Malaysian Journal of Computer Science, Vol. 15 No. 2, December 2002, pp. 17-27

DECISION FUSION COMPARISON FOR A BIOMETRIC VERIFICATION SYSTEM USING FACE AND SPEECH

Andrew Beng Jin Teoh
Faculty of Information, Science and Technology
Multimedia University
Malacca, Malaysia
email: bjteoh@mmu.edu.my

Salina Abdul Samad
Electrical, Electronics and System Engineering Department,
Engineering Faculty
Universiti Kebangsaan Malaysia
Bangi, Malaysia
email: salina@eng.ukm.my

Aini Hussain
Electrical, Electronics and System Engineering Department,
Engineering Faculty
Universiti Kebangsaan Malaysia
Bangi, Malaysia
email: aini@eng.ukm.my

ABSTRACT

This paper presents several fusion decision techniques comparison for a bimodal biometric verification system that makes use of face images and speech utterances. The system is essentially constructed by a face expert, a speech expert and a fusion decision module. Each individual expert has been optimised to operate in automatic mode and designed for security access application. Fusion decision schemes considered are the voting technique, ordinary and modified k-Nearest Neighborhood classifier and linear Support Vector Machine. The aim is to obtain the optimum fusion module from amongst these five techniques best suited to the target application.

Keywords: ***Biometrics, Face verification, Speech verification, Fusion decision***

1.0 INTRODUCTION

In today's wired information society, there are an increasing number of situations, for example when accessing a computer account, an automatic teller machine or even a website, which require an individual as a user to be verified electronically. Traditionally, a user can be verified using their ID card and/or password, but these approaches have several drawbacks. The cards can be stolen or misplaced while the passwords can be forgotten.

The alternative is biometrics which is the automatic identification of a person based on his or her physiological or behavioral characteristics. Biometrics examples include using the facial image, facial thermogram, fingerprint, hand geometry, hand vein, iris, retinal pattern, signature and voice print [1]. However, a major problem with biometrics is that the physical appearance of a person tends to vary with time. In addition, correct authentication may not be guaranteed due to sensor noise and limitations of feature extractor and matcher.

One solution to cope with these limitations is to combine several biometrics in a multi-modal identity verification system [2]. Some works on multi-modal biometric identity verification systems have been reported in literature. Brunelli and Falavigna have proposed in [3] a person identification system based on acoustic and visual features, where they use a HyperBF network as the best performing fusion module. Dieckmann et al. have proposed in [4] a decision level fusion scheme, based on a *2-out-of-3* majority voting, which integrates face and voice, analysed by three different experts: face, lip motion, and voice. Duc et al. proposed in [5] a simple averaging technique and compared it with the Bayesian integration scheme presented by Bigun et al. in [6]. In this multi-modal system the authors use a face identification expert, and a text-dependent speech expert. Kittler et al. proposed in [7] a multi-modal person verification system, using three experts: frontal face, face profile, and voice. The best combination results are obtained for a simple sum rule. Hong and Jain proposed in [8] a multi-modal personal identification system which integrates face and fingerprints that complement each other. The fusion algorithm operates at the expert *(soft)* decision level, where it combines the scores from the different experts under statistically independence hypothesis, by simply multiplying them. Ben-Yacoub proposed in [9] a multi-modal data fusion approach for person authentication, based on Support Vector Machines (SVM) to combine the results obtained from a face identification expert, and a text-dependent speech expert. Pigeon proposed in [10] a multi-modal person

authentication approach based on simple fusion algorithms to combine the results coming from the frontal face, face profile, and voice modal. Choudhury et al. proposed in [11] a multi-modal person recognition using unconstrained audio and video. The combination of the two experts is performed using a Bayes net.

A bimodal biometric verification system based on facial and vocal modalities is described in this paper. It differs from the systems that are mentioned above in the sense that this system is targeted for applications involving automatic verification using personal computers and their multimedia capturing devices. Thus each module of the system has been fine-tuned to deal with the problems that may occur in this type of application, such as poor quality images obtained from using a low cost PC camera and the problem of using various types of microphones that may cause channel distortion or convolution noise. In addition, the system is designed to keep the rate as low as possible for the case when an imposter is accepted as being a genuine client. Each module of the system, i.e. the face and voice, is developed separately and several fusion decision schemes are compared with the aim to obtain the optimum technique for this application.

2.0 VERIFICATION MODULES

2.1 Face Verification

In personal verification, face recognition refers to static, controlled full frontal portrait recognition. There are two major tasks in face recognition: (i) face detection and (ii) face verification.

In our system as shown in Fig. 1, the Eigenface approach [12] is used in the face detection and face recognition modules. The main idea of the Eigenface approach is to find the vectors that best account for the distribution of face images within the entire image space and define as the face-space. Face-spaces are eigenvectors of the covariance matrix corresponding to the original face images, and since they are face-like in appearance they are so called eigenfaces as shown in Fig. 2.

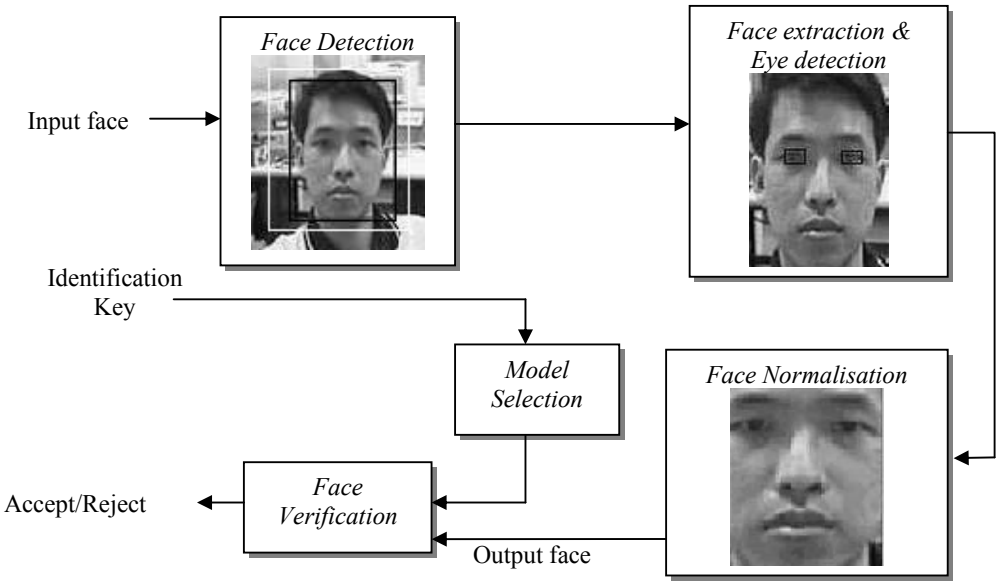

Fig. 1: Face verification system

Now let the training set of face images be $i_1, i_2, \ldots, i_M$, the average face of the set is defined as

$$\bar{i} = \frac{1}{M}\sum_{j=1}^{M} i_j \tag{1}$$

where M is the total number of images.

Fig. 2: Eigenfaces

Each face differs from the average by the vector $\phi_n = i_n - \bar{i}$. A covariance matrix is constructed where:

$$C = \sum_{j=1}^{M} \phi_j \phi_j^T = AA^T \tag{2}$$

where $A = [\phi_1 \ \phi_2 \ldots \phi_M]$.

Then, eigenvectors, v_k and eigenvalues, λ_k with symmetric matrix C are calculated. v_k determine the linear combination of M difference images with ϕ to form the eigenfaces:

$$u_l = \sum_{k=1}^{M} v_{lk} \phi_k \qquad l=1, \ldots, m \tag{3}$$

From these eigenfaces, $K(<M)$ eigenfaces are selected to correspond to the K highest eigenvalues.

Face detection is accomplished by calculating the sum of the square error between a region of the scene and the Eigenface, a measure of Distance From Face Space (DFFS) that indicates a measure of how face-like a region. If a window, ϕ is swept across the scene, to find the DFFS at each location, the most probable location of the face can be estimated. This will simply be the point where the reconstruction error, ε has the minimum value.

$$\varepsilon = \left\| \phi - \phi_f \right\| \tag{4}$$

where ϕ_f is the projection into face-space.

From the extracted face, eye co-ordinate will be determined with the hybrid rule based approach and contour mapping technique [13]. Based on the information obtained, scale normalisation and lighting normalisation are applied for a *head in box* format.

The Eigenface-based face recognition method is divided into two stages: (i) the training stage, (ii) the operational stage. At the training stage, a set of normalised face images, $\{i\}$ that best describe the distribution of the training facial images in a lower dimensional subspace (eigenface) is computed by the operation:

$$\varpi_k = U_k (i_n - \bar{i}) \tag{5}$$

where $n = 1, \ldots, m$ and $k=1, \ldots, K$.

Next, the training facial images are projected onto the eigenspace, Ω_i, to generate the representations of the facials images in eigenface.

$$\Omega_i = [\varpi_{n1}, \varpi_{n2}, \ldots, \varpi_{nK}] \tag{6}$$

where i=1,2, . . . ,M.

At the operational stage, an incoming facial image is projected onto the same eigenspace and the similarity measure which is the Mahalanobis distance between the input facial image and the template is, thus, computed in the eigenspace.

Let $\boldsymbol{\varphi}_1^0$ denotes the representation of the input face image with claimed identity $\boldsymbol{C}$ and $\boldsymbol{\varphi}_1^C$ denotes the representation of the *Cth* template. The similarity function between $\boldsymbol{\varphi}_1^0$ and $\boldsymbol{\varphi}_1^C$ is defined as follows:

$$F_1(\boldsymbol{\varphi}_1^0, \boldsymbol{\varphi}_1^C) = \| \boldsymbol{\varphi}_1^0 - \boldsymbol{\varphi}_1^C \|_m \quad (7)$$

where $\|\bullet\|_m$ denotes the Mahalanobis distance.

2.2 Speaker Verification

Anatomical variations that naturally occur amongst different people and the differences in their learned speaking habits manifest themselves as differences in the acoustic properties of the speech signal. By analysing and identifying these differences, it is possible to discriminate among speakers [15]. Our front end of the speech module aims to extract the user dependent information.

The system includes three important stages: endpoint detection, feature extraction and pattern comparison. The endpoint detection stage aims to remove silent parts from the raw audio signal, as this part does not convey speaker dependent information.

Noise reduction techniques are used to reduce the noise from the speech signal. Simple spectral subtraction [16] is first used to remove additive noise prior to endpoint detection. Then, in order to cope with the channel distortion or convolution noise that is introduced by a microphone, the zero'th order cepstral coefficients are discarded and the remaining coefficients are appended with delta feature coefficients [17]. In addition, the cepstral components are weighted adaptively to emphasize the narrow-band components and suppress the broadband components]. The cleaned audio signal is converted to a 12th order linear prediction cepstral coefficients (LPCC), using the autocorrelation method that leads to a 24 dimensional vector for every utterance.

Fig. 3 shows the process used in front end module.

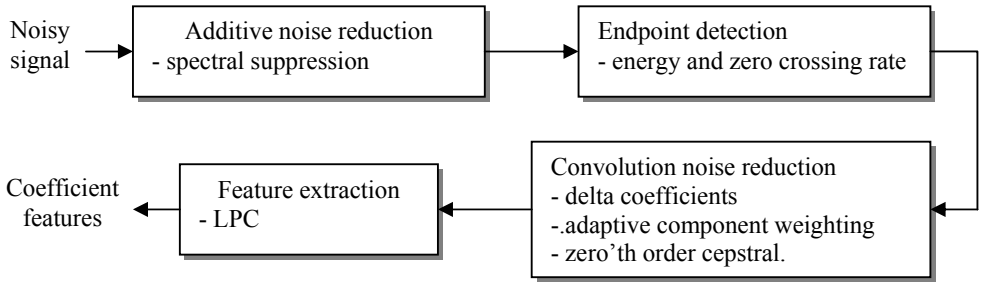

Fig. 3: The front-end of the speaker verification module

As with the face recognition module, the speaker verification module also consists of two stages: (i) the training stage and (ii) the operational stage. At training phase, two sample utterances with the same words from the same speaker are collected and trained using the modified *k*-Mean algorithm [19]. The main advantages of this algorithm are the statistical consistency of the generated templates and their ability to cope with a wide range of individual speech variations in a speaker-independent environment.

At the operational stage, we opted for a well-known pattern-matching algorithm – Dynamic Time Warping (DTW) [20] to compute the distance between the trained template and the input sample.

Let $\boldsymbol{\varphi}_2^0$ represent the input speech sample with the claimed identity $\boldsymbol{C}$ and $\boldsymbol{\varphi}_2^C$ *Cth* template. The similarity function between $\boldsymbol{\varphi}_2^0$ and $\boldsymbol{\varphi}_2^C$ is defined as follows:

$$F_1(\boldsymbol{\varphi}_2^0, \boldsymbol{\varphi}_2^C) = \|\boldsymbol{\varphi}_2^0 - \boldsymbol{\varphi}_2^C\| \quad (8)$$

where $\|\bullet\|$ denotes the distance score result from DTW.

3.0 FUSION DECISION MODULES

3.1 Voting Techniques

Voting techniques are classical empirical techniques where the global decision rule is obtained simply by fusing the hard decisions made by the 2 experts. A hard decision is a score that only returns either a 0 or a 1. This technique accepts the identity claimed by the person under test if at least *k*-out-of-2 experts decide that the person is genuine [21]. They do not take into account the soft decision nor do they analyse earlier behavior of the expert, for example by calculating some statistical moments.

When $k = 1$, this is the called the *OR* rule. The identity claimed is accepted if at least one of the 2 experts decides that the person under test is genuine. Intuitively, this strategy leads to a very indulgent fusion scheme, which means that the acceptance will be fairly easy. While $k = 2$, this is the called the *AND* rule. The identity claimed is accepted only if both the experts decide that the person under test is genuine. Intuitively, this technique leads to a very severe fusion scheme, which means the acceptance will be rather difficult.

3.2 *k* – NN (Nearest Neighborhood) Classifier

The *k*-NN classifier [22] is a simple classifier that needs no specific training phase. All that are required are reference data points for both classes representing the genuine and the imposter. An unknown test data point y is then attributed with the same class label as the label of the majority of its k nearest reference neighbors.

To find the k nearest neighbors, the Euclidean distance between the test point and all the reference points is calculated. The distances are then ranked in ascending order and the reference points corresponding to the k smallest Euclidean distances are taken. The exhaustive distance calculation step during the test phase leads rapidly to large computing time, which is the major drawback of this otherwise very simple algorithm. A special case of the *k*-NN classifier can be obtained when $k = 1$. In this case, the classifier is called the Nearest Neighbor (NN) classifier.

The number k should be large enough to minimise the probability of misclassifying y and small enough with respect to the number of samples so that the points are close enough to x to give an accurate estimate of the true class of y.

3.3 Linear Support Vector Machine Classifier

Support Vector Machines (SVM) is a type of machine learning technique that learns the decision surface to separate the two classes through a process of discrimination. It has good generalisation characteristics. SVMs have been proven to be successful classifiers on several classical pattern recognition problems [23].

In conventional pattern classification problem, empirical risk minimisation (ERM) is the most commonly used optimisation procedure in machine learning. In this regime, the goal is to arrive at a parameter setting that gives the smallest value called risk, R_{emp}. The risk computation can take other forms such as the sum-squared error. Neural network training, back-propagation in particular, is a direct consequence of a similar optimisation process. There are no probability computations involved in the definition of risk.

Another form of risk commonly used is the expected risk or estimated risk, R. Vapnik [24] proved that bounds exist for this expected risk such that,

$$R \leq R_{emp} + f(h) \tag{9}$$

where h is Vapnik Chervonenkis (VC) dimension. Finding a learning machine with the minimum upper bound on the estimated risk leads to a method of choosing an optimal machine for a given task. This is the essential idea of Structural Risk Minimisation (SRM). The SVM is based on the principle of SRM.

3.3.1 Linearly Separable Data in Linear SVM

Fig. 4 shows a typical 2 - class classification example where the classes are perfectly separable using a linear decision region. Let $\boldsymbol{w}$ be normal to the decision region and let the $\boldsymbol{N}$ training examples be represented as the pairs $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$, $i = 1,2,\ldots,N$ where $-1 \leq y_i \leq 1$. The points that lie on the hyper plane to separate the data satisfy,

$$\boldsymbol{w} \cdot \boldsymbol{x} + \gamma = 0 \tag{10}$$

where γ is the distance of the hyper plane from the origin. Let the margin of the SVM be defined as the distance between closest positive and negative example from the hyper plane. The SVM looks for the separating hyper plane, which gives the maximum margin. Once the hyper plane is obtained, all the training examples satisfy the following inequalities:

$$\boldsymbol{w} \,.\, \boldsymbol{x}_i + \gamma \geq +1 \qquad \text{for } y_i = +1 \tag{11}$$

$$\boldsymbol{w}_i \,.\, \boldsymbol{x}_i + \gamma \leq -1 \qquad \text{for } y_i = -1 \tag{12}$$

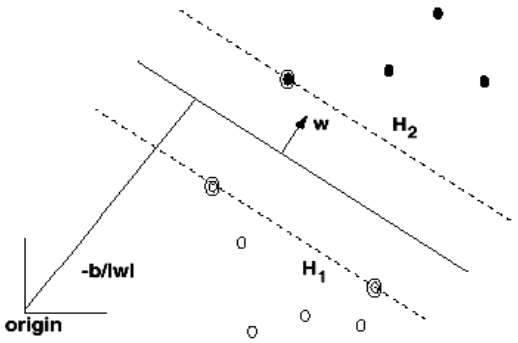

Fig. 4: Linear separating hyper planes for the separable case

Looking at the above equations with respect to Fig. 4, the distance between H_1 and H_2 is $2/\|\boldsymbol{w}\|$. Note that for a completely separable data set, no points fall between H_1 and H_2. Thus for maximising the margin, we need to minimise $\|\boldsymbol{w}\|^2$. In Fig. 4, they are indicated by concentric circles. This leads to a Quadratic Programming problem, which can make use the theory of Lagrange multipliers.

3.3.2 Nonlinearly Separable Data in Linear SVM

Most of the classification problems in real world involve non-separable data. The optimal-margin classifier can be extended to this non-separable case by using a set of slack variables. In this situation, the inequality constraints become,

$$\boldsymbol{w} \,.\, \boldsymbol{x}_i + \gamma \geq +1 - \xi_i \qquad \text{for } y_i = +1 \tag{13}$$

$$\boldsymbol{w} \,.\, \boldsymbol{x}_i + \gamma \leq -1 + \xi_i \qquad \text{for } y_i = -1 \tag{14}$$

$$\xi_i \geq 0 \qquad \forall_i \tag{15}$$

A close look the above inequalities (15) shows that for an error to occur, the corresponding ξ_i needs to be greater than 1. This implies that the upper bound on the number of errors on the training data is $\Sigma_i \xi_i$. In addition, the optimisation process in the new data setting needs to minimise this quantity. The new term that is added to the objective is

$$C(\Sigma_i \xi_i)^2 \tag{16}$$

where C is used to control the penalty for a training error.

4.0 EXPERIMENTS AND DISCUSSION

4.1 Distance Score Normalisation

The similarity measures values from equation (7) and (8) have different ranges and hence cannot be fused directly. They have to be mapped into a common score interval between [0, 1].

A high score indicates the person is genuine, while a low opinion suggests the person is an imposter. The opinions from the modality experts are used by a fusion stage also referred to as a decision stage. It considers the opinions and makes the final decision to either accept or reject the claim. The bimodal biometric system here is designed as shown in Fig. 5.

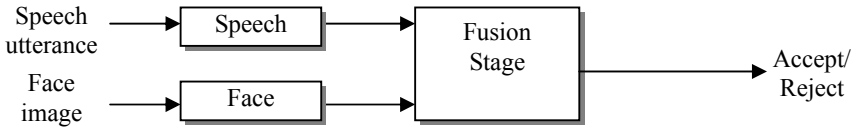

Fig. 5: The building blocks of the bimodal face and speech verification system

From the distance scores, x that is produced by the speech and face databases, the mean (μ) and the variance (σ^2) of the distance values of the speech and face expert, respectively are found by performing validation experiments on the database. The distance score is then normalised by mapping it to the range [-1, 1] using,

$$y = \frac{x - \mu}{\sigma} \tag{17}$$

The [-1,1] interval corresponds to the approximately linear changing portion of the sigmoid function

$$f(y) = \frac{1}{1 + \exp(-y)} \tag{18}$$

used to map the values to the [0, 1] interval. Fig. 6 shows the distribution plot for the genuine and imposter reference points obtained for the system.

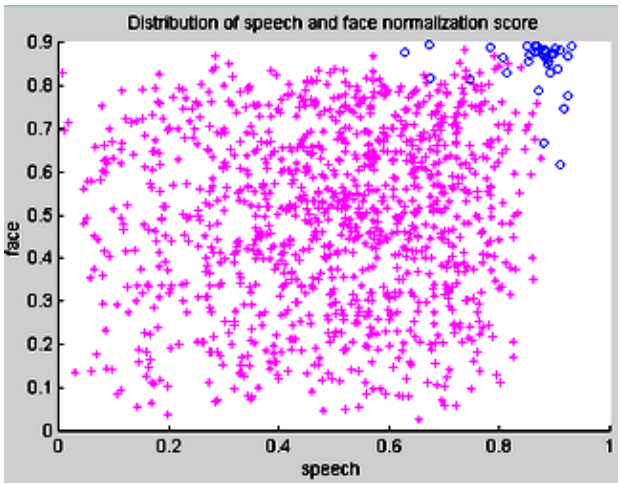

Fig. 6: The distribution plot for the genuine and imposter reference points

4.2 Performance Criteria

The basic error measure of a verification system are false acceptance rate (FAR) and false rejection rate (FRR) as defined in equations (19) and (20).

$$\text{FAR} = \frac{\text{Number of accepted imposter claims}}{\text{total number of imposter accesses}} \text{ x } 100\% \tag{19}$$

$$\text{FRR} = \frac{\text{Number of rejected genuine claims}}{\text{total number of genuine accesses}} \text{ x } 100\% \tag{20}$$

A unique measure can be obtained by combining these two errors into the total error rate (TER) or total success rate (TSR) where

$$\text{TER} = \frac{\text{FAR} + \text{FRR}}{\text{Total number of accesses}} \text{ x } 100\% \tag{21}$$

$$\text{TSR} = 100 - \text{TER} \tag{22}$$

In this targeted application, the Minimum Total Misclassification Error (MTSE) criterion is used, which means it always tries to minimise ε as shown in equation (23).

$$\varepsilon = \min(FA + FR) \times 100\% \tag{23}$$

In order to apply this criterion, we set FAR <0.1% while keeping the FRR to a minimum possible value.

4.3 Experimental Setup

All experiments are performed using a face database obtained from Olivetti Research Lab [25] and speech database contributed by Otago speech corpus [26].

Three sessions of the face database and speech database are used separately. The first enrollment session is used for training. This means that each access is used to model the respective genuine, yielding 34 different genuine models. In the second enrollment session, the accesses from each person are used to generate the validation data in two different manners. The first is to derive a single genuine access by matching the shot or utterance template of a specific person with his own reference model, and the other is to generate 34 impostor accesses by matching it to the 33 models of the other persons of the database. This simple strategy thus leads to 34 genuine and 1122 impostor accesses, which are used to validate the performance of the individual verification system and to calculate the thresholds for the equal error rate (EER) criterion. The third enrollment session is used to test these verification systems, using the thresholds calculated with the validation data set.

4.4 Experimental Results

The performance of the speech and face expert is as shown in Table 1.

Table 1: Individual performance of the face and speech expert

Expert	FAR	FRR	TSR
Speech	8.38%	0.00%	**91.87%**
Face	8.02%	8.82%	**91.96%**

From the values TSR in Table 1, we can observe that the experts are working equally well individually.

Table 2 shows the results obtained with the voting fusion technique described in Section 3.1 by combining the two experts. The continuous decision score from each expert has been converted to a binary number 0 or 1 according to their respective threshold value.

Table 2: Result from voting techniques fusion scheme

	FAR	FRR	TSR
AND rule	4.28%	23.53%	**95.16%**
OR rule	1.16%	0.00%	**98.88%**

From Table 2, it can be observed that the AND rule makes acceptance difficult. This is unfavourable to the genuine access but good with respect to the protection against imposters. On the other hand, the OR rule leads to acceptance being fairly easy. This is good for the genuine access case since it means that the FRR will tend to be small, but is not good with respect to the protection against potential imposters since the FAR will tend to be higher. For the application that is targeted here, the AND rule is undesirable since it introduces a high FAR.

The results shown in following Table 3 are obtained by applying the *k*-NN technique.

From Table 3, it can be observed that the FRR will be increased but FAR decreases when k increases due to the unbalance between the imposter number (1122) compared to genuine number (34). According to P. Verlinde [27], the number of imposter can be reduced using the *k*-means algorithm since this operation induces a loss of information. It will create the cluster *center* points which can replace the actual imposter reference points. Thus, we can create varying R clusters and range k. The best experimental result is obtained at k=3 for the application targeted and the results are as shown in Table 4.

Table 3: Results from k-NN fusion scheme

k	**FAR**	**FRR**	**TSR**
1	0.178%	0.00%	**99.82%**
2	0.178%	0.00%	**99.82%**
3	0.00%	2.94%	**99.91%**
4	0.00%	2.94%	**99.91%**
5	0.00%	5.88%	**99.83%**
10	0.00%	5.88%	**99.83%**
50	0.00%	5.88%	**99.83%**
100	0.00%	100%	**97.06%**

Table 4: Result for modified *k*-NN fusion technique

R	**FAR**	**FRR**	**TSR**
50	1.96%	0.00%	**98.10%**
100	0.18%	0.00%	**99.05%**
500	0.00%	0.00%	**100%**
1132	0.00%	2.94%	**99.91%**

From Table 4, it is observed that the FRR increases, thus FAR decreases with *R* as is be expected. The optimal number of imposter prototypes *R* depends on the cost-function as specified by the application. In experiments, we found that *R*=500 gives the best performance.

The results that are obtained using the linear support vector machine is shown in Table 5.

Table 5: Result for linear SVM fusion technique

	FAR	**FRR**	**TSR**
Result	0.00%	5.88%	**99.83%**

From Table 6, all the fusion techniques are outperformed both single modal experts. Among the techniques considered, modified *k*-NN performed the best in this particular case as it introduces zero FAR and low FRR. Linear SVM and *k*-NN are also adequate for this application by refer to the MTSE criteria that states FAR<0.1% while keeping the FRR to a minimum possible value.

Table 6: Comparisons among the 5 techniques

Techniques	**FAR**	**FRR**	**TSR**
Speech	8.38%	0.00%	**91.87%**
Face	8.02%	8.82%	**91.96%**
AND rule	4.28%	23.53%	**95.16%**
OR rule	1.16%	0.00%	**98.88%**
k-NN (*k*=3)	0.00%	2.94%	**99.91%**
Modified *k*-NN (R=500, *k*=3)	0.00%	0.00%	**100%**
Linear SVM	0.00%	5.88%	**99.83%**

5.0 CONCLUSION

The paper has shown fusion decision technique comparisons for a biometric verification system. The system consists of a speech and face experts developed separately and are targeted for applications involving automatic verification using personal computers and their multimedia capturing devices. In addition, the system is designed to keep the rate as low as possible for the case when an imposter is accepted as being a genuine client. The fusion decision schemes considered are the voting techniques, ordinary and modified k-Nearest Neighborhood classifier and linear Support Vector Machine.

From the experiments, it is found that with the voting technique, the AND rule is undesirable since it introduces a high FAR. On the other hand, the OR rule leads to acceptance being fairly easy. This is good for the genuine access case since it means that the FRR will tend to be small, but is not good with respect to the protection against potential imposters since the FAR will tend to be higher. More complex scheme such as linear SVM and *k*-NN, gives good results that are adequate for this application. The best result is obtained using the modified k-NN technique as it introduces zero FAR and low FRR.

REFERENCES

[1] A. Jain, R. Bolle and S. Pankanti, *BIOMETRICS: Personal Identification in Networked Society*. Boston: Kluwer Academic Publishers., 1999.

[2] L. Hong, A. Jain and S. Pankanti, "Can Multibiometrics Improve Performance?", in *Proceedings AutoID'99*, *Summit, NJ, Oct 1999*, pp. 59-64.

[3] R. Brunelli, and D. Falavigna, "Personal Identification Using Multiple Cues". *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 17 No. 10, 1995, pp. 955-966.

[4] U. Dieckmann, P. Plankensteiner, and T. W. Sesam, "A Biometric Person Identification System Using Sensor Fusion". *Pattern Recognition Letters,* Vol. 18 No. 9, 1997, pp. 827–833.

[5] B. Duc, G. Maýtre, S. Fischer and J. Bigun, "Person Authentication by Fusing Face and Speech Information", in *Proceedings of the First International Conference on Audio and Video-based Biometric Person Authentication, March, 12-24, Crans-Montana,* Switzerland 1997, pp. 311-318.

[6] E. Bigun, J. Bigun, B. Duc and S. Fisher, "Expert Conciliation for Multi Modal Person Authentication Systems by Bayesian Statistics", in *Proceedings of the First International Conference on Audio and Video-based Biometric Person Authentication, March, 12-24, Crans-Montana,* Switzerland 1997, pp. 327–334.

[7] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On Combining Classifiers", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20 No. 3, 1998, pp. 226–239.

[8] L. Hong and A. Jain, "Integrating Faces and Fingerprints for Personal Identification", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20 No. 12, 1998, pp. 1295–1307.

[9] S. B. Yacoub, "Multi-Modal Data Fusion for Person Authentication Using SVM", in *Proceedings of the Second International Conference on Audio and Video-Based Biometric Person Authentication, March, 22-23, Washington D.C.,* USA, 1999*,* pp. 25-30.

[10] S. Pigeon and L. Vandendorpe, "Multiple Experts for Robust Face Authentication", in *SPIE, Editor, Optical Security and Counterfeit Deterrence II 3314, 1998*, pp. 166-177.

[11] T. Choudhury, B. Clarkson, T. Jebara and A. Pentland, "Multimodal Person Recognition Using Unconstrained Audio and Video", in *Second International Conference on Audio- and Video-based Biometric Person Authentication, March, 22-23, Washington D.C.,* USA, 1999, pp. 176–181.

[12] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Detection". *5th International Conference on Computer Vision, June, Cambridge,* MA, 1995, pp. 786-793.

[13] S. A. Samad, A. Hussein and A. Teoh, "Eye Detection Using Hybrid Rule Based Method and Contour Mapping", in *Proceeding of the Sixth International Symposium on Signal Processing and Its Applications, Kuala Lumpur,* Malaysia, August 2001, pp. 631-634.

[14] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces". *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86.

[15] J. Campbell, "Speaker Recognition, A Tutorial", in *Proceeding of the IEEE,* Vol. 85, No. 9, 1997, pp. 1437 – 1462.

[16] R. Martin, "Spectral Subtraction Based on Minimum Statistics". *Proc. Seventh European Signal Processing Conference, September, Edinburgh,* Scotland, 1995, pp. 1182-1185.

[17] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Prentice-Hall International, Inc.,1993.

[18] M. S. Zilovic, R. P. Ramachandran and R. J. Mammone, "A Fast Algorithm for Finding the Adaptive Component Weighting Cepstrum for Speaker Recognition", in *IEEE Transactions on Speech & Audio Processing,* Vol. 5, 1997, pp. 84-86.

[19] J. G. Wilpon and L. R. Rabiner, "A Modified *k*-Means Clustering Algorithm for Use in Isolated Word Recognition", in *IEEE Trans, Acoustics Speech, Signal Proc.*, ASSP, Vol. 33, No. 3, 1985, pp. 587-597.

[20] H. Sakoe and S. Chiba, "A Dynamic Programming Approach to Continuous Speech Recognition", in *Proc. 7th Int. Congress Acoustics,* Vol. 20, No. C13. 1971.

[21] B. V. Dasarathy, "Decision Fusion", in *IEEE Computer Society Press,* 1994.

[22] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* John Wiley & Sons, 1973.

[23] C. J. C Burges, "A Tutorial on Support Vector Machines for Pattern Recognition*"*. *Bell Laboratories, Lucent Technologies, Data Mining and Knowledge Discovery,* Vol. 2, No. 2, 1998, pp. 121-167.

[24] V. N. Vapnik, *Statistical Learning Theory*. John-Wiley & Sons. 1995.

[25] Database of Faces,
http://www.cam-orl.co.uk/facedatabase.html

[26] Otago Speech Corpus,
http://kel.otago.ac.nz/hyspeech/corpusinfo.html

[27] P. Verlinde and G. Chollet, "Comparing Decision Fusion Paradigms Using K-NN Based Classifiers, Decision Trees and Logistic Regression in a Multi-Modal Identity Verification Application", in *Second International Conference on Audio and Video-Based Biometric Person Authentication (AVBPA)*, Washington D. C., USA, March 1999.

BIOGRAPHY

Andrew Beng Jin Teoh obtained his BEng (Electronic) in 1999 from Universiti Kebangsaaan Malaysia (UKM). He is currently working as a tutor at the Faculty of Information Science and Technology, Multimedia University and also as a Ph.D. student in the Department of Electrical, Electronic and System Engineering in UKM. His research interest is in multimodal biometrics and Internet technology.

Salina Abdul Samad has a BSEE from University of Tennessee and a Ph.D. from Nottingham University. Her research area is in the field of signal processing. She is now employed by Universiti Kebangsaan Malaysia as an associate professor.

Aini Hussain received the B.Sc (Electrical) from Louisiana State University, USA; M.Sc. (Systems & Control) from UMIST, England and Ph.D. from Universiti Kebangsaan Malaysia in 1985, 1989 and 1997, respectively. She is currently an Associate Professor in the EESE Dept. at Universiti Kebangsaan Malaysia. Her research interests include signal processing, pattern recognition and soft computing.