Nearest Neighbourhood Classifiers in a Bimodal Biometric Verification System Fusion Decision Scheme

Andrew Teoh

Faculty of Information Science and Technology (FIST), Multimedia University
Jalan Ayer Keroh Lama, Bukit Beruang, 75450, Melaka, Malaysia
bjteoh@mmu.edu.my

S. A. Samad and A. Hussain

Electrical, Electronic and System Engineering Department
Faculty Engineering National University of Malaysia, Malaysia
salina@eng.ukm.my aini@eng.ukm.my

Identity verification systems that use a mono modal biometrics always have to contend with sensor noise and limitations of feature extractor and matching. However combining information from different biometrics modalities may well provide higher and more consistent performance levels. A robust yet simple scheme can fuse the decisions produced by the individual biometric experts. In this paper, k-Nearest Neighbourhood (k-NN) based classifiers are adopted in the decision fusion module for the face and speech experts. k-NN rule owes much of its popularity in pattern recognition community to its simplicity and good performance in practical application. Besides that, k-NN may also provide a ternary decision scheme which is rarely found in other classifiers. The fusion decision schemes considered are voting-, modified- and theoretic evidence of k-NN classifiers based on Dempster-Shafer theory. The performances of these k-NN classifiers are evaluated in both balanced and unbalanced conditions and compared with other classification approaches such as sum rule, voting techniques and Multilayer Perceptron on a bimodal database.
Keywords: Biometrics, face verification, speaker verification, k-NN classifiers.
CR Categories: I.5.4, I.5.1, I.4.9, J.0

1. INTRODUCTION

The authentication of humans for example, in financial transactions, access control or computer access, has mostly been conducted by using ID numbers, such as a PIN or a password. The main problem with those numbers is that they can be used by unauthorised persons. Instead, biometric verification systems use unique personal features of the user himself to verify the identity claimed.

However, a major problem with biometrics is that the physical appearance of a person tends to vary with time. In addition, correct verification may not be guaranteed due to sensor noise and limitations of the feature extractor and matcher. One solution to cope with these limitations is to combine several biometrics in a multi-modal identity verification system. By using multiple biometric traits, a much higher accuracy can be achieved. Even when one biometric feature is

somehow disturbed, for example in a noisy environment, the other traits still lead to an accurate decision.

Some work on the multi-modal biometric identity verification system has been reported in literature. Brunelli and Falavigna (1995) have proposed 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* (1997) proposed a decision level fusion scheme, based on a 2-out-of-3 majority voting, which integrates face and voice data which is analysed by three different experts: face, lip motion, and voice. Duc *et al* (1997) proposed a simple averaging technique and compared it with the Bayesian integration scheme presented by Bigun *et al* (1997). In this multi-modal system the authors use a face identification expert, and a text-dependent speech expert. Kittler *et al* (1998) proposed a multi-modal person verification system, using three experts: frontal face, face profile, and voice. The best combination results are obtained from a simple sum rule. Hong and Jain (1998) proposed 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. Yacoub (1999) proposed 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 and Vandendorpe (1998) proposed 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* (1999) proposed a multi-modal person recognition using unconstrained audio and video. The combination of the two experts is performed using a Bayes net. Ross *et al* (2001) combine the matching of face, fingerprint and hand geometry to enhance the performance of the system. Three different techniques, the sum rule, decision tree and linear discriminant analysis, are used to combine the matching scores. Most recently, Wang *et al* (2003) proposed to combine face and iris together, with Fisher's discriminant analysis and radial basis function network employed in the fusion module.

A bimodal biometric verification system based on facial and vocal modalities is described in this paper. In real world applications, the fusion module needs to be simple yet robust. Generally, fixed rule approaches, such as majority voting and sum rule, fulfill the simplicity requirement and perform well for 'balanced systems', which are ensembles of classifiers exhibiting similar accuracy. However for 'unbalanced systems', where the classifiers have different accuracy, the performance degrades drastically. This is usually the case for many multi-modal biometric systems (Roli *et al*, 2002). Trained rule based classifiers, such as those based on Bayes, neural networks and SVMs, are more effective than fixed rule based classifiers in unbalanced systems but are difficult to optimise. A good compromise, in order to have the simplicity of the fixed rule approach yet avoid the difficulty of the trained rule approach, is the k-Nearest Neighbourhood (k-NN) based classifier. This paper will show the effectiveness of using k-NN based classifiers, namely voting, modified and theoretic evidence k-NN (te$k$-NN) (Denoeux, 1994) in both balanced and unbalanced systems. Well-known fixed and trained rule based approaches are used for comparisons. In addition, the te$k$-NN that uses a ternary decision scheme, accept, reject, inconclusive, is shown to provide a more secure protection for the systems.

2. 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.
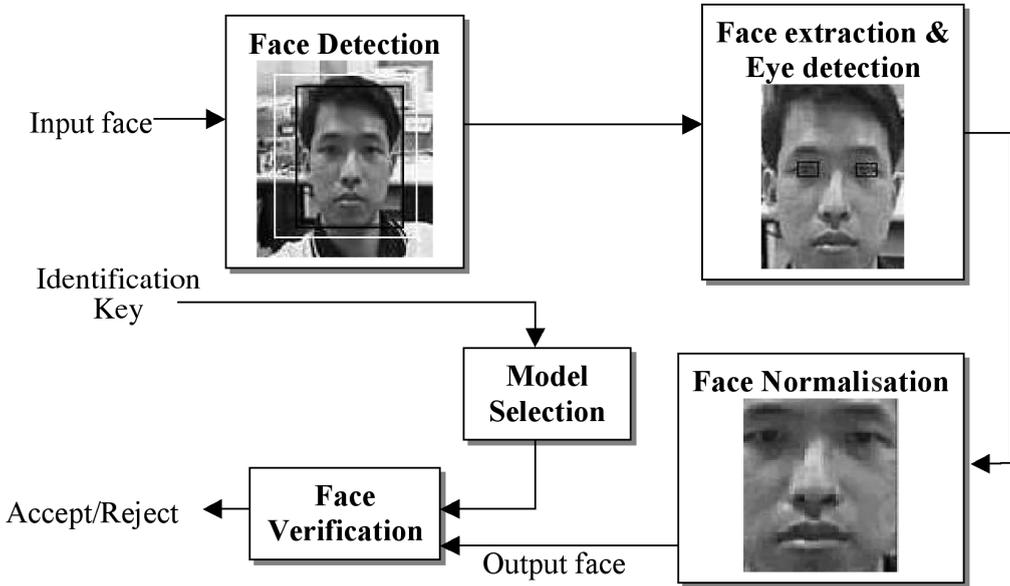
Figure 1: Face Verification System

In our system as shown in Figure 1, the Eigenface approach (Moghaddam and Pentland, 1995) 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 facial features within the entire image space and define this 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 are called eigenfaces.

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.

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

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

Then, eigenvectors, v_k and eigenvalues, λ_k with symmetric matrix C are calculated. Let v_{lk} determine the linear combination of M difference images with ϕ_k to form the eigenfaces, u_l:

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

From these eigenfaces, K out of 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 is. If a window, ψ 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.

$$\varepsilon = \left\| \psi - \psi_f \right\|$$ (4)

where ψ_f is the projection onto the face-space.

From the extracted face, eye co-ordinates will be determined with the hybrid rule based approach and contour mapping technique (Samad *et al*, 2001a). 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 and (ii) the operational stage (Turk and Pentland, 1991). 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:

$$\varpi_{nk} = u_k (i_n - \overline{i})$$ (5)

where $n = 1, \dots, M$ and $k=1, \dots, K$

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

$$\Omega_n = [\varpi_{n1}, \varpi_{n2, \dots}, \varpi_{nK}]$$ (6)

where $n=1,2, \dots, M$.

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 which is computed in the eigenspace.

Let φ_1^0 denote the representation of the input face image with claimed identity C and φ_1^C denote the representation of the Cth template. The similarity measure between φ_1^0 and φ_1^C is defined as follows:

$$F_1(\varphi_1^0, \varphi_1^C) = \| \varphi_1^0 - \varphi_1^C \|_m$$ (7)

where ‖•‖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 (Campbell, 1997). 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 (Martin, 1994) is first used to remove additive noise prior to endpoint detection. Then, in order to cope with the 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
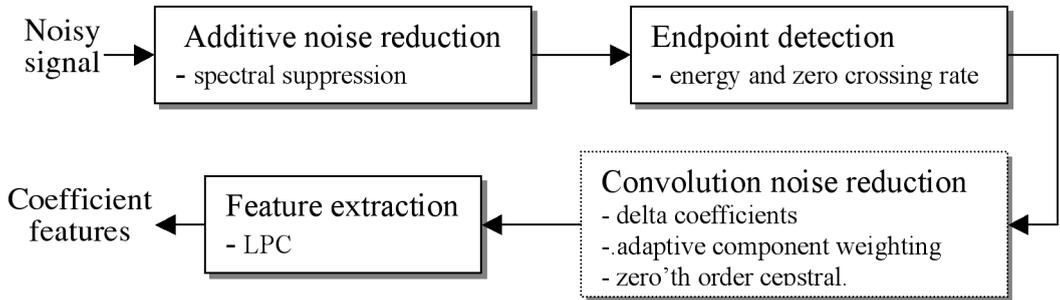
Noisy signal →	**Additive noise reduction** - spectral suppression	→	**Endpoint detection** - energy and zero crossing rate

Coefficient features ←	**Feature extraction** - LPC	←	**Convolution noise reduction** - delta coefficients - adaptive component weighting - zero'th order cepstral

Figure 2: The front-end of the speaker verification module

coefficients (Rabiner and Juang, 1993). In addition, the cepstral components are weighted adaptively to emphasise the narrow-band components and suppress the broadband components (Zilovic *et al*, 1997). 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. The significant improvement of verification rate by using this combination can be found in Samad *et al*, (2001b). Figure 2 shows the process used in front end 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 the training phase, two sample utterances with the same words from the same speaker are collected and trained using the modified k-Mean algorithm (Wilpon and Rabiner, 1985). 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) (Sakoe and Chiba, 1971) to compute the distance between the trained template and the input sample.

Let φ_2^0 represent the input speech sample with the claimed identity C and φ_2^C Cth template. The similarity function between φ_2^0 and φ_2^C is defined as follows:

$$\mathrm{F}_1(\varphi_2^0, \varphi_2^C) = \|\varphi_2^0 - \varphi_2^C\| \tag{8}$$

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

3. FUSION DECISION MODULE

3.1 Voting k – NN (Nearest Neighbourhood) Classifier

The voting k-NN classifier (Duda and Hart, 1973) is a simple classifier that needs no specific training phase. The only requirement is that there be reference data points for both classes representing the genuine and the imposter. A pattern x^s to be classified is then attributed with the same class label as the label of the majority of its k nearest reference Neighbours. To find the k nearest Neighbours, 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.

Cover and Hart (1967) have provided a statistical justification of this procedure by showing that, as the number N of samples and k both tend to infinity such a manner that $k/N \rightarrow 0$, the error rate of the k-NN rules approached Bayes error rate. However, the main drawback of the voting k-NN

rule is that it implicitly assumes the k nearest neighbours of a data point x^s to be contained in a region of relatively small volume, so that sufficiently good resolution in the estimates of the different conditional densities can be obtained. In spite of that, in practice the distance between x and one of its closet neighbours is not always negligible, and even become very large outside the regions of high densities.

3.2 Theoretic Evidence *k*-NN Classifier based on Dempster-Shafer Theory
3.2.1 *Short Review of Dempster-Shafer Theory*
The Dempster-Shafer (D-S) theory of evidence (Shafer, 1976) is a powerful tool for representing uncertain knowledge. In Dempster-Shafer theory, a problem is represented by a set of θ of mutually exclusive and exhaustive hypotheses called the frame of discernment. A basic belief assignment (BBA) is a function $m:2^\theta \rightarrow [0\ 1]$ verifying m(ø) = 0 and $\Sigma_{A\subseteq\Theta}.m(A) = 1$ where ø is the empty set. The quantity $m(A)$ can be interpreted as the measure of the belief that is committed exactly to A, given the available evidence. Two evidential functions derived from the BBA are the credibility function *Bel* and the plausibility function *Pl* defined respectively as $Bel(A) = \Sigma_{B\subseteq A}. m(B)$ and $Pl(A) = \Sigma_{A\cap B\neq\phi} m(B)$ for all $A \subseteq \theta$. Any subset A of θ such that $m(A)>0$ is called a focal element of m. Given two BBAs m^1 and m^2 representing two independent sources of evidence, the Dempster-Shafer's rule defined a new BBA $m = m^1 \oplus m^2$ verifying $m(\phi) = 0$ and

$$m(A) = \frac{1}{K} \sum_{A_1 \cap B_2 = A} m^1(A_1).m^2(A_2) \tag{9}$$

where K is defined by

$$K = \sum_{A_1 \cap A_2 \neq \phi} m^1(A_1).m^2(A_2) \tag{10}$$

for all $A \neq \phi$.

3.2.2 *Implementation*
For a bimodal biometric verification system, a two categories classification problem is considered. The set of classes is denoted by $\Lambda = \{\lambda_{genuine}, \lambda_{impostor}\}$. The available information is assumed to consist of a training set, T = $\{(x^{(1)}, \lambda^{(1)}), \ldots, (x^N, \lambda^N)\}$ of two dimensional patterns $x^{(i)}$, $i = 1, \ldots, N$ and their corresponding class labels λ^i, i=genuine, impostor, taking values in Λ. The similarity between patterns is assumed to be correctly measured by a certain distance function $d(\cdot,\cdot)$.

Let x be the test input to be classified on the basis of the information contained in T. Each pair (x^i, λ^i) constitutes a distinct item of evidence regarding the class membership of x. If x is close to x^i according to the relevant metric d, then one will be inclined to believe that both vectors belong to the same class. On the contrary, if $d(x,x^i)$ is very large, then the consideration of x^i will leave us in a situation of almost complete ignorance concerning the class of x. Consequently, this item of evidence may be postulated to induce a basic belief assignment (BBA) $m(\cdot|x^i)$ over Λ defined by:

$$m(\{\lambda_q\}|x^i) = \alpha\phi_q(d^i) \tag{11}$$

$$m(\Lambda|x^i) = 1 - \alpha\phi_q(d^i) \tag{12}$$

$$m(A|x^i) = 0, \quad \forall A \in 2^\Lambda \setminus \{\Lambda, \{\lambda_q\}\} \tag{13}$$

where $d^i = d(x,x^i)$, λ_q is the class of x^i ($\lambda^i = \lambda_q$), α is a parameter such as $0<\alpha<1$ and ϕ_q is a decreasing function verifying $\phi_q(0) = 1$ and $\lim_{d\to\infty} \phi_q(d) = 0$. When d denotes Euclidean distance, a rational choice for ϕ_q to be

$$\phi_q(d) = exp(-\gamma_q d^2) \tag{14}$$

γ_q being a positive parameter associated to class λ_q.

As a result of considering each training pattern in turn, N BBAs can be combined by using Dempster's rule of combination to form a resulting BBA m synthesising one's final belief regarding the class of x.

$$m = m(\cdot|x^1) \oplus \cdots \oplus m(\cdot|x^N) \tag{15}$$

Since those training patterns situated far from x actually provide very little information, it is sufficient to consider the k nearest neighbours of x in this sum. An alternative definition of m is therefore

$$m = m(\cdot|x^{(i}_1)) \oplus \cdots \oplus m(\cdot|x^{(i}_k)) \tag{16}$$

where $I_k = \{i_1, \ldots, i_k\}$ contains the indices of the k nearest neighbours of x in T.

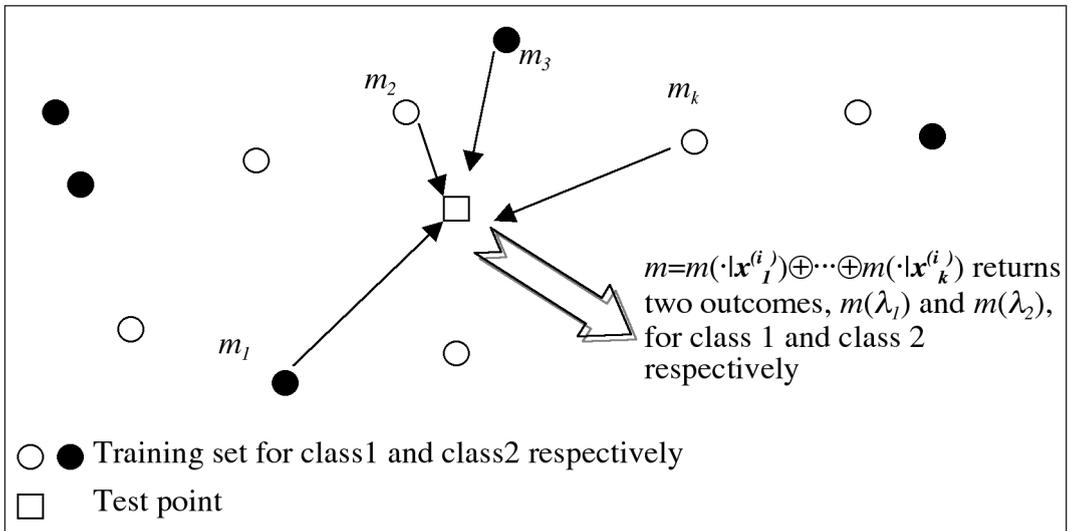This can be illustrated in Figure 3 below:

Figure 3: Theoretic evidence k-NN Fusion Module

The k nearest neighbour of x can be regarded as k independent sources of information, each one represented by a BBA. Note that the calculation of m involves the combination of k simple BBAs, and therefore can be performed in linear time with respect to the number of two classes.

We adopt the definition in equation (11), (12) and (16), m can be shown to have the following expression:

$$m(\{\lambda_q\}) = \frac{1}{K}(1 - \prod_{i \in I_{k,q}}(1 - \alpha \exp(-\gamma d_i^2)))\prod_{r \neq q}\prod_{i \in I_{k,r}}(1 - \alpha \exp(-\gamma d_i^2)_r) \tag{17}$$

$\forall q \in \{genuine, impostor\}$

$$m(\Lambda) = \frac{1}{K}\prod_{r \neq 1}\prod_{i \in I_{k,r}}(1 - \alpha \phi_r(d^i)) \tag{18}$$

where $I_{k,q}$ is the subset of I_k corresponding to those Neighbours of x belonging to class λ_q and K is a normalising factor. Hence the focal elements of m are singletons and the whole frame Λ. Consequently, the credibility and the plausibility of each class λ_q are respectively equal to:

$$bel(\{\lambda_q\}) \qquad = \qquad m(\{\lambda_q\}) \tag{19}$$

$$pl(\{\lambda_q\}) \qquad = \qquad m(\{\lambda_q\}) + m(\Lambda) \tag{20}$$

A decision is made by assigning a pattern to the class λ_q with the maximum value of $m(\{\lambda_q\})$ where $0 < m(\{\lambda_q\}) < 1$.

For real world applications, an inconclusive option may need to be established to postpone decision-making when the conditional error of making a decision given x is high. This situation typically arises in regions of the feature space where there is a strong overlap between classes. Hence, this can be viewed as a problem of conflicting information. In te$k$-NN, this option will be characterised by a BPA, m that will be uniformly distributed among several classes. As a consequence, both the maximum $Bel(\{\lambda_{qmax}\})$ and $Pl(\{\lambda_{qmax}\})$ will take a relatively low value. Given a vector x to be classified, inconclusive state can be decided by comparing $Bel(\{\lambda_{qmax}\})$ to a user defined threshold τ ($0 \leq \tau \leq 1$) if

$$Bel(\{\lambda_{qmax}\}) < \tau \text{ for both genuine and impostor classes} \tag{21}$$

This indicates when $Bel(\{\lambda_{qmax}\})$ for both classes, $q=1,2$ are smaller than τ, then x will be rejected and not placed into any class.

3.3 Other Decision Fusion Schemes

3.3.1 Sum Rule
The sum rule method of integration takes the weighted average of the individual score values. This strategy is applied to all possible combinations of two or more biometrics module. Equal weights are assigned to each modality, as the bias of each matcher is not available.

3.3.2 Voting Techniques
Voting techniques are classical empirical techniques where the global decision rule is obtained simply by fusing the hard decisions made by two biometrics modules. 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 investigation if at least k-out-of-2 modules decide that the person is genuine. When $k = 1$, this is called the *OR* rule. The identity claimed is accepted if at least one of the two experts decides that the person under investigation is genuine. While $k = 2$, this is called the *AND* rule. The identity claimed is accepted only if both the experts decide that the person under test is genuine.

3.3.3 Multilayer Perceptron (MLP)

An MLP is a neural classifier that separates the training data of the several classes by implementing a separation surface, which can have any arbitrary flexible shape. The flexibility of the separating surface is determined by the complexity of the architecture. In this paper, an MLP with two neurons on the input layer (two scores coming from two module biometrics), three neurons on the hidden layer and one neuron (two classes) on the output layer, sigmoidal activation functions for all neurons and the Backpropagation training algorithm is adopted. Using sigmoidal activation functions, the value of the output neuron lies in the interval [0, 1], and the optimal decision threshold is fixed at 0.5.

4. EXPERIMENTS AND DISCUSSION

4.1 Distance score normalisation

The similarity measure values from equations (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].

From the distance scores x, that are 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 separately by performing validation experiments on the database. The distance score is then normalised by mapping it to the range [-a, b], where a, b $\in \Re$ by using

$$y = \frac{x - \mu}{\sigma} \tag{22}$$

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

$$f(y) = \frac{1}{1 + \exp(y)} \tag{23}$$
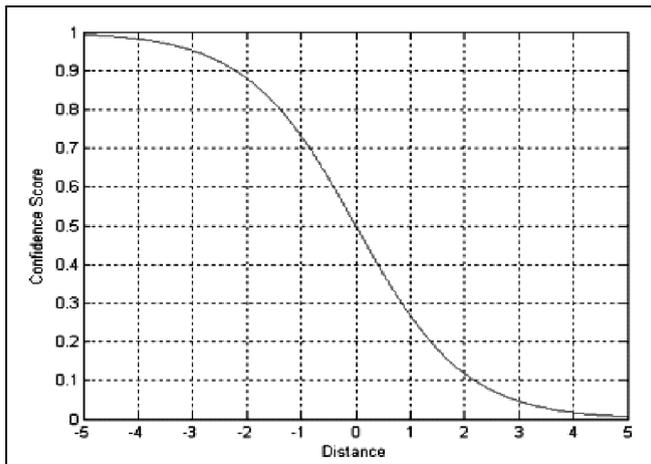
used to map the values to the [0,1] interval.

Figure 4: Sigmoid function

The nature of equation (23) will convert the distance measure into the confidence score as shown in Figure 4, whereby a high score indicates the person is genuine, while a low opinion suggests the
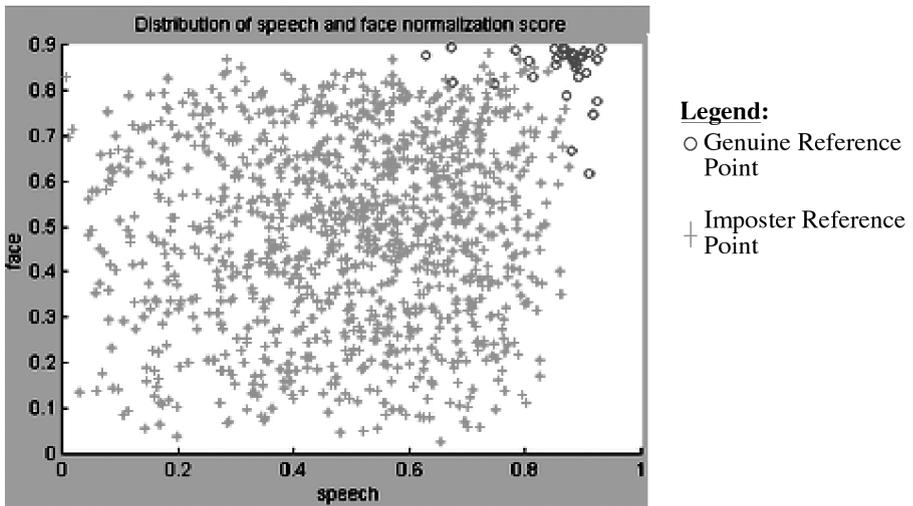
Figure 5: The distribution plot for the genuine and imposter reference point

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. Figure 5 shows the distribution plot for the genuine and imposter reference points obtained for the mapping technique.

4.2 Performance Criteria
The basic error measures of a verification system are false acceptance rate (FAR) and false rejection rate (FRR) as defined in equations (24) and (25).

$$FAR = (\text{Number of accepted imposter/total number of imposter accesses}) \times 100\% \qquad (24)$$

$$FRR = (\text{Number of rejected genuine/total number of genuine accesses}) \times 100\% \qquad (25)$$

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

$$TER = (FAR + FRR)/(\text{Total number of accesses}) \times 100\% \qquad (26)$$

$$TSR = 100\% - TER \qquad (27)$$

For the k-NN based decision fusion modules that are discussed in this paper, note that there is no continuous decision threshold but only a discrete number k, thus it cannot be used to trace a Receiver Operating Curve (ROC).

For the targeted application that uses the bimodal biometric verification system, the Minimum Total Misclassification Error (MTME) criterion is used, which means the system always tries to minimise ε as shown in equation (28)

$$\varepsilon = \min(FAR + FRR) \qquad (28)$$

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

4.3 Experimental Setup

All experiments are performed using a face database obtained from Olivetti Research Lab (ORL) (2000) and Digit multi-modal Database (Sanderson, 2000). Obviously, ORL face database does not come with corresponding speech samples, so to each face image, an arbitrary, but fixed speech class from Otago speech database (Otago, 2000) is assigned. For the speech samples, noise representing additive and convolution noise is added to the samples. Three sessions of the face database and speech database are used separately. The first enrolment session is used for training. This means that each access is used to model the respective genuine, yielding 60 different genuine models. In the second enrolment 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 59 impostor accesses by matching it to the 59 models of the other persons of the database. This simple strategy thus leads to 60 genuine and 3540 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, when FAR=FRR) criterion and other parameters that are required for decision fusion, ie. k value in various k-NN classifiers. 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 speech and face expert is as shown in Table 1.

	FRR(%)	FAR(%)	TSR(%)
Speech	8.33	6.50	93.47
Face	5.00	6.02	94.00

Table 1: Individual performance of the face and speech expert

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

4.4.1 Voting and Modified k-NN

The results shown in following Table 2 are obtained by applying the voting k-NN technique.

k	FRR(%)	FAR(%)	TSR(%)
1	8.33	0.17	99.69
2	8.33	0.17	99.69
3	8.33	0.14	99.72
4	10.00	0.14	99.69
5	10.00	0.11	99.72
10	10.00	0.11	99.72

Table 2: Results from voting k-NN fusion techniques

From Table 2, it can be observed that FRR increases but FAR decreases when k increases due to the unbalance sample number in between the imposter (3540) compared to genuine population (60). Despite of the overall performance increase compare to the face and speech expert alone, the large

No. of Cluster R, $k=3$	FRR(%)	FAR(%)	TSR(%)
100	1.67	1.02	98.97
300	3.33	0.76	99.19
500	3.33	0.34	99.61
1000	3.33	0.40	99.56
1500	8.33	0.17	99.69
2000	8.33	0.17	99.69
3540 (same to voting k-NN)	8.33	0.14	99.72

Table 3: Results for modified k-NN fusion technique

FRR (8.33%) and small FAR (0.14%) for selected $k=3$ indicates that the probability of a genuine being rejected as an impostor is high, which is undesirable in real world applications. According to Verlinde and Chollet (2000), the number of imposters can be reduced using k-means algorithm in order to avoid the above mentioned problem. It will create the cluster centre points which can replace the actual imposter reference points. Thus, we can create varying R clusters but fixed at $k=3$. The experimental result is shown in Table 3.

From Table 3 it is observed that the FRR increases and FAR decreases with R as expected. The optimal number of imposter prototypes R depends on the cost-function as specified by the application. In this experiment, $R=500$ gives the relatively good performance in terms of MTME criteria. Although the TSR is lesser than the voting k-NN, but it achieves a much lower FRR.

4.4.2 te*k-NN Based on Dempster-Shafer Theory*

By using the validation data that are described in section 4.4, it can be observed that $m(\{\lambda_q\})$ in (17) for respective classes mostly are clustered at the interval value in between 0.9 to 1 as shown in Figure 6(a) and Figure 6(b), therefore, $\tau=0.5$ in (21) is a reasonable threshold for the reject option.
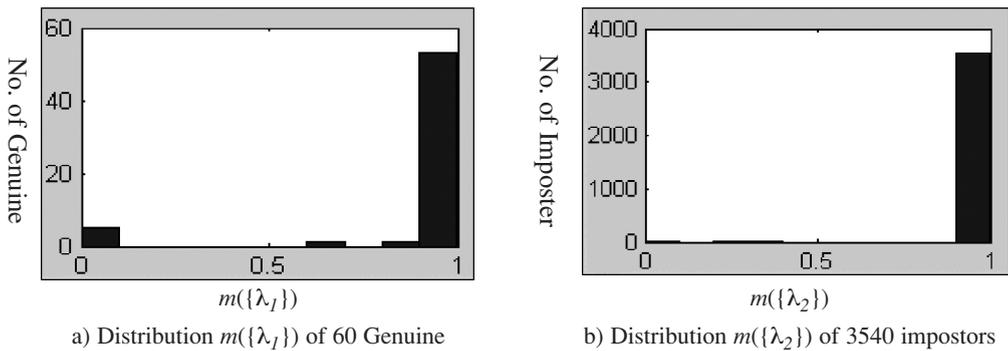

a) Distribution $m(\{\lambda_1\})$ of 60 Genuine   b) Distribution $m(\{\lambda_2\})$ of 3540 impostors

Figure 6: Histogram of m($\{\lambda_q\}$) distribution for genuine and impostor class

The results are shown in Table 4.

From Table 4, it can be seen that the results are insensitive to the choices of the value k. However, the optimal FRR and FAR values which fulfilled the MTME criterion, of FAR<0.1% (0.09%) and a minimum FRR (3.39%), are obtained when $k=2$. The rejection rate is low because both individual biometric modules perform equally well and thus no significant overlapping occurred between the two classes.

k	FRR(%)	FAR(%)	TSR(%)	Note
1	8.48	0.14	99.72	1 rejected
2	3.39	0.09	99.86	3 rejected
3	5.08	0.11	99.80	1 rejected
4	8.48	0.11	99.75	1 rejected
5	8.48	0.14	99.72	1 rejected

Table 4: Results of te$k$-NN fusion based on Dempster-Shafer theory

Method	FRR(%)	FAR(%)	TSR(%)
Voting k-NN (k=3)	8.33	0.14	99.72
Modified k-NN (R=500, k=3)	3.33	0.34	99.61
te$k$-NN (k=2)	3.39	0.09	99.86
Sum Rule (fixed rule)	1.67	1.16	98.83
AND (fixed rule)	23.53	4.28	95.16
OR (fixed rule)	1.67	2.82	97.19
MLP (trained rule)	8.33	0.14	99.72

Table 5: Results comparison for selected k-NN based with other fixed- and trained- fusion methods

4.4.3 Comparison with Others Fixed- and Trained Rule Techniques

By using the other fixed- and trained- rule fusion methods described in section 3.3, the comparison results with selected k-NN based fusion modules are shown in Table 5.

All the fusion techniques in Table 5 exhibit the performance improvement over each individual biometric expert. However, it can be observed that k-NN based fusion modules outperform all the fixed rule approaches, and performance of voting k-NN itself is comparable to the trained rule method, i.e MLP, though the training for MLP is relatively complicated, with consideration given to choosing the number of hidden layers, weight initialisation, minimum error goal setting, etc. compare to k-NN family fusion techniques. Moreover, te$k$-NN shows the best performance and also provides a ternary decision scheme {yes, no, inconclusive} which can be used to provide more secure protection.

4.4.4 Unbalanced System

To mimic real world conditions, an unbalanced version of multimodal biometric system was created with the two experts performing at unequal TSR. Since the speech expert requires extensive pre-processing stage compared to the face expert, it was modified to create an unbalanced case. This can be done by eliminating the pre-processing steps for convolution noise reduction that have been applied in the speech module as indicated in Figure 2. This causes the speech module performance to decrease significantly as shown in Figure 7. The fusion modules are tested again but fixed at $k = 2$ for te$k$-NN and $k = 3$ for voting and modified k-NN. The results are shown in Figure 6.

In Table 6, the performances of all fusion methods reported in this paper were degraded, with particular significance for fixed rule method. However, k-NN based fusion method and MLP still shows good results, even though their performances are decreased. Among all, te$k$-NN exhibits the best result and fulfills the MTME criterion but with an increase in rejection rate. This is due to a larger overlap in between classes and thus more patterns are located near the class boundaries as can be visualised in Figure 7.
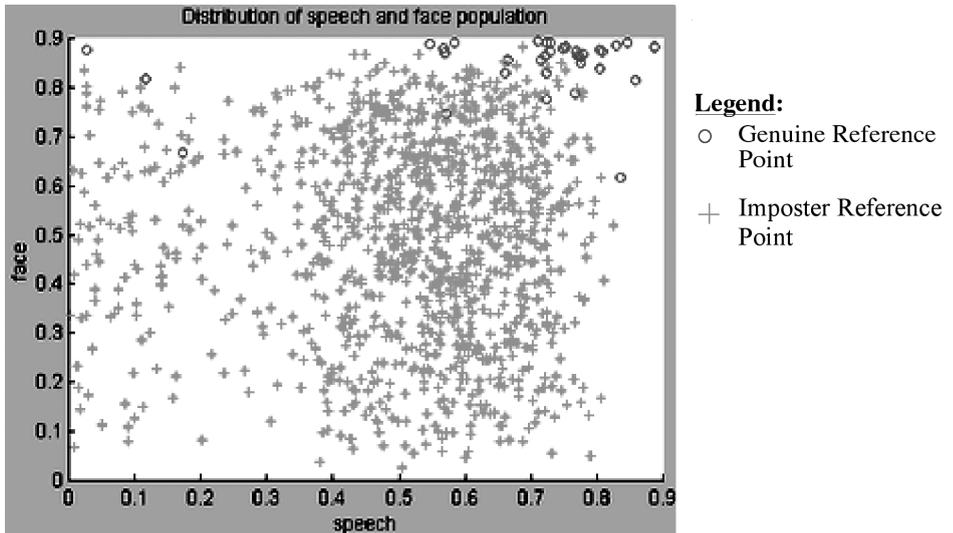
Figure 7: Distribution plot for the genuine and imposter reference point for unbalanced system

	FRR(%)	FAR(%)	TSR(%)	Note
Speech	29.59	5.88	71.11	–
Face	8.33	8.81	91.96	–
Tek-NN (k = 2)	13.79	0.09	99.69	7 rejected
k-NN (k = 3)	21.67	0.25	99.39	–
Modified k-NN (k = 3, R = 500)	15.00	0.65	99.11	–
Sum Rule (fixed rule)	6.67	6.78	93.22	–
AND (fixed rule)	28.33	7.35	92.31	–
OR (fixed rule)	6.67	8.47	91.56	–
MLP (trained rule)	13.33	0.82	98.97	–

Table 6: Result for individual biometric modules and the decision fusion scheme for the unbalanced case

CONCLUDING REMARKS

The paper has shown fusion decision technique comparisons based on k-NN classifiers family and other fixed- and trained rule approaches for bimodal biometric verification systems. From the experiments, it was found that both fixed- or trained rule fusion methods show good results in a balanced system, but degraded when one of the biometric expert exhibits the poor performance, and this is particularly significant for the fixed rule approach. To this extent, k-NN classifiers offer its advantages in terms of the simplicity and robustness. The simplicity results from the fact that it does not undergo complicated parameter tuning processes as a normally trained rule method does, such as MLP but is still able to achieve a good performance, both in balanced and unbalanced conditions. The best result is obtained using the theoretic evidence k-NN classifier as it introduces low FAR and FRR compared to both the voting and modified k-NN classifier. It also provides a ternary decision scheme which can improve the system performance effectively.

REFERENCES

BIGUN, E., BIGUN, J., DUC, B. and FISHER, S. (1997): Expert conciliation for multi modal person authentication systems by Bayesian statistics. *Proceedings of the first international conference on Audio and Video-based Biometric Person Authentication,* 327–334.

BRUNELLI, R. and FALAVIGNA, D. (1995): Personal identification using multiple cues. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 17(10): 955–966.

CAMPBELL, J. (1997): Speaker recognition, A tutorial. *Proceeding of the IEEE* 85(9): 1437–1462, September.

CHOUDHURY, T., CLARKSON, B., JEBARA, T. and PENTLAND, A. (1999): Multimodal person recognition using unconstrained audio and video. *Second International Conference on Audio- and Video-based Biometric Person Authentication,* 176–181.

COVER, T. M. and HART, P. E. (1967): Nearest neighbour pattern classification. *IEEE Trans. Information Theory*, IT 13(i): 21–27.

DENOEUX, T. (1994): A *k*-nearest neighbour classification rule based on Dempster-Shafer theory. *IEEE Trans Syst. Man. Cybern*. SMC-25(5): 804–813.

DIECKMANN, U., PLANKENSTEINER, P. and SESAM, T. W. (1997): A biometric person identification system using sensor fusion. *Pattern recognition letters* 18(9): 827–833.

DUC, B., MAYTRE, G., FISCHER, S and BIGUN, J. (1997): Person authentication by fusing face and speech information. *Proceedings of the First International Conference on Audio and Video-based Biometric Person Authentication* LNCS-1206:311–318.

DUDA, R. O. and HART, P. E. (1973): *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons.

HELLMAN, M. F. (1970): The k nearest neighbour classification rule with a reject option. *IEEE Trans Syst. Man. Cybern*., SMC-6(3): 155–165.

HONG, L. and JAIN, A. (1998): Integrating faces and fingerprints for personal identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 20(12): 1295–1307.

KITTLER, J., HATEF, M., DUIN, R. P. W. and MATAS, J. (1998): On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3): 226–239.

MARTIN, R. (1994): Spectral subtraction based on minimum statistics. *Proc. Seventh European Signal Processing Conference*, 1182–1185.

MOGHADDAM, B. and PENTLAND, A. (1995): Probabilistic visual learning for object detection. *5th International Conference on Computer Vision,* 786–793.

OLIVETTI RESEARCH LAB (2000): Database of faces. http://www.cam-orl.co.uk/facedatabase.html. Accessed 21 April.

OTAGO (2000): Otago Speech Corpus.http://kel.otago.ac.nz/hyspeech/corpusinfo.html. Accessed 1 MAC..

PIGEON, S and VANDENDORPE, L. (1998): Multiple experts for robust face authentication. *SPIE, editor, Optical security and counterfeit deterrence II*, 3314: 166–177.

RABINER, L. and JUANG, B.H. (1993): *Fundamentals of speech recognition*. United State: Prentice-Hall International, Inc.

ROLI, F., JOSEF, K., GIORGIO, F. and DANIELE, M. (2002): An experimental comparison of classifier fusion rules for multimodal personal identity verification systems. *Proceeding of Third International Workshop*, MCS 2002, Cagliari, Italy, June 24–26.

ROSS, A., JAIN, A. K., and PANKANTI, S. (2001): Information fusion in biometrics. In *Proceeding of the Third International Conference on Audio and Video-based Biometric Person Authentication*, (AVPBA'01) 354–359.

SAKOE, H. and CHIBA, S.A. (1971): Dynamic programming approach to continuous speech recognition. *Proc. 7th Int. Congress Acoustics* 20: C13.

SAMAD, S. A., HUSSEIN, A. and TEOH, A. (2001a): Eye detection using hybrid rule based method and contour mapping. *Proceeding of the Sixth International Symposium on Signal Processing and Its Applications*, 631–634, Kuala Lumpur, Malaysia, August.

SAMAD, S. A., HUSSEIN, A. and TEOH, A. (2001b): Increasing robustness in a speaker verification system with template training and noise reduction techniques. *Proceedings of the International Conference on Information Technology and Multimedia*, UNITEN, August.

SANDERSON, C. (2001): Digit database 1.0 – multimodal database for speaker identification/recognition. http://spl.me.gu.edu.au/digit/. Accessed 1 May 2001.

SHAFER, G. (1976): *A mathematical theory of evidence*. Princeton: Princeton University Press.

TURK, M. and PENTLAND, A. (1991): Face recognition using eigenfaces. *Journal of Cognitive Neuroscience,* 3(1): 71–86.

VERLINDE, P. and CHOLLET, G. (1999): Comparing decision fusion paradigms using k-NNbased 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'99)*, Washington D. C., USA, March.

WANG, Y. H., TAN, T. N. and JAIN, A. K. (2003): Combining face and iris biometrics for identity verification. *Proceedings of the Fourth International Conference on Audio and Video-based Biometric Person Authentication (AVBPA'03)*, Guiford, U.K. June.

WILPON J.G. and RABINER, L. R. (1985): A modified K-means clustering algorithm for use in isolated word recognition. *IEEE Trans, Acoustics Speech, Signal Proc*. 33(3): 587–597.

YACOUB, S.B. (1999): Multi-modal data fusion for person authentication using SVM. *Proceedings of the Second International Conference on Audio and Video-based Biometric Person Authentication* (AVBPA'99) 25–30.

ZILOVIC, M. S., RAMACHANDRAN, R. P. and MAMMONE, R. J. (1997): A fast algorithm for finding the adaptive component weighting Cepstrum for speaker recognition. *IEEE Transactions on Speech & Audio Processing,* 5: 84–86.

BIOGRAPHICAL NOTES

Andrew Teoh Beng Jin obtained his BEng (Electronic) in 1999 and PhD in 2003 from the National University of Malaysia. He is currently a lecturer in the Faculty of Information Science and Technology, Multimedia University. He held the post of co-chair (Biometrics Division) in the Centre of Excellence in Biometrics and Bioinformatics at the same university. His research interests are in multimodal biometrics, pattern recognition, multimedia signal processing and Internet security.

Andrew Teoh Beng Jin

Salina Abdul Samad obtained her BSc in Electrical Engineering from the University of Tennessee, USA and a PhD from the University of Nottingham, England. Her research interests are in the field of digital signal processing, from algorithm design to software and hardware implementation. She is now employed by Universiti Kebangsaan, Malaysia as an associate professor.

Salina Abdul Samad

Aini Hussain received the BSc (Electrical) from the Louisiana State University, USA; MSc (Systems & Control) from UMIST, England and PhD from the Universiti Kebangsaan, Malaysia in 1985, 1989 and 1997, respectively. She is currently an associate professor in the EESE Department at Universiti Kebangsaan, Malaysia. Her research interests include signal processing, pattern recognition and soft computing.

Aini Hussain