Fuzzy Logic Decision Fusion in a Multimodal Biometric System

Chun Wai Lau, Bin Ma, Helen M. Meng, Y.S. Moon and Yeung Yam***

Human-Computer Communications Laboratory
Department of Systems Engineering and Engineering Management
* Department of Computer Science and Engineering
** Department of Automation and Computer-Aided Engineering
The Chinese University of Hong Kong, Hong Kong SAR, China
{cwlau, bma, hmmeng}@se.cuhk.edu.hk, ysmoon@cse.cuhk.edu.hk, yyam@acae.cuhk.edu.hk

Abstract

This paper presents a multi-biometric verification system that combines speaker verification, fingerprint verification with face identification. Their respective equal error rates (EER) are 4.3%, 5.1% and the range of (5.1% to 11.5%) for matched conditions in facial image capture. Fusion of the three by majority voting gave a relative improvement of 48% over speaker verification (i.e. the best-performing biometric). Fusion by weighted average scores produced a further relative improvement of 52%. We propose the use of fuzzy logic decision fusion, in order to account for external conditions that affect verification performance. Examples include recording conditions of utterances for speaker verification, lighting and facial expressions in face identification and finger placement and pressure for fingerprint verification. The fuzzy logic framework incorporates some external factors relating to face and fingerprint verification and achieved an additional improvement of 19%.

1. Introduction

Multimodality forms the core of human-centric interfaces and extends the accessibility of computing to a diversity of users and usage contexts. As computing permeates our everyday lives, security that safeguards proper access to computers, communication networks and private information becomes an issue of prime importance. Classical user authentication relies on tokens and passwords that may be easily lost or forgotten. This problem can be overcome by the use of biometric authentication that verifies the user's identity based on his/her physiological or behavioral characteristics such as facial features, voice and fingerprints. User authentication should be transparent to human-computer interaction to maximize usability. In this regard, multimodal human inputs to the computer offer multiple biometric information sources for user authentication. Hence multimodality and multi-biometrics go naturally in tandem.

Performance in biometric verification is often affected by external conditions and variabilities. These are often related to mismatched conditions between enrollment and verification sessions, e.g. handsets/microphones for recording speech, cameras for capturing facial images and fingerprint readers. In addition, the user's speech may vary according to ambient noise conditions, the speaker's health (e.g. contracting a cold) or speaking styles. The user's facial images may vary due to changes in backgrounds, illumination, head positions and expressions. While none of the biometrics alone can guarantee absolute reliability, they can reinforce one another when used jointly to maximize verification performance. This motivates multi-biometric authentication [1,2], where decisions based on individual biometrics are *fused*. Fusion techniques in previous work include majority voting, sum or product rules, different classifier types like SVM, Bayesian classifier, decision trees and k-NN [3-5]. In this work, we developed a speaker verification system, a face identification system and a fingerprint verification system. We also propose a fusion technique based on fuzzy logic in order to incorporate effects of external conditions that affect the confidence in a biometric verification decision. This fuzzy logic fusion technique is compared with other simple techniques such as fusion by majority voting or weighted average scores.

2. Speaker Verification

For the speech modality, we authenticate with a bilingual text-independent speaker verification system [6]. Utterances were collected from 16 subjects in the form of spoken responses to computer prompts for personalized information, e.g. "What is your favorite color?" or "你最喜欢什么颜色?" Each subject provided three (short, medium and long) versions of each spoken response in order to train the data to achieve better text independence, e.g. "Purple," "我喜欢紫色," "My favorite color is purple.". Each subject participated in three enrollment sessions spaced out with one-week intervals as well as a verification session that took place several days after the last enrollment session. In total, each subject recorded 252 utterances for enrollment (42 each in English and Chinese) as well as 30 utterances for verification (15 for each language).

During the enrollment process, we developed a 512 Gaussian-mixture model for each subject and trained it with bilingual data. During verification, each subject is treated in turn as the claimant and the other subjects as imposters. Hence we have in total 480 testing utterances from the true speakers and 7200 from the imposters. We applied cohort normalization in calculating the likelihood ratio scores (see Equation 1).

$$P_{norm}(X \mid \lambda_i) = \frac{P(X \mid \lambda_i)}{\frac{1}{K}\sum_{k=1}^{K} P(X \mid \lambda_k)} \quad (1)$$

where λ_i is the i th claimant's model, λ_k s are the cohort speaker models. K (=4) is the number of selected cohort members. The likelihood ratio scores P_{norm} are compared with a global threshold θ. ($P_{norm} < \theta$) causes the system to reject the subject as an imposter. Otherwise, the system accepts the subject as the claimant. An equal error rate EER_{speech} of 4.35% is obtained for our speaker verification data.

3. Face Identification

An off-the-shelf software, the FaceIt Verification SDK from Identix, is used for face identification in our experiments. FaceIt uses Local Feature Analysis (LFA) [7] to encode facial images. It can automatically detect the face in an image and enroll the face image into a template as well as verify a face image against a template. We recorded videos of the faces of the 16 subjects (same group as in speaker verification). Each subject was recorded with two cameras, capturing facial movements from up to down, left to right and in rotation in separate video clips respectively. In this way we try to capture almost all face orientations. The two cameras include a high-quality webcam for desktop PCs (EagleTec model ET-VCCD) and a low-quality camera for pocket PC (Pretec model CompactCamera OCCAV). Videos were also shot indoors as well as outdoors to incorporate variability in lighting conditions. In total there are 24 videos (12 for enrollment and 12 for verification) per subject and each video is of 5 to 10 seconds in duration. Hence the enrollment (training) and verification (testing) frames are extracted from different video filming sessions.

During the enrollment process, we used FaceIt to automatically select 12 frames per video clip and organized them as four types of enrolled face templates per subject:

Type 1 template: high quality webcam, indoors (WI)
Type 2 template: high quality webcam, outdoors (WO)
Type 3 template: low quality PocketPC camera, indoors (PI)
Type 4 template: low quality PocketPC cam., outdoors (PO)

During verification, we randomly select 30 frames per subject, distributed across different face orientations and video recording conditions. Similar to the case in speaker verification, each subject is treated in turn as the claimant and the other subjects as imposters. Hence we have 480 facial images from the true claimant and 7200 from the imposters.

Testing Conditions	Training Conditions (Type of Enrolled Templates)			
	WI	WO	PI	PO
WI	**5.11**	16.67	17.36	22.15
WO	19.87	**6.99**	26.55	17.00
PI	16.64	28.06	**11.25**	34.58
PO	20.91	16.96	31.95	**11.46**

Table 1: Face identification performance measured in equal error rates (EER_{face}%) for different enrollment and verification conditions, i.e. the camera may be a webcam (W) or PocketPC camera (P); and the lighting conditions may be indoors (I) or outdoors (O).

FaceIt generates a verification score for each trial and compares it with a threshold. If the score falls below the threshold, the system rejects the subject as an imposter; otherwise, it accepts the subject as the claimant. Table 1 shows the verification results in terms of equal error rates (EER_{face}) for all testing conditions (i.e. camera type and lightning conditions) against all types of enrolled face templates. We see that the best performance is obtained when the testing conditions match with the enrollment conditions.

4. Fingerprint Verification

We adopt a direct gray-scale minutiae detection approach [8,9] to extract features for fingerprint verification. Fingerprints from the same 16 subjects (as in previous biometrics) were captured by an optical device, SecureTouch 2000 from Biometric Access Corporation. For each subject, we collect 20 fingerprint images for each of two fingers by asking the subject to remove and replace the finger on the capture device multiple times. Hence we have 40 fingerprint images per subject, of which 10 (i.e. five images per finger) are used as enrollment templates and the remaining 30 images (i.e. 15 per finger) are used during verification. We ran the verification tests in a way similar to the other biometrics, with a total of 480 fingerprint images from the true claimant and 7200 from the imposters. We obtained an equal error rate EER_{finger} of 5.07% based on our verification set.

5. Fusion by Majority Votes

In preparation for our fusion experiments, we randomly grouped one speech utterance, one fingerprint image and one face image for every subject. This generates 480 data groups from the true claimant and 7200 from the imposters. We fused the verification results of individual biometrics by means of majority votes and tabulated the overall verification performance values in Table 2, which shows a marked improvement of 48% relative to speaker verification only. (Note that speech is the best-performing biometric among the three in terms of individual equal error rates).

Testing Conditions	Training Conditions (Type of Enrolled Templates)			
	WI	WO	PI	PO
WI	**1.15**	1.82	1.62	2.40
WO	2.52	**0.98**	3.17	2.27
PI	2.16	2.89	**1.39**	3.36
PO	2.32	2.43	3.34	**1.92**

Table 2: Verification results from fusion by majority voting.

6. Fusion by Weighted Average Scores

This is another simple method of fusion. We scaled the verification scores obtained from the spoken utterances, facial images and fingerprint images to the same range of values by min-max normalization. A fixed weight w_i is assigned to each biometric i. These weights are normalized according to equation (2) to generate W_i which is used in the linear combination of the verification scores S_i to give the fusion score S (see Equation 3).

$$W_i = \frac{w_i}{\sum_{i=1}^{3} w_i} \qquad (2) \qquad\qquad S = \sum_{i=1}^{3} W_i \cdot S_i \qquad (3)$$

6.1 Weight Assignment by Cross-Validation

Testing Conditions	Training Conditions (Type of Enrolled Templates)			
	WI	WO	PI	PO
WI	**0.63**	0.89	1.06	1.42
WO	1.04	**0.54**	1.50	1.25
PI	0.86	1.00	**0.50**	1.46
PO	0.94	1.11	1.56	**0.84**

Table 3: Verification performance with fusion by weighted average scores.

Since we do not have a development test set, the weights w_i are assigned by three-fold cross-validation. The verification set is divided into three equal portions. Each portion is used in turn for testing while the other two are used for optimizing the weights. The weights are varied within the [0,1] range in steps of 0.1 to find values that gave the best performance. The equal error rates of the three testing blocks are then averaged and results are shown in Table 3. There is an improvement of 52% relative to fusion by majority voting.

7. Fusion by Fuzzy Logic Decision

The verification performance based on a biometric is affected by external conditions. For example, face identification performance may degrade when the lighting is too bright or too dark, or when the input facial image for verification is posed at an angle or carries an expression that differs from the enrollment images (see Figures 1a to 1d). Similarly, fingerprint verification performance may degrade if the input fingerprint image is off-centered, faded due to dry fingers or pressing too lightly, or smudged due to sweat or pressing too hard (see also Figures 1e to 1h). Speaker verification performance may also degrade if the input utterances are drowned out by ambient noise, if the speaker's voice characteristics have changed since enrollment (e.g. due to a sore throat or cold) or if the speaking styles between the enrollment and verification utterances are different. It may be difficult to precisely quantify these external conditions and their effects on verification performance. Hence we attempt to incorporate these conditions by the use of a fuzzy logic framework [10,11] for multi-biometric fusion. Fuzzy logic enables us to process imprecise information in a way that resembles human thinking, e.g. big versus small, high versus low, etc., and allows intermediate values to be defined between true and false by partial set memberships. As an initial step, we consider fuzzy variables and fuzzy sets in a fuzzy inference system for face and fingerprint images. Application to speech will be pursued as a next step.

7.1 Fuzzy Inference System

The fuzzy inference system adjusts the weighting for each biometric as affected by the external conditions described above. There are 2 *output* fuzzy variables, w_{face} and w_{finger}, which correspond to the weightings for face and fingerprint verification respectively. Their values range from 0 to 1, with higher values implying higher confidence. The fuzzy sets of both output variables are triangular membership functions (see Figure 2) that define three levels of output weighting (high/medium/low) for each biometric. Defuzzification uses a standard centroid-of-area technique.

There are 6 *input* fuzzy variables – two for the face biometric and four for the fingerprint. Each input variable has a fuzzy set that defines the *favored external condition* for each variable. As seen in Figure 3, the fuzzy sets are either linear or Gaussian combination membership functions *f(x)*. The latter combines two Gaussian functions to determine the shape of the left-most and right-most curves and involve such parameters as the means (m) and variances (σ) of the data, as well as the boundary points c_1 and c_2 which may be set at set using m-0.5σ and m+0.5σ respectively (see Equation 4). The *unfavored external condition* for each input fuzzy variable is can be represented by the fuzzy set *1-f(x).* We list the six input fuzzy variables as follows (see Figures 3a to 3f):

(i) FaceFindingConf is the face finding confidence obtained from FaceIt and has five discrete levels at (0, 2.5, 5, 7.5, 10). Higher input levels represent higher confidence in face detection. A triangular membership function is applied seek high confidence in face finding.

(ii) Illuminance measures the average intensity of the face image. High/low input values are caused by bright/dark environments. The Gaussian combination membership function in Figure 3b defines medium brightness as a favored condition for face images captured indoors by a webcamera.

(iii) CorePosX is the x-coordinate of the fingerprint image core obtained from the fingerprint verification software. The membership function in Figure 3c defines a centrally placed fingerprint image which is favored. High/low values for CorePosX implies an off-centered image.

(iv) CorePosY is the y-coordinate of the fingerprint image. Other properties are similar to (iii).

(v) Darkness measures the proportion of dark pixels with intensities ≤30. Larger values imply darker images due to smudging. Small values are favored as normal images.

(vi) Low-clarity measures the proportion of light pixels with intensities between 110 and 160. Larger values imply faded images and therefore low values are favored by the membership function for clarity. Non-uniform pressure in the fingerprint image may result in high values for *Darkness* and *Low-clarity*.

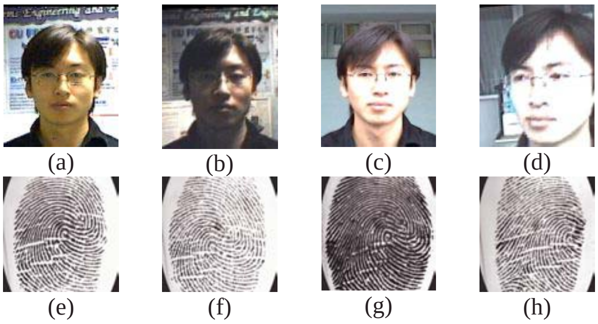


Figure 1: Face identification may be adversely affected by different lighting conditions between enrollment and verification, e.g. (a) medium brightness indoors; (b) dark environment indoors; (c) medium brightness outdoors; (d) bright environment with angled pose, outdoors. Fingerprint identification may also be adversely affected by mismatches in conditions under which the fingerprint image is captured, e.g. (e) a normal image; (f) faded image due to dryness or low pressure; (g) smudged image due to sweat or high pressure; (h) off-centered image due to improperly placed finger.

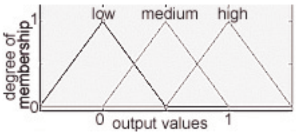
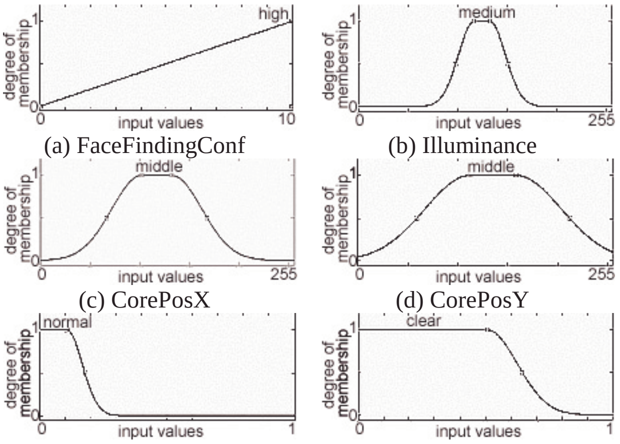


Figure 2: Fuzzy sets for the output fuzzy variables, W_{face} and $W_{fingerprint}$, corresponding to the weightings of the face and fingerprint biometrics.

$$f(x) = \begin{cases} e^{\frac{-(x-c_1)^2}{2\sigma^2}} & ,\text{if } x < c_1 \\ 1 & ,\text{if } c_1 \le x \le c_2 \\ e^{\frac{-(x-c_2)^2}{2\sigma^2}} & ,\text{if } x > c_2 \end{cases} \qquad (4)$$



(a) FaceFindingConf (b) Illuminance

(c) CorePosX (d) CorePosY

(e) Darkness (f) Low-clarity

Figure 3: Fuzzy sets defined for the input variables.

7.2 Fuzzy Rules

The conditions that comprise the fuzzy logic are formulated by two groups of fuzzy IF-THEN rules (20 in all). One group controls the output variable w_{face} (i.e. weighting for the face biometric) according to values of the input variables *FaceFindingConf* and *Illuminance*. The other group controls the output variable w_{finger} (i.e. weighting of fingerprint verification) according to the values of the input variables *CorePosX*, *CorePosY*, *Darkness* and *Low-clarity*. Main properties in the fuzzy rules are:

- if all external conditions (input variables) are favorable, the output variable is set to high;
- if one of the conditions are unfavorable, the output variable is set to medium;
- multiple unfavorable conditions will map the output to low.

An example fuzzy rule for face identification is:

- *IF (FaceFindingConf is high) and (Illuminance is medium) THEN (w_{face} is high)*

7.3 Experiments with Fuzzy Logic Fusion

The experimental setup is the same as previous fusion experiments (see Sections 5 and 6). Again, we used three-fold cross-validation based on the verification data to optimize parameter values of the Gaussian combination membership functions in the fuzzy sets. This procedure generates values for w_{face} and w_{finger} to capture effects due to external conditions. We current do not have corresponding data for the speech biometric, hence weighting for speaker verification is set according to the relative performance among the three biometrics (see Equation 5):

$$w_{speech} = 1 - \frac{EER_{speech}}{EER_{speech} + EER_{face} + EER_{fingerprint}} \qquad (5)$$

Again, the weights w_i are assigned by three-fold cross-validation. The verification set is divided into three equal portions. Each portion is used in turn for testing while the other two are used for optimizing the weights. The weights w_{speech} ,w_{face} and w_{finger} are then normalized (see Equation 2) and combined as in Equation (3) to produce the overall verification result averaging the equal error rates across the three testing blocks . Table 4 shows further improvement of 19% relative to fusion by weighted average scores. This is statistically significant according to a paired t-test (*p=0.05*).

Testing Conditions	Training Conditions (Type of Enrolled Templates)			
	WI	WO	PI	PO
WI	**0.56**	0.86	0.72	1.23
WO	0.81	**0.31**	1.08	0.83
PI	0.75	0.82	**0.42**	1.15
PO	0.87	0.85	1.25	**0.81**

Table 4: Verification performance with fuzzy logic fusion.

8 Conclusions

This paper presents a multi-biometric verification system that combines speaker verification, fingerprint verification with face identification. Their respective equal error rates (EER) were 4.3%, 5.1% and the range of (5.1% to 11.5%) for matched conditions in facial image capture. Fusion of the three by majority voting gave a relative improvement of 48%, which corresponds to an EER range of (0.98% and 1.92%). Another fusion method by weighted average scores produced additional relative improvement of 52%, which corresponds to EER range of (0.50% and 0.84%). We proposed the use of fuzzy logic decision fusion, in order to account for external conditions that affect verification, such as finger placement, pressure and sweat in fingerprint verification; and lightning conditions and head positioning in face identification. Fuzzy logic fusion generated a further improvement of 19% relative to fusion by weighted average scores, which corresponds to an EER range of (0.31% to 0.81%).

9 Acknowledgments

This work is partially supported by the Central Allocation Grant of the Research Grants Council of the Hong Kong Special Administrative Region (Project No.CUHK 1/02C).

10 References

[1] Jain, A., Hong, L. and Kulkarni, Y., "A Multimodal Biometric System Using Fingerprint, Face, and Speech", *Proceedings of AVBPA*, pp.182-197, 1999.

[2] Brunelli, R. and Falavgna, D., "Person Identification Using Multiple Cues", *IEEE PAMI*, 17(10):995-966, 1995.

[3] Teoh, A., Samad, S.A. and Hussian, A., "Theoretic Evidence k-Nearest Neighbourhood Classifiers in a Bimodal Biometric Verification System", *Proceedings of AVBPA*, pp. 778-786, 2003.

[4] Verlinde, P., and Acheroy, M., "A Contribution to Multi-Modal Identify Verification Using Decision Fusion", *Proceedings of PROMOPTICA*, 2000.

[5] Kittler, J., Hatef, M., Duin, R.P.W. and Matas, J., "On Combining Classifiers", *IEEE PAMI*, 20(3):226-239,1998.

[6] Ma, B. and Meng H., "English-Chinese Bilingual Text-Independent Speaker Verification", *Proc. ICASSP*, 2004.

[7] Peney and Atick, "Local feature analysis: A general statistical theory for object representation", *Network: Computation in Neural Systems*, 7(3):477-500, 1996.

[8] Chan, K.C., Moon, Y.S. and Cheng, P.S., "Fast Fingerprint Verification using Sub-regions of Fingerprint Images", *IEEE Transactions on Circuits and Systems for Video Technology*, Nov. 2003.

[9] Mario, D. and Maltoni, D., "Direct Gray-Scale Minutiae Detection in Fingerprints", *IEEE PAMI*, 19(1):27-40,1997.

[10] Zadeh, L.A., *Fuzzy Sets, Information and Control*, 1965.

[11] Zadeh, L.A., "Making computers think like people", *IEEE Spectrum*, pp. 26-32, 8/1984.