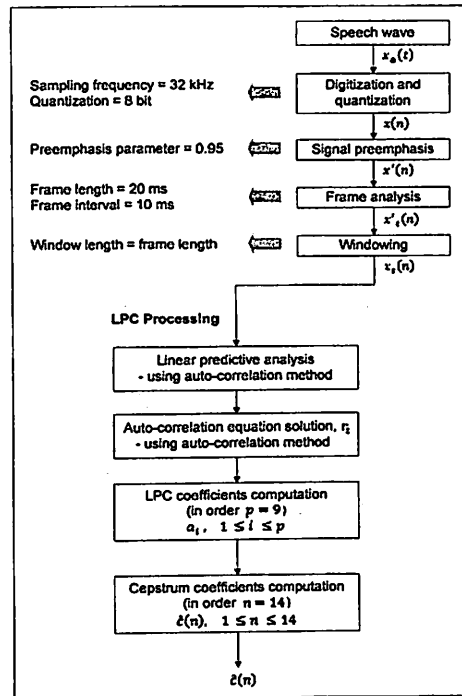


Fig. 4: Implementation of LPC processing.

For the purpose of speech classification, Support Vector Machine (SVM) is used as a classifier. It is a learning method used for classification with the basic idea is to find a hyperplane in order to separate the d-dimensional data into two classes. In the original form, SVM is a training algorithm for linear classification. However, since example data is often linearly separable, SVM introduces the notion of a kernel induced feature space to transform the data into a higher dimensional space where the data is separable (Boswell, 2002). SVM also tunes the capacity of the classification function by maximizing the margin between the training patterns and the decision boundary. According to Gunn (1998) and Vapnik (1995), the solution of linearly separable case can be done by considering a problem of separating the set of training vectors of two separate classes as in equation (1) and (2)

$$D = \{(x^1 y^1), \dots, (x^L y^L)\}, \quad x \in \{-1, 1\} \tag{1}$$

with a hyperplane

$$\langle w, x \rangle + b = 0 \tag{2}$$

where w and b characterize the direction and position in space and w is normal to the plane. For each direction, w the hyperplane has the same distance from the nearest points where each class of the margin is twice this distance. The hyperplane is optimal when it is able to separate the set of vectors from both classes without error and the distance between the closest vectors to the hyperplane is maximal. For non-linear boundary case, the data sets of input space are mapped into higher dimensional feature space, by constructing an optimal separating hyperplane in the new higher dimensional space as described in Boswell (2002). In practice, the mapping is achieved by replacing the value of dot products between the data points in input space with the resultant value when the same dot product is performed in the feature space.

ASR System Development:

The development of the ASR system consists of four types of approaches, which are data dependent-speaker dependent system, data dependent-speaker independent system, data independent-speaker dependent system and data independent-speaker independent system. For each system, two influenced factors are manipulated in term of word and speaker dependability. The data dependent means the testing data uses the same word as in the model data while data independent denotes the testing data uses different word from the model data.

Consequently, speaker dependent signifies the speaker for the modeling and testing process is the same person while speaker independent involves different person for the modeling and testing process.

In this study, the experiment is conducted through two phases as illustrated in Figure 5. The main intention of the first phase is for validation. All the four systems are evaluated using MFCC features in this phase and a set of words i.e. down, happy and milk are used for modeling and testing. While, for the second phase, only the most applicable systems toward real application are evaluated. For this experiment, a new set of words i.e. ball, down, happy, milk, pie and play are employed for modeling and testing. Apart from that, two types of features i.e. MFCC and LPC are experimented. In this phase, the words and feature that give consistent performance are observed.

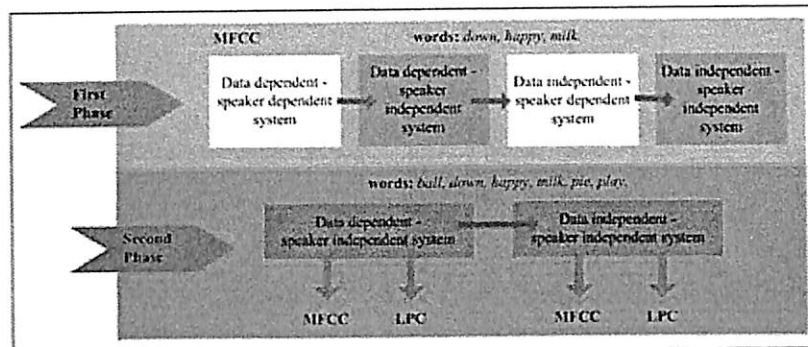


Fig. 5: Summary of system development.

After recording session, the collected audio data files are edited using the editor software to acquire the normal and apraxia data samples. A total of 4680 words are obtained from the editing process. The developed system is fully software based, implemented using Matlab programming. Three training data is used to train the ASR system and the system is executed through the following procedures:

1. Read the WAV files of model mfcc folder. Perform data collection using Hamming window with sampling frequency of 32 kHz to the 240 N data samples. Store the processing data in *datasp_all_data* array.
2. Load the *datasp_all_data* array. Extract model data using MFCC and LPC methods. Store the extracted features in *wordsp_mfcc_3_data* array and *wordsp_lpc_3_data* array.
3. Load the *datasp_test_data* array. Extract test data using MFCC and LPC methods. Store the MFCC features in *wordsp_mfcc_test_data* array and LPC features in *wordsp_lpc_test_data* array.
4. Load the *wordsp_mfcc_3_data* array and *wordsp_lpc_3_data* array. Set the number of model data as 3. Train the model features using SVM method. Store the training model in *wordsp_mfcc_model_3_data* array and *wordsp_lpc_model_3_data* array, accordingly.
5. Load the *wordsp_mfcc_test_data* array with its corresponding *wordsp_mfcc_model_3_data* array or *wordsp_lpc_test_data* array with its corresponding *wordsp_lpc_model_3_data* array. Set the number of training data as 20. Perform SVM model to match the data pattern between the two arrays. Evaluate the matching score using percentage of accuracy and MSE parameters. Store the data results in *wordsp_result_mfcc_data* array and *wordsp_result_lpc_data* array, accordingly. Display the evaluation parameters.

RESULTS AND DISCUSSIONS

Two types of speech signal patterns have been obtained which are normal and Apraxia along with their spectrograms. The spectrograms are used to show the spectral density of a signal varies with the time which indicated by the darkness of the plot of the frequency analysis. During the period of voiced sound, frequency energy of spectral is seen in the spectrogram while in silence period, the spectral cannot be observed. The speech signal pattern and its spectrogram of normal and Apraxia for word *down* and word *happy* are depicted in Figure 6 and 7, respectively. The dissimilarity between normal and apraxia for word *down* is the characteristic of unusual pausing pattern for Apraxia word as shown in Figure 6. For word *happy*, there is an obvious difference between the normal and Apraxia utterances in the two syllables of word *happy* as shown in the following Figure 7.

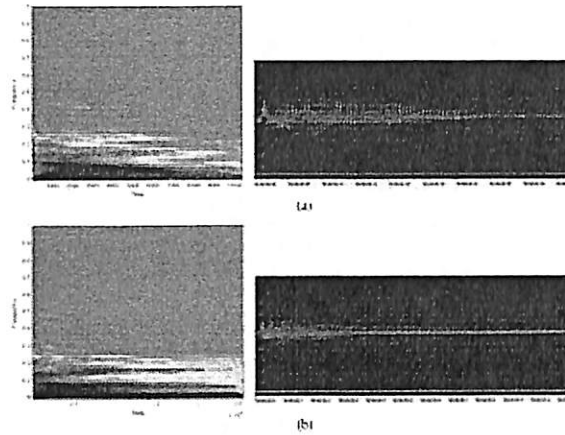


Fig. 6: Speech signal pattern and spectrogram of (a) normal word *down* and (b) Apraxia word *down*.

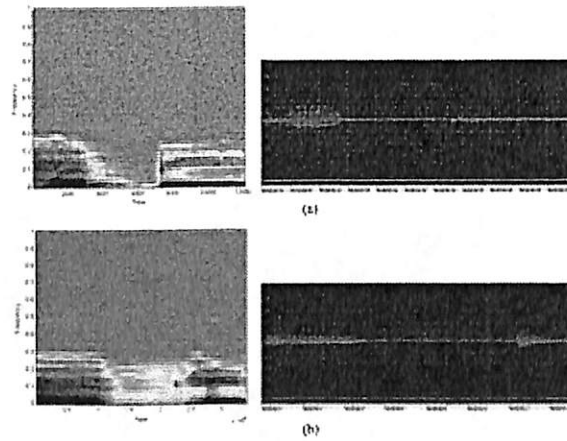


Fig. 7: Speech signal pattern and spectrogram of (a) normal word *happy* and (b) Apraxia word *happy*.

Whereas, Figure 8 displays the representation of normal word *milk* and Apraxia word *milk*. When comparing the speech signals and spectrograms, the utterance of word *milk* affects the vowel resulting continuous lines at the end of the speech as in the Apraxia sample.

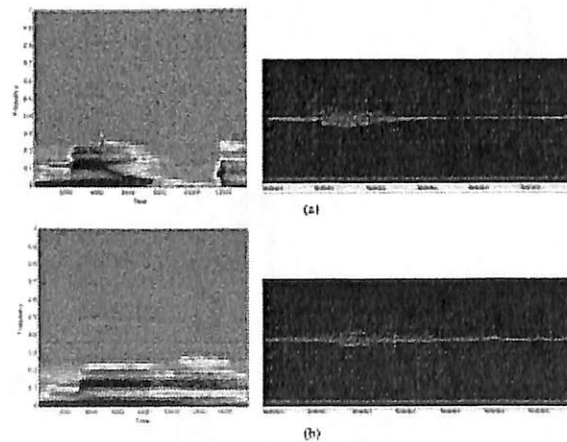


Fig. 8: Speech signal pattern and spectrogram of (a) normal word *milk* and (b) Apraxia word *milk*.

The word *ball* is utilized in the system since the word is not clearly spoken by the Apraxia candidates. The speech signal pattern and its spectrogram of normal word *ball* and Apraxia word *ball* are given in Figure 9. The difference between the normal and Apraxia utterance can be seen in the figure where the Apraxia sample holds longer at the end of the speech.

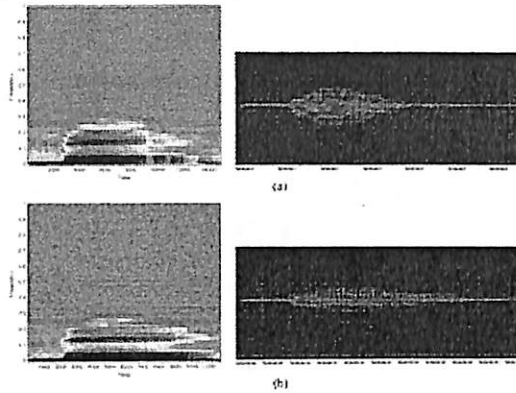


Fig. 9: Speech signal pattern and spectrogram of (a) normal word *ball* and (b) Apraxia word *ball*.

The following Figure 10 shows the speech signal pattern and corresponding spectrogram of normal word *pie* and Apraxia word *pie*, respectively. The word *pie* is used in the system due to its characteristic of unusual stress pattern for Apraxia of speech problem. From the figure, it can be seen that the Apraxia *pie* sample is displayed as two absurd syllables.

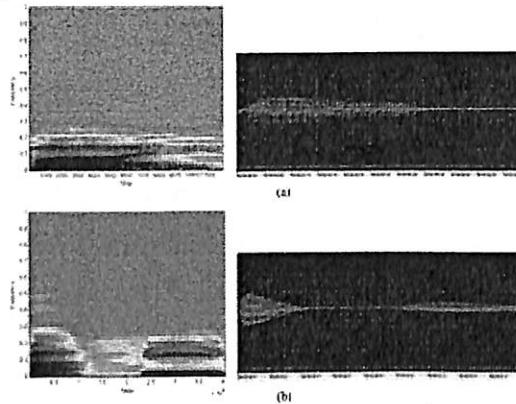


Fig. 10: Speech signal pattern and spectrogram of (a) normal word *pie* and (b) Apraxia word *pie*.

Meanwhile, the speech signal pattern and spectrogram of normal word *play* and Apraxia word *play* are displayed in Figure 11. The Apraxia characteristic of the word *play* is unusual intonation pattern, where the Apraxia word *play* is spoken like a single absurd syllable as shown in the corresponding figure.

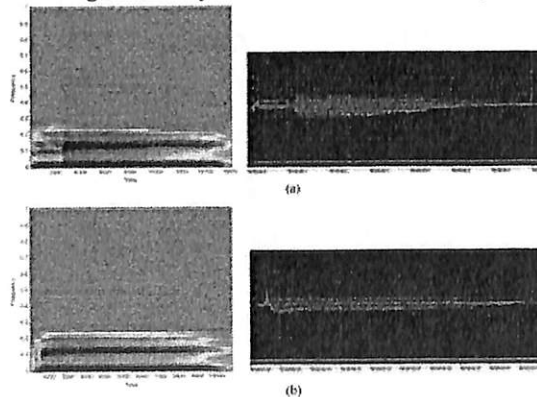


Fig. 11: Speech signal pattern and spectrogram of (a) normal word *play* and (b) Apraxia word *play*.

Performance of the First Phase System Development:

From the experimental result, the performance of data dependent-speaker dependent system gives the average of 99.86% accuracy and 0.01 MSE. Except for the word *happy* from speaker sp1, the systems are able to achieve 100% of accuracy as shown in Figure 12. This may be due to speech signals produced by this speaker are shaky and sometimes unstable.

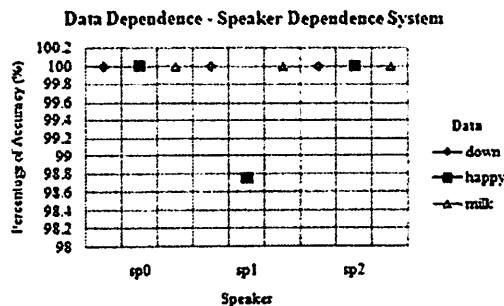


Fig. 12: Performance of data dependent - speaker dependent system.

For the data dependent-speaker independent system, the average accuracy is 81.20% and the average MSE is 0.75. The testing speech samples are taken from different speakers, independently from the speaker of the model samples. Figure 13 shows that among three types of words, the *happy* words are indicated as lower percentage of accuracy than the other words since the word has two syllables sound that may affect the accuracy consistency. Besides, the accuracy is also decreased for the input speech samples of speaker sp2 due to the speaker sp2 has the lowest pitch sound compared to the speaker sp0 and speaker sp1.

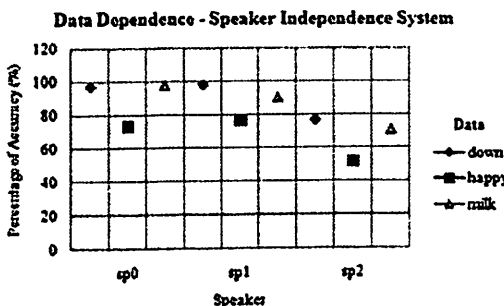


Fig. 13: Performance of data dependent - speaker independent.

For data independent-speaker dependent system, the input of word samples is different from the model samples. Compared to the first system, this system reaches the average accuracy of 95.28% and average MSE of 0.19. Otherwise, the accuracy is also decreased for the input speech samples of speaker sp1 when comparing with the other speakers as in the second approach as shown in Figure 14.

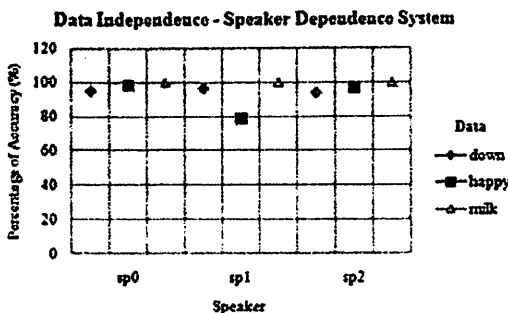


Fig. 14: Performance of data independent - speaker dependent system.

For the last system in the first phase of ASR system, the performance of data independent-speaker independent system is the lowest compared to the other three systems due to the effect of system complexity. The average accuracy is 71.04% with average MSE of 1.16. The testing speech samples are taken from different speakers, which independent from the speaker of the model samples. Besides, the input speech samples are also different from the model samples. The following Figure 15 indicates that among three types of words, the *happy* words are indicated as lower percentage of accuracy than the other words since the word has two syllables sound that can deteriorate the accuracy consistency.

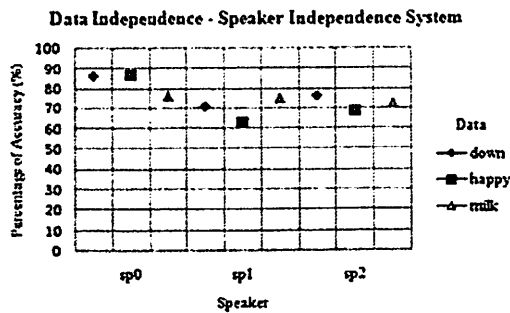


Fig. 15: Performance of data independent - speaker independent system.

Performance of the Second Phase System Development:

The second phase of ASR system is implemented using different methods of feature extractions for performance comparison. Besides, this phase is also conducted to seek out the consistent word samples that are preferable to be used in real application of ASR system. The numbers of words are increased to six types of words. Two systems, which are data dependent-speaker independent and data independent-speaker independent, are experimented. These systems are selected since the manipulation of independent speaker is the most realistic approach to be applied for the real application.

Performances of data dependent-speaker independent system using MFCC and LPC methods are shown in Figure 16. From the results, the word *ball* and word *play* give lower result at the range below 60% of accuracy for both MFCC and LPC methods. The reason of this outcome is probably due to the Apraxia characteristic for these words are not clearly spoken and contain unusual intonation pattern that affect the data accuracy. The overall performances of the system based on words and feature extraction methods can also be found in Table 1. In general, yellow color indicates good performances while low performances are specified by grey color.

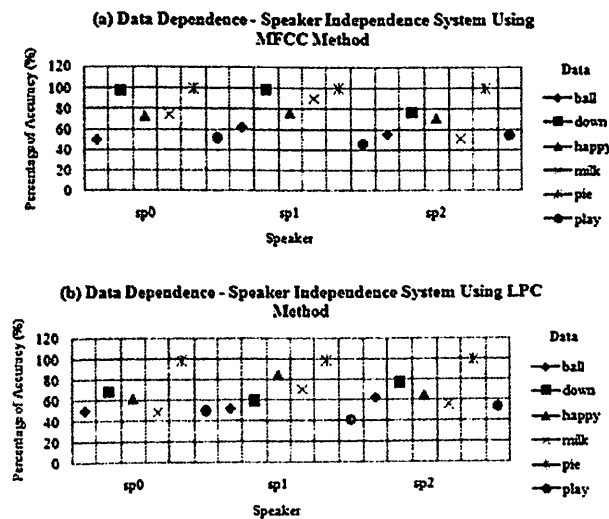


Fig. 16: Performance of data dependent - speaker independent system in the second phase of ASR system using (a) MFCC method (b) LPC method.

Consequently, Figure 17 presents the performances of data independent-speaker independent system using the MFCC and LPC methods. The word *play* and word *happy* give poor result compared to the other data samples with accuracy below 60% for both MFCC and LPC methods. The Apraxia characteristic of unusual pattern for the word *play* is probably one of the factors that will affect the data accuracy. On the contrary, the word *happy* which is observed as lower performance is due to the word that has two syllables sound that may stray the accuracy consistency. The overall performances for this system can also be referred in Table 2 which yellow and grey colors indicate good and low performances, respectively.

Table 1: Result of data dependent - speaker independent approach.

Speaker		Word	Percentage of accuracy (%)		Mean squared error (MSE)	
Input	Model		MFCC	LPC	MFCC	LPC
sp1, sp2	sp0	ball	50	50.42	2	1.98
		down	97.08	67.92	0.12	1.28
		happy	72.92	61.67	1.08	1.53
		milk	74.58	48.33	1.02	2.07
		pie	100	99.17	0	0.03
		play	51.25	49.58	1.95	2.02
sp0, sp2	sp1	ball	62.08	52.08	1.52	1.92
		down	97.92	59.17	0.08	1.63
		happy	76.25	84.58	0.95	0.62
		milk	90	69.58	0.4	1.22
		pie	100	98.75	0	0.05
		play	45.83	40.83	2.17	2.37
sp0, sp1	sp2	ball	55.42	62.92	1.78	1.48
		down	76.67	77.08	0.93	0.92
		happy	70.83	64.58	1.17	1.42
		milk	51.67	55.83	1.93	1.77
		pie	99.58	100	0.02	0
		play	53.75	52.92	1.85	1.88

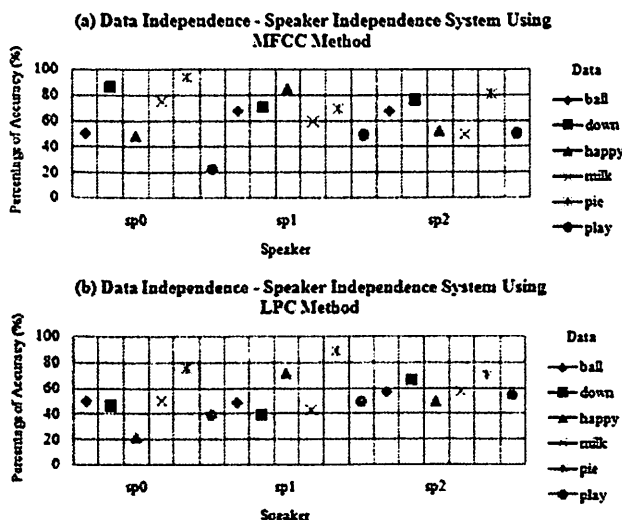


Fig. 17: Performance of data independent - speaker independent system of the second phase of ASR system using (a) MFCC method (b) LPC method.

To sum up, obviously the performance of MFCC method in ASR system is better than the LPC method in term of accuracy as shown in the following Figure 18. From observation of the analysis graph, most of the data samples have reached above than 60% of accuracy for implementation of MFCC method. For other comparison, the word *pie* and word *down* are the most accurate data samples that reach above 90% of accuracy. Hence, these words are meaningful to be implemented as model data in real application since the data are consistent for processing of larger input testing data. For word *pie*, the system gives up to 99.86% and 81.88% of accuracy using MFCC feature extraction. While, by using the LPC feature extraction method, 99.31% and 78.54% of accuracy percentages are observed. Furthermore, the result of word *pie* data samples has also proved that MFCC method is better than LPC method

Table 2: Result of data independent - speaker independent approach.

Speaker		Word	Percentage of accuracy (%)		Mean squared error (MSE)	
Input	Model		MFCC	LPC	MFCC	LPC
sp1, sp2	sp0	ball	50.63	50	1.98	2
		down	86.25	46.25	0.55	2.15
		happy	48.75	21.88	2.05	3.125
		milk	75	50	1	2
		pie	94.38	76.25	0.23	0.95
		play	22.5	38.75	3.1	2.45
sp0, sp2	sp1	ball	67.5	48.75	1.3	2.05
		down	70.63	38.75	1.18	2.45
		happy	85.63	71.88	0.58	1.13
		milk	60	43.13	1.6	2.28
		pie	70	88.75	1.2	0.45
		play	49.38	50	2.03	2
sp0, sp1	sp2	ball	67.5	56.88	1.3	1.73
		down	76.25	66.25	0.95	1.35
		happy	52.5	50	1.9	2
		milk	50	56.88	2	1.73
		pie	81.25	70.63	0.75	1.18
		play	50	55	2	1.8

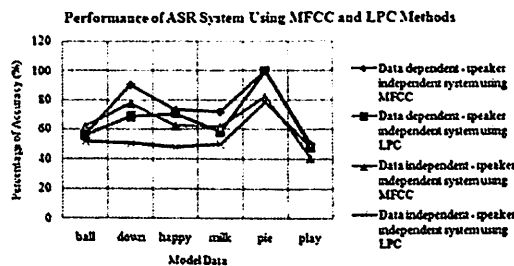


Fig. 18: Comparison between data dependent - speaker independent and data independent - speaker independent systems using different feature extraction methods.

Conclusions:

In a nutshell, the experimental results have proved that the implementation of ASR system for early detection of Apraxia can be a viable system which gives up to 99.86% and 81.88 of accuracy percentages for data dependent-speaker independent and data independent-speaker independent systems, respectively. For modeling the most outstanding apraxia word, the experimental results have proved that the word *pie* and word *down* are the most preferable that give more than 90% of accuracy. Meanwhile, the implementation of MFCC as features to the ASR system achieves better performance compared to the implementation of the LPC feature extraction. For future research, a real time ASR system for early detection of apraxia speech disorder with the integration of speech therapy facilities will be developed.

ACKNOWLEDGMENT

This research is supported by the following research grants: Research University Grant, Universiti Sains Malaysia, 1001/PELECT/814098 and Incentive Grant, Universiti Sains Malaysia.

REFERENCES

Boswell, D., 2002. Introduction to Support Vector Machines: University of California, San Diego.
 Bowen, C., 2009. Children's Speech Sound Disorders, Oxford Wiley-Blackwell.
 Chulhee, L., H. Donghoon, C. Euisun, G. Jinwook and L. Chungyong, 2003. Optimizing Feature Extraction for Speech Recognition. Speech and Audio Processing, IEEE, 11: 80-87.
 Cleuren, L., 2003. Speech Technology in Speech Therapy? State of the Art and Onset to the Development of a Therapeutic Tool to Treat Reading Difficulties in the First Grade of Elementary School. SLT Internship at ESAT-PSI Speech Group, 2-3.
 Danubianu, M., S.G. Pentiuc and T. Socaciu, 2009. Towards the Optimized Personalized Therapy of Speech Disorders by Data Mining Techniques. In: Computing in the Global Information Technology. ICCGI, 09: 20-25.

- Flipsen, P.J., 2006. Measuring the Intelligibility of Conversational Speech in Children. *Clinical Linguistics and Phonetics*, 20(4): 303-312.
- Georgopoulos, V.C. and G.A. Malandraki, 2005. A Fuzzy Cognitive Map Hierarchical Model for Differential Diagnosis of Dysarthrias and Apraxia of Speech. In: *Engineering in Medicine and Biology Society. IEEE-EMBS, 27th Annual International Conference*, pp: 2409-2412.
- Gordon-Brannan, M. and B.W. Hodson, 2000. Intelligibility/Severity Measurements of Prekindergarten Children's Speech. *American Journal of Speech-Language Pathology*, 9: 141-150.
- Gudi, A.B., H.K. Shreedhar and H.C. Nagaraj, 2010. Signal Processing Techniques to Estimate the Speech Disability in Children. *IACSIT International Journal of Engineering and Technology*, 2(2).
- Gunn, S.R., 1998. *Support Vector Machines for Classification and Regression*. Faculty of Engineering, Science and Mathematics, University of Southampton.
- Hosom, J.P., A.B. Kain, X. Niu, J.P.H. Van Santen, M. Fried-Oken and J. Staehely, 2007. Improving the Intelligibility of Dysarthric Speech. *Speech Communication*, 49: 743-759.
- Kesarkar, M.P., 2003. *Feature Extraction for Speech Recognition: Electronic Systems Group, EE. Dept, IIT Bombay*.
- Kondoz, A.M., 2000. *Digital Speech: Coding for Low Bit Rate Communication Systems*, John Wiley & Son Ltd.
- Kumin, L., 2006. Speech Intelligibility and Childhood Verbal Apraxia in Children with Down Syndrome. *Down Syndrome Research and Practice*, 10(1): 10-22.
- Lim Sin, C., A. Ooi Chia, M. Hariharan and S. Yaacob, 2009. MFCC based recognition of repetitions and prolongations in stuttered speech using k-NN and LDA. In: *Research and Development (SCORED) IEEE*. pp: 146-149.
- Ogar, J., H. Slama, N. Dronkers, S. Amici and M.L. Gorno-Tempini, 2005. *Apraxia of Speech: An overview*. Neurocase. Taylor & Francis LLC.
- Ramli, D.A., N.C., Rani and K.A., Ishak, 2010. A Multi-Instance Speech Signal Data Fusion Approach for Biometric Speaker Authentication System Enhancement. *World Applied Sciences Journal (IDOSI Journal)*, 10(7): 847-852.
- Ramli, D.A., N.C., Rani and K.A., Ishak, 2011. Performances of Weighted Sum-Rule Fusion Scheme in Multi-Instance and Multi-Modal Biometric Systems. *World Applied Sciences Journal (IDOSI Journal)*, 12(11): 2160-2167.
- Shriberg, L.D., K.J. Ballard, J.B. Tomblin, J.R. Duffy, K.H. Odell and C.A. Williams, 2006. Speech, Prosody, and Voice Characteristics of a Mother and Daughter With a 7;13 Translocation Affecting FOXP2. *Journal of Speech, Language and Hearing Research*, 49: 500-525.
- Tian-Swee, T., L. Helbin, A.K. Ariff, T. Chee-Ming and S.H. Salleh, 2007. Application of Malay Speech Technology in Malay Speech Therapy Assistance Tools. *Intelligent and Advanced Systems, ICIAS*, pp: 330-334.
- Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc.



ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com



Development of Multibiometric Verification System Based on Speech and Palm Print Information

Lau Su Ching, Noor Salwani Ibrahim and DzatiAthiarRamli

¹Intelligent Biometric Group (IBG), School of Electrical & Electronic Engineering, USM Engineering Campus, 14300, NibongTebal, Penang MALAYSIA

ARTICLE INFO

Article history:

Received 19 September 2014

Received in revised form

19 November 2014

Accepted 22 December 2014

Available online 2 January 2015

Keywords:

single biometrics, multibiometrics, palm print, MFCC, LPC, SVM classifier.

ABSTRACT

Biometrics is science and technology of measuring and analyzing biological data. But, unibiometrics systems are prone to be lack of accuracy, non-universality and spoof attacks. However, these limitations can be minimized by developing a multibiometric system. Biometric fusion is the combination of information from multiple sources of sensors, modalities or biometric algorithms. In this project, a multibiometric based on speech and palm print information is developed. In speech biometric system, Mel Frequency Cepstrum Coefficient (MFCC) and Linear Predictive Coding (LPC) techniques are used to extract audio features while Region of Interest (ROI) is used to extract visual features of palm print. In pattern-matching, Support Vector Machine (SVM) is used as classifier. Subsequently, for multibiometric system, score level fusion with sum-rule fusion and weighted sum-rule fusion techniques are used to increase the system performances. Receiver Operation Characteristic (ROC) curve of Genuine Acceptance Rate (GAR) versus False Acceptance Rate (FAR) is plotted and Equal Error Rate (EER) is calculated to evaluate the performance of the biometric systems. Lastly, the multibiometric system is developed by using GUI. In this project, it has been proven that the multibiometric system with MFCC technique together with the weighted sum-rule score fusion gives the best performance with the lowest EER percentage i.e. 0.2046% compared to 2.1678%, 1.0088% and 6.8515% for single biometric system based on MFCC, LPC and palm print.

© 2015 AENSI Publisher All rights reserved.

ToCite ThisArticle:Lau Su Ching., Noor Salwani Ibrahim &DzatiAthiarRamli. Development Of Multibiometric Verification System Based on Speech and Palmprint Information. *Aust. J. Basic & Appl. Sci.*, 9(27): 112-118, 2015

INTRODUCTION

Information technology field becomes advance and widely used nowadays. As the internet becomes common for information exchange, hence a reliable authentication system is very important to ensure its privacy and confidential. The applicability of traditional methods such as pin identification, password or token based (ID cards) arrangement has many limitations and restricted. Furthermore, password or pin identification can be lost, forgotten or stolen. As a consequent, biometrics approach has become a good alternative for the identity authentication due to its uniqueness to each individual and cannot be lost, recreated or forgotten (Teoh *et al.*, 2003, Bhattacharyya *et al.*, 2009).

Biometrics is the science of measuring human's characteristics for the purpose of authenticating or identifying the identity of an individual. There are two main classes in biometric characteristics which are physiological characteristic and behavioral characteristic. Physiological characteristic is a biometric feature that measures the parts of body and

it varies from person to person. The examples of physiological characteristics are face recognition, palm print, hand geometry, fingerprint and iris recognition. On the other hand, behavioral characteristic measures the action that is performed by human such as signature and voice (Yih *et al.*, 2008).

However, unibiometric systems are prone to be affected by problems such as lack of accuracy, non-universality, noise sensor data and spoof attacks due to its identifier only use single source of biometric information. These limitations of the unibiometric systems can be reduced by combining multiple sources of biometric information. Accordingly, a system which consolidates information from multiple sources i.e. multibiometric systems are developed in order to enhance the accuracy and the performance of unibiometric systems (Kiskuet *al.*, 2011). The performance of multimodal biometric system can be increased by information fusion as reported in (Liau and Isa, 2011).

Multimodal biometrics can overcome the limitations possessed by single biometric trait and

Corresponding Author: DzatiAthiarRamli, Intelligent Biometric Group (IBG), School of Electrical & Electronic Engineering, USM Engineering Campus, 14300, NibongTebal, Penang MALAYSIA

give better classification accuracy (Ross and Jain, 2003). This paper proposes an audio-visual system based on fusion at matching score level using support vector machine (SVM). The support vector machine-based fusion method also gave very promising results. Speech biometrics is proposed in this research due to the ease of data collection which is natural and non-obstructive. Furthermore, the hardware used is cost effective and only a simple arrangement is needed for setting up for the data collection process.

Subsequently, the advantages of the use of palm print biometrics are large palm area for feature extraction is available, easy of capturing and high user acceptability (Sun *et al.*, 2005). The cost of palm print acquisition device is less compared to other biometric trait like iris and fingerprint scanning device. Besides, palm print is harder to imitate than fingerprint. It is more acceptable than face recognition system that may cause privacy issues (Shu and Zhang, 1998). As reported in Kong (2000), human palm has line features including minutiae points as in the case of fingerprints. In order to obtain the palm features, the palm can be scanned so as to get an abundance of ridges and minutiae information which form the finer details of the palm. In this research, these finer details are obtained by scanning the palm image a high resolution of 1000 DPI. In Ibrahim and Ramli (2013), instead of focusing on the final details, the higher level textural information

presents on the palm in the form of major lines and small wrinkles are extracted. The reason for choosing to work with these features is that the higher level details can be captured by using a generic web camera at low resolution (Ibrahim *et al.*, 2014).

Methodology:

The methodology in this study is divided into 4 modules i.e. data acquisition, feature extraction, classification, fusion and development of GUI for the developed palm print biometric system.

Data Acquisition and Feature Extraction:

a. Speech Biometric:

This database consists of the digitized speech signals of the recording voices of 37 speakers stored as monophonic 16 bit, 32 kHz and in WAV format. There are 60 audio data for each speaker obtained. Hence, it consists of 2220 data. A speaker verification system is developed by using mel-frequency cepstral coefficients (MFCC) and linear predictive coding (LPC) as feature extraction and SVM as classifier

MFCC processing consists of signal pre-emphasis, windowing, spectral analysis, filter bank processing, log energy computation and mel frequency cepstrum computation as shown in Fig. 1. There are 12 melcepstrum coefficients, one log energy coefficient and three delta coefficients per frame have been set in the experiments.

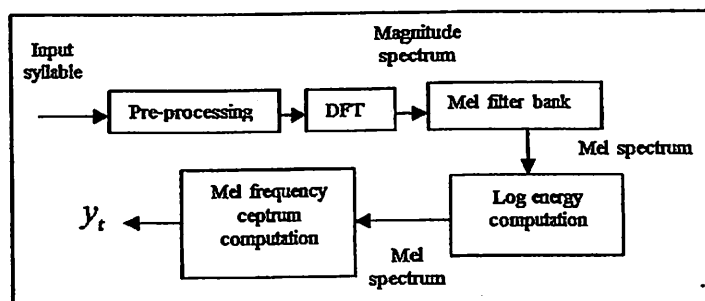


Fig. 1: Typical MFCC process

On the other hand, LPC feature extraction models the process of speech production and is defined as a digital method for encoding an analogue signal in which a particular value is predicted by a

linear function of the past values of the signal (Rabiner & Juang, 1993; Furui, 1981) where the process of the LPC is shown in Fig. 2.

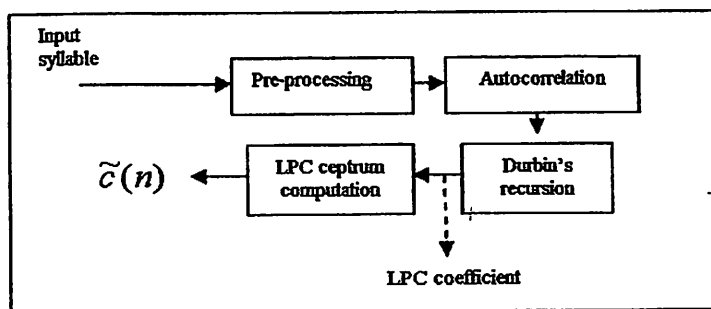


Fig. 2: Typical LPC process

b. Palm print Biometric:

Palm print database contains 60 images right hand images collected from 37 individuals by using Canon optical scanner. The acquired images of palm print are in BMP format. Each speaker consists of 60 sequences of palm print images (20 sequences from each session) hence in total of 2220 images from entire speakers. The visual data of 2220 images from 37 speakers are converted into gray scale with the size of 351 x 351 pixels. The gray scale images are stored for the feature extraction process. For each

speaker, first 20 images are selected as model or training set, and then the rest 40 images of each speaker in this database are used in verification system or testing set. Hence, there are 740 data are stored in training set and 1480 data as testing set.

The ROI of palm print image with gray scale level is used as features in this study. The appropriate coordinate is chosen on the palm print image in order to crop the image that contains the principal lines. A region of ROI is determined to get the cropped image.

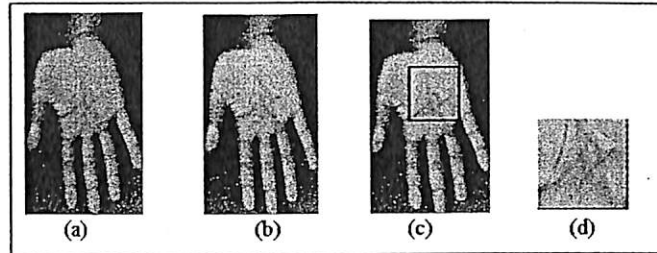


Fig. 3: Extraction of palm print biometric modality (a) Acquired image from scanner (b) Gray scale level of palm print image (c) ROI region to be cropped (d) Cropped image

Pattern Matching:

After extracting features from the speech signal and palm print images, the features are fed to the classifier and proceed with the pattern matching process. In pattern matching process, the similarity of the testing data and training data is measured.

In SVM classifier, 20 data of each speaker is used as the training data and 40 data of each speaker is used as testing data. The SVM classifier calculates the score of the pattern matching between training data and testing data. Several sets of parameters are determined in polynomial Kernel to compare the performance of the system. Consequently, the FARs, GARs and EERs values are calculated and the ROC curve is plotted in order to evaluate the performance of the multibiometric system.

Score Fusion:

After pattern matching via SVM classifier is executed, the scores obtained from two single biometric systems which are speech biometrics and palm print biometrics will proceed to fusion task. The score fusion techniques used in this study are sum rule technique and weighted sum-rule technique.

In sum rule technique, the new set of score is obtained as shown in Eq. (1).

$$\text{Scorefusion} = \frac{\text{score}_{\text{speech}} + \text{score}_{\text{palmprint}}}{2} \quad (1)$$

The combined matching score can also be computed as a weighted sum-rule as given in the Eq. (2) and Eq. (3).

$$\text{Scorefusion} = w_1 \cdot \text{score}_{\text{speech}} + w_2 \cdot \text{score}_{\text{palmprint}} \quad (2)$$

$$w_1 + w_2 = 1 \quad (3)$$

The weights, w_1 and w_2 are varied from range of 0.1 to 0.9 in steps 0.1. ROC curve based on the new set of fusion scores is plotted to analyze the system performances.

GUI development:

Finally, a GUI of multibiometric system is developed. GUI is a visual object that enhances interaction between a computer and a user. The GUI layout is designed so as to support the process of data collection, user-friendly between users and interface as well as to assist administrative work.

RESULTS AND DISCUSSION

To evaluate the performance of the system, a ROC curve of GAR versus FAR is plotted. A total of 1,480 genuine data (40 genuine data from each speaker) and 53,280 imposter data are used to plot the ROC curve. Several ROC curves are plotted to compare the performances of single biometric system and multibiometric system which have been undergone the score fusion.

FAR is the percentage of wrongly accepted individuals over the total number of wrong matching. FRR is the percentage of number of wrongly rejected individuals over the total number of correct matching. GAR is the percentage of the number of correctly accepted individuals divided by the number

of identification attempts. EER is the value when FAR is equal to FRR.

Performances for Single Biometric System:

Fig. 4 shows the ROC curve for three distinct cases in single biometric systems, (i) speech biometrics with LPC features, (ii) speech biometrics with MFCC features, and (iii) palm print biometric system with ROI features. When FAR is 1%, the GAR values of palm print system with ROI features, speech system with LPC features and speech system

with MFCC features are 80%, 95% and 99% respectively. When FAR achieves 10%, speech biometrics with LPC and MFCC features approximately approach 100% of GAR where palm print biometrics with ROI features is 95%. Table 1 indicates the EER performances for three different biometric systems. Palm print biometrics with ROI features has the highest value of EER where speech biometrics with MFCC features has the lowest value of EER.

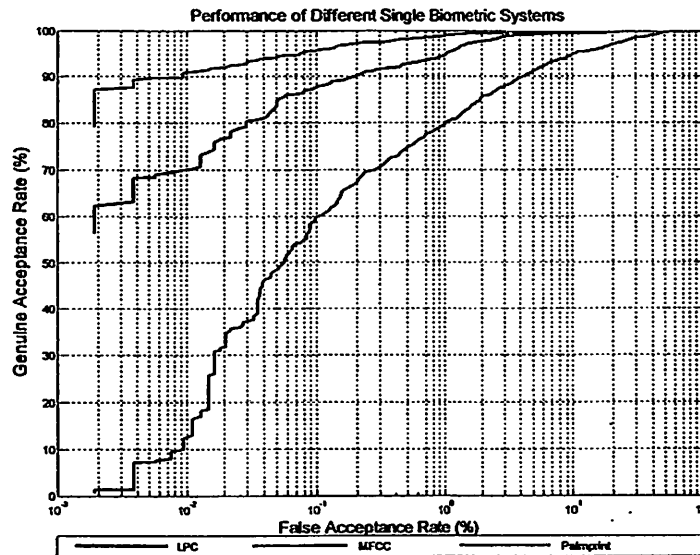


Fig. 4: Comparison of ROC curves of speech biometric and palm print biometric

Table 1: EER percentages for single biometric systems

LPC	MFCC	Palm print
2.1678	1.0088	6.8515

Performances for Sum-Rule Fusion of Multibiometric Systems and Single Biometric System:

Fig 5 shows the ROC curves of sum-rule score fusion of multibiometric system with different single biometric system (a) speech biometrics with LPC features; (b) speech biometrics with MFCC features and (c) palm print biometrics with ROI features. When FAR is 0.1%, the GAR value of LPC and palm print sum-rule score fusion is 98%. For MFCC and palm print sum-rule score fusion, the GAR percentage is 99%. When FAR is 1%, the GAR values of score fusion of speech biometrics with LPC

features and palm print biometrics is achieved 99.5%. Else, the GAR values of score fusion of speech biometrics with MFCC features and palm print biometrics is 100%. Table 2 shows the EER performances of single biometric system and multibiometric system with sum-rule score fusion. The multibiometric system with combination of speech biometrics with MFCC as features and palm print biometrics with ROI features has the smallest value of EER. Hence, the fusion multibiometric system has the higher performance than the single biometric system.

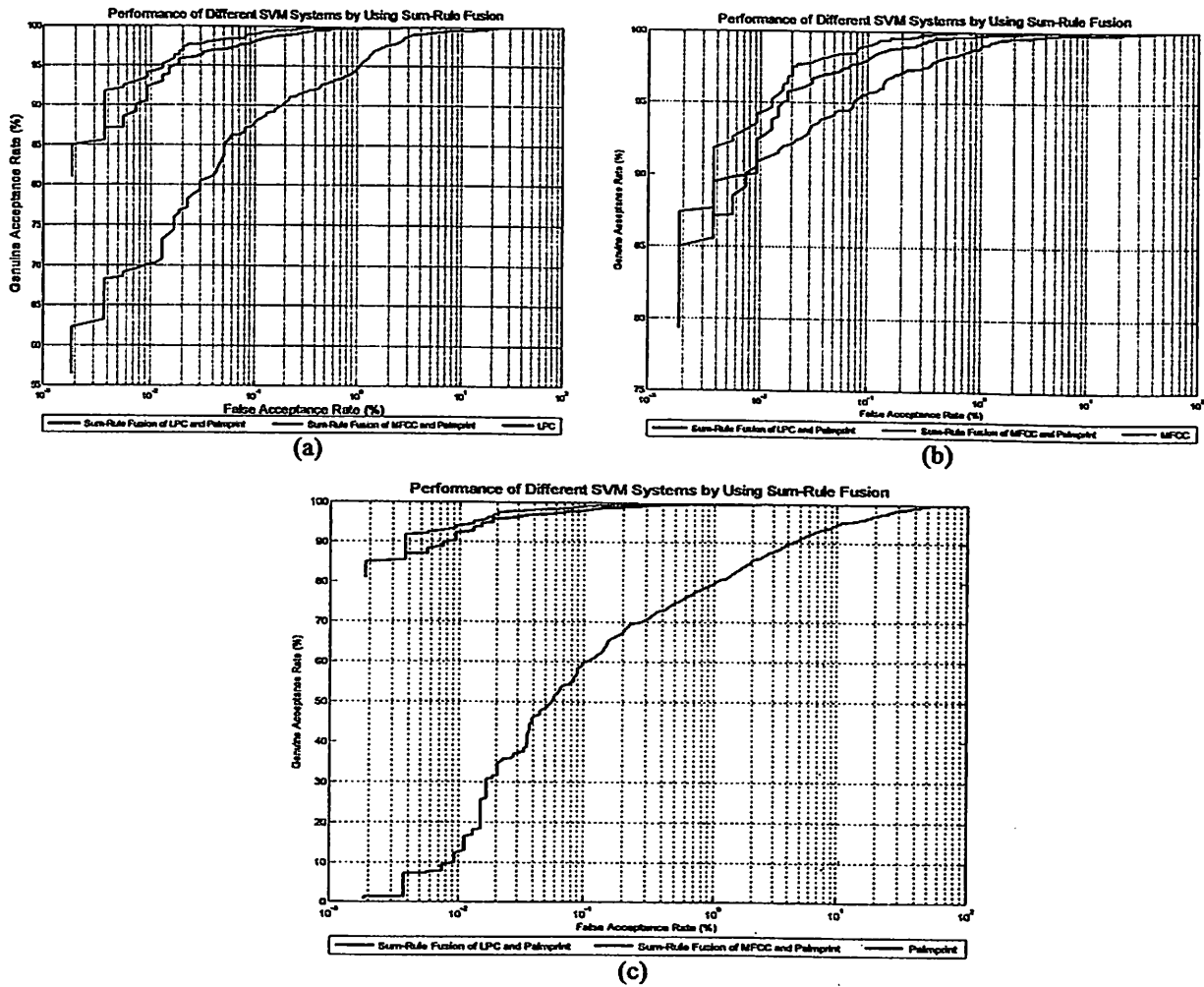


Fig. 5: Comparison of ROC curves of sum-rule score fusion of multi-biometric system with respective single biometric system (a) speech biometrics with LPC as feature extraction, (b) speech biometrics with MFCC as feature extraction and (c) palmprint biometrics

Table 2: EER percentages for single biometric system and multi-biometric system

LPC and Palm print	MFCC and Palm print	LPC	MFCC	Palm print
0.4673	0.3087	2.1678	1.0088	6.8515

The intersection point of the FAR and FRR rate in the multi-biometric system (speech biometrics with LPC feature and palm print biometric system) by using weighted sum-rule score fusion is 0.4730 which represents EER rate for the system as shown in Fig. 6. The desired threshold value obtained is 0.7469.

Fig 7 shows the FAR and FRR percentages with different threshold values for weighted sum-rule fusion of speech biometric system with MFCC feature and palm print biometric system. The intersection point of the FAR and FRR rate is 0.2027 which represents EER rate for the system. The desired threshold value obtained is 0.7175.

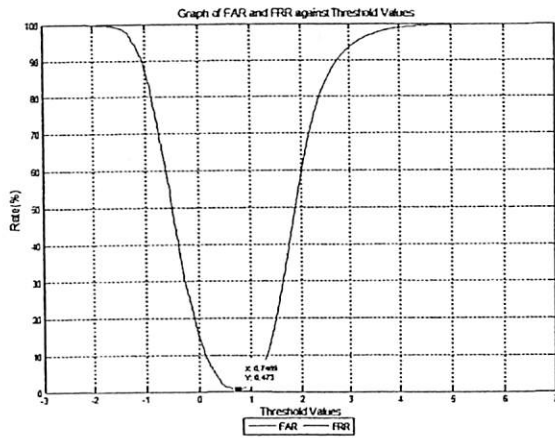


Fig. 6: Graph of FAR and FRR percentages versus different threshold values of multibiometric system (weighted sum-rule fusion of speech biometrics with LPC feature and palm print biometrics)

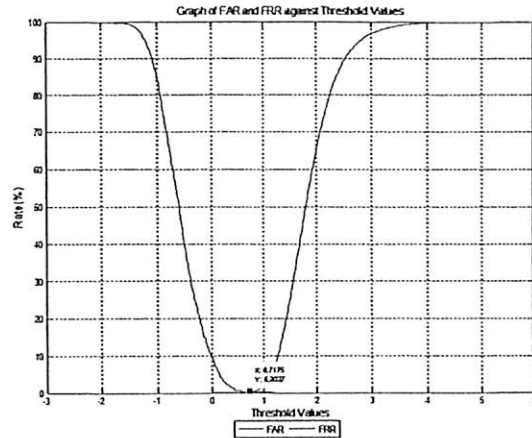


Fig. 7: Graph of FAR and FRR percentages versus different threshold values of multibiometric system (weighted sum-rule fusion of speech biometrics with MFCC feature and palm print biometrics)

Implementation of GUI:

A GUI for multibiometric system as shown in Fig 8 is developed to increase the user friendliness. First of all, the user needs to key in their ID and press the enter button. Once the valid ID is obtained by the system, the system will proceed to the next step. The user can choose either speech biometrics, palmprint biometrics or both as shown in Fig 8(a). Next, the system requests for the biometric trait(s) enrollment. For this case, the user enrolls his/her biometric trait(s) by browsing the database. The enrolled biometric trait will pop up as in Fig 8(b). After the system obtains the enrolled biometric trait(s), the

user can choose the types of feature extraction to be used in the system. In speech biometric system, there are two available feature extraction techniques which are MFCC and LPC. In palmprint biometric system, the feature extraction technique used is ROI. After the feature extraction, the system will match the enrolled trait with the model by using SVM classifier. Fig 8(c) indicates the pattern matching in process. Once the score is obtained, the system will proceed to decision. If the score is below the threshold value, the system will show accept as in Fig 8(d), else the system will show reject which means that user is an imposter.

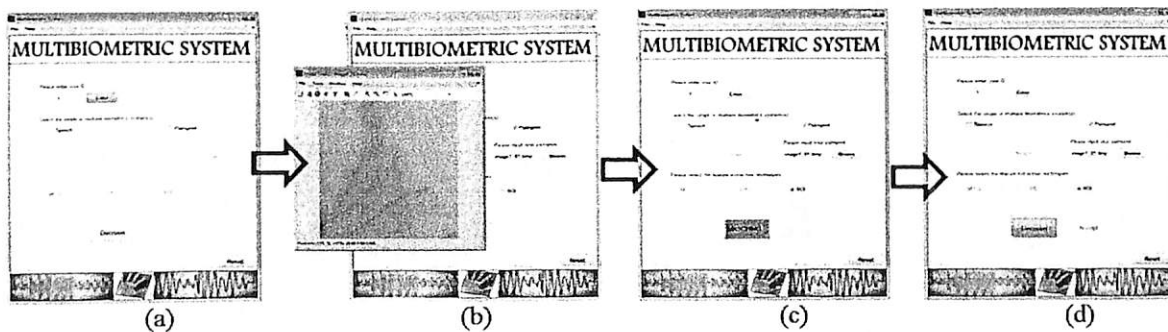


Fig. 8: Layout of the GUI

Conclusion:

At the end of this research, a reliable multibiometric system is successfully developed. Based on the analysis result, weighted sum-rule score fusion of multibiometric system has the best performance. These projects successfully develop 2 single modal systems based on speech and palm print using SVM classifier. Speech biometric system and palm print biometric system are developed for verification. SVM classifier is used during the

pattern-matching process. Palm print biometric system has the lowest performance compared to speech biometric system. Next, a multibiometric system is developed by combining the speech and palm print biometric. Score level fusion is applied to obtain a better performance of the system. In this project, the score level fusion techniques used are sum-rule fusion and weighted sum-rule fusion. The performance of the system is analyzed by plotting

ROC curve. The sum-rule fusion has the best performance.

ACKNOWLEDGMENT

The authors would like to thank the financial support provided by Universiti Sains Malaysia Research University Grant 1001/PELECT/814161 for this project.

REFERENCES

- Bhattacharyya, D., R. Ranjan, F. Alisherov and M. Choi, 2009. Biometric authentication: A review. *International Journal of u-and e-Service, Science and Technology*, 2: 13-28.
- Furui, S., 1981. Cepstral analysis technique for automatic speaker verification. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29: 254-272.
- Ibrahim, S. and D.A. Ramli, 2013. Evaluation on Palm-Print ROI Selection Techniques for Smart Phone Based Touch-Less Biometric System, *American Academic & Scholarly Research Journal*, 5(5): 205-21.
- Ibrahim, S., H. Jaafar, and D.A. Ramli, 2014. Robust Palm Print Verification System Based On Evolution Kernel Principal Component Analysis, *IEEE International Conference on Control System, Computing and Engineering*.
- Kisku, D.R., P. Gupta and J.K. Sing, 2011. Multibiometrics Feature Level Fusion by Graph Clustering. *International Journal of Security and Its Applications*, 5.
- Kong, W.K., D. Zhang and W. Li, 2000. Palm print feature extraction using 2-D Gabor filters. *Pattern Recognition*, 36: 2339-2347.
- Liau, H.F., and D. Isa, 2011. Feature selection for support vector machine-based face-iris multimodal biometric system. *Expert Systems with Applications*.
- Lu, G., D. Zhang and K. Wang, 2003. Palm print recognition using eigenpalms features. *Pattern Recognition Letter*, 24: 1463-1467.
- Rabiner, L.R., and B.H. Juang, 1993. *Fundamental of speech recognition liveness verification in audio-video speaker authentication*, Prentice-Hall International: United State.
- Ross, A. and A. Jain, 2003. Information fusion in biometrics. *Pattern Recognition Letters*, 24: 2115-2125.
- Ross, A., K. Nandakumar and A. Jain, 2008. Introduction to multibiometrics. *Handbook of Biometrics*, 271-292.
- Shu, W. and D. Zhang, 1998. Automated personal identification by palm print. *Optical Engineering*, 37: 2359.
- Sun, Z., T. Tan, Y. Wang and S.Z. Li, 2005. Ordinal palm print representation for personal identification. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE*, 1: 279-284.
- Teoh, A., S.A. Samad and A. Hussain, 2003. An Internet based speech biometric verification system. In: *IEEE*, 1: 47-51.
- Yih, E.W.K., G. Sainarayanan and A. Chekima, 2008. Palm print Based Biometric System: A Comparative Study on Discrete Cosine Transform Energy, Wavelet Transform Energy and Sobel Code Methods. *Intl. J. on Biomedical Soft Comp. and Human Science*, 14: 11-19.

Data reduction on MFCC features based on kernel PCA for speaker verification system

Mohd Azha Mohd Saleh*, Noor Salwani Ibrahim, Dzati Athiar Ramli

Intelligent Biometric Group (IBG), School of Electrical & Electronic Engineering, USM Engineering Campus 14300, Nibong Tebal, Penang Malaysia

Abstract Introduction: Mel Frequency Cepstrum Coefficients (MFCC) is one of the most widely used feature extraction techniques for speech recognition and produce MFCC features as input for the classification phase. In this study the reduction of feature dimension on MFCC features is studied due to large data size affects computational time which leads to slower verification speed. So, implementation of data reduction techniques so as to retain the most important feature parameters is evaluated in this study. In this study, an investigation of data reduction based on principal component analysis is proposed. Two approaches of Kernel Principal Component Analysis (KPCA) techniques i.e. Gaussian and Polynomial KPCA and PCA are evaluated and compared. The features based on MFCC and the reduced dimensions based on KPCA and PCA are then classified using two types of Support Vector Machine (SVM) classifiers i.e. linear and polynomial SVM. A set of clean data samples with three different dimensions of principle components i.e. 80, 117 and 180 are used for system evaluation. For performance evaluation, Equal Error Rates (EER) and verification time (VT) are employed in this study. The best system performance is observed for MFCC-KPCA Gaussian feature extraction technique with 117 features dimensions using linear SVM as classifier. This study proves that the use of data reduction technique can speed up verification time tremendously and improve system performances as well.

Key words: *Data reduction; MFCC; Kernel PCA; Speaker verification*

1. Introduction

Speech recognition systems have been developed successfully by utilizing many types of feature extraction methods. One of the most popular methods is Mel Frequency Cepstrum Coefficients (MFCCs) which provide a compact parametric representation of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a Mel frequency scale. In 1980, Davis and Mermelstein (1980) compared parametric representations for monosyllabic word recognition in continuously spoken sentences. This study utilized parametric representation based on Mel Frequency Cepstrum, Linear Frequency Cepstrum, Linear Prediction Cepstrum and Linear Prediction Spectrum. Based on the experimental results, this study concluded that MFCC possess significant advantage over the other methods. The superior performance of the MFCC may be attributed by the fact that they are better in representing the perceptually relevant aspects of the short-term speech spectrum.

Lee, Fang, Hung and Lee (2001) have presented a new feature extraction approach that designs the shapes of the filters in the filter bank. The study applied PCA approach on the FFT spectrum of the training data. As a result the conventional MFCC features have been improved by the PCA-optimized

filter bank. The proposed features are robust to additive noise for speech recognition while providing the same result for clean speech. It is claimed due to the PCA-optimized filter bank has maximized both the SNR variance ratio and the variation of the features.

A comparative evaluation on various MFCC implementations has been implemented by Ganchev et al. (2005). The implementation differs from other researches mainly in the number of filters, the shape of the filters, the way the filters are spaced, the bandwidth of the filters, and the manner in which the spectrum is warped. In addition, the frequency range of interest, the selection of actual subset and the number of MFCC coefficients employed in the classification are also evaluated. As a result, this study reported that the speaker verification performance does not vary vastly when different approximations of the non-linear pitch perception of human are used. However, some observations suggested that regardless of the specific filter bank design, a larger number of filters favour the speaker detection performance. Beside the number of filters in the filter bank, the overlapping among the neighbouring filters also proved as a sensitive parameter.

Chen and Luo (2009) in their paper have proposed a study on the use of MFCC and SVM for text-dependent speaker verification. The MFCCs used in this paper are extracted from the voiced password spoken by the user. These parameters are

* Corresponding Author.

then normalized and then used as the speaker features for training a claimed speaker model via SVM. By using speech signals selected from the Aurora-2 database, experimental results shown the performance of the proposed speaker verification algorithm yields an average accuracy rate of 95.1% with 22-order MFCCs.

Although speech recognition systems have been developed successfully with great performances and features, this success depends much on the extracted speech features, which has an important role in the whole recognition system (Amaro et al; 2004). If the speech features are not well extracted or come with an extreme data size, it will cost much computational time to the speaker verification system, which then will affect its performance and speed.

Further research on MFCC applying data reduction in speaker recognition system has been done by Hasan, Jamil and Rahman (2004). This study presented a security system based on speaker identification by utilizing MFCC as feature extraction method while Vector Quantization (VQ) technique as data reduction method. The study revealed that, when the number of centroids increased, the identification rate of the system also increases. They also found that combination of Mel frequency and Hamming window leads to better performance. This study also observed that the linear scale can improve system accuracy if comparatively higher number of centroids is used. However, the recognition rate using a linear scale would be much lower if the number of speakers increased.

The use of MFCC and Vector Quantization in speaker recognition has also been carried out by Mishra and Agrawal (2012). This study implemented an enhanced MFCC with silence removal. The silence present before and after the voiced part is removed to improve the performance of classifier. Based on their findings, this research suggested an effective normalization algorithm can be adopted on extracted parametric representations in order to improve the identification rate. Apart from that, a combination of features i.e. MFCC, LPC, LPCC, Formant etc. may be used so as to obtain a robust parametric representation for speaker identification.

The combination of MFCC and PCA has also been presented by Ittichauchareon, Suksri and Yingthawornsuk (2012). This paper described an approach of speech recognition using the MFCC features extracted from speech signal of spoken words. PCA is employed as the supplement in feature dimensional reduction state, prior to training and testing speech samples via Maximum Likelihood Classifier (ML) and SVM. It is found that the combination of MFCC-PCA-SVM with more MFCC samples have shown the improvement in recognition rates significantly compared to MFCC-PCA-ML.

Motivated by all of these researches, this study comes out with the proposed system of MFCC-KPCA-SVM model for speaker verification system where KPCA is used as data reduction technique on MFCC features. PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight

their similarities and differences (Rodriguez, de Paz et al; 2008). Since patterns are hard to find in high dimensional data, PCA is a powerful tool for analyzing data.

According to Leitner, Pernkopf and Kubin (2011), linear PCA refers to orthogonal transformation of the space containing the data samples. The transformed space is spanned by the eigenvectors that are found by eigenvalue decomposition of the covariance matrix estimated from the data samples. The coordinates of the data samples after transformation are referred as principal components. Normally, few principal components capture most of the characteristics of the data. The directions of these components are given by the eigenvectors corresponding to large eigenvalues, as a large eigenvalue means that its eigenvectors covers relevant information of the data.

Kernel PCA (KPCA) is one of the kernel algorithms that have been known from mid-nineties. It performs the PCA in the feature space, so it looks for directions of largest variance that yields nonlinear directions in the input space. KPCA was introduced after PCA to merit the performance of PCA (Huang et al; 2009). KPCA is a non-linear extension of PCA which data is first mapped and PCA is applied to the mapped data. KPCA make it possible for us to represent the speech features in a higher dimensional space which can possibly generate more distinguishable speech features (Amaro et al; 2004). It can extract up to n (number of samples) nonlinear principal components without expensive computations. It also can give a good re-encoding of the data when it lies along a non-linear manifold.

KPCA involves calculation of the eigenvalues decomposition or singular value decomposition of centered kernel data and is in search for orthogonal functions that optimize the kernel data scatter. In the linear case, it is well known that the classical PCA is not robust against data contamination and a small portion of outliers can give disturbance to the resulting principal components (Huang et al; 2009). PCA is a powerful technique for extracting structure from possibly high-dimensional data sets. But it is not effective for data with nonlinear structure. In KPCA, the input data with nonlinear structure is transformed into a higher dimensional feature space with linear structure, and then linear PCA is performed in the high-dimensional space.

So, in next section, discussion on feature extraction method using MFCC is presented. Subsequently, PCA and KPCA as data reduction techniques are described. Finally, explanation on SVM classifier as classification method is then given.

2. Material and methods

2.1. Research framework

The overall research framework of this study is summarized as in Fig. 1 below.

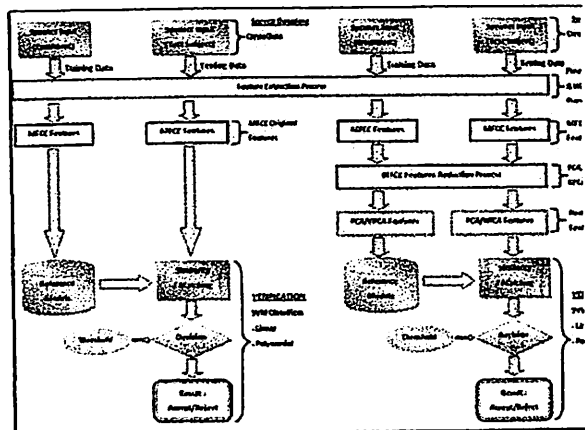


Fig. 1: Research framework

2.2. Database

The digitized audio signals from the Audio Visual Digit Database (Sanderson and Paliwal; 2001); which is monophonic, 16 bit, 32 kHz in WAV format have been used for performance evaluation in this research. This database consists of video and corresponding audio recordings of 37 speakers (21 males and 16 females). The recordings are done in three sessions. In each session, each speaker performed 20 repetitions of digit zero to nine hence 60 audio data for each speaker from all sessions. In

total, 2220 audio data from entire speakers have been used for this research.

2.3. Feature Extraction

The entire process of extracting MFCC features is illustrated in Fig. 2. In this research, we utilize a feature set consists of 12 mel cepstrum coefficients, one log energy coefficient, 13 delta coefficients and 13 delta² coefficients per frame which in total 39 coefficients. The entire frames are then resized with data in interpolation technique to 64x64 matrix. This feature matrix is then reshaped to 1x4096 as feature vector to represent each voice sample. A matrix of 740x4096 dimensions based on 20 voice samples and 37 speakers is then constructed.

For the purposes of comparison and evaluation with the proposed method of PCA and KPCA, we resize the feature matrix of 64x64 dimensions to 10x10, 12x12 and 14x14 using data interpolation technique. This entire new matrix is then reshaped to 1x100, 1x144 and 1x196 respectively which will represent each voice samples. As a results, for training data with 20 voice samples and 37 speakers, four sets of feature dimensions i.e. 740x100, 740x144, 740x196 and 740x4096 are used.

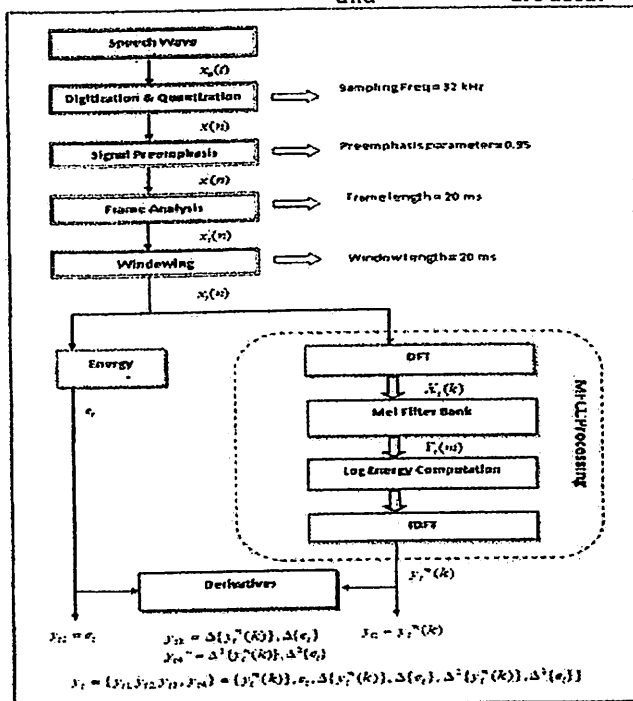


Fig. 2: Feature extraction process

3. Principal Component Analysis (PCA) Processing

For PCA technique, a set of eigenvoices (eigenvectors) from the training data is constructed. This eigenvoices are then used to generate the projection of new training and testing data. Our

training data consist of MFCC features of matrix in $N^2 \times M$ dimension where N^2 is the feature sizes and M is sample size hence $x = 4096 \times 740$.

The eigenvoices for training data are computed as the following steps.

Step 1: Computation of mean and subtraction of mean from each data point of all samples.

Define the data average vector:

$$\mu = \frac{1}{M} \sum_{i=1}^M x_{ij} \quad (1)$$

where x_{ij} represent each data point in matrix x for $i=1,2,3,\dots,M$ and $j=1,2,3,\dots,N^2$

Step 2: Compute the covariance matrix

$$XX^T = \frac{1}{M-1} \sum_{i=1}^M (x_{ij} - \mu)(x_{ij} - \mu)^T \quad (2)$$

Determine the Eigenvoices and Eigenvalues

In order to get the eigenvectors and eigenvalues, we need to compute a covariance matrix which characterizes the distribution of all samples.

$$C = XX^T \quad (3)$$

By using equation (3), a very large matrix with the dimension of $N^2 \times N^2$ is produced i.e. 4096×4096 . The large size matrix will also add a large computational burden for the analysis. This can be avoided by using an optimum method introduced by Turk and Pentland (1991) as below:

$$C = X^T X \quad (4)$$

where a smaller matrix of $M \times M$ dimensions (740×740 in this research) is used.

Then, the Eigenvector ϕ_i and Eigenvalue ψ_i can be calculated as below:

$$C\phi_i = \psi_i \phi_i \text{ for } i=1,2,3,\dots,M \quad (5)$$

$$\psi_i = \frac{1}{M} \sum_{j=1}^M (\phi_i^T \phi_j) \quad (6)$$

while the eigenvalue in equation (6) is subjected to the following conditions:

$$\phi_j^T \phi_i = \begin{cases} 1 & i=j \\ 0 & \text{elsewhere} \end{cases} \quad (7)$$

Step 3 : Project the original data into eigenspace.

Then the Eigenvoices is created by multiplying matrix A with each column of the Eigenvector above using the following equation,

$$v_k = A\phi_k \quad (8)$$

where ϕ_k is the Eigenvector column with k^{th} elements. As a result we get the Eigenvoices as follows,

$$V = [v_1, v_2, v_3, \dots, v_Q] \quad (9)$$

The size of Eigenvoice matrix, V is $N^2 \times Q$, as a result of multiplying A with dimension of $N^2 \times M$ and ϕ_k with dimension of $M \times Q$.

As the Eigenvoice matrix, V has been created; the projection of training data into the eigenspace can be done. The projection is implemented using the equation below,

$$Y_r = V^T (x_i - \mu_x) \text{ for } i=1,2,3,\dots,M \quad (10)$$

Y_r is the projected training vector with the $Q \times M$ dimensions.

Similarly, the projected testing data into the eigenspace can be done as follow:

$$Y_s = V^T (r_i - \mu_x) \text{ for } i=1,2,3,\dots,P \quad (11)$$

where r_i is the i^{th} testing data and P is the number of testing data. Y_s is the projected testing vector with $Q \times P$ dimensions. Y_r and Y_s is used as an input features to the SVM classifier for the verification process.

Based on the three selected threshold values, three sets of principle components with sizes of 740×80 , 740×117 and 740×180 are used as training data while 1480×80 , 1480×117 and 1480×180 of principle component sizes are used for testing data.

4. Kernel Principal Component Analysis (KPCA) Processing

KPCA technique is the result of applying the kernel function to PCA in order to obtain the representation of PCA in a higher dimensional space. In order to perform KPCA, the training samples, x needs to be projected into the high dimensional feature space F as follows:

$$\Phi: x \rightarrow F \quad (12)$$

In this research, two types of kernel are applied i.e. Gaussian and polynomial kernel. The Gaussian and Polynomial kernel are given as in equation (13) and equation (14) respectively:

$$k(x_i, x_j) = \exp\left(-\frac{1}{2s^2} \|x_i - x_j\|^2\right)$$

s is Gaussian parameter (13)

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^p \quad ; p \text{ is polynomial parameter (14)}$$

Except for utilizing the kernel trick, KPCA perform the same process as PCA in projecting Y_r and Y_s . The output of the KPCA process is also a matrix in dimension of $Q \times M$ for training samples and $Q \times P$ for testing samples. The projected training and testing data for KPCA technique are set to the same dimensions as in the PCA technique.

5. Classification

In this research, the multi-class classifier is performed for the verification process. Several methods can be used to implement SVM classifier for multi-classes such as one against one, one against all and Directed Acyclic Graph Support Vector Machine (DAGSVM) method. This research uses the one against all method. Here, for N class classification,

SVM requires the N training data to be built as a reference model, where each model is used to isolate one class from the remaining N classes.

Two types of SVM classifiers i.e. linear SVM and polynomial SVM have been used in this study. Linear SVM is the original optimal hyperplane algorithm that widely used as classifiers in a linearly separable case. Meanwhile, Polynomial SVM is a way to create nonlinear classifiers by applying the kernel trick.

The speaker verification system in this research uses four types of feature reduction methods i.e. MFCC, MFCC-PCA, MFCC-KPCA_Gaussian and MFCC-KPCA_Polynomial. Each type of features is evaluated using two types of SVM i.e. linear and polynomial.

6. Results and discussion

In this research, system performances will be evaluated in term for Equal Error Rate (EER) and verification time (VT). According to Kung et al. (2005), the accuracy of biometric system is evaluated using false rejection rate (FRR) and false acceptance rate (FAR) which respectively corresponds to sensitivity and specificity. FRR which is also known as miss probability is the rejection percentage of authorized individuals while genuine acceptance rate (GAR) is the percentage of authorized individuals accepted by the verification system.

FAR which is also known as impostor pass rate is the percentage of unauthorized individuals is

accepted by the verification system. FRR and FAR values can help us to determine the level of sensitivity and specificity. High FRR specify low sensitivity, while high FAR specify low specificity. A good verification system supposed to have a low FRR (high sensitivity) and low FAR (high specificity). Consequently, verification time, VT is also determined where it is the time taken by SVM classifier to verify testing data samples. This verification time is calculated to evaluate a significant time saving for verification process.

6.1. Performances based on different feature extraction techniques and different feature dimensions using linear SVM classifier

Table 1 shows the performances of speaker verification system using linear SVM based on different feature extraction techniques. The EER percentage and verification time for MFCC technique before the dimension reduction (dimension=4096) is 1.0163% and 37.4093s, respectively. The MFCC-KPCA_Gaussian technique gives the best EER performance at feature dimension of 117 with EER value equals to 0.8146% and verification time equals to 2.3385s. It is observed that the EER values for MFCC technique of smaller dimensions decrease the system performance but it can speed up the verification time.

Table 1: EER and verification time performances of different feature extraction techniques based on linear SVM classifier

Features Extraction Techniques	Features Dimension							
	D=100		D=144		D=196		D=4096	
MFCC	EER(%)	VT(s)	EER(%)	VT(s)	EER(%)	VT(s)	EER(%)	VT(s)
	5.4608	2.1413	4.5965	2.8167	3.8786	3.6361	1.0163	37.4093
Features Extraction Techniques	Features Dimensions							
	D=80		D=117		D=180			
MFCC-PCA	EER(%)	VT(s)	EER(%)	VT(s)	EER %	VT(s)		
	1.2294	2.2465	1.159	2.5117	1.0698	3.0604		
MFCC-KPCA Gaussian	EER(%)	VT(s)	EER(%)	VT(s)	EER %	VT(s)		
	1.1571	2.0806	0.8146	2.3385	0.8643	2.8530		
MFCC-KPCA Polynomial	EER(%)	VT(s)	EER(%)	VT(s)	EER %	VT(s)		
	3.6924	2.2414	2.6642	2.4330	1.8253	2.9464		

6.2. Performance based on different feature extraction techniques and different features dimension using Polynomial SVM classifier

Table 2 shows the performances of speaker verification system using polynomial SVM based on different feature extraction techniques. The EER percentage and verification time before the dimension reduction (dimension=4096) is 0.9685% and 42.5254s, respectively. It is observed that the data reduction techniques based on PCA and KPCA significantly improved the verification time and have surpassed the EER performances of MFCC method for the same category of size dimensions except for MFCC-KPCA Gaussian with 100 feature dimension.

6.3. Receiver Operating Curve based on GAR and FAR performances for selected feature dimension

Fig. 3 shows the performances of different feature extraction techniques based on selected feature dimensions according to their best EER performances using linear SVM as classifier. 100% of GAR performance for MFCC with 4096 dimensions is found at FAR equals to 40%. At the same FAR percentage, GAR performances for MFCC with 196 dimensions, MFCC-PCA with 180 dimensions, MFCC-KPCA_Gaussian with 117 dimensions and MFCC-KPCA_polynomial with 180 dimensions are 99.7%, 99.9%, 99.9% and 99.85% respectively.

Table 2: EER and verification time performances of different feature extraction techniques based on polynomial SVM classifier

Features Extraction Techniques	Features Dimension							
	D=100		D=144		D=196		D=4096	
	EER(%)	VT(s)	EER(%)	VT(s)	EER(%)	VT(s)	EER(%)	VT(s)
MFCC	4.3956	2.2326	4.0756	2.8417	3.4797	3.5331	0.9685	42.5254
Features Extraction Techniques	Features Dimensions							
	D=80		D=117		D=180			
	EER(%)	VT(s)	EER(%)	VT(s)	EER(%)	VT(s)		
MFCC-PCA	1.7258	1.5817	1.6441	3.5346	1.5888	4.5867		
MFCC-KPCA Gaussian	1.4621	1.5526	1.2885	3.0872	1.5559	4.0749		
MFCC-KPCA Polynomial	4.4125	1.6109	3.2245	3.2089	2.5375	4.1237		

Meanwhile, at FAR equals to 1%, the GAR performances are 98.99%, 90.95%, 98.92%, 99.39% and 97.43%, respectively. Table 3 shows the EER performances based on the selected features

dimensions. The MFCC-KPCA_Gaussian with 117 dimensions achieves a significant improvement and outperforms the other approaches.

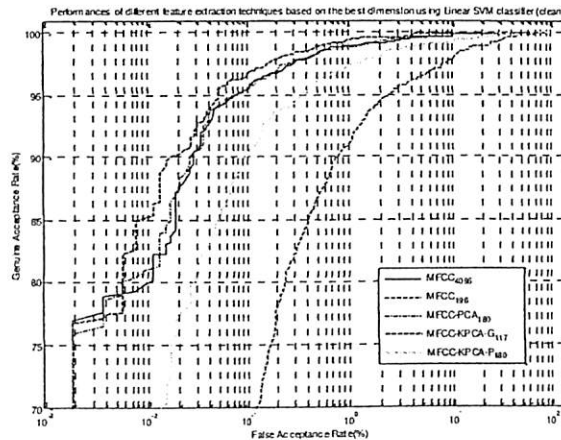


Fig. 3: Receiver operating curve (ROC) based on the selected feature dimensions for linear SVM classifier

Table 3: EER performances based on the selected feature dimensions for linear SVM classifier

Feature Extraction Technique	MFCC ₄₀₉₆	MFCC ₁₉₆	MFCC-PCA ₁₈₀	MFCC-KPCA-Gauss ₁₁₇	MFCC-KPCA-Poly ₁₈₀
EER (%)	1.0163	3.8786	1.0698	0.8146	1.8253

Subsequently, Fig. 4 shows the performances of different feature extraction techniques based on the selected features dimensions according to their best EER performances using polynomial SVM as classifier. 100% of GAR performance for MFCC with 4096 dimensions is observed at FAR equals 25.28%. At the same FAR percentage, GAR performances for MFCC with 196 dimensions, MFCC-PCA with 180 dimensions, MFCC-KPCA_Gaussian with 117 dimensions and MFCC-KPCA_polynomial with 180 dimensions are 99.26%, 99.66%, 99.93% and 99.26%, respectively. Meanwhile, at FAR equals to 1%, the GAR performances are 99.05%, 93.45%, 97.64%, 98.38% and 96.35%, respectively. Table 4 shows the EER performances based on the selected features dimensions. The MFCC with baseline dimensions of 4096 shows the best EER results. However, this is unfavourable due to the long processing time as discussed in the previous section.

speech signal data. This study reveals that by executing the right method for data reduction can really improve the time taken for verification process and at the same time can maintain the system accuracy. The performances of MFCC-SVM, MFCC-PCA-SVM, MFCC-KPCA_Gaussian-SVM and MFCC-KPCA_Polynomial-SVM system have been evaluated for this purpose. Based on EER evaluation, the best performance has been observed using KPCA_Gaussian by using linear SVM as the classifier. For future research, the improvement based on speed up algorithm on kernel calculation should be considered.

Acknowledgments

The authors would like to thank the financial support provided by Universiti Sains Malaysia under Research University Grant 814161 for this project.

7. Conclusion

As the processing time is critical in running the real time system, this study evaluated the data reduction based on principle component analysis for

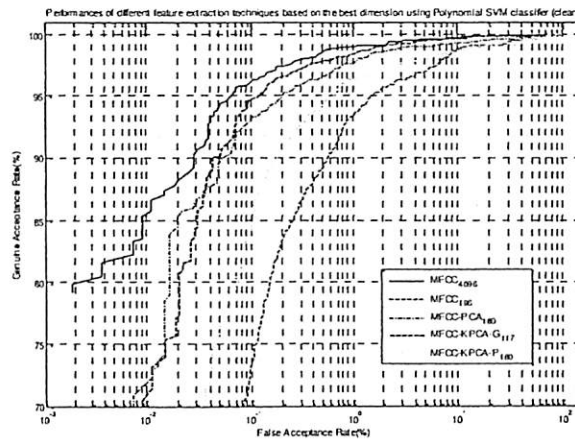


Fig. 4: Receiver operating curve (ROC) based on the selected feature dimensions for linear SVM classifier

Table 4: EER performances based on the selected feature dimensions for linear SVM classifier

Feature Extraction Technique	MFCC _{49%6}	MFCC _{19%6}	MFCC-KPCA ₁₀	MFCC-KPCA-Gauss ₁₁₇	MFCC-KPCA-Poly ₁₈₀
EER (%)	0.9685	3.4797	15888	1.2885	2.5375

References

Amaro L., Heiga Z., Nankaku Y., Miyajima C., Tokuda K., and Kitamura T.; 2004. On the use of kernel PCA for feature extraction in speech recognition. *IEICE Transactions on Information and Systems*, 87(12), pp. 2802-2811.

Chen S.H. and Luo Y.R.; 2009. Speaker verification using MFCC and support vector machine. *Proceedings of the International Multi Conference of Engineers and Computer Scientists*: 18-20.

Davis S. and Mermelstein P.; 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustic Speech and Signal Processing*, 28(4), pp. 357-366.

Ganchev T., Fakotakis N. and Kokkinakis G.; 2005. Comparative evaluation of various MFCC implementations on the speaker verification task. *Proceedings of the SPECOM 1*: 191-194.

Hasan M. R., Jamil M. and Rahman M.G.R.M.S.; 2004. Speaker identification using Mel frequency cepstral coefficients. *3rd International Conference on Electrical & Computer Engineering (ICECE 2004)*: 565-568.

Huang S.Y., Yeh Y.R. and Eguchi S.; 2009. Robust kernel principal component analysis. *Neural Computation*, 21(11), pp. 3179-3213.

Ishak K.A., Samad S.A. and Hussain A.; 2006. A face detection and recognition system for intelligent vehicles. *Information Technology Journal*, 5(3), pp. 507-515.

Ittichaichareon C., Suksri S. and Yingthawornsuk T.; 2012. Speech recognition using MFCC. *International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012)*: 135-138.

Lee S.M., Fang S.H., Hung J.W. and Lee L.S.; 2001. Improved MFCC feature extraction by PCA-optimized filter-bank for speech recognition. *IEEE Workshop on Automatic Speech Recognition and Understanding*: 49-52.

Leitner C., Pernkopf F. and Kubin G.; 2011. Kernel PCA for speech enhancement. *INTERSPEECH*: 1221-1224.

Mishra P. and Agrawal S.; 2012. Recognition of voice using Mel cepstral coefficient & Vector Quantization. *International Journal of Engineering Research and Applications (IJERA)* 2012, 2(2), pp. 933-938

Rodríguez J.M.C., de Paz F., Rocha M.P. and Riverola F.F.; 2008. Improving a leaves automatic recognition process using PCA. *2nd International Workshop on Practical Applications of Computational Biology and Bioinformatics (IWPACBB 2008)*: 243-251.

Sanderson C. and Paliwal K.K.; 2001. Noise compensation in a multi-modal verification system. *International Conference on Acoustics, Speech, and Signal Processing*: 157-160.

Turk M.A. and Pentland A.P.; 1991. Face recognition using eigenfaces. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1991*: 586-591.



Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 00 (2015) 000–000

Procedia

Environmental Sciences

www.elsevier.com/locate/procedia

International Conference on Environmental Forensics 2015 (iENFORCE2015)

Robust Syllable Segmentation Of The Automatic Frog Calls Identification System

Haryati Jaafar^a, Dzati Athiar Ramli^a

^a*School of Electrical and Electronic, Universiti Sains Malaysia Engineering Campus, Nibong Tebal, Pulau Pinang 14300, Malaysia*

Abstract

The automatic frog sound identification system is one of the most useful approaches to assist experts in identifying frog species and to replace manual techniques claimed to be costly and time-consuming. However, to execute an automatic system in a noisy environment due to background noise is a challenging task. Hence, more robust syllable segmentation techniques are required. In this paper, a combination of enhanced starting and end points detection namely short time energy (STE) and short time average zero crossing rates (STAZCR) is proposed to improve the syllable segmentation. There were fifteen frog species from the Malaysian forest were employed in this study. To validate the performance of the STE and STAZCR, a comparison of the syllable segmentation techniques based on time-frequency domain i.e. sinusoidal modelling (SM) and time domain i.e. Energy and Zero Crossing Rate (E+ZCR) were employed. The experimental results demonstrated that the STE+STAZCR technique is able to obtain 96.27% performance compared to the other techniques i.e. SM and E+ZCR which only achieved 88.53% and 89.97% respectively.

© 2015 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of organizing committee of Environmental Forensics Research Centre, Faculty of Environmental Studies, Universiti Putra Malaysia.

Keywords: Frog calls identification system; syllable segmentation; time-frequency domain; time-domain; peak finding algorithm

1. Introduction

Frogs are often the most abundant, diverse group of vertebrate organisms in forested or high trophic levels, and in many systems, they are considered as the top predators¹. This amphibian is also considered to be a bio-indicator of environment stress. The health of the frog population indicates the health of the whole ecosystem due to their bi-phasic life². These phenomena which can be observed in our surroundings can become signs to the environmental disturbances. Frogs have survived for the past 250 million years in countless ice ages, asteroid crashes and other environmental disturbances but yet, one-third of these amphibian species are on the verge of extinction nowadays.

1878-0296 © 2015 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of organizing committee of Environmental Forensics Research Centre, Faculty of Environmental Studies, Universiti Putra Malaysia.

So, this should be served as an alarm call to humans that if there is a presence of disturbance in our environment. Because of their importance to the ecosystems and the ability to indicate the environmental stress, research involving frog identification systems based on their calls is warranted to preserve the world from frog species elimination. Although a few frog species are flourishing in human environments and has been adapting to the noise, many species have suffered dramatic population declines. Therefore, studies on frog species recognition have become crucial as frogs also play an important role in the ecological system.

Generally, frog uses acoustic communication for a wide range of essential functions, not only for territorial defence and mating ritual, but also for navigation, nurturing, detection of predators and foraging¹. Their communication can occur over varying distances that allows an obstructive detection of their existence³. Therefore, identifying frog species based on their calls is more effective for environmental monitoring. Commonly, the frog calls are represented as a sequence of syllables. A syllable is basically a sound that frog produces with a side blow of air from the lungs. Compared to the human, the syllables of frog seem to be slightly less complex than the human's due to non-vowel-consonant syllables and less intricate grammar⁹. However, it typically has much more variety sounds even among the same species to show their mating readiness, defend their territory and communicate with each other. Hence, a syllable segmentation is considered to be the beginning of the processing step in frog calls identification system³.

So far, there have been many techniques in signal segmentation are proposed in the literature, which can be broadly categorized as time-frequency and time domains. One of the best-known time-frequency techniques is the spectrogram and was used in various applications⁴. However, it can be successfully implemented in clear condition and it is also highly demanded in data storage and computation⁷. To increase the discriminative power of the features under noisy condition, many approaches have been proposed. For example, Harma⁷ proposed the Sinusoidal Modelling (SM) technique to segment the syllables of continuous bird song. Following on the successful SM technique in the bird song segmentation, this technique was applied in different bioacoustics application^{8,9}. Meanwhile, the Energy-based segmentation technique has been used mostly in the time domain due to their simplicity and easiness in the segmentation process. To improve the noise robustness, the energy technique was combined with Zero Crossing Rate (ZCR)¹⁰. This method has been employed to segment the sound in animal sound^{11,12}.

Nonetheless, identifying particular frog calls becomes challenging in a case where background noise often interferes with the process. In the real condition, the recordings may be corrupted by stationary and non-stationary background noise. Since the frog sounds are recorded in a real environment that is normally corrupted by non-stationary background noise, these methods lead to detection errors¹³. For the case of the stationary background noise, the existing techniques are still reliable. However, to deal with non-stationary noise, these techniques are no longer appropriate. This is because these techniques can only achieve good accuracy in high signal to noise ratio (SNR) environments but they will fail in low SNR environments¹⁴. Therefore, an improvement from the previous segmentation methods has been investigated to give greater performance in syllable detection in the cases of both stationary and non-stationary background noises. This paper proposes two techniques, namely Short Time Energy (STE) and Short Time Average Zero Crossing Rate (STAZCR) to overcome the above problem. The STE was used to estimate the initial syllable boundaries while the STAZCR was employed to refine these boundaries. Both STE and STAZCR were combined together to determine the starting and end point detections to detect the region of interest for the syllables. At the same time, it should be able to exclude the background noise and syllables which were not in the same group of the syllables of interest. This paper is outlined as follows. In Section 2, the architecture of the frog identification system contains the background of the study. The proposed syllables segmentation technique is explained in Section 3. Section 4 describes the experimental results and discussion. Finally, conclusions are summarized in Section 5.

2. Data acquisition

The frog sounds were collected from two sites in the Malaysian forest in the state of Kedah. The first site is located in Sungai Sedim, Kulim and the sounds were recorded next to a running stream from 8.00 pm to 12.00 pm. The second site is located in Baling where the frog sounds were recorded in a swampy area from 6.00 pm to 10 pm.

The recordings were made using a Sony Stereo IC Recorder ICD-AX412F supported by a Sony electric condenser microphone 32-bit, 32kHz sampling frequency with WAV format. They were fifteen species are obtained from five families which are *Hylarana glandulosa* (rough sided frog), *Kaloula pulchra* (asian painted bullfrog), *Kaloula baleata* (flower pot toad), *Microhyla heymonsi* (Taiwan rice frog) and *Microhyla butleri* (painted chorus frog) from Microhylidae family, *Phrynoidis aspera* (river toad), *Duttaphrynus melanostictus* (black-spectacled toad) and *Genus ansonia* (stream toad) from Bufonidae family, *Odorrana hossi* (poisonous rock frog) and *Hylarana labialis* (white-lipped frog) from Ranidae family, *Polypedates leucomystax* (common tree frog), *Philautus mjobergi* (bubble-nest frog), *Rhacophorus appendiculatus* (frilled tree frog) and *Philautus petersi* (kerangas bush frog) from Rhacophoridae family and *Fejervarya limnocharis* (grass frog) from Dicroglossidae family⁶.

3. Syllable segmentation

The proposed technique is first implemented by the framing process where the signals are converted into frames where each frame had the same number of samples with a frame size of 20ms and each frame overlaps by 10ms. Therefore, the number of samples in a frame was set to 640 and number of samples for frame shift was set to 320 samples. The windowing process is then applied to minimize the signal discontinuities at the beginning and end of each frame by zeroing out the signal outside the region of interest. In this study, the Hamming window is used as the window function due to the side lobes of this window being lower compared to other windows. Moreover, the hamming window gives much attenuation outside the bandpass than other comparable windows¹⁵. The window is defined by the expression below:

$$w(k) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi k}{N-1}\right) & k = 0, \dots, N-1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $w(k)$ is the window function and N is the length of each frame.

The signal after framing and windowing process is given as:

$$x_w(m) = x(k)w(m-k) \quad (2)$$

where $x_w(m)$ is the input signal in one frame, m is the temporal length of each frame and the operator $w(m-k)$ represents a frequency shifted window sequence, whose purpose is to select a segment of the sequence $x(k)$ in the neighbourhood of sample $m=k$.

By introducing the framing and windowing processes, the STE function is defined by the following expression:

$$E_m = \frac{1}{N} \sum_{k=1}^N [x(k)w(m-k)]^2 \quad (3)$$

where E_m is the function which measures the change of voice signal amplitude.

If the STE of the incoming frame is high, the frame is classified as a voiced signal frame and if the STE of the incoming frame is low, it is classified as an unvoiced signal frame.

On the other hand, the STAZCR is defined as:

$$Z_m = \frac{1}{2N} \sum_{k=1}^N |\text{sgn } x(k) - \text{sgn}[x(k-1)]| w(m-k) \quad (5)$$

where Z_m is the function which defines the zero crossing count.

If the STAZCR is high, the frame is considered to be an undesired signal and if it is low, the frame is considered to be a desired signal frame.

In order to ensure that the start and end point detection of the syllable performs well, threshold levels need to be set. Here, the peak finding algorithm is proposed to iteratively narrow the searching numbers of local minima and maxima before determine the threshold level. In this algorithm, the potential points are firstly determined. This is done by calculating the first derivative of E_m and Z_m . The potential points can be detected by considering the sign of the difference. A change from negative to positive number corresponds to a local maximum and a change from positive to negative corresponds to a local minimum. Subsequently, a selective point is used to ensure the local maxima is selected at least $\frac{1}{4}$ of the range of the data. For the STE, this point is given as:

$$E_{ms} = \frac{E_{m\max} - E_{m\min}}{4} \quad (6)$$

where E_{ms} is the selective point for the STE, $E_{m\max}$ is the maximum value of the STE and $E_{m\min}$ is the minimum value of the STE.

Meanwhile, the selective point for the STAZCR is given as:

$$Z_{ms} = \frac{Z_{m\max} - Z_{m\min}}{4} \quad (7)$$

where Z_{ms} is the selective point for the STAZCR, $Z_{m\max}$ is the maximum value of the STAZCR and $Z_{m\min}$ is the minimum value of the STAZCR.

The next step is to decide whether the potential points can be selected as local maxima or otherwise. The potential point is considered as local maxima if the point is satisfied with the condition written in Equations (8) and (9).

$$E_{mp} > E_{ms} + E_{mr} \quad (8)$$

where E_{mp} is the value of the STE at current test point and E_{mr} is the value at the reference point.

$$Z_{mp} > Z_{ms} + Z_{mr} \quad (9)$$

where Z_{mp} is the value of the STAZCR at current test point and Z_{mr} is the value at the reference point.

Initially, the minimum value of the STE and STAZCR is set as reference point. However, if the local maxima is found, the new reference point is selected. After determining the local maxima, the threshold level is given as:

$$T_h = \begin{cases} 1, E_m \geq T_E \text{ and } Z_m \geq T_Z \\ 0, \text{otherwise} \end{cases} \quad (10)$$

where T_E and T_Z are the thresholds for STE and STAZCR, respectively and they are defined as:

$$T_E = \frac{W(E_{m\max,1}) + E_{m\max,2}}{W + 1} \quad (11)$$

$$T_Z = \frac{W(Z_{m\max,1}) + Z_{m\max,2}}{W + 1} \quad (12)$$

where W is the weight parameter, $E_{m\max,1}$ and $Z_{m\max,1}$ are the first local maximum values while $E_{m\max,2}$ and $Z_{m\max,2}$ are the second local maximum values of frequency distribution, and T_E and T_Z are the threshold level for the STE and STAZCR respectively. We observe that the large values of W obviously lead to the threshold values being closer to the maximum values of STE and STAZCR. In this study, the best value of W obtained was 5.

Fig. 1 illustrates the outputs of the segmentation by using the STE and STAZCR techniques. The red line in Figs. 1(a) and 1(b) is the threshold level for the STE and STAZCR respectively. In the meantime, the boundary line in Figure 3.8(c) is marked to indicate that the signal is detected. It can be observed the threshold level plays an important role to determine the starting and ending point of the STE and STAZCR.

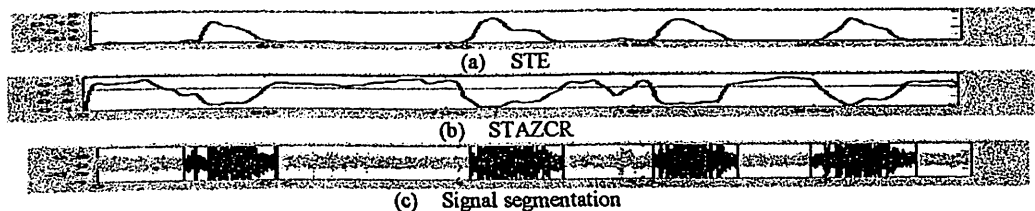


Fig. 1. Signal segmentation process

4. Experimental results

The proposed methods have been implemented in Matlab R2009 (b) and have been tested in Intel Core i7, 2.1GHz CPU, 2G RAM and the Windows 8 operating system. The results of STE+STAZCR technique was

compared with the SM⁷ and E+ZCR⁹ techniques. They were ran in two experiments. The first experiment discusses results of signal segmentation on the audio signal based on subjective and objective evaluations. The subjective evaluation is investigated by visual interpretation for each technique while the objective evaluation is implemented by comparing the results based on five parameters which are CORRECT, Front End Clipping (FEC), Mid Speech Clipping (MSC), OVER, Noise Detected as Speech (NDS)⁵. In this experiment, there were 10 calls from each species were tested. The second experiment discusses the performance comparison of the SM, E+ZCR and STE+STAZCR techniques in the classification results. In this part, the audio features were extracted by using Mel frequency cepstral coefficient (MFCC) and the feature dimension size was fixed at 4096 (64×64). The k nearest neighbor (kNN) classifier was used and the value of k was set to 5. For each species, 20 syllables have been used for training and 25 syllables have been used for testing. Thus, the total 300 and 375 syllables from 15 different species were used for the training and testing purposes respectively. This experiment was evaluated in terms of Classification Accuracy (CA) as in Equation (13).

$$CA = \frac{N_c}{N_T} \times 100\% \tag{13}$$

where N_c is the number of syllables which is recognized correctly and N_T is the total number of test syllables.

4.1. Performance results based on the subjective and objective evaluations

Fig. 2 presents the signal segmentation results for a *Polypedates leucomystax* in the real environment. In the SM technique, the results indicate that the majority of the frog calls were detected within the insertion and clipping errors. Meanwhile, the insertion errors i.e. OVER were present when the signal was segmented by E+ZCR technique. The results reveal that the STE+STAZCR can reduce both the insertion and clipping errors compared to the SM and E+ZCR techniques.

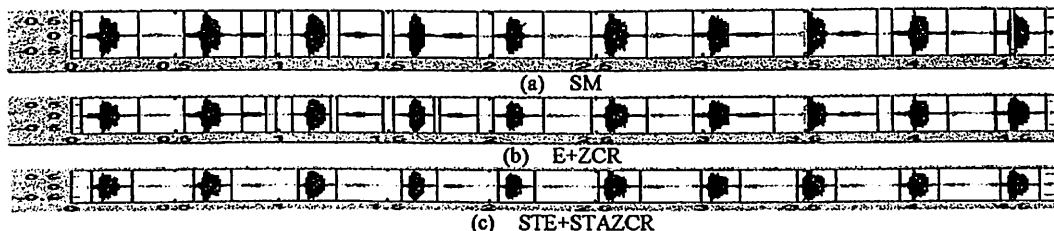


Fig.2. Comparison of syllable segmentation for a *Polypedates leucomystax*

A comparison of the performances of the SM, E+ZCR and STE+STAZCR techniques based on objective evaluation are presented in Table 1. This table reveals the STE+STAZCR has achieved the highest correct detection rate compared to other techniques. As observed in the table, the SM exhibited the highest clipping errors (FEC+MSC) of 1.775% followed by the E+ZCR of 0.67%. They revealed that the STE+STAZCR had the lowest clipping errors rate compared to SM and E+ZCR with 0% was achieved. Meanwhile, the insertion errors (NDS+Over) were found in all techniques. Similar results were observed in the SM where this technique achieves the highest error rate with 5.115% followed by the E+ZCR with 5.11%. On top of these, the STE+STAZCR was able to reduce the effect of the clipping errors compared to SM and E+ZCR techniques with 0% was achieved.

Table 1. Signal segmentation performances based on objective evaluation

Error	Parameter	SM	E+ZCR	STE+STAZCR
	CORRECT (%)	86.22	88.44	94.45
Clipping errors	FEC (%)	1.33	0.67	0
	MSC (%)	2.22	0.67	0
Insertion errors	NDS (%)	4.67	2.67	0.44
	OVER (%)	5.56	7.55	5.11

4.2. Performance results based on the classification accuracy

The comparison analyses in term of CA performances on various noises in different levels of SNRs are shown in

Table 2. By using the CA calculation, it was found that the SM and E+ZCR were unable to achieve more than 90% of the CA rates. Overall, the CA rates for both techniques exhibited slight similarity. The STE+STAZCR provided better CA rate compared to the SM and E+ZCR techniques with the CA rate was 96.27%. It concluded that despite in both stationary and non-stationary environments, the proposed technique performed well in terms of CA.

Table 2. Signal segmentation performances based on the CA

Technique	SM	E+ZCR	STE+STAZCR
CA(%)	88.53	89.97	96.27

5. Conclusion

Frogs play an important role in ecosystems since this species controls the insect population, and is a food source for certain animals. Since frogs bring many advantages to ecosystems, an automatic syllable segmentation for frog calls is needed. In this paper, the frog sound identification system based on syllable segmentation has been studied and successfully implemented. This study also introduced the combination of the STE and STAZCR to detect the syllables, which can increase the accuracy of the start and end point detection of the syllables. The results were evaluated based on the objective evaluation and CA. The proposed method has been compared with the baseline method which is the combination of energy and ZCR. It was found the proposed STE and STAZCR outperform other technique and was correctly detect edthe syllables with 94.45% and classified the frog sound with 96.67%.

Acknowledgements

The authors would like to express their gratitude for the financial support provided by Universiti Sains Malaysia Research University Grant 814161 and Research University-Post Graduate Grant Scheme 8046019 for this project.

References

1. Beebe TJ, Griffiths RA. The amphibian decline crisis: a watershed for conservation biology?. *Biological Conservation* 2005;125:3, 271-285.
2. Southerland MT, Jung RE, Baxter DP, Chellman IC, Mercurio G, Volstad JH. Stream salamanders as indicators of stream quality in Maryland, USA. *Applied Herpetology* 2004; 2:1, 23-46.
3. Acevedo MA, Corrada-Bravo CJ, Corrada-Bravo H, Villanueva-Rivera LJ, Aide TM. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics* 2009; 4:4, 206-214.
4. Costa DC, Lopes GAM, Mello CA, Viana HO. Speech and phoneme segmentation under noisy environment through spectrogram image analysis. In: *IEEE International Conference on Systems, Man, and Cybernetics* 2012; 1017-1022.
5. Beritelli F, Casale S, Ruggeri G, Serrano S. Performance evaluation and comparison of G. 729/AMR/fuzzy voice activity detectors. *IEEE Signal Processing Letters* 2002; 9:3, 85-88.
6. Jaafar H, Ramli, DA, Rosdi BA, Shahrudin S. Frog identification system based on local means k-nearest neighbors with fuzzy distance weighting. In: *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications* 2014; 153-159.
7. Harma, A. Automatic identification of bird species based on sinusoidal modeling of syllables. In: *Proceedings of Acoustics, Speech, and Signal Processing* 2003; 5, 545-548.
8. Neal L, Briggs F, Raich R, Fem, XZ. Time-frequency segmentation of bird song in noisy acoustic environments. In: *IEEE International Conference on Acoustics, Speech and Signal Processing* 2011; 2012-2015.
9. Huang CJ, Yang YJ, Yang DX, Chen YJ. Frog classification using machine learning techniques, *Expert System with Applications* 2009; 36, 3737-3743.
10. Chen WP, Chen SS, Lin C. Automatic recognition of frog calls using a multi-spectrum average spectrum. *Computer and Mathematics with Application* 2012; 64:5, 1270-1281.
11. Ganchev T, Mporas I, Jahn O, Riede K, Schuchmann KL, Fakotakis N. Acoustic bird activity detection on real-field data. *Artificial Intelligence: Theories and Applications* 2012; 190-197.
12. Agranat, I. Method and apparatus for automatically identifying animal species from their vocalizations. *U.S. Patent No. 7,454,334*, Washington, DC: U.S. Patent and Trademark Office; 2008.
13. Guo C, Li R, Fan M, Liu, K. Research on voice activity detection in burst and partial duration noisy environment. In: *International Conference on Audio, Language and Image Processing* 2012; 991-995.
14. Shin, JW, Chang JH, Kim NS. Voice activity detection based on a family of parametric distributions. *Pattern recognition letters* 2007; 28:11, 1295-1299.
15. Rabiner LR, Schafer RW. Introduction to digital speech processing. *Foundations and trends in signal processing* 2007; 1:1, 1-194.

The Local Histogram Equalization And Adaptive Thresholding for Hand-Based Biometric Systems

Haryati Jaafar, Salwani Ibrahim and Dzati Athiar Ramli

Abstract— Hand-based biometric systems are the emerging type of biometrics that attracts researchers in biometrics area. As compared to the other biometric traits such as face and iris, the image quality of a hand-based system are robust with more information can be employed even though it is in low resolution. A new approach image enhancement and segmentation called the local histogram equalization and adaptive thresholding (LHEAT) was proposed to improved the quality of image taken. It was firstly obtained to ensure an equal distribution of the brightness levels. The useful information of the image was then extracted and the foreground from the nonuniform illumination background was separated. The sliding neighborhood operation was also applied such that the computation is much faster. Three hand-based biometric databases i.e. the fingerprint, finger vein and palm print databases were employed and evaluated based on the quality of image and classification accuracy (CA). Experimental evaluation based on quality of image shows that the proposed LHEAT has better performance than local histogram equalization (LHE) and local adaptive thresholding (LAT) with more than 45 of peak-signal-to-noise ratio (PSNR). The results also shows that the proposed LHEAT is able to achieved more than 90% in term of CA. This shows that the proposed LHEAT is able to enhance and segmented the images effectively.

Keywords— Hand-based biometric system, LHEAT, LHE, LAT, sliding neighborhood

I. INTRODUCTION

TODAY'S complex demands for reliable authentication and identification methods are increasing rapidly. Initially, the traditional technologies such as personal identification number (PIN), smart cards and passwords were introduced [1]. However, they had a number of inherent disadvantages such as duplication, misplacing and hacking. Therefore, biometrics were introduced in the late 90s to

recognize a person based on the physiological or biological characteristics [2]. The biometric technology is inherently more reliable. It is capable to provide a level of assurance for the preventions of duplication, stealing and hacking. Due to the specific physiological or behavioral characteristics that are possessed by the users, this technology is able to be implemented in various fields such as door access controls, criminal investigations, logical access points and surveillance applications [3].

There are various kinds of modalities of the biometric systems that are either widely used or developed such as the fingerprint, iris, face, hand geometry, palm print, gait, voice and signature [1]. Among the available biometrics, hand-based systems such as the finger vein, fingerprint and palm print are found to be the most popular due to their high user acceptance and excellent advantages in their application [4].

The features of the finger vein are inside the skin surface, which makes it difficult to be duplicated. Thus, it is more secure compared to other modalities and leads to the high recognition accuracy. In addition, as the veins are located inside the body, it is less likely to be influenced by changes in the weather or physical condition of the individual. Moreover, the rushes, cracked and rough skin does not affect the result of recognition [11]. On the other hands, the images quality of a fingerprint and palm print are robust because of its multiple lines, wrinkles and ridges while the palm print covers even more information and the ridge structures remain unchanged throughout the life, except for a change in size [12]. Currently, they are offering low costs for data acquisition and the possibility of acquiring the data easily. The image can be collected in the real environment where the acquisition devices had no pegs holding the finger or palm.

However, the main problem with hand-based images is that they are of low quality due to several reasons such as the movement of hands, use of low resolution capturing devices and environmental factors. These factors obscure image details and create noise which badly effect object detection and recognition. Commonly, the problem of suppression of noise in these images is solved by a smoothing technique [5]. However, this process has the potential to blur all sharp edges containing an important information about the image [6]. In order to overcome this problem, a combination of image enhancement and segmentation techniques is found to be more appropriate in such ways. Hence, this paper proposed a contrast image enhancement and image segmentation by

This work was supported in part by Research University Grant 814161 and Research University-Post Graduate Grant Scheme 8046019.

H. Jaafar, is with Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia (e-mail: haryati.jaafar@yahoo.com).

S. Ibrahim, is with Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia (e-mail: salwani.ibrahim@gmail.com).

D.A. Ramli is with Intelligent Biometric Group, School of Electrical and Electronic, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia (corresponding author to provide phone: +604-5996028; e-mail: dzati@usm.my).

introducing the local histogram equalization and adaptive thresholding (LHEAT) technique. This technique is an improved version of the local histogram equalization (LHE) and local adaptive thresholding (LAT) techniques [7, 8]. In the LHEAT, the LHE was firstly obtained to ensure an equal distribution of the brightness levels. The LAT was employed to extract the useful information of the image that had been enhanced by the LHE and separated the foreground from the nonuniform illumination background. In addition, the sliding neighborhood operation was applied such that the computation is much faster. This is an advantage of the LHEAT on reducing the time processing of the image enhancement stage compared to the baseline LHE and LAT techniques. The rest of this paper is organized as follows: The proposed LHEAT technique is described in Section II. The experimental set up and results are explained in Section III, and this paper is concluded in Section IV.

II. THE PROPOSED LOCAL HISTOGRAM EQUALIZATION AND ADAPTIVE THRESHOLDING

The flowchart of the LHEAT techniques is shown in Fig. 1.

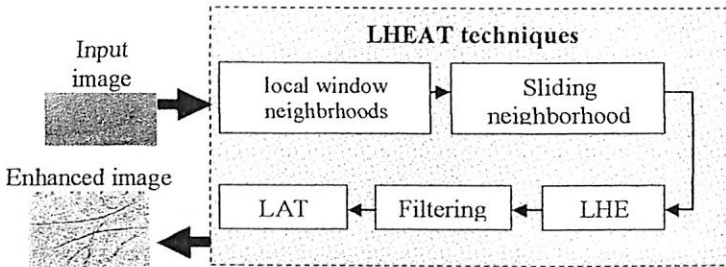


Fig. 1 A flowchart of the proposed fingerprint enhancement algorithm.

An input image was first broken into small blocks or local window neighborhoods containing a pixel. This was similar in the LHE, LAT and LHEAT. Each block was surrounded by a larger block. The input image was defined as $X \in R^{H \times W}$, with dimensions of $H \times W$ pixels, and the enhanced image was defined as $Y \in R^{H \times W}$, with $H \times W$ pixels. The input image was then divided into the block $T_i = 1, \dots, n$ of window neighborhoods with the size $w \times w$, where $w < W, w < H$ and $n = \left\lceil \frac{H \times W}{w \times w} \right\rceil$.

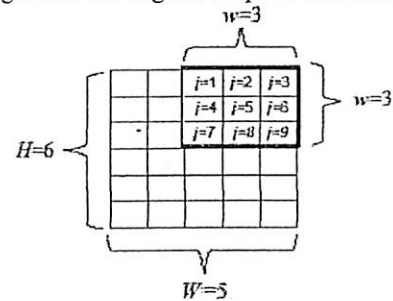
Each pixel in the small block was calculated using a mapping function and threshold. The size of w should be sufficient to calculate the local illumination level, both objects and the background [9]. However, it led to a complex computation which can be reduced by employing the sliding neighborhood. This operation can also decrease the acceleration of the computation. Fig. 2 shows an example of the sliding neighborhood operation. An image with a size of 6×5 pixels was divided into blocks of window neighborhoods

with a size of 3×3 pixels. It is shown in Fig. 2(a). The 6×5 image matrix was first rearranged into a 30 column ($6 \times 5 = 30$) of temporary matrix, as shown in Fig. 2(b). Each column contained the value of the pixels in its nine rows ($3 \times 3 = 9$) window. The temporary matrix was then reduced by using the local mean (M_i):

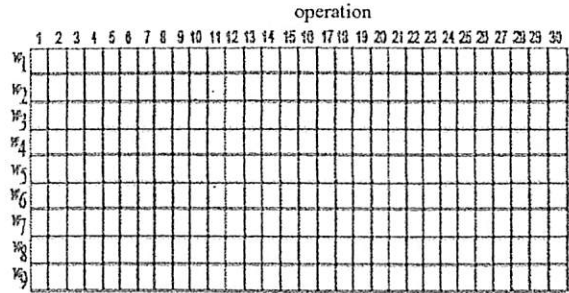
$$M_i = \frac{1}{N} \sum_{j=1}^n w_j \tag{1}$$

where w was size of window neighborhoods, j was the number of pixels contained in each neighbourhood, i was the number of column in temporary matrix and N was the total number of pixels in the block.

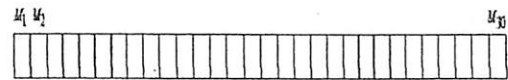
After determining the local mean in Equation (1), there was only one row left as shown in Fig. 2(c). Subsequently, this row was rearranged into the original shape as shown in Fig. 2(d).



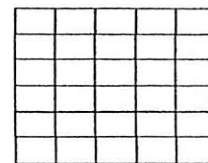
(a) Original image with window neighborhoods



(b) Temporary matrix



(c) One row matrix



(d) Rearranged row into the original shape

Fig. 2 The sliding neighborhood

The LHE was then obtained to ensure an equal distribution of the brightness levels. There are three major steps in the LHE technique. There are the probability density (PD), the cumulative distribution function (CDF) and the

mapping function. The probability distribution of image PD for each block can be expressed as:

$$P(i) = \frac{n_i}{N} \text{ for } i = 0, 1, \dots, L-1 \quad (2)$$

where n_i is the input pixel number of level, i is the input luminance gray level and L is gray level, which was 256 in the investigated case.

The LHE uses an input-output mapping that is derived from CDF of the input histogram as defined in below:

$$C(i) = \sum_{i=0}^n P(i) \quad (3)$$

Although the image has been enhanced, it remains mildly degraded because of the background noise and variation in contrast and illumination. Hence, the 2D median filter, containing a 3×3 mask was applied over the grayscale image to reduce the effect of salt and pepper noise and the blur of the edge of the image. Given an input vector is $x(n)$ and $y(n)$ is the output median filter of length l where l defines the number of samples over which median filtering takes place. When l is odd, the median filter can be defined as stated in Equation (4).

$$y(n) = \text{median}\{x(n-k) : n+k, k = (l-1)/2\} \quad (4)$$

When l is even, the mean of the two values at the center of the sorted samples list is used.

Once it was filtered, the image was segmented using the LAT technique to separate the foreground from the background by converting the grayscale image into binary form. The Sauvola's technique was applied here due to its promising effects on the degraded images. By using the Sauvola's technique, the following formula for the threshold is:

$$T_h(i) = M \left[1 + k \left(\frac{Z}{R} - 1 \right) \right] \quad (5)$$

where T_h is the threshold, k is a positive value parameter with $k = 0.5$, R is the maximum value of the standard deviation, which was set at 128 for grayscale image and Z is the standard deviation which can be found as:

$$Z = \sqrt{\frac{1}{N-1} \sum_{j=1}^n (w_j - M)^2} \quad (6)$$

The binarization results can be denoted as follows $y(i)$ as in Equation (7)

$$y(i) = \begin{cases} 1 & \text{if } q(i) > T_h(i) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

As mentioned before, a suitable value for window size, w was greatly affected the image. If the window size is too small, the image resulted segmented regions appear was less visible. Meanwhile the large window size had caused the important details of the image was disappeared. In this study, the best value of w is obtained by the $w=11$ for the finger vein and fingerprint databases and $w=9$ for the palm print database.

III. EXPERIMENTAL RESULTS

In this section, a comparative study on the performance of LHEAT technique on the hand-based biometric database had been investigated and compared with the LHE and LAT techniques. The experiments were implemented using Matlab R2010 (b) and were tested in Intel Core i5, 2.1GHz CPU, 6G RAM and Windows 7 operating system.

A. Data Acquisition

The finger vein database was provided by the IBG, USM. It is available for downloading from the following website: <http://blog.eng.usm.my/fendi/>. The capturing device was comprised of three units of Near-Infrared-light emitted diode (NIR-LED) of wavelength = 850 nm and a Sony PSEye camera with an IR passing filter. The NIR-LEDs were placed in a row on the top section while the camera was attached to the bottom side of the capturing device. To reduce the user's discomfort, the users were simply asked to place their fingers on the acquisition devices and there had no pegs holding the finger. The spatial and depth resolutions of the images were set at 640×480 pixels and 256 grey levels, respectively. The images were then segmented into the region of interest (ROI). A few examples of the ROI of the finger vein images are shown in Fig. 3.

The database was obtained from 123 volunteers who were staffs and students (83 males and 40 females) from University Sains Malaysia (USM). The range of age of the users was from 20 to 52 years old. Each user contributed four of their fingers which were the left index, left middle, right index and right middle fingers resulting in 492 finger classes for this investigation. The images were acquired in two sessions with a time gap of by more than two weeks. Each finger was captured six times in every session. There were 2952 samples extracted from the first and second sessions were used as the training and testing samples, respectively.

The fingerprint database was obtained from the Fingerprint Verification Competition 2006 (2006FVC). The image was collected by using an optical sensor with the resolution of the sensor is 569 dpi in the image format of BMP, 256 gray-levels size of 400×560 pixels. [10]. This database was collected from 150 volunteers who were randomly selected including the manual workers and elderly people. They were simply asked to place their fingers on the acquisition device. There was no constraint was enforced to guarantee the highest quality of the captured images. The final databases were selected from a larger database by choosing the fingers that were more difficult to be evaluated according to a quality index. This was done to make the benchmark sufficiently difficult for a technology evaluation. Each user had provided 12 samples per finger. Thus, the final databases collected were 1800 fingerprint images 1800 samples from 150 users and they were then partitioned in half (900 as the training samples and the other 900 for the testing samples). Some examples of the fingerprint images are shown in Fig. 4

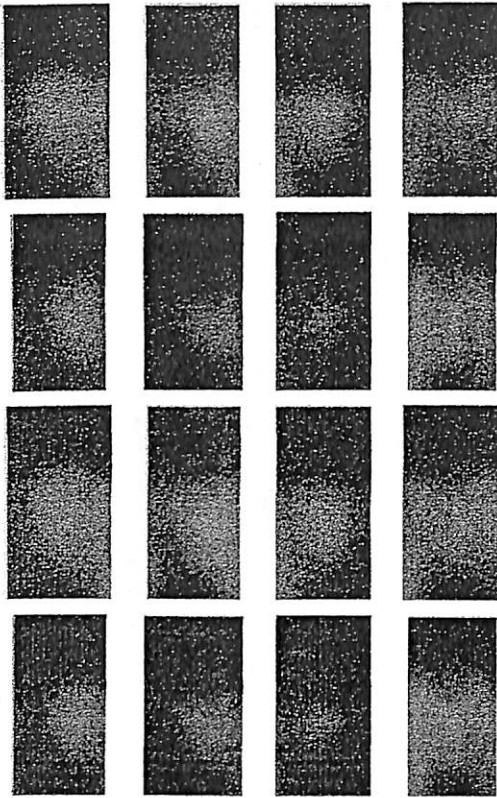


Fig. 3 The example of extracting ROI finger vein image collection

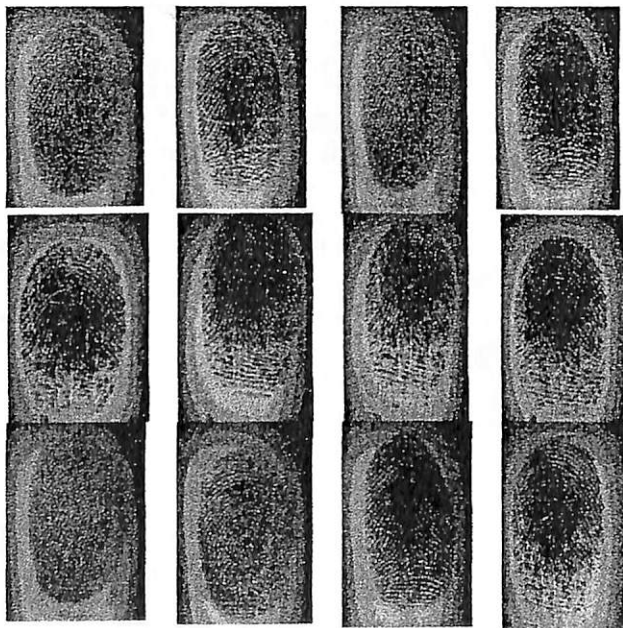


Fig. 4 The examples of fingerprint image collection

The palm print database was provided by the IBG, USM. The image was taken using a HTC One X android mobile phone with the image resolution of 8 megapixels of image resolution at a fixed background the files were saved in JPEG format. It

was obtained by collecting the palm print images from 40 users, who were the students from School of Electrical and Electronic, USM. The age range of the users were from 19 to 23 years old. The database were comprised of 60 palm print images from every user in which 20 of them were used as the training samples and the other 40 images were applied as the testing samples. The original image was then transformed into a gray scale image and extracted into the ROI. A few examples of the ROI of the palm print images are shown in Fig. 5.

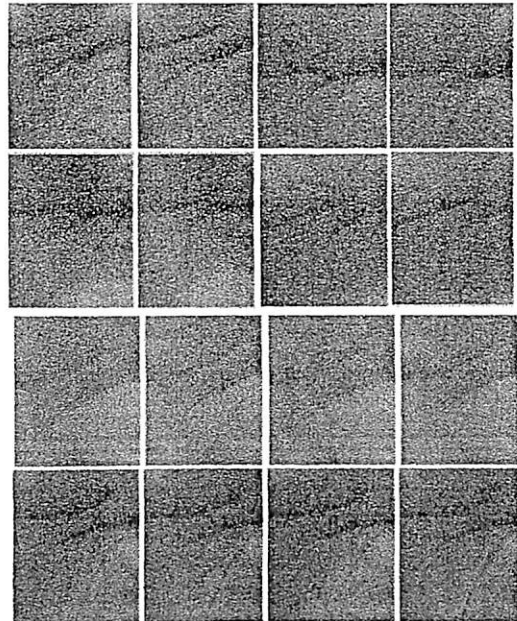


Fig. 5 The examples of extracted ROI palm print image collection

B. Performance Evaluation

In this study, the evaluation of the performance of the hand-based biometric system is based on quality of image and the CA. In the quality of image, the performance of the proposed technique was evaluated in two evaluations subjectively and objectively. The perception of an image quality improvement in the human visual system is a subjective evaluation while the perception of quantitative measures is an objective evaluation. The objective evaluation is determined based on the PSNR computation. The higher value of the PSNR the more improved is an image. PSNR is calculated using:

$$PSNR = \frac{10 \log_{10} (L-1)^2}{MSE} \quad (8)$$

where MSE can be calculated as:

$$MSE = \sqrt{\frac{\sum_{i=1}^Y \sum_{j=1}^X (m-Y)^2}{R \times C}} \quad (9)$$

where X is the original image, Y is an enhanced image, m is the intensity of the pixel at position (i,j) , R and C are the row and column of the image size.

The duration of the processing time of each method is also compared to investigate the complexity of the enhancement

approaches. The enhancement process is expected to be computed with a minimum period of run time.

In order to investigate the effectiveness of proposed technique based on the CA, the k nearest neighbor (kNN) classifier with k=5 was employed to calculate the score of the pattern matching between training and testing data of the databases. The experiments were evaluated in terms of CA such that:

$$C_A = \frac{N_c}{N_A} \times 100\% \quad (10)$$

where N_c was the correct identified number of samples and N_A was the total number of test samples.

Table I, II and III show the comparison of output results based on the quality of images for the finger vein, fingerprint and palm print, respectively. It was observed the LHEAT technique attains the best result in all conditions and exhibits the highest quality results according to visual inspection, PSNR and processing time.

For the subjective evaluation, the details in the enhanced images using LHEAT were clearer and sharper, especially in the fine details like the ridges in which they were became more visible. For the objective evaluation, the LHEAT obtained the highest value of PSNR with more than 45 compared to LHE and LAT techniques. The LHEAT gives another advantage over other methods in term of its simplicity in computation. In the proposed LHEAT technique, the time complexity is $O(n^2)$ because the sliding neighborhood is only used to obtain local mean (M) and local standard deviation (Z). Hence, the time required for LHEAT is much closer to global techniques.

TABLE I. COMPARISON OF THE LHE, LAT AND LHEAT FOR THE FINGER VEIN IMAGE

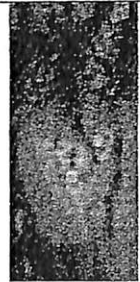


	LHE	LAT	LHEAT
Image			
Time (s)	0.436	0.976	0.141
PSNR	33.81	38.89	49.55

TABLE II. COMPARISON OF THE LHE, LAT AND LHEAT FOR THE FINGERPRINT IMAGE






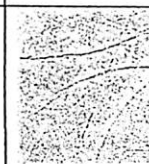
	LHE	LAT	LHEAT
Image			
Time (s)	0.551	4.766	0.133
PSNR	42.81	40.79	49.93

TABLE III. COMPARISON OF THE LHE, LAT AND LHEAT FOR THE PALM PRINT IMAGE

	LHE	LAT	LHEAT
Image			
Time (s)	2.847	3.596	0.531
PSNR	40.98	41.44	45.37

To further investigate the superiority of the proposed LHEAT, the analytical results of LHE, LAT and LHEAT in term of CA are also presented in Table IV. It was observed that the LHEAT achieves the highest CA compares to the LHE and LAT, yielding a CA of 90.93%, 93.26% and 92.6% for finger vein, fingerprint and palm print databases, respectively. It can be concluded in the LHEAT yield promising results since the brightness levels has been enhanced by distributing the brightness equally and recovered original images that were over- and under-exposed.

TABLE IV. COMPARISON OF THE LHE, LAT AND LHEAT BASED ON THE CA

	LHE	LAT	LHEAT
Finger vein	78.6%	81.07%	90.93%
Fingerprint	88.31%	83.72%	93.26%
Palm print	87.2%	82.41%	92.66%

IV. CONCLUSION

This paper focused on image enhancement and segmentation of the quality of the hand-based biometric images such as the finger vein, fingerprint and palm print. To enhance images, we propose the LHEAT technique. Because the sliding neighborhood operation is applied in the LHEAT technique, the computation was much faster compared with

previous techniques, such as LHE and LAT. Moreover, this method works well in in the real environments.

Extensive experiments were performed to evaluate the performance of the system in terms of image enhancement and image classification. The proposed system exhibits promising results. In terms of quality of the image, the PSNR with the LHEAT technique was more than 45, and the processing time was three-fold lower than with the LHE and LAT techniques. In addition, the proposed LHEAT was achieved more than 90% in term of CA. The proposed LHEAT technique is convenient and able to manage in the real environment.

ACKNOWLEDGMENT

This work was financially supported by Research University Grant 814161 and Research University-Post Graduate Grant Scheme 8046019.

REFERENCES

- [1] H. Jaafar and D. A. Ramli, "A Review of Multibiometric System with Fusion Strategies and Weighting Factor," in *International Journal of Computer Science Engineering (IJCSSE)*, vol.2, no.4, 2013, pp. 158-165.
- [2] J. P. Campbell, D. A. Reynolds and R. B. Dunn, "Fusing High-And Low-Level Features for Speaker Recognition," in *INTERSPEECH*, 2003.
- [3] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition. Circuits and Systems for Video Technology," *IEEE Transactions*, vol. 14. No. 1, 2004, pp. 4-20.
- [4] L. Zhang, D. Zhang, and H. Zhu, "Online finger-knuckle-print verification for personal authentication," *Pattern Recognition*, vol. 43, pp. 2560-2571, 2010.
- [5] D. Luo, *Pattern recognition and image processing*, Horwood Publishing Limited, West Sussex, England, 1998.
- [6] M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*. Third Ed. Thompson Corporation, USA, 2008.
- [7] H. Zhu, F. H. K. Chan and F. K. Lam, "Image Contrast Enhancement by Constrained Local Histogram Equalization," *Comput Vis Image Underst*, vol. 73, 1999, pp. 281-290.
- [8] T. R. Singh, S. Roy, O. I. Singh, T. Sinam and K. Singh, "A new local adaptive thresholding technique in binarization," *International Journal of Computer Science Issues*, vol. 8, no 2, 2011, pp. 271-277.
- [9] Y. T. Pai, Y. F. Chang and S. J. Ruan, "Adaptive thresholding algorithm: Efficient computation technique based on intelligent block detection for degraded document images," *Pattern Recognition*, vol. 43, 2010, pp. 3177-3187.
- [10] J. Fierrez, J. Ortega-Garcia, D. Torre Toledano and J. Gonzalez-Rodriguez, "Biosec baseline corpus: A multimodal biometric database," *Pattern Recognition*, vol. 40, no. 4, 2007, pp. 1389-1392.
- [11] G. K. O. Michael, C. Tee, A. T. Jin, "Touch-less palm print biometrics: Novel design and implementation," *Image Vis Comput*. Vol. 26, pp. 1551-1560, 2008.
- [12] M. S. M. Asaari, S. A. Suandi and B. A. Rosdi, "Fusion of Band Limited Phase Only Correlation and Width Centroid Contour Distance for finger based biometrics" in *Expert Systems with Applications*, 41, 2014, pp. 3367-3382.

Evaluation on palm-print ROI selection techniques for smart phone based touch-less biometric system

Salwani Ibrahim^a Dzati Athiar Ramli^a

^a Intelligent Biometric Group (IBG), School of Electrical & Electronic Engineering, USM Engineering Campus, 14300, Nibong Tebal, Penang, MALAYSIA
dzati@eng.usm.my

Abstract. There are many methods have been carried out for human recognition such as personal identification number (PIN), password or ID card but all of these methods can be guessed, hacked or stolen. Palm-print verification system is a biometric technology which is developed to authenticate person based on individual palm-print pattern. This paper presents an initial effort to perform touch-less palm-print recognition system by considering the effective way to extract the palm-print region of interest (ROI). The system starts with hand image collection using smart phone device. This project proposes two hand tracking algorithms i.e. two point method and canny method so as to detect the peak and valley of the palm. Afterwards, the desired ROI is selected and the palm-print ROI is stored in database for the evaluation of their appropriateness to be used for the touch-less palm-print recognition data.

Keywords: palm-print, biometrics, region of interest (ROI), touch-less, smart phone.

1 INTRODUCTION

Biometrics refers to an automatic verification or identification of a person based on his/her physiological and behavioral characteristics (Yih et al., 2009), (Zhang, 2004). Many types of biometric systems have been developed based on traits such as speech, face, fingerprint and many more. A palm-print verification or identification system is one of the biometric systems that use palm-print trait as features to authenticate or identify individuals (Goh et al., 2010).

Palm is the inner surface of a hand between the wrist and the fingers. The palm itself consists of principal lines, wrinkles (secondary lines) and epidermal ridges (Tabejammatt and Kangarloo, 2007). It differs to a fingerprint in that it also contains other information such as texture, indents and marks which can be considered as informative features when comparing one palm to another (Yih et al., 2009). It serves as a reliable human identifier because the print patterns are not duplicated in other people (Goh et al., 2010). Most of the palm-print biometric systems utilize scanner or Charge Couple Device (CCD) camera as the input sensor (Tabejammatt and Kangarloo, 2007), (Kasturika and Misra, 2011). However, these devices should be handling in controlled or semi-controlled environment. Furthermore, many types of equipment are involved during data collection thus these devices are limited to be used at specific places only (Goh et al., 2010), (Goh et al., 2008). Another weakness of using touch based device is the users must touch the sensor to capture their hand images. Due to the sanity issue, people are concerned to put their hand on the same sensor that may spread virus or bacteria through the device. So, in this study, a touch-less device is proposed for palm image capturing as it will be more comfortable for the users of the system, less equipment are involved during data collection and can be implemented without restriction of certain places.

Today, there are many technology devices that can be used to implement this system. In this project, android smart phone camera is used to capture the palm image where the user does not need to touch any panel or screen to avoid the hygiene and sanity issue (Julio and Shu, 2009), (Meraomia et al., 2011). For this purpose, an android application is developed and some guidelines will be displayed on the smart phone screen. So that the user will be assisted

in term of hand positioning and the correct distance between the user hand and smart phone camera. The application has been built as a user friendly tool for palm image capturing and once captured, the image will be sent via internet to server for database collection. Since this technology device is widely used nowadays, it is easy to be executed by installing the application on the android device (Julio and Shu, 2009). The main requirement is an android smart phone and wireless or 3G network will be used for server connection (Ismail and Sabri, 2010).

The entire process in developing palm-print recognition system is illustrated as in Figure 1. However, this paper only focuses on the data acquisition part for the ROI selection for touch-less palm-print recognition system. The first objective of this paper is to collect palm-print images using smart phone camera for data collection. The second objective is to implement two types of hand tracking algorithms i.e. two point methods and canny method to the palm-print images so as to detect the peak and valley of the palm. Finally, the last objective is to select the ROI of the palm based on the obtained hand peak and valley as the reference point.

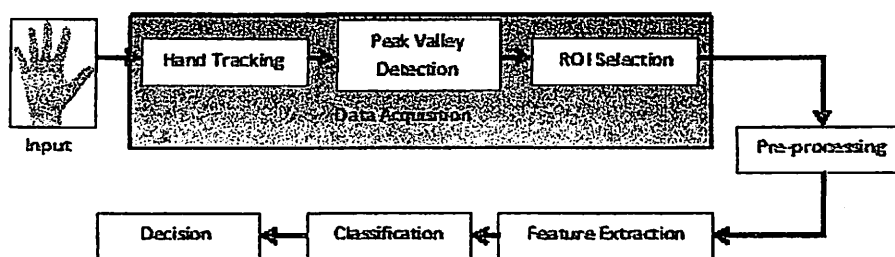


Fig. 1. The processes of the proposed palm-print recognition system.

2 METHODOLOGY

This paper proposes a touch-less palm-print recognition systems and will use phone camera for capturing hand image. On the display screen itself, we provide two points for user to control their hand, align hand into the area and capturing hand image perfectly. After that, ROI will be selected by using peak-valley detection method (Al-Kutabi et al., 2012). Some experiments have been carried out for the evaluation of the propose technique performances. There are a few challenges and limitations that need to take into account before developing the data acquisition module as discussed in the following items:

Distance between hand and input device – In order to capture good quality images, the distance between the hand and device must be in the right distance. This is because if the distance is too far, the images may come out unclear or blurred. This will cause problem while extracting the features from the images. The developed module should provide a function which is able to control the distance during image capturing. In this study, this problem is solved via developing a smart phone application by preparing hand template on the display screen. So users can adjust their hand to fit into the region.

Hand position – Position of the hand is one of the importance steps to ensure the hand tracking and peak and valley detection can be well executed. To overcome this limitation, the user is requested to put their palm in vertical position with the help of reference points on the smart phone application.

Chromatic color background – Due to the data are collected either in indoor or outdoor which may be influent by bad illumination, so the chromatic color background also one of the challenges. So that the developed module must be able to properly differentiate the skin to

background color. Moreover, a busy background should be avoided and set as a limitation to the system.

2.1 Data collection

In this experiment, the lighting and distance between hand and smart phone is set as constant. User only needs to make sure hand is aligned well and follow the point area in phone screen. Sample of 40 individuals hand images has been captured with 60 images for each subject. The experiment set up for the developed smart phone application data acquisition is shown in Figure 2.

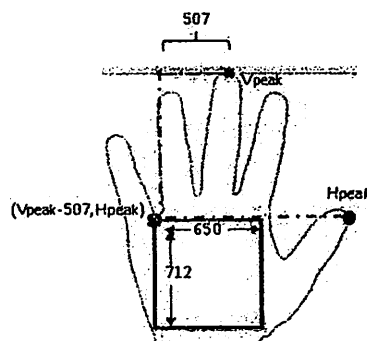


Fig.2. Experiment set up for the smart phone application.

On the smart phone screen, 2 reference points i.e. Vpeak and Hpeak are displayed for the user to control the hand and fit into the point region. This will control distance and ignore the chromatic color background. This task also requires the user to spread apart the fingers. Subsequently, the hand image obtained from smart phone camera is sent via server to the database. A standard PC with Intel Core i5 processor (2.50 Ghz) and 8.00GB random access memories are used to run MATLAB code program for pre-processing steps.

2.2 Hand tracking, peak and valley detection and ROI selection

This step consists of two stages i.e. hand image tracking and peak-valley detection in order to locate the ROI. In this study, two methods i.e. two point method and canny method have been experimented as discussed as follows:

2.2.1 Two point method

Figure 3 shows the whole process for the two point method. Before the process starts, there are few pre-conditions need to be considered:

- The lighting during capturing the image should be saturated.
- Hand position is in straight position and thumb finger should be aligned near to the palm area.

Two point method steps:

1. Originally, the hand image is represented using Red-Green-Blue (RGB) format. The image will be threshold to get a binary image (only 0's and 1's) of class logical and the 'hole fill' function is applied to the small hole so as to get the perfect image.
2. Use the 'bw boundaries' function to get the connected boundaries from the image and plot the image.
3. Get the horizontal highest peak (Hpeak) and vertical highest peak (Vpeak). The two points is plotted. To get the reference point (Vpeak-507, Hpeak) for cropping the

ROI, the calculation as in figure 2 is followed. Size of the ROI area is fixed to 650x712 pixels and result will same size for all samples.

4. From the calculation, square shape is drawn on the original image and ROI is cropped.

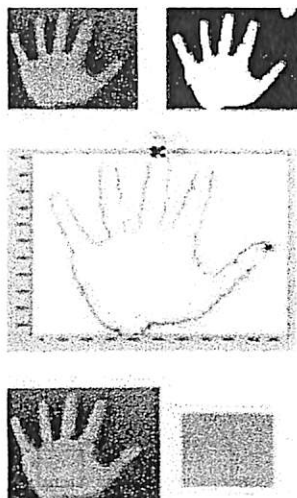


Fig.3. Two point method process.

2.2.1 Canny Method

Canny method only has one pre-condition i.e. all tips must be captured. This is to ensure that 5 peaks and 4 valleys are inside the ROI. Figure 4 and figure 5 show the whole process for the canny method.

Canny method steps:

1. First the RGB image is converted to gray scale then to binary image. Noise is then removed.
2. Use the Canny edge detection algorithm to identify edges in the image. The set of connected pixels with the largest area in the image is then hypothesized to correspond to the hand.
3. Trace hand boundary from the canny processed image. To get smoother hand boundary, the binary image need to filter out the strength noise.
4. Find and mark the peak valley (5 peaks and 4 valleys)
5. To find the peak and valley, use local minima and local maxima method. This method is accurate if the hand boundary image is smooth. Plot the peak and valley in the hand boundary image.
6. Sort peak and valley and name the points. This variable will be used in next step to calculate the ROI of the palm area.
7. For the ROI area calculation, T1, T4, P1 and P3 are used as the reference point. The ROI is located based on the intersection of tangent line drawn between T1 and P1 (first reference point) and between T4 and P3 (second reference point). Draw square shape based on the reference point. As the size of ROI varies from hand to hand (depending on the width of the hand), all images are fixed into 700 x 700 pixels.
8. Auto-crop the ROI and save the image into new database

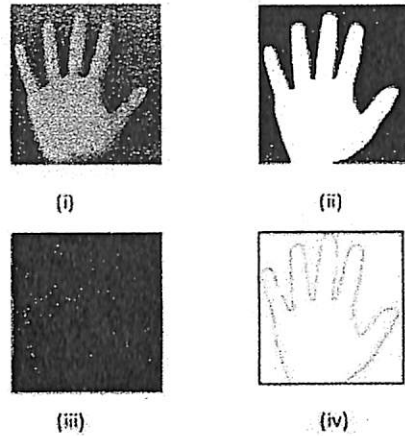


Fig.4. Hand Image Detection, (i) Original hand image, (ii) Binary image, (iii) Hand contour with Canny method, (iv) Perfect hand boundary plot.

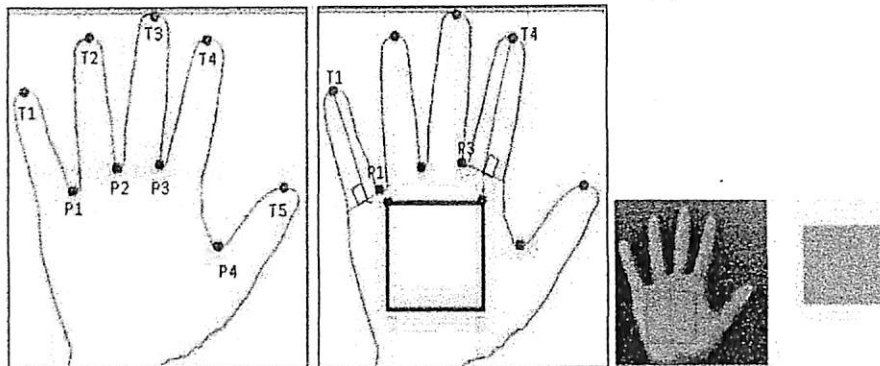


Fig.5. Five peaks and four valleys that represent the tips and roots of the fingers and the ROI

3 RESULTS AND DISCUSSIONS

Problem of the two point method:

Problem 1: If the lighting of the image is not saturated, the threshold image is missing more information due to only the brighter area is captured. So the boundaries of the image will be also affected.

Problem 2: The hand boundary image is not smooth enough and for some image, it is hard to get the perfect two points.

Problem 3: Besides, these two points method also have problem with hand positioning. If the position of the hand is too wide or the thumb finger does not align well, the wrong ROI is found.

The examples of wrong ROI obtained as discussed above are illustrated in figure 6.

Problem of the Canny method:

Problem 1: if the image captured is missing one or more fingers due to image is too closed, it will give an error because this technique requires 5 peaks and 4 valleys to be detected As an example, figure 7 below shows the ROI obtained when only 3 peaks and 2 valleys are successfully detected by the system.

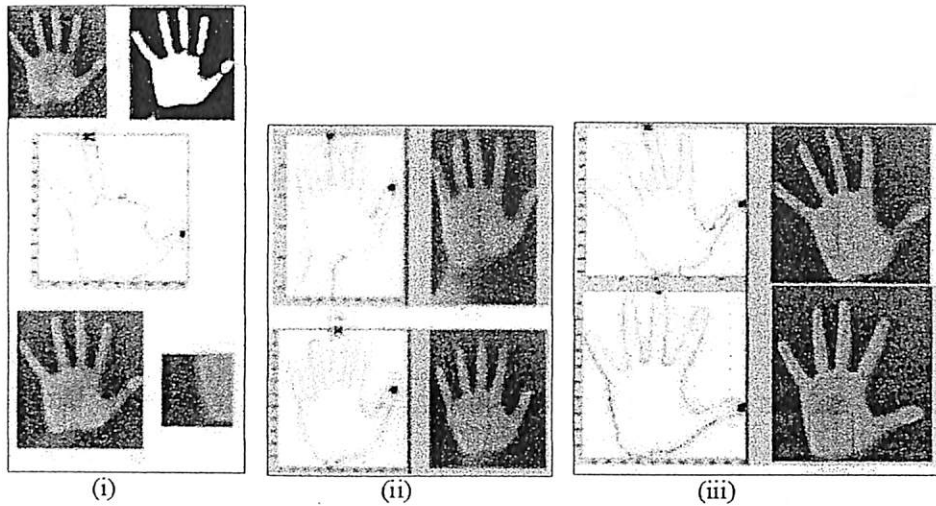


Fig. 6. Problem with Two Point Methods (i) Problem 1 (ii) Problem 2 (iii) Problem 3

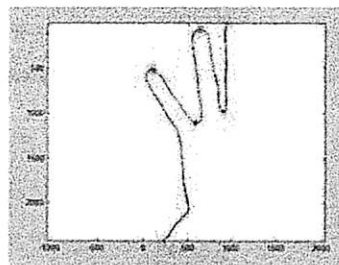


Fig. 7. Problem with Canny method.

Consequently, as observed from the experimental results, the major different between these two methods is the hand boundary image processed by canny method is smoother than the two point method. This is because the two point method is directly traced the boundary from the binary image. But for canny method, the boundary is traced from the set of connected pixels with the largest area in the image. Figure 8 shows the different of the hand boundary image between the two point method and canny method. Evaluation on the overall data collection, the two point method gives 70 % while for the canny method is 89 %.

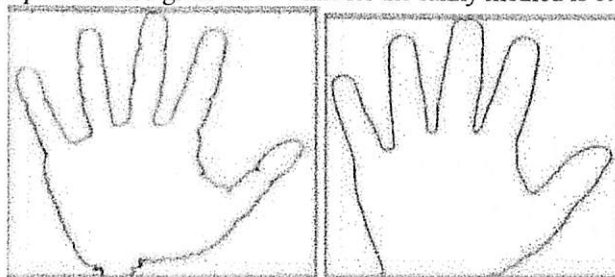


Fig. 8. Image selected using the two point method (left) and canny method (right).

4 CONCLUSION

This paper presented step by step process in developing the automated palm-print ROI selection for touch-less palm-print recognition system. Two methods namely two point method and canny method have been implemented and evaluated. Due to the capability of the canny method to trace the hand boundary image smoothly hence can detect the 5 peaks and 4 valleys correctly; the desired palm-print ROI can be obtained without facing many problems. So that, the canny method can be a viable technique in selecting the accurate ROI for the use of touch-less palm print biometric system. Future research will be devoted to the complete process of the implementation of palm-print based touch-less biometric system using smart phone device.

Acknowledgments

The authors would like to thank the financial support provided by Universiti Sains Malaysia Short Term Grant, 304/PELECT/60311048, Research University Grant 814161 and Research University Grant 814098 for this project.

References

- Yih, E. W. K., Sainarayanan, G. & Chekima, A. (2009). Palmprint based biometric system: a comparative study on discrete cosine transform energy, wavelet transform energy and sobelcode methods. *Biomedical Soft Computing and Human Sciences*, 14(1), 11-19.
- Zhang, D. D. (2004). *Palmprint authentication* (Eds). Massachusetts, USA: Kluwer Academic Publishers.
- Goh, M. K. O., Tee, C. & Teoh, A. B. J. (2010). Bi-Modal Palm-print and Knuckle Print Recognition System", *Journal of IT in Asia*, 3, 53-66.
- Al-Kubati, A. A. M., Saif, J. A. M., & Taher, M. A. A. (2012). Evaluation of canny and otsu image segmentation. *International Conference on Emerging Trends in Computer and Electronics Engineering (ICETCEE'2012)*, March 24-25, 2012 Dubai.
- Julio, A. & Shu, W. (2009). Exploring touch-screen biometrics for user identification on smart phones. *Proceeding of the 2009 IEEE/IFIP International Conference on Dependable System*, 1-14.
- Tabejamaat, M. & Kangarloo, K. (2011). A Multibiometric system based on hand geometry and palmprint features. *2011 International Conference on Signal, Image Processing and Applications with Workshop of ICEEA*, 21, 111-115.
- Goh, M. K. O., Tee, C. & Teoh, A. B. J. (2008). Touch-less palm-print biometrics: novel design and implementation. *Image and Vision Computing*, 26, 1551-1560.
- Meraomia, A., Chitroub, S. & Bouridane, A. (2011). Fusion of finger-knuckle-print and palmprint for an efficient multi-biometric system of person recognition", *IEEE Communication Society for IEEE ICC 2011*, 1-5.
- Ismail N. & Sabri, M. I. M. (2010). Mobile to server face recognition: A system overview, *World Academy of Science, Engineering and Technology*, 69, 2010, 761-765.
- Kasturika, B. R. & Misra, R. (2011). Palmprint as a biometric identifier. *International Journal of Electronics and Communication Technology*, 2(3), 2011, 12-16.

A Review of Multibiometric System with Fusion Strategies and Weighting Factor

Haryati Jaafar

Intelligent Biometric Group, School of Electrical and Electronic Engineering,
USM Engineering Campus,
14300 Nibong Tebal, Pulau Pinang, Malaysia
haryati_jaafar@yahoo.com

Dzati Athiar Ramli

Intelligent Biometric Group, School of Electrical and Electronic Engineering,
USM Engineering Campus,
14300 Nibong Tebal, Pulau Pinang, Malaysia
dzati@eng.usm.my

Abstract—Biometric is a technology for verification or identification of individuals by employing a person's physiological and behavioural traits. Although these systems are more secured compared the traditional methods such as key, smart card or password, they also undergo with many limitations such as noise in sensed data, intra-class variations and spoof attacks. One of the solutions to these problems is by implementing multibiometric systems where in these systems, many sources of biometric information are used. This paper presents a review of multibiometric systems including its taxonomy, the fusion level schemes and toward the implementation of fixed and adaptive weighting fusion schemes so as to sustain the effectiveness of executing the multibiometric systems in real application.

Keywords- *biometric, multibiometric, level of fusions, fixed weighting, adaptive weighting.*

I. INTRODUCTION

In the modern world, there is a high demand to authenticate and identify individuals automatically. Hence, the development of technology such as personal identification number (PIN), smartcard or passwords have been introduced. However, those technologies are inadequate since they are disclosable and transferable. For example, PIN and smart card can be duplicated, misplaced, stolen or lost, long password can be hard to remember by client and short password can be guessed easily by the imposter [1,2].

In order to overcome these problems, biometric-based authentication and identification methods are introduced in late 90s. By applying biometric systems, it is possible to identify the person, or to validate a claimed identity. Hence, the biometric systems have become an active research since these systems can be implemented as security protection systems (e.g., access control), criminal investigations, logical access points (e.g. computer login) and surveillance applications (e.g., face recognition in public spaces).

A biometric system is essentially a pattern-recognition system that recognizes a person based on a feature vector derived from a specific physiological or behavioural characteristic the person possessed for authentication or identification purposes [3]. It differs from classical user authentication system which is based on something that one has (e.g., identification card, key) and/or something that one knows (e.g., password, PIN). Hence, a number of physiological and behavioural traits can be utilized in the biometric systems such as fingerprint, iris, face, hand geometry, palm print, finger vein structure, gait, voice, signature. Depending on the context of applications, biometric systems may operate in two modes i.e. verification or identification [4,5]. Biometric verification is the task of authenticating the test biometric sample with its corresponding pattern or model according to the claim given by user. Whereas, biometric identification is the task of associating a test biometric sample with one of number of patterns or models that are available from a set of known or registered individuals [6].

Most biometric systems deployed in real-world applications are unimodal. These systems suffer with problems such as noise in sensed data, non-universality, upper bound on identification accuracy and spoof attacks [7]. In order to overcome the problem, Hong et al. [8] examined the possible performance improvement of biometric systems by using multiple biometrics. This paper showed that by integrating with other multiple biometric sources, the performance was indeed improved. Such systems, known as multibiometric systems can improve the matching accuracy of biometric systems and deterring spoof attacks [2].

Mutibiometric systems can also improve other limitations faced by biometric systems. For example, the multibiometric system can address the non-universality problem encountered by biometric systems. If a person cannot be enrolled in the fingerprint system, this person can aid the problem using other biometric traits such as voice, face or iris. The multibiometric systems can also reduce the effect of noise data. If the quality biometric sample obtained from one sources is not sufficient, the other samples can provide sufficient information to

enable decision-making. Another advantage of multibiometric over single biometric systems is that, they are more resistant to spoof attacks since it is difficult to simultaneously spoof multiple biometric sources. The multibiometric systems are able to incorporate a challenge-response mechanism during biometric acquisition by acquiring a subset of the trait in some random order [9].

However, the multibiometric systems also have major drawbacks compared with single biometric systems. For example, the cost for the implementation of multibiometric systems is more expensive since these systems require many sensors. Furthermore, such a system may also increase the user inconvenience and required the user to interact with more than one sensor. For example, in a multibiometric system, both fingerprint and iris images of a person are required. Therefore, a user not only needs to touch the fingerprint scanner, but also needs to work together with an iris imaging system. Such activity gives impact on the raising of computation, memory and storage. Moreover, this also increases the operating time during enrollment and verification process [9].

In order to describe the current scenario of multibiometric systems, this review paper is organized as the following. Section II describes the taxonomy of multibiometric systems which explained the different roles of multibiometric systems in term of multi-sensor, multi-algorithm, multi-instance, multi-sample and multimodal systems. Section III provides detailed explanation for the level of fusion techniques that used in the combination phase for the fusion of different sources of biometric information. Finally, a review toward to the implementation of fixed and adaptive weighting fusion schemes is then discussed in the Section IV.

II TAXONOMY OF MULTIBIOMETRIC SYSTEM

Based on the nature of the sources of biometric information, a multibiometric system can be classified into five categories which are multi-sensor, multi-algorithm, multi-sample, multi-instance and multi-modal systems. The scenario of multibiometric systems is depicted as in Fig. 1.

Multi-sensor systems: Multi-sensor systems employ multiple sensors to capture single biometric trait of an individual. The example of this system is reported in [10] where multiple 2D cameras are used to capture the image of subject. Subsequently, in [11], an infrared sensor and visible-light sensor are applied to acquire the information of a person's face while in Rowe and Nixon [12] and Pan et al. [13], a multi spectral camera has been employed to acquire images of iris, face or finger. The application of multi-sensors in the researches is able to enhance the recognition ability of the biometric systems. For instance, the infrared and visible-light images of person's face can present different types of information which can enhance the matching accuracy based on the nature of illumination due to ambient lighting.

Multi-algorithm systems: multi-algorithm systems combine the output of multiple methods such as feature extraction or/and classification algorithms for the same biometrics data [7]. In other words, the supplementary information by more than one algorithm helps to improve the performance. So, utilization of new sensor is not required thus it is cost effective. However, this system has a drawback due to many feature extraction and matching modules can cause complexity of system computation. Example of this system can be found in Lu et al. [14] where three different feature extraction schemes which are Principle Discriminate Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminate Analysis (LDA) have been combined to improve a face recognition system. Another researcher has also combined multiple algorithms such as Iterative Closet Point (ICP), PCA and LDA to perform 3D face recognition [15]. In Imran et al. [16], three subspace algorithms such as PCA, Fisher Linear Discriminant (FLD) and ICA are applied for palm print and face separately in order to determine the best algorithm performance. The result shows that the ICA algorithm performs well for both individual modalities.

Multi-sample systems: multi-sample systems use multiple samples derived from the same biometrics acquired by a single sensor. The same algorithm processes each of the samples and the individual results are fused to obtain an overall recognition results. The advantage of using multiple samples is to avoid poor performance due to the slack properties of sample if only one sample is used. This system has been studied in Chang et al. [17] for face recognition where 2D face image has been applied as a baseline in order to compare the performance of multi-sample 2D + 3D face in speech recognition. Another research has proposed multi-sample approach to UMACE filter classifier by combining scores from several samples from lipreading features and spectrographic features [18].

Multi-instance systems: In this system, the biometric information has been extracted from the multiple instances of the same body trait. For example, the left and right index finger and iris of an individual is proposed in Jang et al. [19] and Prabhakar and Jain [20], respectively.

Multi-modal systems: multi-modal systems use the evidence of multiple biometric traits to extract the biometric information of an individual. These different biometric traits can come from a variety of modalities [9]. The multi-modal system is reliable due to the presence of multiple independent biometrics. However, the drawback of this system is due to the substantial cost because of the requirement of many sensors. The example of this system has been reported by Brunelli and Falagivna [21] where a person identification system using face and speech is presented. This research showed that by combining three biometrics i.e. frontal face, face profile and

voice using sum rule combination scheme, the system performance has been improved [22]. Another combination such as fingerprint, face and finger vein has been presented in Hong et al. [8] while Ramli et al. [23], and Lip and Ramli [24] used the speech signal as a biometric trait to the biometric verification system and lipreading image as a second modality to assist the performance of the single modal system in the multibiometric systems.

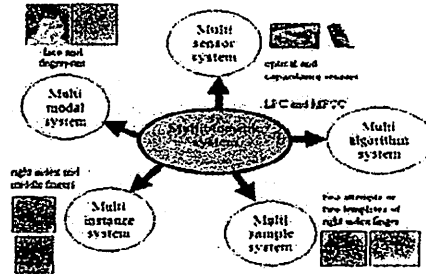


Figure 1. Scenarios in a multibiometric system

III. LEVEL OF FUSION

The important issue to designing multibiometric system is to determine the sources of information and combination strategies. Depending on the type of information to be fused, the fusion scheme can be classified into different levels. According to Sanderson and Paliwal [25], the level of fusion can be classified into two categories, fusion before matching (pre classification) and fusion after matching (post classification) as shown in Fig. 2.

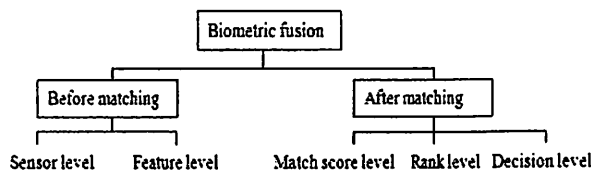


Figure 2. Level of fusion

For fusion before matching, the integration of information from multibiometric sources in this scheme includes fusion at the sensor level and fusion at the feature level. Meanwhile, fusion after matching can be divided into two categories which are fusion at the match score level and fusion at the decision level.

A. Fusion Before Matching

- Sensor Level Fusion

In this level, the raw data from the sensor are combined together as shown in Fig. 3. However, the source of information is expected to be contaminated by noise such as non-uniform illumination, background clutter and other [26]. Sensor level fusion can be performed in two conditions i.e. data of the same biometric trait is obtained using multiple sensors; or data from multiple snapshot of the same biometric traits using a single sensor [27, 28].

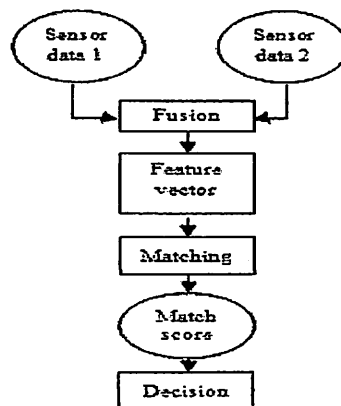


Figure 3. Sensor level fusion process flow

- Feature level fusion

In feature level fusion, different feature vectors extracted from multiple biometric sources are combined together into a single feature vector as depicted in Fig. 4. This process undergoes two stages which are feature normalization and feature selection. The feature normalization is used to modify the location and scale of feature values via a transformation function and this modification can be done by using appropriate normalization schemes [2]. For instance, the min-max technique and median scheming have been used for hand and face [9] and the mean score from the speech signal and lipreading images scores have been employed in the feature level fusion [24]. Another research has implemented Scale Invariant Feature Transform (SIFT) to obtain features from the normalized fingerprint and ear [29]. Consequently, feature selection is executed to reduce the dimensionality of a new feature vector in order to improve the matching performance of the feature vector by accepting more authentic as true accept. There are several feature selection algorithms have been applied in the literature for instances Sequential Forward Selection (SFS), Sequential Backward Selection (SBS) and Partition About Medoids [30]. The advantage of the feature level fusion is the detection of correlated feature values generated by different biometric algorithms, and, in the process, identifying a salient set of features that can improve recognition accuracy [2]. However, in practice, fusion at this level is hard to accomplish due to the following reasons i.e. the feature sets to be joined might be incompatible and the relationship between the joint feature set of different biometric sources may not be linear [31]. Moreover, concatenating two feature vectors yield a new feature vector which gives larger dimensionality compared to the original once thus leads to the dimensionality problem. Large feature variance affects the system accuracy and also increases the processing time. Hence, only few researchers have focused on the feature level scheme compared to the other levels of fusions such as score level and decision level.

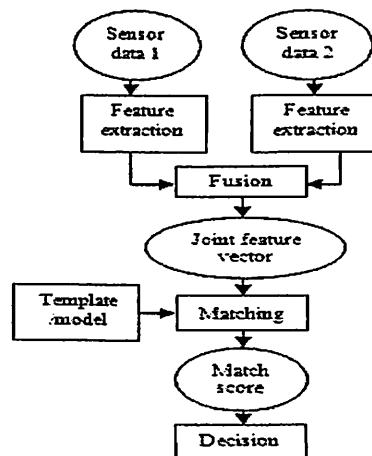


Figure 4. Feature level fusion process flo

B. Fusion After Matching

- Score level fusion

In score level fusion, the match outputs from multiple biometrics are combined together to improve the matching performance in order to verify or identify individual as shown in Fig. 5 [32]. The fusion of this level is the most popular approach in the biometric literature due to its simple process of score collection and it is also practical to be applied in multibiometric system. Moreover, the matching scores contain sufficient information to make authentic and imposter case distinguishable [6]. However, there are some factors that can affect the combination process hence degrades the biometric performance. For example, the matching scores generated by the individual matchers may not be homogenous due to be in the different scale/range or in different probability distribution. In order to overcome this limitation, three fusion schemes have been introduced i.e. density-based schemes; transformation-based scheme; and classifier-based scheme [7]. The density-based scheme is based on score distribution estimation and has been applied in well-known density estimation models such as Naive Bayesian and Gaussian Mixture Model (GMM) [33]. This scheme usually achieves optimal performance at any desired operation point and estimate the score density function accurately. However, this scheme requires a large number of training samples in order to perfectly approximate the density functions. Moreover, it requires more time and effort for the operational setting compared to the other schemes. On the other hand, the transformation-based scheme is commonly applied for the score normalization process. This process is essential to change the location and scale parameters of the

underlying match score distributions in order to ensure compatibility between multiple score variables [7]. This scheme can be applied using various techniques such as sum rule, product rule, min rule and max rule techniques [34]. In the classifier-based scheme, the scores from multiple matchers are treated as a feature vector and a classifier is constructed to discriminate authentic and imposter score [33]. From the literatures, various types of classifiers such as SVM, neural network and *multi-layer perceptron* (MLP) [34] have been implemented to classify the match vector in this scheme. However, this scheme has some drawbacks such as unbalanced training set and misclassification problems.

- Decision level fusion

Fusion at the decision level is executed after a match decision has been made by the individual biometric source as depicted in Fig. 6. So far, there are many different methods have been applied to join the distinct decision into a final decision such as “AND” and “OR” rules [24], majority voting, weighted majority voting, Bayesian decision fusion, Dempster-Shafer theory of evidence and behaviour knowledge space [7]. On the other hands, Ramli et al., [35] implemented the proposed decision fusion by using the spectrographic and cepstrumgraphic as features extraction and UMACE filters as classifiers in the system to reduce the error due to the variation of data.

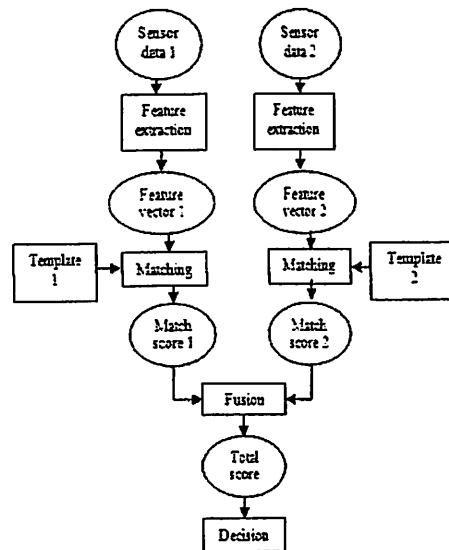


Figure 5. Score level fusion process flow

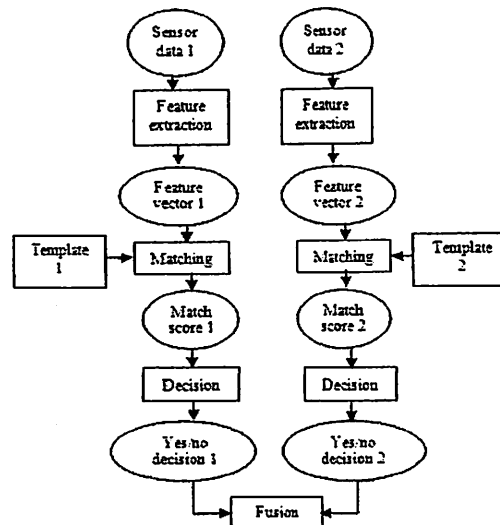


Figure 6. Decision level fusion process flow

IV. FIXED AND ADAPTIVE WEIGHTING IN BIOMETRIC

Multibiometric systems are found to be useful and exhibit robust performance over the single biometric systems. However, in uncontrolled conditions, the reliability of the multibiometric systems drops severely. As the results, the systems are poorly executed in uncertain condition. Therefore, it is imperative to assign different

weighting in fusion scheme to each biometric trait in order to vary the importance of matching scores of each biometric trait since the optimum weight can maximize the performance of multibiometric system.

In general, multibiometric systems can be divided into two categories of weighting scheme which are fixed and adaptive weighting. In the fixed weighting, the fusion weight is fixed for each training data set. Otherwise, retraining the optimum weight is needed. Research of fixed weighting fusion has been done as reported in Parviz and Moin [34]. This study presented fusion of score produced independently by speaker recognition system and face recognition system using weighted merged score. The result shows that the identification of 51% was achieved for the speech only system and 92% for the face system. Subsequently, performance of the integration system using the optimal weight is observed up to 95%. In another study was done in Brunelli and Falavigna [35], the weighting product is applied to fuse two voice features i.e. static and dynamic and three face features i.e. eye, nose and mouth. This research used tan-estimators for score normalization and weighted geometric average was used for score combination. The results showed the correct identification percentage of the integrated system is 98% which represents a significant improvement compared to 88% and 91% rates provided by the single systems i.e. speaker and face based system respectively. The EER performance of face recognition, voice recognition and the integrated face and voice recognition are obtained as 3%, 3.4% and 1.5% from this experiment respectively. Imran et al. [16] has presented the score level fusion of palm and face modalities using weighted sum rule for different algorithms (PCA, FLD and ICA). The results showed that the performance of fusion of face and palm with ICA, FLC and PCA are 75.52%, 73.69% and 66.60%, respectively. In additional, Ramli et al., [36] used the weighting factor for combination of audio and visual scores and the min-max normalization technique in fusion scheme to determine the performances of speech based biometric systems at different levels of signal to noise ratio i.e. clean, 30dB, 20dB and 10dB. The results show the EER performance of the integration system in clean, 30dB, 20dB and 10dB SNRs are observed as 0.0019%, 0.0084%, 0.9356% and 5.0160%, respectively compared to the EER performances of 1.1599%, 2.5113%, 19.3423% and 39.8649% for audio only system.

The second approach of weighting in fusion scheme is an adaptive weighting where the fusion weight is adaptable according to the current system condition. Two methods which are reliability estimation and reliability information can be applied in an adaptive weighting. The reliability estimation is performed either relying on the statistic-based measure or directly based on the quality of signal. Two methods have been proposed for the statistics based reliability measure i.e. entropy of posteriori probabilities and dispersion of posteriori probabilities. In the quality of signal, the weight for fusion scheme is adapted corresponding to the quality of the current input signal instead of using the optimum weight estimated from the available training set. On the other hand, the reliability information can be obtained by the shape of posteriori probabilities [37].

Study on the adaptive weighting can be found in Gurban and Thiran [38] where the audio visual phonetic classification accuracy using GMM entropy has been studied and 54.44% accuracy has been achieved. In another research, the entropy of a posteriori probabilities using MLP states has been applied [17]. The reliability information can be obtained by the shape of a posteriori probabilities distribution of HMM states and the results showed that the audio visual speech recognition performance at 10dB SNR using inverse entropy and negative entropy are obtained as 93.35% and 94.30%, respectively. According to Soltane et al. [39], GMM based Expectation Maximization (EM) estimated algorithm for score level data fusion based on face and speech modalities is proposed. The database obtained from eNTERFACE 2005 contained 30 subjects was used for the experiments. The result shows that EER performance for face and voice are 44.94% and 2.690% respectively. In order to reduce the EER performance for face mode, the combination of face-voice with different weighting has been applied. The result shows that combination of face-voice is able to reduce the percentage of EER to 8.73%. Kisku et al. [29] presents a robust feature level fusion technique of fingerprint and ear. In this paper, the reliability of each fused matching score has been increased by applying adaptive Doddington's user-weighting scheme. The proposed adaptive weighting scheme is to decrease the effect of imposter users rapidly. In this scheme, the adaptive weights has been computed by using tan hyperbolic weight for each matcher by assigning weights to individual matching scores. The identification rate for the proposed system are obtained as 98.71% while that for fingerprint and ear biometrics are found as 95.02% and 93.63%, respectively.

The comparison of fixed weighting and adaptive weighting can also be found in Lau et al. [40]. This paper presents a multibiometric verification system that combines speaker, fingerprint and face biometrics and fusion has been done in score level using GMM entropy. Their respective equal EER are 4.3%, 5.1% and the range of 5.1% to 11.5% for matched conditions in facial image capture. Fusion by majority voting gave a relative improvement of 48% over speaker verification. In another experiment, a fixed weight is assigned to each biometric trait. The weights are varied within the [0,1] range in steps of 0.1 to find values that gave the best performance. There is an improvement of 52% additional relative improvement of 52%, which corresponds to EER range of (0.50% and 0.84%). The weighting for each biometric has then been adjusted by using the fuzzy logic framework in order to account the external conditions that affect verification, such as finger position,

facial geometry and lightning conditions. The result shows fuzzy logic fusion generated a further improvement of 19% which corresponds to an EER range of 0.31% to 0.81%.

V. CONCLUSION

Multibiometric systems are expected to alleviate many limitations of biometric systems by combining the evidence obtained from different sources using an effective fusion scheme. In this paper, the sources of biometric information were presented. The description regarding the level of fusions was also presented in this paper. From the study, it reveals that, performance of multibiometric systems can be further improved if an appropriate fusion strategy is used especially for the system which executed in uncontrolled environment. Hence, a different weighting in fusion is applied to maximize the performance of multibiometric system. Based on the review, the most promising recent research that can be implemented is fusion at the score level involving adaptive weighting. This approach have great potential to get rid the uncertain problem such as noise in sensed data, non-universality, upper bound on identification accuracy and spoof attacks.

ACKNOWLEDGMENT

The authors would like to thank the financial support provided by Universiti Sains Malaysia Short Term Grant, 304/PELECT/60311048, Research University Grant 814161 and Research University Grant 814098 for this project.

REFERENCES

- [1] G. Williams, "More than a pretty face", Biometrics and SmartCard Tokens. SANS Institute reading room, 2002, pp. 1-16.
- [2] A. K. Jain, K. Nandakumar, U. Uludag, and X. Lu, "Multimodal Biometrics", from *Augmenting Face With Other Cues*, in W. Zhao, and R. Chellappa, (eds) *Face Processing: Advanced Modelling and Methods*, Elsevier, New York, 2006. pp. 675-705.
- [3] J.P. Campbell, D.A. Reynolds, and R.B. Dunn, "Fusing High And Low Level Features for Speaker Recognition", in *Proceeding of EUROSPEECH*, 2003, pp. 2665-2668.
- [4] J. Ortega-Garcia, J. Bigun, D. Reynolds, and J. Gonzalez-Rodriguez, "Authentication Gets Personal With Biometrics", in *IEEE Signal Processing Magazine*, Vol. 21, 2004, pp. 50-62.
- [5] A.K. Jain, A. Ross, and S. Prabhakar, "An Introduction To Biometric Recognition", in *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No.1, 2004, pp. 4-20.
- [6] M.X. He, S.J. Horng, P.Z. Fan, R.S. Run, R.J. Chen, J.L. Lai, M.K. Khan and K.O. Sentosa, "Performance Evaluation of Score Level Fusion in Multimodal Biometric Systems", *Journal of Pattern Recognition*, Vol. 43, No. 5, 2010, pp. 1789-1800.
- [7] A. Ross, and A.K. Jain, "Fusion Techniques in Multibiometric Systems", from *Face Biometrics for Personal Identification*. In. R.I. Hammound, B.R. Abidi and M.A. Abidi (eds.), *Publisher Springer Berlin Heidelberg*, 2007, pp. 185-212.
- [8] L. Hong, A.K. Jain and S. Pankanti, "Can Multibiometrics Improve Performance?", *Proc. AutoID '99*, 1999, pp. 59-64.
- [9] A. Jaina, K. Nandakumara, A. Ross, and A. Jain, "Score Normalization in Multimodal Biometric Systems", *Journal of Pattern Recognition*, Vol. 38, 2005, pp. 2270.
- [10] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju, "Finding Optimal Views for 3D Face Shape Modeling", in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2004, pp. 31-36.
- [11] A. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi, "Recent Advances in Visual and Infrared Face Recognition - A Review", *Computer Vision and Image Understanding*, Vol. 97, No.1, 2005, pp. 103-135.
- [12] R.K. Rowe, and K.A. Nixon, "Fingerprint Enhancement Using a Multispectral Sensor", in *Proceedings of SPIE Conference on Biometric Technology for Human Identification II*, Vol. 5779, 2005, pp. 81-93.
- [13] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face Recognition in Hyperspectral Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, 2003, pp. 1552-1560.
- [14] X. Lu, Y. Wang, and A.K. Jain, "Combining Classifiers for Face Recognition", in *IEEE International Conference on Multimedia and Expo (ICME)*, Vol. 3, 2003, pp. 13-16.
- [15] K. Chang, K. Bowyer, and P. Flynn, "Face Recognition Using 2D And 3D Faces", *Workshop on Multi Modal User Authentication (MMUA)*, 2003, pp. 25-32.
- [16] M. Imran, A. Rao, and G.H. Kumar, "Multibiometric Systems. A comparative study of multi-algorithmic and multimodal approaches", *Procedia Computer Science*, Vol. 2, 2010, pp. 207-212.
- [17] K. I. Chang, K.W. Bowyer, and P.J. Flynn, "An Evaluation of Multimodal 2D+3D Face Biometrics", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 4, 2005, pp. 619-624.
- [18] S.A. Samad, D.A. Ramli and A. Hussain, "A Multi-Sample Single Source Model using Spectrographic Features for Biometric Authentication", in *IEEE International Conference on Information, Communications and Signal Processing (ICICS 2007)*, 2007.
- [19] J. Jang, K.R. Park, J. Son, and Y. Lee, "Multi-unit iris recognition system by image check algorithm", in *Proceedings of International Conference on Biometric Authentication (ICBA)*, 2004, pp. 450-457.
- [20] S. Prabhakar, and A.K. Jain, *Decision-Level Fusion in Fingerprint Verification*, Technical Report MSU-CSE-00-24. Michigan State University, 2000.
- [21] R. Brunelli, D. Falavigna, L. Stringa and T. Poggio, "Automatic Person Recognition by Using Acoustic and Geometric", *Machine Vision & Applications*, Vol. 8, 1995, pp. 317-325.
- [22] J. Kittler, M. Hatef, R.P. Duin, and J.G. Matas, "On Combining Classifiers", *IEEE Trans. PAMI*, Vol. 20, No. 3, 1998, pp. 226-239.
- [23] D.A. Ramli, S.A. Samad, and A. Hussain, "A Multibiometric Speaker Authentication System with SVM Audio Reliability Indicator", *IAENG International Journal of Computer Science*, Vol. 36, No.4, 2009, pp. 313-321.
- [24] C.C. Lip, and D.A. Ramli, "Comparative Study on Feature, Score and decision Level Fusion Schemes for Robust Multibiometric Systems", *Advances in Intelligent and Soft Computing*, Vol. 133, 2012, pp. 941-948.
- [25] C. Sanderson, and K.K. Paliwal, "Noise compensation in a person verification system using face and multiple speech features", *Pattern recognition*, Vol. 2, 2003, pp. 293-302.
- [26] S.S. Iyengar, L. Prasad, and H. Min, *Advances in Distributed Sensor Technology*, Prentice Hall, 1995.
- [27] R. Singh, M. Vatsa, A. Ross, and A. Noore, "Performance Enhancement of 2D Face Recognition via Mosaicing", in *Proceedings of the 4th IEEE Workshop on Automatic Identification Advanced Technologies (AutID)*, 2005, pp. 63-68.
- [28] A. Ross, and R. Govindarajan, "Feature Level Fusion Using Hand And Face Biometrics", in *Proceedings of SPIE Conference on Biometric Technology for Human Identification*, Vol. 5779, 2005, pp. 196-204.

- [29] D.R. Kisku, P. Gupta, and J.K. Sing, "Feature level Fusion Of Biometrics Cues: Human Identification with Doddington's Caricature", in International Conference of Security Technology, Communications in Computer and Information Sciences, 2010, pp. 157-164.
- [30] A. Kumar, and D. Zhang, "Personal Authentication Using Multiple Palmprint Representation", Pattern Recognition, Vol. 38, No.10, 2005, pp. 1695-1704.
- [31] A. Ross, and A. Jain, "Information Fusion in Biometrics", Pattern Recognition, Vol. 24, 2003, pp. 2115-2125.
- [32] A.K. Jain, and A. Ross, "Multibiometric Systems. Communications of The ACM", Special Issue on Multimodal Interfaces, Vol. 47, No. 1, 2004, pp. 34-40.
- [33] K. Nandakumar, Y. Chen, C. Dass, and A.K. Jain, "Likelihood Ratio Based Biometric Score Fusion", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, pp. 1-9.
- [34] M. Parviz, and M.S. Moin, "Boosting Approach For Score Level Fusion In Multimodal Biometrics Based On AUC Maximization", Journal of Information Hiding and Multimedia Signal Processing, Vol. 2, No.1, 2011, pp. 51-60.
- [35] R. Brunelli, and D. Falavigna, "Person Identification Using Multiple Cues", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No.10, 1995, pp. 955-966.
- [36] D.A. Ramli, S.A. Samad, and A. Hussain, "Performances of Speech Signal Biometric Systems Based on Signal to Noise Ratio Degradation", Advances in Intelligent and Soft Computing, Vol. 85, 2010, pp. 73-80.
- [37] D.A. Ramli, S.A. Samad, and A. Hussain, "A Multibiometric Speaker Authentication System with SVM Audio Reliability Indicator", IAENG International Journal of Computer Science, Vol. 36, No.4, 2009, pp. 313-321.
- [38] M. Gurban, and J.P. Thiran, "Using Entropy as a Stream Reliability Estimate for Audio-Visual Speech", in 16th European Signal Processing Conference, 2008, pp. 25-29.
- [39] M. Soltane, N. Doghmane, and N. Guersi, "Face and Speech Based Multi-Modal Biometric Authentication", International Journal of Advanced Science and Technology, Vol. 21, 2010, pp. 41-56.
- [40] C.W. Lau, B. Ma, H.M. Meng, Y.S. Moon, and Y. Yam, "Fuzzy Logic Decision Fusion In A Multimodal Biometric System", in Proceedings of the 8th International Conference on Spoken Language Processing, 2007.

AUTHORS PROFILE

Haryati Jaafar is a student in Universiti Sains Malaysia (USM) attached to the School of Electrical and Electronic. Currently, she continues her PhD in the field of Software Engineering.

Dzati Athiar Ramli holds a PhD from the Universiti Kebangsaan Malaysia (UKM) and is attached with the School of Electrical and Electronic at USM. His research is mostly related to biometric technology systems.

Chapter 18

Frog Identification System Based on Local Means K-Nearest Neighbors with Fuzzy Distance Weighting

Haryati Jaafar, Dzati Athiar Ramli, Bakhtiar Affendi Rosdi
and Shahriza Shahrudin

Abstract Frog identification based on the vocalization becomes important for biological research and environmental monitoring. As a result, different types of feature extractions and classifiers have been employed. Yet, the k-nearest neighbor (kNN) is one of the popular classifiers and has been applied in various applications. This paper proposes an improvement of kNN in order to evaluate the accuracy of frog sound identification. The recorded sounds of 12 frog species obtained in Malaysia forest have been segmented using short time energy and short time average zero crossing rate while the features are extracted by mel frequency cepstrum coefficient. Finally, a proposed classifier based on local means kNN and fuzzy distance weighting have been employed to identify the frog species. Comparison of the system performances based on kNN, local means kNN and the proposed classifier i.e. fuzzy kNN with manual segmentation and automatic segmentation is evaluated. The results show the proposed classifier outperforms the baseline classifier with accuracy of 94.67 % and 98.33 % for manual and automatic segmentation, respectively.

Keywords Frog identification · kNN · Local means KNN · Fuzzy kNN · Distance weighting

H. Jaafar (✉) · D. A. Ramli · B. A. Rosdi
School of Electrical and Electronic Engineering, USM Engineering Campus, 14300,
Nibong Tebal, Pulau Pinang, Malaysia
e-mail: haryati_jaafar@yahoo.com

D. A. Ramli
e-mail: dzati@eng.usm.my

B. A. Rosdi
e-mail: eebakhtiar@eng.usm.my

S. Shahrudin
School of Pharmacy Sciences, USM, 11800, Minden, Pulau Pinang, Malaysia
e-mail: shahriza18@usm.my

H. A. Mat Sakim and M. T. Mustaffa (eds.), *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications*,
Lecture Notes in Electrical Engineering 291, DOI: 10.1007/978-981-4585-42-2_18,
© Springer Science+Business Media Singapore 2014

18.1 Introduction

Frogs are unique creatures that have been living in this planet for more than 250 million years. Over a decade, these amphibians become crucial since their impact of whole ecosystem is great as bio indicators. In addition, their bodies may keep the important key for new discoveries in medical research. The chemical compounds in their skin may provide antimicrobial peptides that used to treat pain and block infections [1]. Commonly, frog relies on their sound to present the presence, behaviors and species. This is because their sound can be received over varying distance that allow and obstructive detection of their existence [2]. Different techniques which involved feature extractions and classifiers have been studied and proposed in order to identify the frog species based on their vocalization automatically [3, 4]. Among of the classifiers, k nearest neighbor (kNN) becomes the most popular nonparametric classifier which has widely been used in pattern classification application and generally archives good result. Nonetheless, this classifier required a large number of training samples to determine desired values of probability of correct classification [5]. Moreover, this classifier suffers from existing outliers particularly in small training sample size situation [6]. Hence, the improvement of kNN has been investigated actively [5, 7–10]. This paper proposes an improvement of kNN by employing local means kNN with fuzzy distance-weighting (LMKNN-FDW). As compared with the previous papers, the distance between query pattern or testing sample and local means vector is assigned using fuzzy algorithm. In addition, the comparative studies with kNN, FKNN and LMKNN are discussed. The various frog sounds in manual segmentation and automatic segmentation based on short time energy (STE) and short time average zero crossing rate (STAZCR) are conducted in this experiments. Consequently, a standard mel frequency ceptrum coefficient (MFCC) is executed as feature extraction in this study. The first objective is to improve kNN classifier by proposing LMKNN-FDW. The second is to compare the performance results between proposed classifier with the baseline classifiers. This paper is outlined as follows. In Sect. 18.2, the methodology of this study is discussed. Section 18.3 describes the proposed classifier in detail and the experimental results are presented in Sect. 18.4 and the conclusion are summarize in Sect. 18.5.

18.2 Methodology

18.2.1 Data Acquisition

All of 12 frogs sound were recorded from locations around Baling and Kulim, Kedah, Malaysia using Sony Stereo IC Recorder ICD-AX412F supported with Sony electret condenser microphone in 32-bit wav files at a sampling frequency of

48 kHz. Consequentially locations were selected based on frog's potential habitat such as next to a swamp, running stream and ponds from 8.00 pm to 12.00 pm.

18.2.2 Syllables Segmentation

The syllables segmentation based on STE and STAZCR were applied where the principle of the techniques is to determine the endpoint of syllable boundaries accurately to detect the syllable signal that has been segmented [11].

18.2.2.1 Short Time Energy

This technique is used to classify voiced and unvoiced parts. The voice part has high energy than unvoiced part due to the periodicity. The STE function is defined by the following expression;

$$E_n = \frac{1}{N} \sum_{m=1}^N [x(m)w(n-m)]^2 \quad (18.1)$$

where E_n is the energy of the sample n of the signal, $x(m)$ is the discrete-time signal and $w[m]$ is a hamming window of size N .

18.2.2.2 Short Time Average Zero Crossing Rate

On the other hands, STAZCR is often used as a part of the front-end processing in automatic speech recognition system. During the frog signal processing, the amplitudes of the unvoiced part normally have higher values and vice-versa. The ZCR is the rate at which signal changes from positive to negative and back and defined as;

$$Z_n = \frac{1}{2N} \sum_{m=1}^N |\text{sgn}x(m) - \text{sgn}[x(m-1)]|w(n-m) \quad (18.2)$$

where

$$\text{sgn}[x(m)] = \begin{cases} 1, & x(m) \geq 0 \\ -1, & x(m) < 0 \end{cases} \quad (18.3)$$

18.2.3 Feature Extraction

MFCC is selected due to the features are robust to noise which is suitable to be implemented in outdoor environment that contains interference of background

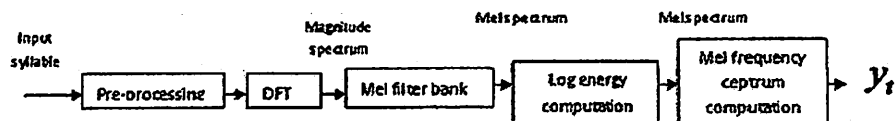


Fig. 18.1 Typical MFCC process

noises such as sound of wind, running water and other animal calls where MFCC processing is shown in Fig. 18.1. There are 12 mel cepstrum coefficients, one log energy coefficient and three delta coefficients per frame have been set in the experiments [12].

18.3 Propose Classifier, LMKNN-FDW

In order to design a simple classifier based on kNN, the following steps are executed. Let $X_i = \{x_n \in R^m\}_{n=1}^N$ be a training sample where m is the number of dimensional in feature space, N is the total number of training sample and $y_n \in \{c_1, c_2, \dots, c_M\}$ denotes the class label for x_n . A query pattern is first determined;

1. Determine the k nearest neighbor from the set X_i for each class y_n by Euclidean distance where $k \leq N$;

$$d(x, x_{ij}^N) = \sqrt{(x - x_{ij}^N)^T (x - x_{ij}^N)} \quad (18.4)$$

2. Search the local mean vector Y_{ik} by applying k nearest neighbor of training sample such that;

$$Y_{ik} = \frac{1}{k} \sum_{j=1}^k x_{ij}^N \quad (18.5)$$

3. In the fuzzy method, the testing data value or query pattern is classified by assigning membership values, $U_{ij}(k)$ in particular class based on percentage of neighbors in that class weighted. Hence, by applying Eq. (18.5) in the fuzzy method, the query pattern, x is classified as follows;

$$u_{ij} = \frac{\sum_{j=1}^k u_{ij} \left[\frac{1}{\|x - Y_{ik}\|^{2/(m-1)}} \right]}{\sum_{j=1}^k \left[\frac{1}{\|x - Y_{ik}\|^{2/(m-1)}} \right]} \quad (18.6)$$

where m is the scaling factor for fuzzy weight. Note the notation m denote the fuzzy weight of the distance or fuzzy relationship. If value m increases, the neighbors are more evenly weighted. This caused the distance between training and query pattern have less effect on each other and vice versa. In this paper, the value of $m = 2$ is used for the proposed classifier.

Table 18.1 Manual segmentation results

Scientific name	kNN	LMKNN	FKNN	LMKNN-FDW
<i>Hylarana glandulosa</i>	25	24	25	24
<i>Kaloula pulchra</i>	25	22	15	23
<i>Odorrana hossi</i>	11	19	22	22
<i>Polypedates leucomystax</i>	25	25	25	25
<i>Kaloula baleata</i>	25	23	25	24
<i>Philautus mjobergi</i>	24	23	24	22
<i>Phrynoedis aspera</i>	19	25	23	25
<i>Microhyla heymonsi</i>	13	20	18	23
<i>Microhyla butleri</i>	23	21	25	25
<i>Rhacophorus appendiculatus</i>	25	23	25	25
<i>Hylarana labialis</i>	25	25	25	25
<i>Philautus petersi</i>	7	14	16	21
Total	247	264	268	284
Percentage (%)	82.33	88	89.33	94.67

18.4 Experimental Result

The experiments are implemented in Intel Core i5, 2.1 GHz CPU, 2G RAM and Window 7 operating system. In this experiment, 540 syllables in total have been extracted with 20 syllables are used for training and 25 for testing while the value of $k = 3$ is used for all classifiers. The experiments have then been divided into two techniques of segmentation i.e. manual and automatic. Gold Wave software has been used to segment the samples manually while endpoint detection techniques have been employed for automatic segmentation. The classification accuracy (C_A) is defined as;

$$C_A = \frac{N_C}{N_T} \times 100\% \quad (18.7)$$

where N_c is the number of syllables which are recognized correctly and N_T is the total number of test syllables.

Table 18.1 lists the analytical results of the manual segmentation. The results show that all of the classifiers have been able to identify with more than 80 % of accuracies. By improving the kNN classifier, all of modified kNN classifiers show the improvement in performances compared to baseline kNN with 88, 89.33 % for LMKNN, FKNN, respectively and the proposed classifier, LMKNN-FDW gives the best performance i.e. 94.67 %. Table 18.2 lists the analytical results of the automatic segmentation. After applying the automatic segmentation, improvement of the results are observed. However, the percentage of accuracy for FKNN is slightly less than basic kNN with 96 % compared than 96.67 % to kNN. Nevertheless, the proposed classifier is the most outstanding classifiers compared to the other classifiers with 98.33 % of accuracy with 8 species can be identified 100 % accurately.

Table 18.2 Automatic segmentation results

Scientific name	kNN	LMKNN	FKNN	LMKNN-FDW
<i>Hylarana glandulosa</i>	24	24	24	24
<i>Kaloula pulchra</i>	25	25	25	25
<i>Odorrana hossi</i>	25	25	25	25
<i>Polypedates leucomystax</i>	24	25	24	24
<i>Kaloula baleata</i>	25	25	25	25
<i>Philautus mjobergi</i>	23	23	23	25
<i>Phrynoidis aspera</i>	25	24	25	25
<i>Microhyla heymonsi</i>	21	22	20	23
<i>Microhyla butleri</i>	25	23	22	24
<i>Rhacophorus appendiculatus</i>	25	25	25	25
<i>Hylarana labialis</i>	24	25	25	25
<i>Philautus petersi</i>	24	25	25	25
Total	290	291	288	295
Percentage (%)	96.67	97	96	98.33

18.5 Conclusion

In this paper, an improvement classifier based on kNN is proposed to overcome the problem of existing outliers particularly in small training sample size situation. The overall accuracy shows that the proposed classifier outperforms the other classifiers with the most outstanding result using the automatic segmentation. By using automatic segmentation, their rates were further improved remarkable. From this experiment, it may be inferred that proposed classifier is effective for frog identification system and is comparable to several state-of-the-art methods regardless of their training sample size and future space dimension.

Acknowledgments The authors would like to thank the financial support provided by Universiti Sains Malaysia Short Term Grant, 304/PELECT/60311048, Research University Grant 814161 and Research University Grant 814098 for this project.

References

1. Bevier CR, Sonnevend A, Kolodziejek J, Nowotny N, Nielsen PF, Conlon JM (2004) Purification and characterization of antimicrobial peptides from the skin secretions of the mink frog *Rana septentrionalis*. *Comp Biochem Physiol* 139(1-3):31-38
2. Obrist MK, Pavan G, Sueur J, Riede K, Llusia D, Márquez R (2010) Bioacoustic approaches in biodiversity inventories. In: *Manual on field recording techniques and protocols for all taxa biodiversity inventories*. *Abc taxa*, vol 8. pp 68-99
3. Huang CJ, Yang YJ, Yang DX, Chen YJ (2009) Frog classification using machine learning techniques. *Expert Syst Appl* 36:3737-3743
4. Han NC, Muniandy SV, Dayou J (2011) Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. *J Appl Acoust* 72:639-645

5. Mitani Y, Hamamoto Y (2006) A local mean-based nonparametric classifier. *Pattern Recogn Lett* 27:1151–1159
6. Fukunaga K (1990) *Introduction to statistical pattern recognition*, 2nd edn. Academic, London
7. Zeng Y, Yang Y, Zhao L (2009) Nonparametric classification based on local mean and class statistics. *Expert Syst Appl* 36:8443–8448
8. Gou J, Yi Z, Du L, Xiong T (2012) A local mean-based k-nearest centroid neighbor classifier. *Comput J* 55(9):1058–1071
9. Zuo W, Wang K, Zhang H, Zhang D (2007) Kernel difference-weighted k-nearest neighbors classification. *ICIC* 2:861–870
10. Jena PK, Chattopadhy S (2012) Comparative study of fuzzy k-nearest neighbor and fuzzy c-means algorithms. *Int J Comput Appl* 57(7):22–32
11. Jaafar H, Ramli DA (2013) Automatic syllables segmentation for frog identification system. In: 2013 IEEE international colloquium on signal processing and its application, vol 9
12. Hasan MH, Jaafar H, Ramli DA (2012) Evaluation on score reliability for biometric speaker authentication system. *J Comput Sci* 8(9):1554–1563

Title: Development of Touch-Less Palm Print Biometric Authentication System to Smart Phone using Android Operating System

Introduction

The emerging of internet and wireless dimension has brought a new era in biometric technology. Instead of operating the biometric system with fixed or static device for example auto teller machine (ATM), mobile biometric system can be implemented and this approach leads to more efficient and reliable implementation [1,2,3]. Mobile biometric system is a biometric system that extends the functionality and capabilities of a static biometrics by allowing user to capture any biometric data out in the field. Mobile devices such as smart phone, tablet, laptop and handheld gadget can be used for this purpose. Mobile biometric device is designed for intuitive operation by integrating a reader, scanner or camera for data capturing. Subsequently, by converting the biometric data to the digital format, the authentication or identification process is done either locally where the database and processing software are stored on the handheld device itself or remotely by sending the captured biometric data to the centralized biometric server [4,5]. For the local implementation, the biometric device consumes more memory for storing the database and processing software inside its space and this requires a high-end device to achieve good performance [6,7]. On the other hand, for remote verification, the mobile biometric device communicates through common wireless technologies such as cellular, Wi-Fi or Bluetooth with the server in which the verification and identification process is run.

In this study, the biometrics characteristics i.e.palm print information are acquired by the mobile device and are sent over the server to be processed and recognized. The developed application and system can be implemented on any smart phone which uses Android operating system. In brief, the application which is running on the smart phone acquires a person's characteristics and the server is the side where the recognition process is accomplished as summarized in Figure 1.

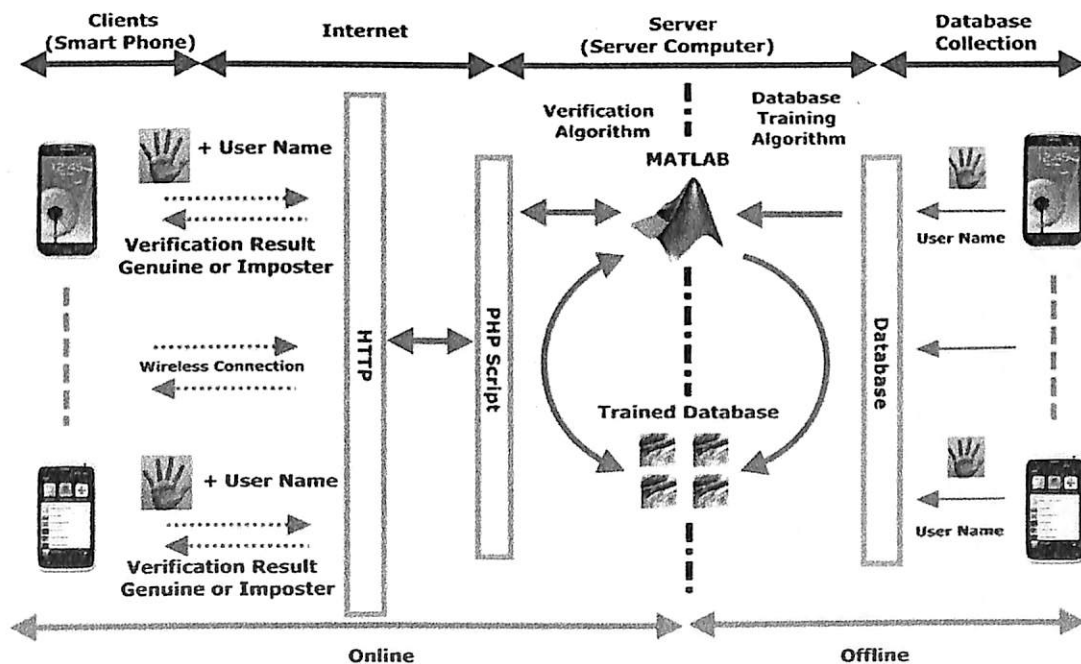


Figure 1: System architecture

Objectives

1. To develop an android application for palm print data collection and to collect data using the application.
2. To develop hand image identification and region of interest (ROI) extraction algorithms and to implement several subspace based feature processing algorithms for fast and accurate verification performances.
3. To develop a mobile palm print biometric system by setting up the client to server and server to client communication. The system performances based on the developed algorithms in objective 2 are then evaluated.

Methods

In real-time verification system for smart phone devices, the whole chain of process in the biometric system is computed until the results are sent back to the smart phone devices. The process is executed by two main processing sides which are the developed Android application and the server. The process starts with the data collection from the testing subject in the smart phone by the developed application. After the input of palm print image and username, the data are sent through the local network or wireless internet to the server. Next, the verification algorithm which runs in MATLAB programming is used to verify the subject's palm print by comparing the palm print pattern with its corresponding model in the database. Once the verification is complete, the results are sent back to the running mobile application through local network or wireless internet. Figure 2 shows the overview of real-time palm print verification system for smart phone while the developed application is shown in Figure 3.

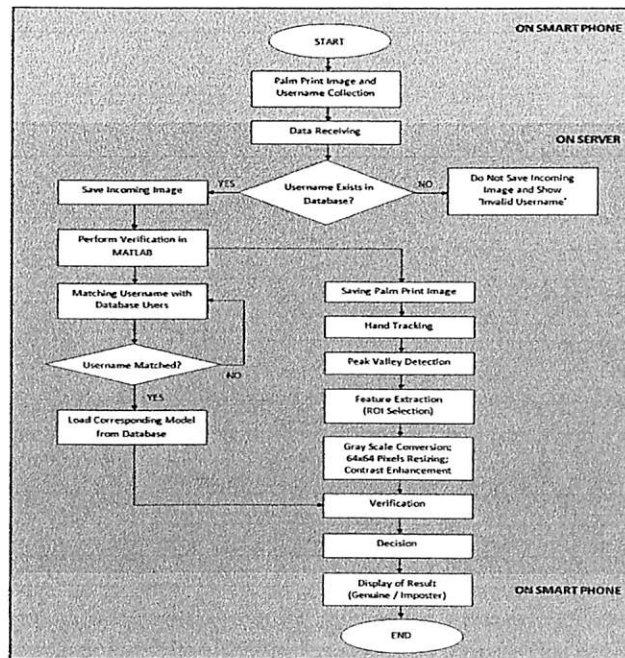


Figure 2: Overview of real-time palm print verification system for smart phone

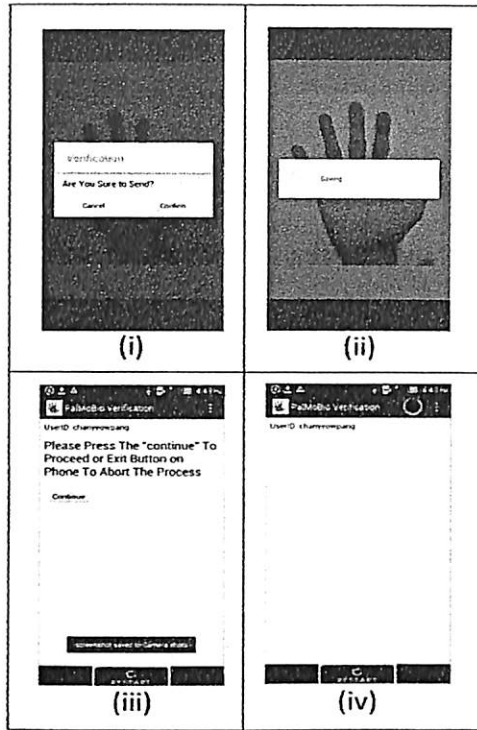


Figure 3: Data sending, (i) Confirmation of sending, (ii) Image saved into phone memory, (iii) Confirmation to proceed, (iv)Results waiting.

Biometric Verification on the Server Side

After receiving the hand image and the username from the smart phone device, the data are fed into the biometric system for verification using MATLAB programming. The hand image undergoes biometric process which includes data preprocessing, feature extraction and classification while the username is used to extract the corresponding trained model. Hand image processing is given as in Figure 4. In the classification, the pre-loaded model and the extracted image undergo pattern matching using the SVM classifier. The classifier yields the result in terms of predict label, decimal value and accuracy. Display result on smart phone is shown as in Figure 5.

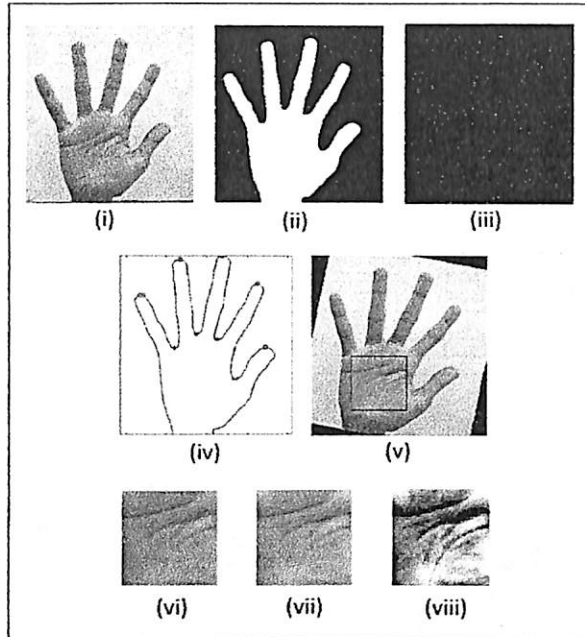


Figure 4: Hand image processing, (i) Captured hand image, (ii) Binarized image, (iii) Tracked hand contour, (iv) Peak valley detection, (v) ROI Selection, (vi) ROI image, (vii) Gray scale image, (viii) Contrast-enhanced-image.

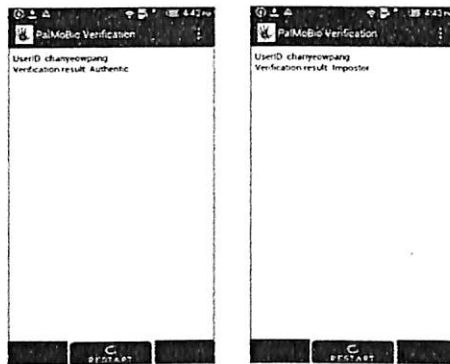


Figure 5: Displays of verification results on the phone

Results and discussion

Verification Activity

The developed Android application result is shown by Graphic User Interface (GUI). The GUIs for enrolment activity for the 3 devices which are HTC One X, Samsung S3 and Samsung Tab 2 are shown in this section. The GUIs for the 3 devices are almost similar but the captured image qualities are different. Figure 6 shows the interface sequence for verification activity. First, the user needs to sign in by keying in the username. Then the hand image is captured.

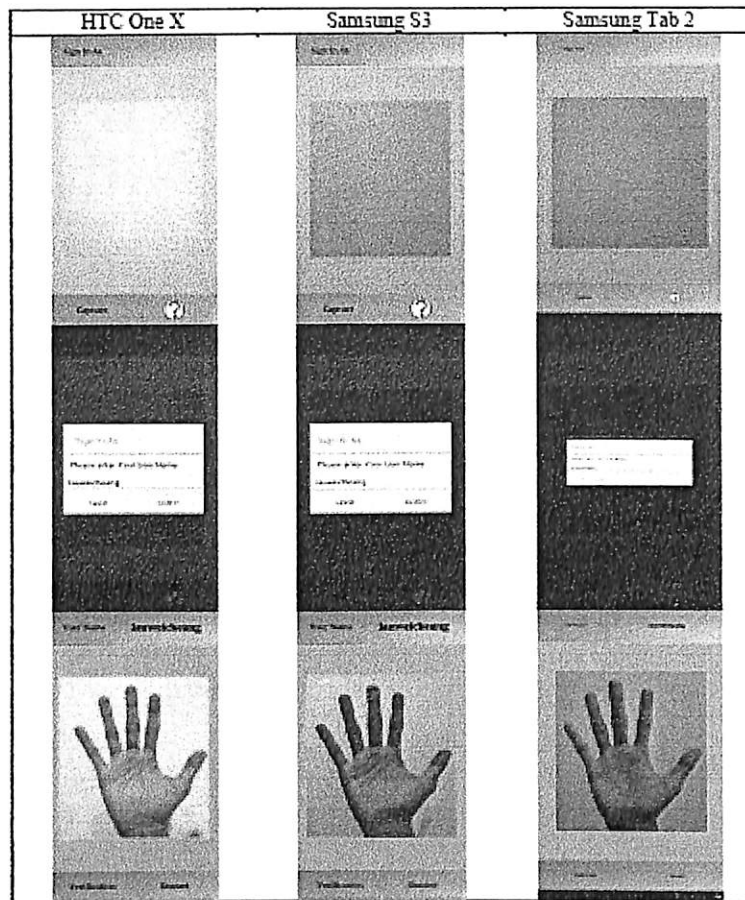


Figure 6: Verification process – sign in process and the captured test image

Once the hand image is captured, user presses the verification button and confirmation dialog box will appear. If the hand image quality is good, the user needs to press 'Confirm' to proceed to the verification stage and the application will be connected to server via internet connection. After the verification process is done in the server, the verification results will be displayed either authentic or imposter. However, if the user name does not exist at the database, "Invalid User ID" is shown.

Collected Palm Print Images

Figure 7 is the sample of hand image in palm print database. The figure shows 3 different subjects (5 sample each subject) captured by 3 different devices. Based on the collected database, the Samsung brand devices, S3 and Tab 2 palm images color appeared darker as compared to HTC One X device although the location and lighting condition is the same. Due to there is an autofocus function in Samsung Galaxy S3 and HTC One X, the captured image is clearer and the quality of the image is higher. The ridges and wrinkles are clearly seen although they are fine. While the Samsung Galaxy Tab 2 image is not clear as compared to Samsung S3 and HTC One X because of the Samsung Tab 2 camera does not have autofocus function. For most of the Samsung Galaxy Tab 2 palm images, only the principle lines can be seen clearly whereas the wrinkles and ridges are hardly seen.

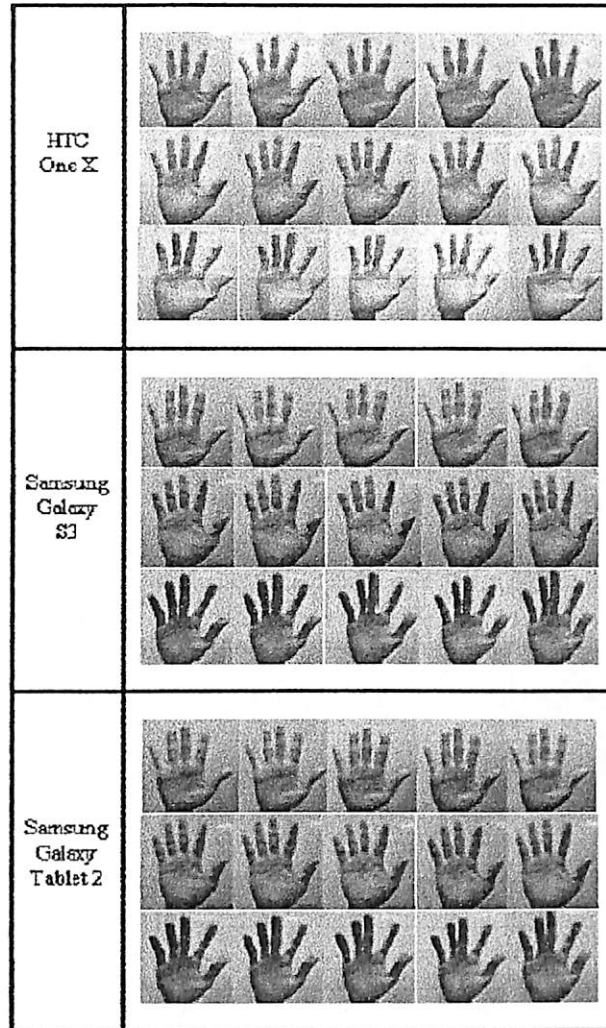


Figure 7: Some of the hand image captured by three different devices that has been stored into palm print database

Performance Result Analysis

Figure 7 to 9 show the overall system performances based on ERR and matching time for HTC One X, Samsung Galaxy S3 and Samsung Galaxy Tablet 2 devices, respectively. From this observation, it can be concluded that RSKPCA feature extraction method gives the best performance by reducing the features dimension and good EER percentage compared with the baseline approaches for the devices.

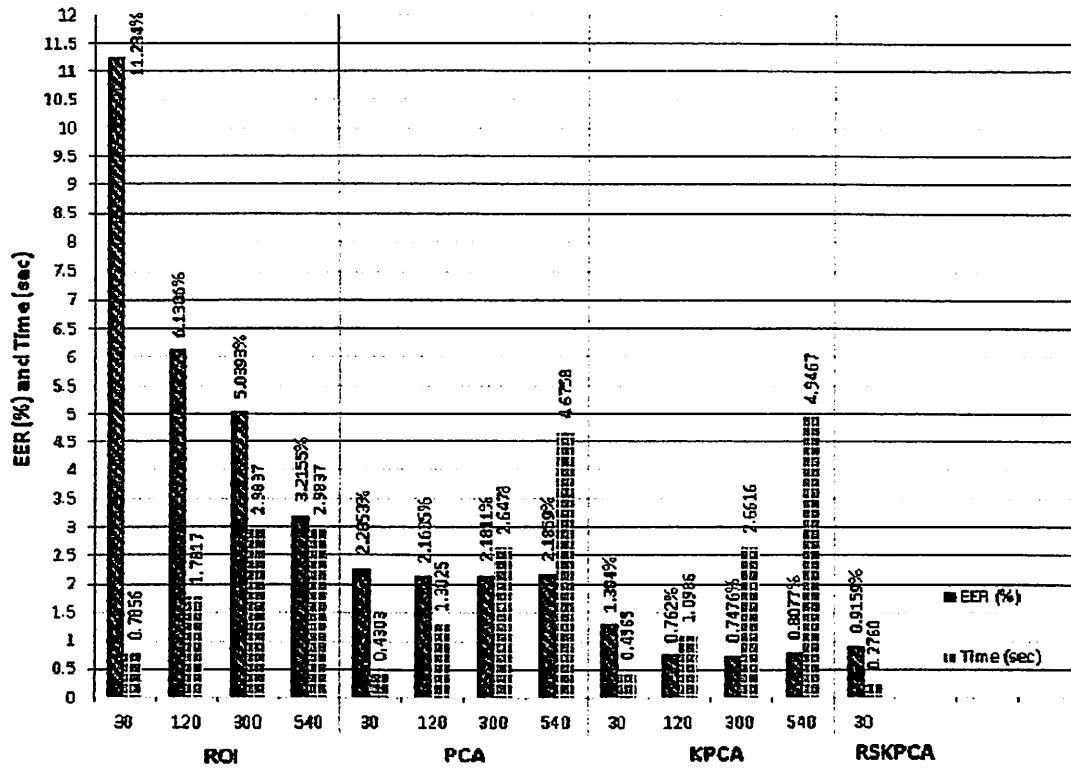


Figure 7: Overall system performances based on EER and matching time for HTC One X

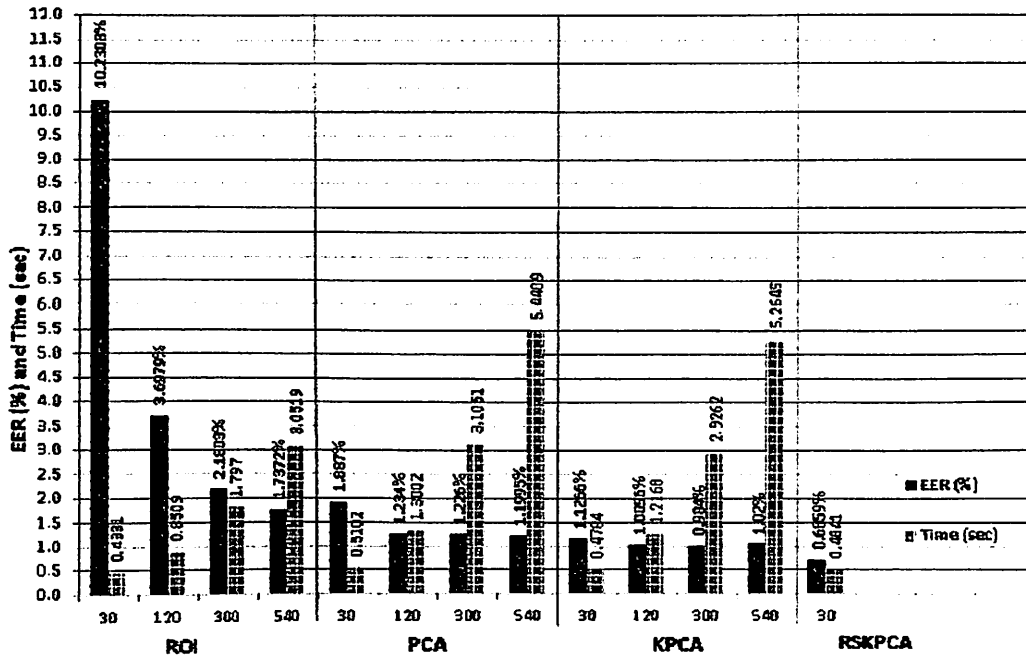


Figure 8: Overall system performances based on EER and matching time for Samsung Galaxy S3

- [5] Zhang, D., Guo, Z., Lu, G., Zhang, L., Liu, Y. & Zuo, W. 2011. Online joint palm print and palmvein verification. *Expert Systems with Applications*, 38, p.2621-2631.
- [6] Zhang, D., Kong, W., You, J. & Wong, M. 2003. On-line palm print identification, *IEEE Transaction on PAMI*, 25 (9), p.1041-1050..
- [7] Caldwell, T. 2010. Biometric access to mobile in pipeline. *Biometric Technology Today*, 2.

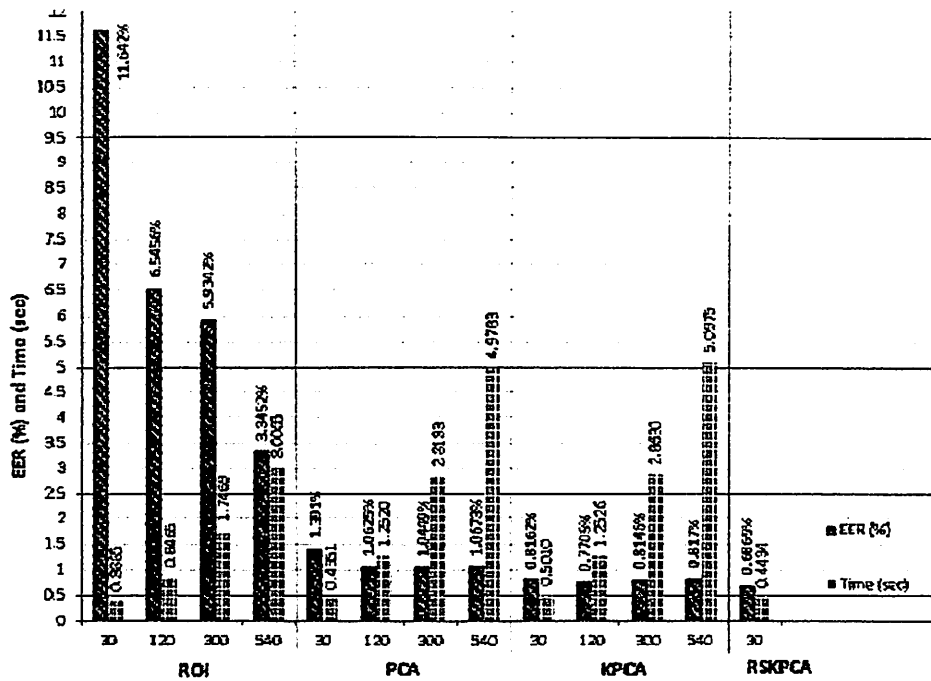


Figure 9: Overall system performances based on EER and matching time for Samsung Galaxy Tablet

Conclusion and Suggestion

The first objective is to collect palm print data by using smartphone device for development of palm print database and develop hand tracing and ROI extraction. The database of 60 palm print images from different 40 subjects is been created. An auto hand tracing and ROI extraction technique has been successfully developed.

The second objective is to evaluate palm print processing algorithm for fast and accurate verification performance. In this project, ROI, PCA, KPCA and RSKPCA approach has been implemented. The results show that the performance was validated by comparing the processing time and EER percentage. The ROC curve was used to scrutinize the trade-off between the FAR and GAR performance. According to the experimental results, RSKPCA approach gives the best performance result which is faster and accurate compared to other approach.

The third or the last objective of this project is achieved by implementing a mobile palm print biometric based on Android operating system which can perform the palm print biometric verification system. The system is able to acquire the palm print image form the Android devices and send the user name and palm print image to the server via the internet. The server is able to receive the data from multiple clients at the same time. The server receives the user name and hand image, performs the verification process and then sends the verification result to the Android devices.

Acknowledgement

We gratefully thank the Malaysian Ministry of Higher Education and Malaysia Ministry of Science, Technology and Innovations as this research which is headed by Dr Dzati Athiar Ramli was supported under Research University (RU) Grant, Universiti Sains Malaysia, 1001/PELECT/814161 and Incentive Grant, Universiti Sains Malaysia.

References

- [1] Jain, A.K., A. Ross and S. Prabhakar, 2004. An introduction to biometric recognition. *IEEE Trans. Circuits Syst. Video Technol.*, 14: 4-20. DOI: 10.1109/TCSVT.2003.818349
- [2] Jain, A., K. Nandakumar and A. Ross, 2005. Score normalization in multimodal biometric systems. *Patt. Recogn.*, 38: 2270-2285. DOI: 10.1016/j.patcog.2005.01.012
- [4] Mir, A.H., Rubab, S. & Jhat, Z.A. 2011. Biometric Verification: A Literature Survey. *Journal of Computing & ICT Research*. 5(2), p.67-80.