

# ADAPTIVE WEIGHT ESTIMATION IN MULTI-BIOMETRIC VERIFICATION USING FUZZY LOGIC DECISION FUSION

Henry Pak-Sum Hui<sup>1</sup>, Helen M. Meng<sup>1</sup> and Man-Wai Mak<sup>2</sup>

<sup>1</sup>Dept. of Systems Engineering and Engineering Management, The Chinese University of Hong Kong

<sup>2</sup>Dept. of Electronic and Information Engineering, The Hong Kong Polytechnic University

<sup>1</sup>{pshui, hmmeng}@se.cuhk.edu.hk, <sup>2</sup>enwmak@polyu.edu.hk

## ABSTRACT

This paper describes a multi-biometric verification system that is fully adaptive to variability in data acquisition using fuzzy logic decision fusion. The system uses fuzzy logic to dynamically alter the weight of three biometrics (face, fingerprint and speech), taking into account the variations during data acquisition (e.g. lighting, noise and user-device interactions). A specific decision boundary can be determined by this dynamic weight assignment to make the authentication decisions. An overall EER improvement of 42.1% relative to weighted average fusion has been achieved.

**Index Terms**—Fuzzy logic, multi-modal, biometrics

## 1. INTRODUCTION

In this paper, we present a multi-biometric authentication system that is adaptive to variability in data acquisition using fuzzy logic. The system uses speaker verification, face verification and fingerprint verification for multi-biometric authentication. Multi-biometric overcame the limitation of single biometric authentication [1]. In multi-biometric fusion, especially in matching score level, weighted average fusion often out-performs other methods such as the product rule, min rule and max rule [2]. However, the weighted average fusion suffers from the lack of adaptability when the quality of testing data samples deviates severely from that of the training data.

To improve the robustness of the multi-biometric authentication system, we investigate the following factors that affect the quality of biometric samples: (1) distortions due to the environment (e.g. ambient noise for speech and lighting conditions for face), (2) user-device interactions (e.g. misplaced fingerprint relative to the capture device) and (3) mismatch between acquisition devices during enrollment and verification (e.g. recordings using the desktop PC or PocketPC). However, the precise relationships between these factors and their influence on verification performance of different biometrics are complex.

Conventional methods such as boolean logic are insufficient for handling the intra-user variations described above, because the use of a single decision threshold cannot adapt to such variations. The use of fuzzy logic fusion proposed in [3] to model the intra-class variations resolved the uncertainties in the quality of biometric samples incorporating human knowledge in terms of fuzzy rules. Different weights are assigned to biometric samples of different qualities using fuzzy logic [3] to create a dynamic decision boundary for determining a user's identity by a dynamically weighted average fusion.

In our previous work [3], we implemented the adaptive weight estimation components for the face biometrics using the user's head pose and image illumination. Similarly, for the finger biometrics we used fingerprint positioning and image clarity. In this paper, we present a fully adaptive biometric authentication system by incorporating an adaptive weight estimation

component for the speech biometrics in using Signal-to-Noise-Ratio (SNR).

## 2. EXPERIMENTAL CORPUS

Our experimental corpus is the CUHK Bilingual Speech Corpus (BSC) [3], which is the pilot data subset of M3 [4]. BSC contains bilingual speech data (English and Cantonese) from 16 subjects. In addition, it also contains frames extracted from face movement video captured by a PC webcam or a Pocket-PC camera, and fingerprint images captured using an optical fingerprint sensor. The data partitioning for enrollment and verification (development and test) are described in Table 1. Each subject acts as a client to his/her own profile and as an imposter to the other subjects' profiles. The data in the enrollment set is used for training speaker models, face master templates and fingerprint master templates. We used 10 and 24 samples of each biometrics per subjects in the development and test sets respectively.

**Table 1.** BSC data partitioning (16 subjects in total)

		Enrollment Set	Verification Set	
			Dev. Set	Test Set
Speech Utterances	Client	4032	160	384
	Imposter	-	2400	5760
Face Movement Video Frames	Client	576	160	384
	Imposter	-	2400	5760
Fingerprint Images	Client	96	160	384
	Imposter	-	2400	5760
Total		-	2560	6144

# of client data = # of subjects × # of samples

# of imposter data = # of profiles × (# of subjects - 1) × # of samples

There are four different device and environment combinations for the face movement videos: (1) Indoor Webcam videos (WI), (2) Outdoor Webcam videos (WO), (3) Indoor PocketPC Camera videos (PI) and (4) Outdoor PocketPC Camera videos (PO). Frames are randomly picked from different videos for enrollment and testing (Fig. 1). The verification performance under mismatched devices and environment are investigated to test whether the proposed system can adapt to these variations.



**Fig. 1.** Example face images recorded by diverse devices and environments, left to right are (1) WI, (2) WO, (3) PI, and (4) PO.

## 3. UNI-MODAL BIOMETRICS

### 3.1. Face Verification

To perform face verification, we use the FaceIt Verification SDK from Identix [5]. It uses local feature analysis to encode facial images for verification [6]. During training, 36 face movement video frames were enrolled into the system to create 3 master templates per user, where each template was created from 12

video frames. During verification, FaceIt first locates the face to generate a face-finding confidence out of 5 possible values {0, 2.5, 5, 7.5, 10}. It then matches the face image with the master templates and computes the matching score for each template. The maximum of the three scores is the output score. If the engine fails to locate the face in the image, the verification score will become zero. The verification result is shown in Table 2.

**Table 2.** Performance of facial verification based on facial images captured by different devices under different environments. Numbers are equal error rates (EER) in percentages.

Testing Type	Template Type (Overall EER = 26.55%)			
	Webcam Indoor (WI)	Webcam Outdoor (WO)	Pocket PC Indoor (PI)	Pocket PC Outdoor (PO)
WI	5.21	16.93	17.09	21.97
WO	19.83	6.19	26.26	16.86
PI	17.32	27.28	11.89	34.64
PO	21.04	17.05	32.59	11.30

### 3.2. Fingerprint Verification

Our fingerprint verification engine uses a direct gray scale minutiae detection approach for feature extraction [7]. The core point of a fingerprint is used as the reference point for alignment. Master templates are the files that contain the features values from the enrollment fingerprint images. All six fingerprint images in the enrollment set are converted into templates. During verification, the maximum among all similarity scores from each template is returned to a fusion server (see Section 6). The resulting EER of the test set is 5.05%.

### 3.3. Speaker Verification

We used the bilingual text-independent speaker verification system developed in [8] to perform speaker verification. English and Cantonese speech utterances from the subjects are used to train the speaker-dependent Gaussian-Mixture Model (GMM) with 512 mixtures. We used 12 mel-frequency cepstral coefficients (MFCC) features and their delta coefficients by 28ms Hamming window with 14ms (50%) window overlapping to form the feature vectors. A Universal Background Model (UBM) trained from all subjects data with 512 mixtures was used for normalization [9]. The EER of the test set is 5.80%.

## 4. MULTI-BIOMETRIC WEIGHTED AVERAGE FUSION

As the scores for the different biometrics lie in different ranges, namely [0, 9.99], [0, 99] and [-inf., +inf.] for face, fingerprint and speech, respectively. We use min-max normalization to scale the scores to the same range and compute the fused score, as shown in Equation (1).

$$fusedScore = (\sum_i w_i S_i) / (\sum_i w_i) \text{ where } S_i = \frac{s_i - \min_i}{\max_i - \min_i} \cdot 10 \quad (1)$$

$S_i$  is the normalized score,  $s_i$  is the matching score from the recognizer,  $\min_i$  and  $\max_i$  are the minimum and maximum scores from the development set and  $w_i$  is the weight for biometrics  $i$ .

Weighted average fusion is used as a baseline for comparison with our system. To obtain the best results in the development set, we vary the weights ( $w_i$ ) of different biometrics (summed to one) within the [0, 1] range in steps of 0.01. The performance of weighted average fusion is shown in Table 3.

**Table 3.** Verification performance of weighted average fusion method on the test set in EER (%)

Testing Type	Template Type (Overall EER = 1.71%)			
	WI	WO	PI	PO
WI	1.56	1.12	1.29	2.00
WO	1.82	0.78	1.35	1.04
PI	1.56	1.97	1.82	2.82
PO	1.82	2.08	2.08	1.56

Weighted average fusion gives promising results in improving the verification performance by using a static weight assigned to

each biometrics. However, we found that variations in the quality of the biometric sample may degrade overall verification performance, because the average weighted fusion methodology is not adaptive to such variations. To improve this situation, we propose a fuzzy logic weight estimation that captures the variation in the sample quality in each biometrics.

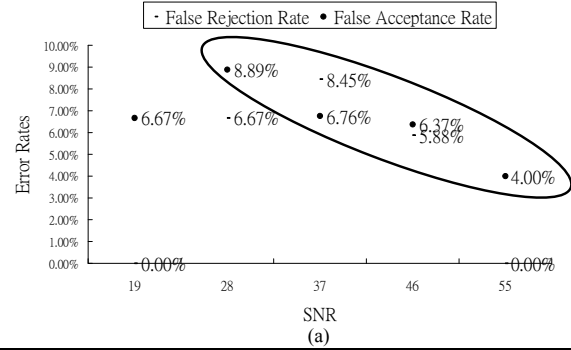
## 5. SNR AND SPEAKER VERIFICATION PERFORMANCE

We investigated the effects of signal-to-noise-ratio (SNR) on SV in order to develop an SNR-adaptive weight estimation methodology. A high level of background noise can harm the performance of speaker verification since the noise masks the vocal tract characteristics of speakers. The SNR measures the ratio between the power of signal and the power of noise in each speech utterance to reflect its quality (see Equation 2):

$$SNR = 10 \log_{10} (P_{signal} / P_{noise}) \quad (2)$$

where SNR is measured in decibel (dB),  $P_{signal}$  is the peak speech power and  $P_{noise}$  is the mean noise power.

The SNR values were obtained by using the NIST tool [10]. The speech data are grouped into five different intervals of SNR, their false rejection rates (FRR) and false acceptance rates (FAR) are shown in Fig. 2a. The amount of data in each SNR interval is shown in Fig. 2b. The EERs are obtained based on the development test set, with an overall EER of 6.88%.



**Fig. 2.** Performance of speaker verification and the number of utterances (trials) distributed over five different SNR intervals

SNR intervals with reasonable amount of data are encircled in Fig. 2a. We ignore the points that do not have a sufficient number of trials<sup>1</sup>. We observe that data with lower SNR tend to have higher FAR. Based on this observation, we define the “high”, “medium” and “low” SNR intervals at 50.5 to 59.5, 32.5 to 50.4 and 23.5 to 32.4 dB respectively in our fuzzy logic fusion methodology, as will be explained later.

## 6. FUZZY LOGIC (FL) WEIGHT ESTIMATION

The weight assigned to each biometrics should reflect the reliability of its use in authentication as well as the quality of the acquired data. We incorporate such factors in the fuzzy logic fusion methodology. When the sample quality is poor, the reliability of the recognition result is low. Measuring the sample quality helps us to estimate biometric weights according to their reliability.

Fig. 3 shows the flow of fuzzy logic weight estimation for each biometrics. Numerical measurements of external factors that reflect the quality of biometric data samples serve as input into the fuzzy system. Each value is converted (fuzzified) into a degree of membership for each fuzzy linguistic variables (e.g.

<sup>1</sup> More than 30 trials are treated as sufficient.

high, medium or low). The set of linguistic variables with the degree of membership is called a fuzzy set and the conversion process is called fuzzification. After fuzzification, the system will infer results from the logical rules (fuzzy rules) using the linguistic variables. Each rule maps to an output fuzzy set. To combine the output fuzzy sets, aggregation will be performed to come up with a single output fuzzy set. Finally, we need to perform defuzzification such that we can get the weight (degree of support) of the biometrics for decision fusion. Details about fuzzy logic system can be found in [12].

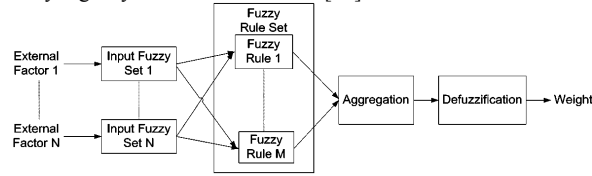


Fig. 3. Procedure of fuzzy logic weight estimation

### 6.1. Illustrative Example of FL Weight Estimation

We illustrate the fuzzy logic (FL) weight estimation produce through computing the weight of an example speech utterance. Details of the face and fingerprint weight estimation components are provided in [11]. The client's speech in this example has a score of -0.17 with a medium SNR of 33dB. Given that the decision threshold is -0.05, we run the risk of having a false rejection if we consider the voice biometrics only. In this case, the fusion methodology should aim to down-weight the contribution of speech in verification and up-weight the other two biometrics.

The input external factor related to sample quality, namely SNR, needs to be fuzzified according to a fuzzy set. A fuzzy set is represented by its fuzzy membership functions. We used SNR as the input to determine the SV weight. The SNR fuzzy set is modeled by three functions:  $SNR_{Low}$ ,  $SNR_{Medium}$  and  $SNR_{High}$ . These functions are mathematically defined using a combination of Gaussian distributions (Equation 3) shown in Fig. 5. Parameters selection for Eq. 3 is described in the previous section. The boundaries parameters  $c_1$  and  $c_2$  are shown in Table 4. Analysis of the influence of SNR on SV performance shows that, we roughly separated the SNR values into three ranges. We varied the parameters within the three ranges with reference to the mean ( $m$ ) and the standard deviation ( $\sigma$ ) of SNR obtained from the development set. The parameter values giving the best performance in the development set are used.

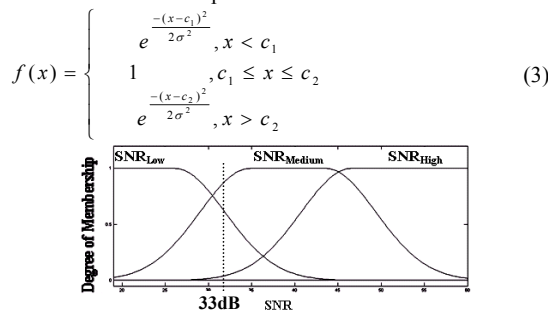


Fig. 5. Membership functions for speech factors for fuzzy logic decision fusion for  $SNR_{Low}$ ,  $SNR_{Medium}$  and  $SNR_{High}$ .

Table 4. The parameters used by SNR membership functions

	$SNR_{Low}$	$SNR_{Medium}$	$SNR_{High}$
$c_1$	$-\infty$	$m - \sigma$ (35dB)	$m + \sigma$ (47dB)
$c_2$	$m - 2.5\sigma$ (26dB)	$m + 0.5\sigma$ (44dB)	$+\infty$

The High, Medium and Low SNR fuzzy membership functions represent a YES (Y) or NO (N) concept at a linguistic level. The degree of membership for  $SNR_{Low}$ ,  $SNR_{Medium}$  and  $SNR_{High}$  in our example are 0.55, 0.92 and 0.06, respectively.

There are five fuzzy rules for speech as shown in Table 5. We have excluded some illogical combinations, e.g., we cannot have both  $SNR_{High}$  and  $SNR_{Low}$  being YES. From the development set used in parameter tuning, speech performs better than face and fingerprint. Therefore we rely more on speech when we have uncertainty on whether speech should be up-weight or not. For example when both  $SNR_{High}$  and  $SNR_{Medium}$  are YES, the weight output is mapped to the higher weight (i.e. High).

Table 5. Fuzzy rules used for the speech

ID	Fuzzy Rules
	IF Cond. 1 (C1) and Cond. 2 (C2) and Cond. 3 (C3) THEN (Consequence)
R1	IF ( $SNR_{High}$ is Y) and ( $SNR_{Medium}$ is N) and ( $SNR_{Low}$ is N) THEN ( $w_{speech}$ is Hi)
R2	IF ( $SNR_{High}$ is Y) and ( $SNR_{Medium}$ is Y) and ( $SNR_{Low}$ is N) THEN ( $w_{speech}$ is Hi)
R3	IF ( $SNR_{High}$ is N) and ( $SNR_{Medium}$ is Y) and ( $SNR_{Low}$ is N) THEN ( $w_{speech}$ is Med)
R4	IF ( $SNR_{High}$ is N) and ( $SNR_{Medium}$ is Y) and ( $SNR_{Low}$ is Y) THEN ( $w_{speech}$ is Med)
R5	IF ( $SNR_{High}$ is N) and ( $SNR_{Medium}$ is N) and ( $SNR_{Low}$ is Y) THEN ( $w_{speech}$ is Lo)

As shown in Table 5, there are three conditions specified in the IF clause. These conditions determine the degree of membership of the consequence (i.e. statement in the THEN clause). The conditions are combined using the fuzzy operation AND which is a minimum function since it can give the maximum agreement from different conditions. If the condition specifies a YES concept, the input will be set to the degree of membership. For a NO concept, the input will be set at one minus the degree of membership. Table 6 shows the combined conditions of the five rules. The combined degree of membership is mapped (implicated) to the designated output fuzzy set. We used the Gödel implication (minimum function), which is an R-implication method [13].

Table 6. Combined conditions for each rule (input values from the example for each condition shown in the parenthesis)

ID	C1: $SNR_{High}$ is	C2: $SNR_{Medium}$ is	C3: $SNR_{Low}$ is	Combine C1-C3: AND(C1,C2,C3)	Map To
R1	Y (0.06)	N (0.08)	N (0.45)	0.06	$w_{Hi}$
R2	Y (0.06)	Y (0.92)	N (0.45)	0.06	$w_{Hi}$
R3	N (0.94)	Y (0.92)	N (0.45)	<b>0.45</b>	$w_{Med}$
R4	N (0.94)	Y (0.92)	Y (0.55)	<b>0.55</b>	$w_{Med}$
R5	N (0.94)	N (0.08)	Y (0.55)	0.08	$w_{Lo}$

The results from the fuzzy rules are aggregated using the maximum function to combine them into a single output fuzzy set. The output fuzzy set is defined by three fuzzy membership functions (Fig. 6). They are triangle membership functions which define  $Low(z)$ ,  $Medium(z)$  and  $High(z)$  weight, corresponding to  $w_{Lo}$ ,  $w_{Med}$  and  $w_{Hi}$  respectively in Table 6. The output of these functions ranges from 0 to 1. The peak value of each function corresponds to the semantic meaning they represented, i.e.  $Low(z)$  peak at 0,  $Medium(z)$  peak at 0.5 and  $High(z)$  peak at 1.

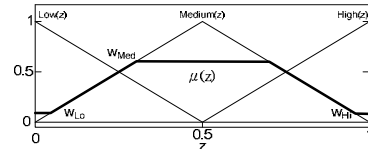


Fig. 6. Fuzzy output sets of fuzzy logic weight estimation.

The last step in applying FL is to defuzzify the fuzzy set to produce a quantifiable result, which is the SV weight in our case. The three values in our aggregated output fuzzy set,  $w_{Lo}$ ,  $w_{Med}$  and  $w_{Hi}$ , are 0.08, 0.55 (maximum of 0.45 and 0.55 in Table 6) and 0.06 respectively. The defuzzification method we use is centroid-of-area (Equation 4), which is the gravity of the area under the bolded line,  $\mu(z)$ , in Fig. 6 [12]. The method provides consistency, section invariance, monotonicity and most importantly simplicity [14]. The computation of the weight for face and fingerprint are similar but with different inputs and fuzzy rules for each biometrics [3].

We use the centroid-of-area method to compute the weight for SV. For our example using input speech data, the weight

computed is 0.51/1.00. This means that the system supports SV with 51% confidence. This is consistent with our objective of reducing the weight of using SV when the speech data do not have good quality.

$$Weight_{SV} = \int_{z=0}^1 z \cdot \mu(z) / \int_{z=0}^1 \mu(z)$$

$$\mu(z) = \text{Max}(\text{Min}(w_{Hi}, \text{High}(z)), \text{Min}(w_{Med}, \text{Medium}(z)), \text{Min}(w_{Lo}, \text{Low}(z))) \quad (4)$$

where  $\mu$  is the output fuzzy set and  $z$  is the value in the output domain  $[0, 1]$ .

## 7. EXPERIMENTAL RESULTS




We use the fuzzy logic weight estimation instead of the static weight in Equation (1) to compute the fusion results. Verification results that incorporate the SNR adaptive weight estimation component are shown in Table 7. We used the 16 EER obtained from fuzzy logic decision fusion and those obtained from weighted average fusion (Table 3) to conduct a paired t-test. The results are shown to be statistically significant with a 95% confidence level.

**Table 7.** Verification Performance of fusion results of fuzzy logic decision fusion on test set in EER (%)




Testing Type	Template Type (Overall EER = 0.99%)			
	WI	WO	PI	PO
WI	1.04	0.78	1.04	0.85
WO	1.04	0.34	1.18	1.16
PI	1.16	1.45	0.59	1.56
PO	1.04	0.78	1.07	0.57

We achieve an overall EER improvement of **42.1%** relative to the weighted average fusion. The relative improvement is 59.6% on matched device and environment between testing and enrollment, and 38.7% on mismatched. The major advantage of fuzzy logic (FL) weight estimation over weighted average (WA) is that it can adapt to varying biometric sample quality. Two scenarios are given to illustrate the adaptability of FL.

**Table 8.** A client trial where both WA and FL make correct decision (Scenario 1)

Biometrics	Speech	Facial image	Fingerprint
			
<b>Input Factors</b>	SNR: 42 dB	Face-finding Confidence: 10 Illuminance: 144	Core position: (135, 109) Darkness: 0.0283 Low-clarity: 0.4587
<b>Score</b>	0.60	9.21	43
<b>Threshold</b>	-0.06	8.32	28
<b>WA weight</b>	0.38	0.38	0.24
<b>FL weight</b>	Fused Score: 7.19 (Threshold: 6.05)	Accepted	Accepted

**Table 9.** A client trial that WA give false rejection while FL gives a true acceptance (Scenario 2)

Biometrics	Speech	Facial image	Fingerprint
			
<b>Input Factors</b>	SNR: 40 dB	Face-finding Confidence: <b>2.5</b> Illuminance: 124	Core position: (138, 134) Darkness: 0.0211 Low-clarity: 0.4756
<b>Score</b>	0.77	<b>6.22</b>	36
<b>Threshold</b>	-0.06	<b>8.32</b>	28
<b>WA weight</b>	0.38	<b>0.38</b>	0.24
<b>FL weight</b>	Fused Score: 5.4 (Threshold: 6.05)	<b>Not accepted</b>	Accepted

All the biometric samples illustrated in scenario 1 (Table 8) have decent qualities. However, when the quality of the biometric samples degrades, weighted average fusion fails to

make the correct decision (as illustrated in Table 9). In this example, the client does not look at the camera and it creates difficulties for the face-finding algorithm (which operates by locating the eyes). The baseline weighted average fusion makes the wrong decision and rejects the client. In contrast, fuzzy logic fusion dynamically reduces the weight of facial verification such that it can mitigate the error for this modality to an overall correct decision. We are making further comparisons between FL fusion and other approaches (such as Support Vector Machine).

## 8. CONCLUSIONS

We have implemented a multi-biometric authentication system (with face, fingerprint and speech) that is fully adaptive to variability in data acquisition using fuzzy logic. Adaptive weight estimation is used for the three different biometrics in order to account for external factors that affect verification performance. FL decision fusion shows a relative improvement of 42.1% on overall performance when compared with weighted average fusion. The improvement is due to the adaptability of FL fusion to variations in sample qualities. Relative improvements of 59.6% and 38.7% are respectively observed on matched device (i.e. using same device and environment for enrollment and test) and mismatched device verification. Mismatched conditions give rise to greater variability and hence smaller improvements when compared with matched conditions. Estimation on the influence of mismatched conditions should be conducted to further improve the system.

## 9. ACKNOWLEDGMENTS

This work is partially supported by the Central Allocation Grant of the Research Grants Council of the HKSAR (Project No. CUHK 1/02C) and affiliated with the Microsoft-CUHK Joint Laboratory for Human-centric Computing and Interface Technologies. The authors also wished to thank Dr. Wai-Kit Lo for his suggestions.

## 10. REFERENCES

- [1] A.K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Trans. on Circuits and Systems for Video Technology*, 14:4--20, 2004.
- [2] J. Kittler, M. Hatef, R. P.W. Duin, and J. Matas, "On combining classifiers," *IEEE Tran. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, March 1998.
- [3] C.W. Lau, B. Ma, H.M. Meng, Y.S. Moon, and Y. Yam, "Fuzzy logic decision fusion in a multi-modal biometric System," *Proc. of the 8th ICSLP*, 2004.
- [4] H.M. Meng et. al., "The Multi-biometric, Multi-device, and Multi-lingual (M3) Corpus," *Proc. of the 2nd MMUA*, 2006.
- [5] *User Manual FaceIt SDK*, Visionics Corp., 2000.
- [6] Peney and Atick, "Local feature analysis: A general statistical theory for object representation," *Network: Computation in Neural Systems* 7(3), pp.477-500, 1996.
- [7] K.C. Chan, Y.S. Moon, and P.S. Cheng, "Fast fingerprint verification using sub-regions of fingerprint Images," *IEEE Tran. on Circuits and Systems for Video Technology*, 2003.
- [8] B. Ma and H.M. Meng, "English-Chinese bilingual text-independent speaker verification," *Proc. of the ICASSP*, 2004.
- [9] D.A. Reynolds, "Comparison of background normalization methods for text-independent speaker verification," *Proc. EUROSPEECH*, pp. 963-966m, 1997.
- [10] Available: <http://www.itl.nist.gov/iaui/894.01/tools/>
- [11] Henry P.S. Hui, "Decision Fusion for Multi-modal Person Authentication," Thesis (M.Phil.), the CUHK, June 2006.
- [12] M. Hellmann, "Fuzzy logic introduction," *Epsilon Nought Radar Remote Sensing Tutorials*, 2001.
- [13] J.K. George and Y. Bo, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, NJ, 1995.
- [14] T.A. Runkler, "Selection of Appropriate Defuzzification Methods Using Application Specific Properties," *IEEE Trans. On Fuzzy Systems*, vol. 5, No. 1, Feb. 1997.