

Multimodal Biometric Recognition System for Cloud Robots

Shuqing Tian¹, Sung Gyu Im¹ and Suk Gyu Lee¹

¹ Department of Electrical Engineering

Yeungnam University

Gyeongsan, Korea 712-749

tianshuqing@gmail.com, tamop87@nate.com, sglee@ynu.ac.kr

Abstract

This paper presents a Multimodal Biometric Recognition System (MBRS) which is capable of integrating various biometric information for person recognition. The MBRS is deployed as a cloud server and provides person recognition service for smart robots. Through the experiments based on multimodal biometric traits, the fact that the multimodal biometric recognition performs better than individual biometric recognition has been proved. In our approach, the implementation of a multimodal biometric recognition system based on face recognition system and voice recognition system is proposed. The MBRS provides the possibility of integrating multi biometric subsystems to do recognition. Even more, since the MBRS is deployed as a cloud server, the public interfaces were provided for the robots to do real-time person recognition. The experimental results show that the MBRS outperforms any individual face recognition subsystem and voice recognition subsystem.

Keywords: Face recognition, Voice recognition, Multimodal biometric recognition system, Cloud robotics

1. Introduction

Biometric recognition is related with many fields such as: security systems [Lai *et al.* 08], surveillance systems [Garibotto, 09], computer interface [Gips *et al.*, 02], and service robot [Kim *et al.*, 05]. In these systems, various biometric traits are used for identification. Such are fingerprint [Hong *et al.*, 97], palm print [Zhang *et al.*, 03], face [Harmon *et al.*, 81], iris [Daugman *et al.*, 93], voice [Campbell *et al.*, 97], hand geometry [Jain *et al.*, 99], gait [Moeslund, 01], gesture [Lee *et al.*, 06], and so on.

However, most of these systems are unimodal. It means that the systems have employed only a single biometric trait. Consequently, a variety of limitations such as noisy data, spoof attacks [Jain, 04], nonuniversality [Golfarelli *et al.*, 97] and unacceptable error rates (false rejection rate and the false acceptance rate) have been realized.

Multimodal biometric systems which fuse two or more biometric traits can overcome the problems caused by unimodal system and decrease the error rates. There are three levels of fusion, fusion at the feature extraction level, fusion at matching score level and fusion at decision level [Ross *et al.* 04]. In our approach, the matching score level fusion was deployed which used the weighted averaging method.

Although multimodal biometric recognition systems show better performance than the unimodal ones, the processed data and calculation also increase undesirably. It is a difficult work for an embedded system in robots to run a multimodal biometric recognition system. With the development of cloud robotics [Hu *et al.*, 12], it is feasible for cloud robots to offload computationally intensive tasks to cloud servers.

Particularly, in our approach, face and voice recognition systems were integrated into a multimodal biometric recognition system (MBRS), and deployed in a high performance

computer as a cloud service. The cloud server provides cloud recognition service for the smart robots. The simulations showed that the performance of the MBRS is better than individual face and voice recognition systems. Using MBRS for a robot is a good implementation of the cloud robotics system.

2. Related Work

The development of multimodal biometric systems grows rapidly. Different combination of biometric traits was employed in recognition systems. Kartik *et al.* presented a multimodal biometric system using speech and signature features [Kartik *et al.*, 08]. The system provided good performance even the biometric data are affected by noise. Also, Lakshmiprabha *et al.* presented a new multimodal biometric approach using face and particular biometric which performed better than other face recognition and individual biometric methods [Lakshmiprabha *et al.*, 11]. A multimodal biometric recognition system based on fusion of palm print, fingerprint and face was introduced by [Chaudhary *et al.*, 11]. The system overcame the limitations of individual biometric systems and also met the response time as well as the accuracy requirements. In [Chang *et al.*, 03], they have investigated the comparison and combination of 2D and 3D face recognition at the matching score level. The recognition rate of their multimodal system was statistically significantly greater than either 2D or 3D alone. Kumar *et al.* [Kumar *et al.*, 03] presented a new approach for the personal identification using hand images which improved the performance of palm print-based verification system. Jain *et al.* [Jain *et al.*, 99] integrated various fingerprint matching algorithms to improve the performance of a finger print verification system. A person identification system based on acoustic and visual features Roberto was presented by [Brunelli *et al.*, 95]. The performance of the integrated system is shown to be superior to that of the acoustic and visual subsystems. Bigün *et al.* develop a statistical framework based on Bayesian statistics to integrate the speech (text dependent) and face data of a user [Bigün *et al.*, 97]. The estimated bias of each classifier is taken into account during the fusion process.

Biometric recognition systems were also deployed in robots. Lim *et al.* presented a voice command system for mobile robots which used text-dependent voice recognition technology [Lim *et al.*, 97]. Commercial humanoid service robot using face and voice recognition is also available [REEM].

3. Implementation of Multimodal Biometric Recognition System

The main title (on the first page) should begin 1 3/16 inches (7 picas) from the top edge of the page, centered, and in Times New Roman 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Please initially capitalize only the first word in other titles, including section titles and first, second, and third-order headings (for example, “Titles and headings” — as in these guidelines). Leave two blank lines after the title.

3.1. System Architecture

The system architecture of the MBRS is shown in Figure 1. It consists of three main components, i) Face Recognition Subsystem (FRS), ii) Voice Recognition Subsystem (VRS) and iii) Weighting Factor Assignment Component (WFAC).

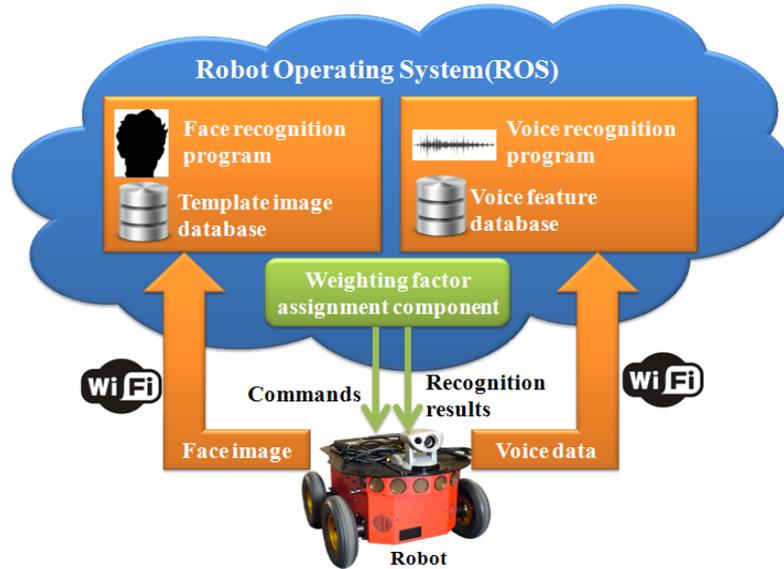


Figure 1. Architecture of Multimodal Biometric Recognition System

Particularly, the proposed MBRS is based on face and voice recognition. However, more biometric recognition systems (finger print, iris and so on) will be integrated into it in the future. In the MBRS, each biometric recognition system is an individual component. It is convenient for a unimodal biometric recognition system to integrate into or disassemble from the MBRS. In addition, the MBRS can be also used as a unimodal biometric recognition system when the other subsystems are set disabled.

The inputs of the MBRS are biometric traits information which captured by sensors. As shown in Figure 1, vision sensor and acoustic sensor are required. Robots in client side with these sensors can easily access this cloud service by using Wifi connection.

3.2. Face Recognition Subsystem (FSR)

The processes of face recognition consist of two main steps, face detection and face recognition. The purpose of face detection is finding a face in the acquired images or videos. Then face recognition algorithm will be applied to match detected face with known faces in the database.

In our approach, the OpenCV library is used to do the face recognition. The Haar-like features was employed for face detection [Mita *et al.*, 05]. The principal component analysis (PCA) was applied for face recognition [Moon, 01]. The Face Recognition Subsystem (FRS) workflow is shown in Figure 2.

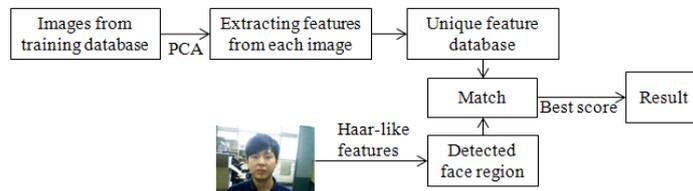


Figure 2. The Workflow of Face Recognition Subsystem

The performance of the subsystem can be tuned by increasing or lowering the threshold value. There is a function named “threshold” in the OpenCV library. The parameters passed to this function have been adjusted several times, and in the end, the performance we reached ranges from FRR of 0.04% with FAR of 24.40% to FRR of 1.65% with FAR of 10.38%.

3.3. Voice Recognition Subsystem (VSR)

There are two types of voice recognition systems: text-independent voice recognition system and text-dependent voice recognition system [Ross *et al.* 04]. In our approach, [Harrington *et al.*, 08] was referenced to implement the text-independent voice recognition system. The mechanism of the system will be described in Section 5. The average approximation method was employed as the voice recognition algorithm. The workflow of voice recognition is shown in Figure 3. In order to get precise test results, we recorded four 5-minute long voice audio files in a quiet environment and used these four audios as the training data instead of real-time speaking.

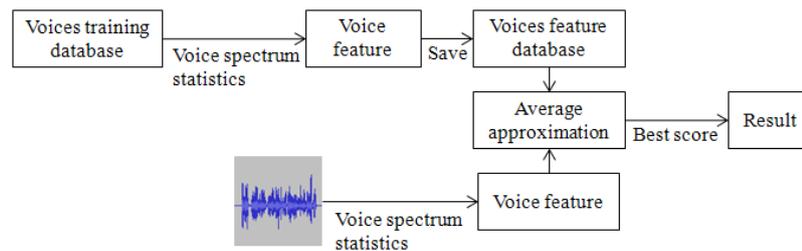


Figure 3. The Workflow of Voice Recognition Subsystem

4. Multimodal Biometric Fusion Method

The proposed MBRS employed the fusion at the matching score level [Ross *et al.* 04]. In our fusion method, the weighted averaging has been used to combine FRR and FAR by face and voice recognition subsystems in different brightness and noise level conditions. The proposed fusion method is as follows. Suppose $M = \{\text{all unimodal biometric recognition systems (face, voice, iris, fingerprint ...)}\}$ and $n = |M|$ (the length of set M).

$$\begin{aligned}
 FRR_m &= \sum_{i=1}^n FRR_i * W_i \\
 FAR_m &= \sum_{i=1}^n FAR_i * W_i \\
 \sum_{i=1}^n W_i &= 1
 \end{aligned}$$

In the equations, the FRR_m refers to FRR by the MBRS. The FAR_m refers to FAR by the MBRS, where W_i is the weighting factor of unimodal biometric system i in the set M . The FRR_i and the FAR_i are FRR and FAR by system i respectively. The purpose of system fusion is to decrease both FRR and FAR as much as possible.

We considered the conditions which robots serviced in. In the dark and quiet condition where the FRR and FAR by the VRS are lower than those in the FRS, weighting factor of VRS should be 1 and weighting factor of FRS is 0. It means the MBRS solely relies on the VRS. Vice verse in very noisy but normal brightness conditions, weighting factor of the FRS should be 1 and weighting factor of VRS is 0.

In some complex conditions, both lightning intensity and noise problems exist. The least-square method was used to calculate the FRR difference and FAR difference in the two subsystems, and the bigger difference would be retained to calculate the weighting factor.

In the source code, the W_F and W_V are the specific W_i . It shows that if the FRR difference is bigger than FAR difference, a smaller weighting factor will be assigned to the subsystem which has the higher FRR, and vice verse. We used this method to

maximally decrease the FRR and FAR. The experimental showed that this method is feasible.

5. Experiment Results

5.1. Hardware System

The proposed MBRS is employed as a cloud server for the robots to do the face and voice recognition. As shown in Figure 5, the robot just has a webcam and an embedded microphone on it. Benefiting by the Cloud Robotics technology, the implemented robot is able to offload computation-intensive tasks to the MBRS server and access to vast amounts of data with the small size of 150mm (length)*100mm (width)*100mm (height).



Figure 4. The Smart Robot based on the MBRS

The robot carries the Wifi receiver and communicates with the MBRS by using the UART communication protocol.

5.2. Performance Comparison between the MBRS and the FRS

We have 25 template face images (640*480 pixels) in face image database (Figure 6) and another 25 test face images. In order to test the recognition rate in different brightness conditions, we modified the brightness (from -90% to 90%) of the test images to simulate face in different lighting intensity (Figure 6). The subsystem snapshot is shown in Figure 7. It's based on the OpenCV library, some user interfaces and communication functions were added.



Figure 5. Template Face Images



Figure 6. The Simulation of Faces in Different Lighting Intensity

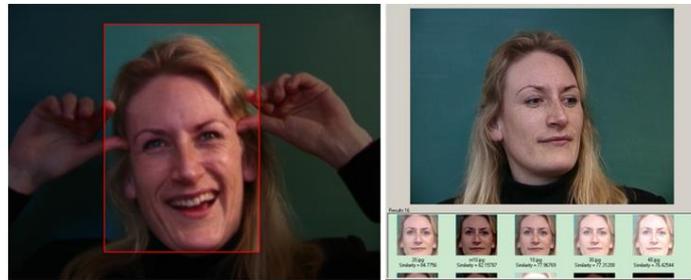


Figure 7. Face Detection (Left) and Face Recognition (Right)

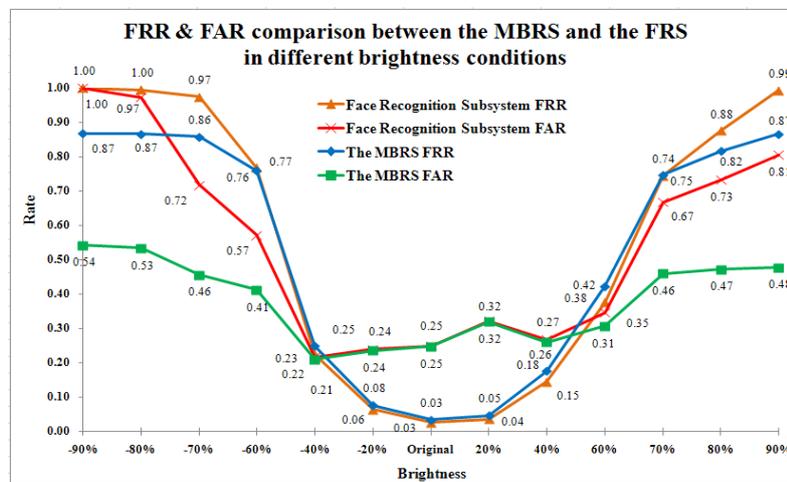


Figure 8. FRR and FAR Comparison (Figure form) between the MBRS and the FRS

The FAR and FRR comparison between the MBRS and the FRS is shown in Fig. 9. It shows that in the normal brightness conditions (from -40% to 40%), error rates (FAR and FRR) by the MBRS and FRS are nearly the same. In the darker (brightness less than -40%) and brighter (brightness more than 40%) conditions, the error rates by the MBRS are lower than the FRS. In these conditions, the MBRS performs better than single FRS. Because in the darker and brighter conditions, the weighting factor of the VRS is bigger which means the MBRS relies on the VRS more in these conditions. Compared with the FRS, the MBRS decreased the FRR and FAR by 22.67% and 10.86 in average.

5.3. Performance Comparison between the MBRS and the VRS

In the comparison test of the MBRS and the VRS, we recorded four five-minute long audio files which called template voices by using professional equipments. In order to test in different noise decibel conditions, we made six fixed decibel noise audio files (-15dB, -20dB, 25dB, -30dB, -35dB and -40dB) and combined each noise audio file with each template voice. Finally 36 test audio files (32 with noise and 4 without noise) were obtained which called test voices (Figure 10). The subsystem is based on the “sndpeek” software. Once the speaker is recognized by the system, the speaker’s name will be printed in the screen. The function calculating the printed name is added. Based on the

function, the FAR and FRR will be calculated. The test result compared with MBRS is shown in Figure 11. In the Figure 11, the FRR and FAR in different noise level is tested. With the increase of noise level, the errors become bigger. Since the MBRS includes the face recognition subsystem, when the noise level is high (-5), the MBRS relies on face recognition subsystem which is capable of increasing the recognition ratio.

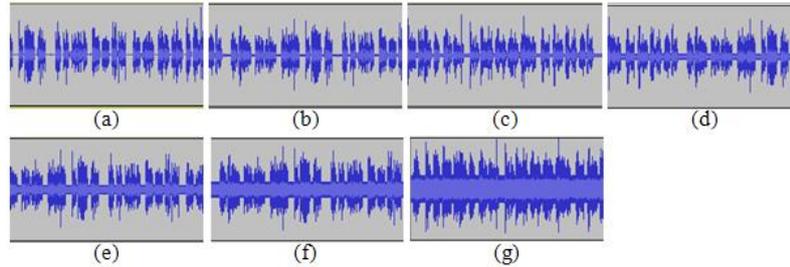


Figure 9 Test Voice Sound Wave; (a) Template voice sound wave, (b) Template voice mixed with -40 decibel noise, (c) Template voice mixed with -35decibel noise, (d) Template voice mixed with -30 decibel noise, (e) Template voice mixed with -20 decibel noise, (f) Template voice mixed with -15 decibel noise, (g) Template voice mixed with -25 decibel noise

	Noise level	FRR	FAR		Noise level	FRR	FAR
VRS	-5	100.00%	100.00%	MBRS	-5	51.58%	50.78%
	-10	100.00%	100.00%		-10	51.58%	50.78%
	-15	100.00%	100.00%		-15	51.58%	50.78%
	-20	99.30%	56.25%		-20	53.54%	41.32%
	-25	97.13%	48.86%		-25	52.78%	37.90%
	-30	86.13%	43.37%		-30	48.68%	34.48%
	-35	69.76%	23.58%		-35	43.01%	23.16%
	-40	64.30%	8.57%		-40	41.27%	14.61%
	No noise	64.44%	7.63%		No noise	42.21%	13.20%

Figure 10. Voice Recognition Subsystem, (a) Program Running Snapshot (b) The VRS Test Result Snapshot

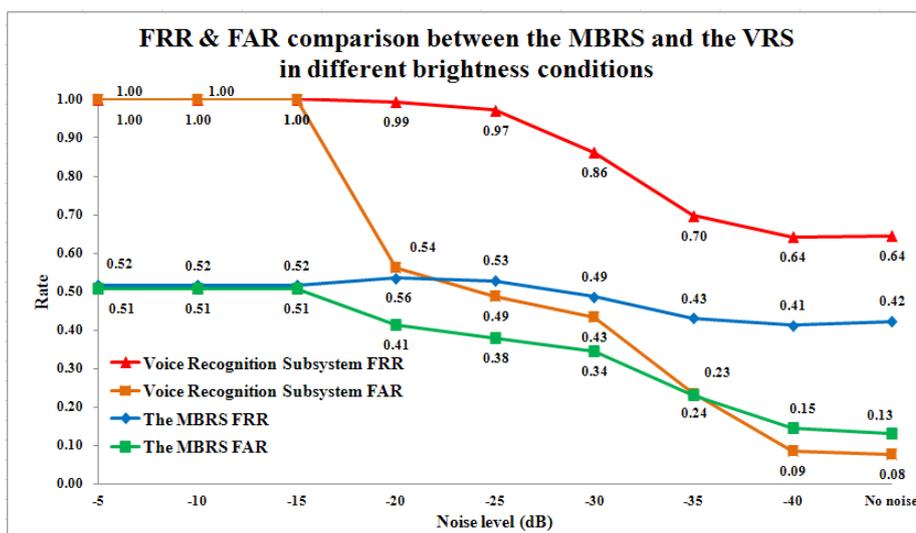


Figure 11. FRR and FAR Comparison between MBS and the VRS

The FAR and FRR comparison between the MBRS and the VRS are shown in Figure 12. It shows that when the noise decibel is bigger than -35dB, the error rates by the MBRS are lower than those in the VRS. However, in the region from -40dB to no noise, the FAR by fusion system becomes higher than the VRS. In the “no noise” condition, the FAR increase is 5%, but the FRR decrease by 22%, the whole performance was also improved. Compared with the VRS, the MBRS decreased the FRR and FAR by 14.27% and 28.74% in average.

6. Conclusions

In this paper, an enhanced multimodal biometric recognition system (MBRS) by integrating two biometric features (*i.e.*, face and voice) is presented. The MBRS is deployed a cloud service and provides person recognition service for the robots. Unimodal face and voice recognition systems will encounter problems in some conditions. For face recognition, the lighting intensity could affect the recognition rate. The recognition rate by voice recognition system will be affected by environmental noise. In order to overcome these problems, we imitated human physiology that added an eye (face recognition system) and an ear (voice recognition system) to robot. The face recognition system can overcome the problems caused by environmental noise and the voice recognition system can overcome the problems by lighting intensity. The proposed multimodal biometric recognition system which integrated face and voice recognition subsystems performs better than individual face and voice recognition systems.

In the experiments, we have evaluated the proposed method which combines noise with template voice to simulate speaker in different noise decibel environment. In this way, the relationship between noise decibel and recognition rate can be recorded precisely. We have also evaluated the face recognition method by changing the brightness of a face image to simulate the person in different lighting intensive environment. Eventually, we have shown that the multimodal biometric recognition system decreased the error rates (the FRR and FRA) which meant the recognition rate was increased.

References

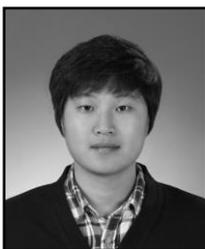
- [1] L. Lai, S-W. Ho and H. V. Poor, “Privacy–Security Trade-Offs in Biometric Security Systems”, Communication, Control, and Computing, 2008 46th Annual Allerton Conference, (2008) September 23-26.
- [2] G. Garibotto, “Video Surveillance and Biometric Technology Applications”, IEEE Advanced Video and Signal Based Surveillance, 2009. AVSS '09, (2009) September 2-4.
- [3] M. B. J. Gips and P. Fleming, “The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities”, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 10, (2002) March.
- [4] D-H. Kim, J. Lee, H. S. Yoon, H. J. Kim, Y. Cho and E. Y. Cha, “A vision-based user authentication system in robot environments by using semi-biometrics and tracking”, Intelligent Robots and Systems, IEEE/RSJ International Conference, (2005) August 2-6.
- [5] A. K. J. L. Hong and S. P. R. Bolle, “An identity authentication system using fingerprints”, Proceedings of the IEEE, vol. 85, (1997), September.
- [6] D. Zhang, W. K. Kong, J. You and M. Wong, “Online palmprint identification”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, (2003) September.
- [7] L. D. Harmon, M. K. Khan, R. Lasch and P. F. Raming, “Machine identification of human faces”, Pattern Recognition, vol. 13, (1981).
- [8] J. G. Daugman, “High confidence visual recognition of persons by a test of statistical independence”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, (1993), November.
- [9] J. Campbell, “Speaker recognition: a tutorial”, Proceedings of the IEEE, vol. 85, (1997) September.
- [10] A. K. Jain, A. Ross and S. Pankanti, “A prototype hand geometry based verification system”, Proceedings of the Second International Conference on Audio- and Video-based Biometric Person Authentication, (1999).
- [11] T. Moeslund and E. Granum, “A survey of computer vision-based human motion capture”, Computer Vision and Image Understanding, vol. 81, (2001), March.
- [12] S. W. Lee, “Automatic gesture recognition for intelligent human-robot interaction”, Automatic Face and Gesture Recognition, 7th International Conference, (2006) April 2-6.

- [13] A. K. Jain and A. Ross, "Multibiometric systems, Communications of the ACM, vol. 47, (2004) January.
- [14] M. Golfarelli, D. Maio and D. Maltoni, "On the error-reject trade-off in biometric verification systems", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, (1997) July.
- [15] P. Kartik, P. S. R. Mahadeva and P. R. V. S. S. Vara, "Multimodal biometric person authentication system using speech and signature features", TENCON 2008 - 2008 IEEE Region 10 Conference, (2008), November 19-21.
- [16] N. S. Lakshmiprabha, J. Bhattacharya and S. Majumder, "Face recognition using multimodal biometric features", Image Information Processing (ICIIP), 2011 International Conference, (2011) November 3-5.
- [17] S. Chaudhary and R. Nath, "A Multimodal Biometric Recognition System Based on Fusion of Palmprint", IEEE Transactions on Fingerprint and Face, Signal Processing, vol. 59, (2011) December.
- [18] S. H. Lim and J. W. Jeon, "Multiple Mobile Robot Interface Using a Low Cost Voice Recognition Chip", Robot and Human Communication, 6th IEEE International Workshop, (1997) 29 September-1 October.

Authors



Shuqing Tian, he received the B.S. degree in information Security Engineering from Yunnan University, China, in 2009, and he received the M.S. degree in Robotics Engineering from Yeungnam University in 2004. His research interests include robotics, intelligent device, biometric recognition.



Sung Gyu Im, he is currently a graduate student in M.S program in Department of Electrical Engineering, Yeungnam, University. His research interests include mobile robotics, and SLAM.



Suk Gyu Lee, he received the B.S. and M.S. degree in Electrical Engineering from Seoul National University in 1979, and 1981 respectively, and he received the Ph.D. degree in Electrical Engineering from UCLA in 1990. His research interests include robotics, SLAM, nonlinear control and adaptive control.

