

Biometrics: New Perspectives on Multimodal and Client-Centered Systems

Vinod Chandran, *Senior Member, IEEE*, and Anthony Nguyen, *Member, IEEE*

Abstract—An introduction to biometrics and an overview of biometric identification technology is presented. The suitability of biometric identification to facilitate secure access is examined with reference to e- developments. Multimodal systems employing combinations of biometrics such as face, voice, fingerprint and signature may see a re-emergence for applications such as smart credit cards, on-line banking, on-line share trading, on-line examinations, etc. False acceptance rates can be dropped significantly using independent information in these systems. It is shown how false rejection rates need not actually be traded off in prompted (or text-dependent) versions of voice and handwriting recognition systems. The true trade-off is the time taken for repeating partial patterns. This translates to a cost to the client and is not transferred to the service provider or to the general tax-payer or to insurance rates. The concept of a client-centered system in the context of biometric identification is explored. Client-centered systems will also be trained by the user and perform most of the processing locally on an embedded system. Advantages of such systems include elimination of the need to collect and maintain large biometric databases, scalability with technology, protection of privacy and customization to individuals.*

Index Terms—Biometric identification, client-centered system, face recognition, fingerprint recognition, handwriting recognition, iris recognition, multimodal system, voice recognition.

I. INTRODUCTION

BIOMETRICS is an area of research and development concerned with the automated processing (with digital computers usually) of biometric data (e.g. iris, fingerprint, face voice and handwritten signatures) for identifying or verifying the identity of living human individuals.

Manuscript received April 1, 2005. A/Prof. Chandran's research on speaker recognition and face recognition has been supported by the Australian Research Council through Grants A00106132 (2001-4) and DP0558415 (2005-7).

V. Chandran is with the School of Engineering Systems, Queensland University of Technology, Brisbane 4001 Australia (phone: +61 7 3864 2124; fax: +61 7 3864 1516; e-mail: v.chandran@qut.edu.au).

A. Nguyen is with the Image and Video Research Laboratory, School of Engineering Systems, Queensland University of Technology, Brisbane QLD 4001 Australia (e-mail: a.nguyen@ieee.org).

The paper presents tutorial material on biometrics and then reexamines the concept of improving biometric identity verification performance by using multiple modalities. This is done with particular reference to online or e-* applications and prompted or text-dependent systems. It also introduces the concept of a client-centered approach to biometrics and explains the principles underlying this approach and the advantages that ensue from it. The paper is organized as follows: Section II presents an overview of Biometric identification – its history, terminology, applications and performance benchmarks. Section III discusses each type of biometric and provides brief accounts of how they are processed. Section IV examines biometrics in the context of internet security. Section V is a discussion of multimodal systems that combine several biometric modalities from a new perspective. Section VI explains the concept of a client-centered approach to biometrics and Section VII is a conclusion summarizing the main contributions of the paper.

II. BIOMETRIC IDENTIFICATION

A. History

The word biometric [1], [2] is derived from the Greek roots – bios for life and metrikos for measurement. It stands for any characteristic that distinguishes living individuals from one another and can be measured such that a comparison is possible. The biometric can be physiological such as a face, fingerprint, voice, iris, retina or hand geometry, or it can be behavioral such handwriting, signature or gait. Any human trait can serve as a biometric provided it is:

- *Universal*: Everyone should have it,
- *Distinctive*: It should not be identical for two different persons,
- *Permanent*: It should be invariant over a period of time, and
- *Collectable*: It should be possible to acquire it for processing.

To be useful for real life applications, a biometric must additionally show *good performance* (accuracy, speed and reasonable resource requirements), be *acceptable* to users (it must be harmless) and *robust to circumvention* or fraudulent methods of impersonation.

Human beings have used the face and voice for recognition from very early stages of social evolution, to identify members of a family or social group. Recognition of identity

by humans is very robust to within-class variations. Face recognition by adult humans for example is quite accurate despite variations in illumination, pose, expression etc. However, this accuracy, in general, does not scale to very large numbers of individuals; it is confined to the order of a few hundred individuals in most cases. Fingerprint and handwriting identification was introduced mainly for the task of criminal investigation more than a century ago. These tasks required visual examination and comparison by experts. Prior to the development of digital computers, the storage and analysis of biometrics were challenging tasks.

With increase in computing power, memory and secondary storage, and decrease in the cost of peripheral devices such as CCD cameras and video capture cards, microphones, sound cards and writing tablets it is now possible to acquire, store and analyze biometrics with relative ease and efficiency. The automated processing of biometric data for recognition or verification is referred to as biometrics. Automated processing usually (but not necessarily) involves digital computers. Some of the earliest papers on automated fingerprint comparison appeared in 1963, on automated speaker verification in 1976, and computer recognition of human faces around 1977. Fingerprint and hand geometry based systems were operational in the mid 1970s and Iris Recognition systems became available in the 1990s [1],[3],[6],[7].

B. Disciplines

Biometrics as an area of research and development, draws from many disciplines such as physics (for sensors that capture biometric information), mathematics and statistics (for algorithms to process the information and learn how to discriminate between individuals), hardware (for platforms that efficiently run the algorithms), software (to implement the algorithms, manage the data and provide appropriate user interfaces), pattern recognition (to provide a systematic basis for classifying and comparing algorithms that learn to discriminate between patterns), neuroscience (to understand how the human brain might process biometrics) etc. The electrical engineering and computer science communities are particularly active in research and development in this area – within such internal subgroups as signal processing, image processing, computer vision, neural networks.

C. Applications

Some research groups around the world have in fact focused their research efforts around biometrics itself. After the September 11 terrorist attacks in the United States, there has been increased research and development funding in this area and many commercial products have been introduced for crew and passenger verification at airports, surveillance of public places, secure access to premises, etc.

However, military uses, criminal investigation and counter-terrorist measures are not the only driving forces behind biometrics. The Internet and the World Wide Web have experienced phenomenal growth in all parts of the world in recent years and spurred many new modes of conducting

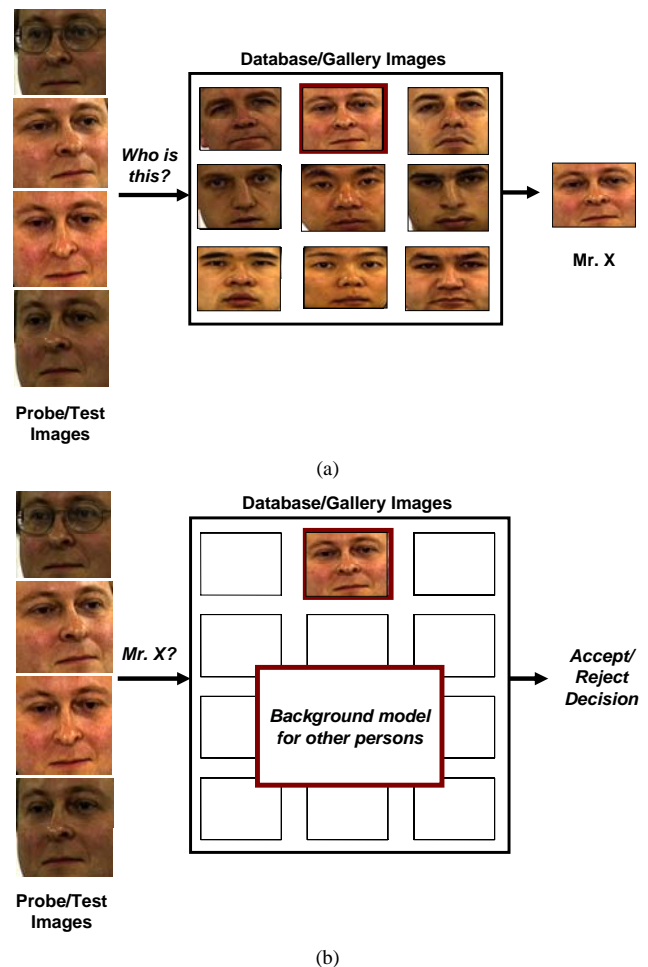


Fig. 1. Task of face (a) recognition or identification, and (b) verification.

business, commerce and other transactions, referred to as e-commerce, e-law, e-education, etc. Many of these on-line applications, at present, require the customer or client to enter an identification number (such as a credit card or bank card number) and a password to confirm the identity of the individual requesting the transactions. Despite the use of secure sockets and encryption these transactions are not entirely safe. Deliberate attacks by computer hackers, Trojans and pirate web servers can steal information from computers including details of bank accounts and passwords. Biometrics such as face, voice or signature can provide added protection from attempts at fraud in electronic transactions over the internet.

Although electronic transactions such as online booking of airline tickets and hotels, online banking and online examinations by educational institutions is becoming popular, they have not reached their potential partly because of lack of user confidence. Biometric verification of identity and endorsement of electronic transactions with electronic handwritten signatures will go a long way to improving user confidence. Very high accuracy or very low false acceptance and false rejection rates, while desirable, need not necessarily stand in the way of their introduction and further

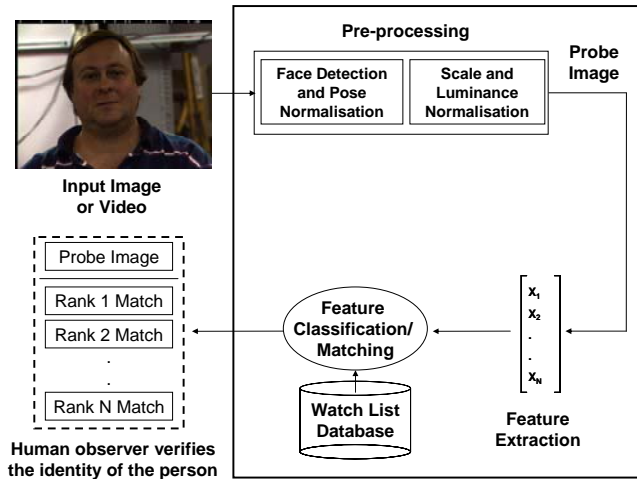


Fig. 2. “Watch List” example – a rank-ordered list of faces is returned rather than only the best match for subsequent decisions by a human in the loop.

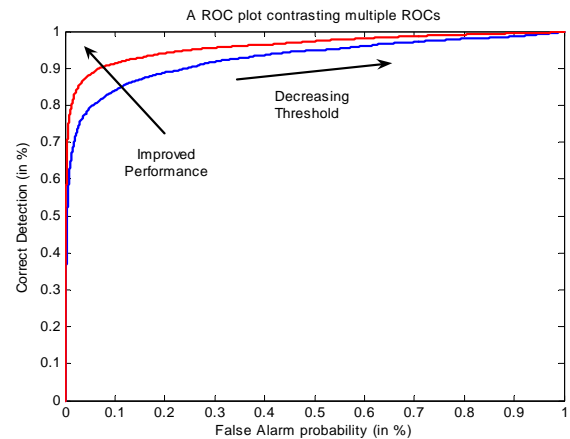
development, in parallel with speed, reliability and security of online transactions using advances in data communication and encryption.

D. Recognition (or Identification) vs. Verification

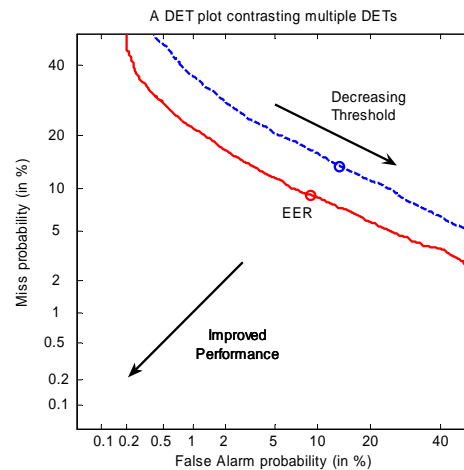
Biometric identification may be considered as a subset of pattern recognition. There are two types of questions that can be posed giving rise to two categories of such problems: Recognition (or identification) and Verification.

In the recognition problem (see Fig. 1(a)), we ask the question: “Who is this?” We compare the unknown biometric (say face) with stored templates or models corresponding to the faces of the individuals registered in the database, and select the one that matches best. If faces that are not in the database are to be permitted, an artificial class of “other” may be created to close the set of faces. For some applications, a rank-ordered list of faces is returned rather than only the best match for subsequent decisions by a human in the loop. The “watch-list” task where an identity is returned when a person is in a watch-list (smaller than the training set or gallery) also requires this. The recognition problem is usually computationally more demanding because of the need to compare to every model in the database which may be quite large for some applications.

In the verification problem (see Fig. 1(b)), we ask the question: “Is this X?” when presented with the biometric from an unknown person and a claim to be X. A comparison is then made to the template or model for X, and possibly a background (or “other”) template or model. If the match is either absolutely or relative to the background, good enough as judged by using a threshold for acceptance, the claim is upheld. Otherwise it is rejected. In the verification problem, there are two types of errors. A false acceptance occurs when someone other than X is accepted to be X. A false rejection occurs when X makes a true claim but is rejected. The probability of these errors varies with the threshold. In



(a)



(b)

Fig. 3. Examples of (a) Receiver Operating Characteristic (ROC), and (b) Detection Error Trade-Off (DET) curves.

general, one type of error decreases at the cost of an increase in the other.

E. System Performances and Benchmarks

A curve depicting the performance of the biometric identification system, showing false acceptance versus verification rate (between 0 and 1) for various values of threshold is known as the “Receiver Operating Characteristic” – terminology derived from the area of radar receivers and aircraft detection. Usually false acceptance and false rejection errors are plotted as percentages or probabilities. The error rate when both types of errors are equal is known as the Equal Error Rate (EER). Often the characteristics around the EER are of particular interest and are emphasized through the use of logarithmic scales and such log-log plots of performance are known as “Detection Error Trade-Off” or DET curves. An ideal system would have zero false rejection and zero false acceptance. Three dimensional curves with rank on a third axis are sometimes used with watch-list tasks.

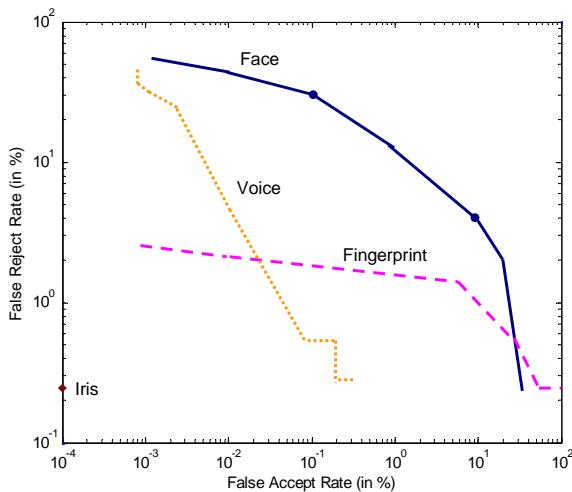


Fig. 4. Detection Error Trade-off curves depicting relative levels of performance of automated identity verification systems based on different biometrics – Iris, Voice, Fingerprint, and Face (adapted from [18])

It is important to realize that the performance of a biometric identification system is not an absolute indicator of the ‘goodness’ of the system or the processing algorithm used in it alone. The performance depends on many other factors such as the size of the database and the conditions under which the biometric data was obtained. In fact, the greatest challenge faced by biometrics is the extent of intra-class variations possible in real-life applications. For example, a face recognition system may perform quite well given frontal shots with good illumination as would be the case when the person stands in front of a well illuminated booth. Its performance may be very poor in outdoor situations such as obtained from an airport surveillance camera or even with images obtained from a web cam where background and illumination can vary widely from room to room. A voice recognition system also may suffer deterioration in performance when the microphone, room acoustics or background audio noise change.

Despite such variations it is possible to benchmark performance using common databases where the other factors that influence performance are controlled or relaxed in a controlled manner and specified along with the performance indicators. Research groups at universities and organizations such as the National Institute of Standards and Technology (NIST), USA, have from time to time undertaken the task of evaluating the performance of algorithms proposed for various biometrics. NIST has conducted competitive, algorithm test programs such as the Fingerprint Verification Competitions (FVC 2000 and 2002) and the Facial Recognition Vendor Tests (FRVT 2000 and 2002). The UK Biometrics Working Group’s Biometric Testing Programme Report [18] compared six different biometrics for verification performance in a normal office environment with cooperative users. An example plot of DET curves showing the relative levels of performance of systems using iris, fingerprint, face and voice,

adapted from this report is shown in Fig. 4.

III. TYPES OF BIOMETRICS

A. Iris

It can be seen that the best performance is obtained using iris. This is owing to the richness of the pattern on the iris and its differences from individual to individual as well as its invariance for the same person with aging. Iris based identification can scale to large numbers of enrolled individuals as well. The drawback with iris recognition is the intrusive nature of data acquisition requiring special imaging devices. It is unlikely that iris based verification of identity will be used for over the counter financial transactions or online electronic transactions at least in the near future.

B. Fingerprint

Fingerprint based systems provide the next best levels of performance and are able to achieve around 2% false reject rates at 0.1% false acceptance. There are a number of different types of fingerprint sensors – visible optical, infrared, static electric etc. and a range of performance levels and reliability in terms of immunity from attempts at copying and presenting fingerprint data from other individuals. Fingerprints are not imaged exactly identically each time – since the portion of the finger imaged can change, the finger may be rotated and there is plastic deformation of the skin. There is usually also some degradation over portions of the print. However, since fingerprints are imaged in direct contact with the imaging device, size normalization is not essential.

Most fingerprint identification systems rely on detecting local features such as ridge endings and ridge bifurcations referred to as minutiae. The minutiae extraction process usually involves the use of a Gabor filter and morphological operations. The Gabor filter is tuned to the periodicity of ridges, of extent determined by the directional stability, and rotated to reveal ridge direction. The geometrical arrangement or “constellation” of the minutiae points is represented as a graph. A matching procedure is then adopted to find the best match to the unknown input. Matching must account for the warping owing to plastic deformation of the skin. The procedure works even if only a subset of minutiae is detected although the reliability is higher with greater coverage of nodes in the graph. Fingerprint based systems are in use at some airports in the USA and Europe.

Fingerprint databases are available with criminal investigation organizations such as the Federal Bureau of Investigation (FBI) in the USA and the Police in many countries. Interestingly, hand geometry based systems have also performed with comparable accuracy although such systems have not been tested on similar large databases, have not seen rapid developments in sensor technology and are less popular. Solid-state fingerprint scanners are available today for about 25 Australian dollars and they are an economically feasible option to use in smart cards for on-line or over-the-

counter transactions. However, fingerprints are inadvertently left on surfaces and sophisticated methods of copying them for fraudulent use have been employed in the past. There is also some stigma associated with fingerprints in the West because it is a commonly used form of identifying criminals. Further, about 5% of the population is believed to have illegible fingerprints.

Some systems for remote user authentication using fingerprint have recently [13-15] been proposed where secret keys for encryption are stored on a smart card accessible only by a user's unique fingerprint. These are attempts at removing the vulnerability of public key encryption systems to impersonation attacks.

C. Voice

Voice based systems have also been implemented to perform at levels comparable to fingerprint recognition. Voice based verification systems are particularly suitable for telephone based transactions and may play an important role with the emergence of third generation multimedia capable mobile telephones and services.

Voice recognition systems can be implemented at a lower cost than fingerprint based systems because microphones are relatively less sophisticated and cheaper than fingerprint sensors. However, better quality microphones, especially stereo and directional ones used for recording can cost in the order of a few hundred Australian dollars, and voice based systems are known to be sensitive to the acoustic channel including the microphone. Another problem associated with these systems is background audio noise, which can be quite high in public places like airports and shopping malls or even within Internet cafes and access points. The use of directional microphones and robust features overcome this problem to some extent. A person's voice can also change when he/she has a cold or throat infection and identification based on voice alone will be unreliable for days in such an event. In terms of acceptability, there is perhaps greater reluctance on the part of a user to speak following prompts than to sign or present ones face.

Most speaker identification systems are based on features extracted from the short-term power spectrum of the speech signal [9]. The bandwidth of the speech signal is roughly 4 KHz for telephone quality speech (sampled at 8 KHz) and 22 KHz for high fidelity microphone speech. Speech data is divided into overlapping (usually 50%) frames. The frame length (around 17 msec) is determined by the correlation of normal speech. The frames are windowed to reduce leakage and a discrete Fourier Transform is applied. Energy normalization is applied in the time or frequency domains. Cepstral processing is then performed because multiplicative channel effects can be separated and removed or compensated by their conversion to additive effects upon logarithmic transformation. Logs of spectral coefficients are taken and the frequency is warped using the "Mel-scale". The "Mel-scale" models human perception. It keeps uniform spacing up to 1000 Hz and above 1000 Hz, it interpolates log energies for

frequency bands increasing by a factor of 1.1 (1100, 1221, etc.). A Discrete Cosine Transform (DCT) is applied to the real warped log energy values. The resulting feature set is referred to as Mel-frequency Cepstral Coefficients. MFCC coefficients (prior to any adaptation procedures) are sensitive not only to the voice but also to the channel and to the spoken text.

The effect of a channel that does not vary during the utterance is in the form of a bias that can be removed through cepstral mean subtraction.

In order to remove the text dependence, features are accumulated from many frames over an utterance (usually 10s of seconds long). This ensures that most of the states of the vocal tract of the person are covered. Features will exhibit clustering or higher probability density around values corresponding to such states. The probability density of the features is modeled using Gaussian Mixture models (GMM) trained using the Expectation Maximization algorithm. The mean values and the covariance matrix corresponding to the different modes in the mixture represent the parametric model for a particular speaker. Given a test utterance from an unknown speaker and a claimed identity, an identical feature extraction procedure is performed on the test data and the likelihood of the data having been generated by the claimed model is computed. For recognition this is computed for every model in the dataset and the model corresponding to the maximum likelihood is selected for the recognized identity. For verification, this may be compared with the likelihood of the data having been generated by a "background speaker model" or generic model and a normalized score compared against a threshold to make an accept or reject decision.

Mel-cepstral feature based systems discard information in the phase of the Fourier spectrum. Recent work at QUT has shown that utilizing the phase information through higher order spectral phase features (see [10] and references therein) can improve performance and provide better robustness to additive noise.

Text-dependent speaker identification systems require the user to utter specified words such as for example, a sequence of digits generated by the system. These systems use Hidden Markov Models (HMM) for speech recognition. HMMs also incorporate the transition probabilities between states unlike GMMs, which ignore the temporal behaviour. The model for each word may be generated specific to each speaker. In that case, an incorrect speaker will encounter word recognition errors in the speech recognition stage itself. If the sequence of words cannot be uttered without such errors, the claim may be rejected straightaway. If a valid sequence is uttered by the speaker, a combined log-likelihood score is generated from all the data collected, background normalized and compared to a threshold. The Vocal Access system developed at QUT is an example of one such system that has been commercialized. NIST evaluations of speaker recognition algorithms have been largely dominated by GMM based methods in recent years and the speech laboratory at QUT has produced top ranking algorithms in certain categories.

Recent research in speaker recognition has moved from using only the acoustics of the speech to “ideolectal” schemes that use idiosyncratic sound utterances, prosody, etc.

D. Face

Computer recognition of faces has come a long way since the first attempts in the mid 1970s and many algorithms have been proposed. Three FERET evaluations (1994, 1995, 1996) and the NIST Face Recognition Vendor Test FRVT in 2000 helped evaluate the technology from early implementations to the first prototype systems. Many commercial implementations also exist today. The NIST Face Recognition Vendor Tests of 2002 were aimed at benchmarking such mature algorithms over large databases. 10 commercial systems were tested with faces from a database of 37,437 people. For frontal indoor images the false rejection rate was around 10% for false acceptance of 1% for the best algorithms but the performance decreases with increase in gallery size and increase in the time elapsed between gallery and probe images.

According to the FRVT 2002 [5] evaluations, three-dimensional morph-able models and normalization of scores improve performance while face recognition from video sequences offers a marginal improvement. Normalization is a post-matching, statistical technique that is aimed at correcting differences in templates among people. Three-dimensional data on facial geometry can be obtained using stereo vision, structured light systems and scanning laser rangefinders. Three-dimensional facial models and recognition systems are currently a hot topic of research in facial recognition and the subject of evaluations such as the Face Recognition Grand Challenge (FRGC) being organized by NIST and the University of Notre Dame, USA.

Details of algorithms used in mature commercial face recognition software are not known but it is widely believed that two of the U.S. companies use algorithms derived from the very popular eigenface method based on principal component analysis developed at the MIT.

Preprocessing steps prior to feature extraction are quite critical for face recognition systems. The image must be segmented to locate the face. Colour information may be used for segmentation but after normalization the image is usually gray-scale. Normalization is performed by detecting the eyes using correlation with a standard eye template, and interpolating or decimating the face to normalize the interocular distance in pixels. Lighting correction is usually accomplished by normalizing the gray-scale histogram.

Feature extraction methods rely on signal processing and a variety of methods have been proposed and implemented. The eigenface method unwraps the image into a vector without regard to physiological facial structures. This vector is projected on to a space of basis vectors called eigenfaces. The basis vectors are the global eigenvectors associated with the largest eigenvalues of a covariance matrix of various training images. They are in some cases modified to minimize spectral leakage against local changes in the facial image as would be

the case with an expression change. The coefficients or weights of the projection comprise a feature vector. Feature vectors from probe images are compared with those in the gallery (or enrolled) images using Euclidean distance or classified using a neural network trained on features from the enrolled images.

Other methods implemented in successful face recognition systems include global two-dimensional Fourier transform coefficients fed into a neural network, and Gabor filter coefficients extracted at important points located using deformable templates.

Model based face recognition is currently popular with researchers in this area. It has potential to overcome the sensitivity to pose and illumination changes. A 3D model [8] is trained on available 2D data by estimating the illumination direction and pose. Physiological features are adjusted accordingly. The model is “corrected” in pose and illumination and a 2D image is obtained that can be compared to enrolled images using image-based methods mentioned above.

Face recognition is already in use at some airports for verification of the identities of crew and/or passengers. Controlled lighting, the use of multiple reasonably good quality cameras for image capture, computational and storage requirements make it somewhat unattractive for use with computer networks for on-line transactions. However webcams cost only in the order of a few hundred Australian dollars and mobile phones now come equipped with CCD cameras. Improvement in the performance of face recognition algorithms from images captured by such low cost devices will make it a viable modality for use in secure on-line electronic transactions.

The current research focus in face recognition at QUT is to (a) use three dimensional models, (b) use hybrid 2D-3D techniques, (c) use stereo images directly for feature extraction without the computationally intensive depth estimation as an intermediate step, (d) to track persons and faces using multiple cameras and depth information, and (e) use techniques such as super-resolution to enhance face segments obtained from video surveillance.

E. Handwritten Signatures

Although there are a number of other biometrics that can be used in this context, I will confine attention to just one more – the handwritten signature. The handwritten signature is not a particularly reliable or accurate biometric. It is known that some individuals have relatively simple signatures that can be easily copied. However, signatures have been used in legal documents and financial transactions for centuries and a person signs his/her name many times shopping with a credit card. It has no stigma attached (unlike fingerprints in western society) and is highly socially acceptable. Manual comparisons of signatures are very error-prone and expert forgeries cannot be detected by untrained individuals. Credit card fraud alone is reported to cost billions of dollars every year worldwide. E-commerce and online banking are

increasing in popularity rapidly. A reliable handwritten signature verification system could be a useful feature for such applications. Pen-tablet systems have become reasonably inexpensive at a few hundred Australian dollars, about the same as a good web-cam or a directional microphone. Further, they are capable of capturing the dynamics of signatures – variations in pen-tip pressure and pen angles in addition to the spatial coordinates. Many different on-line signature verification algorithms [11] have been proposed by research groups around the world and some commercial products are also available. However, the first benchmarking exercise (Signature Verification Competition, SVC2004) was only conducted recently at the International Conference on Biometric Authentication, Hong Kong, 2004. It was conducted over a small database of about 40 individuals. The equal error rate for signature verification is probably around 10% owing to the large intra-class variations. Some commercial products have reported equal error rates of about 2%. One commercial system has reportedly been tested at public trials with around 8500 signatures [11]. In fact, no two genuine signatures of the same person are exactly identical. Identical signatures are only produced by tracing forgery, and only of the static type. Most forgeries are rather clumsy attempts and when analyzed reveal many of the natural handwriting characteristics of the forger. Algorithm developers have handled intra-class variations through deformable template techniques, reliance on the dynamics and global characteristics rather than local, highly variable characteristics, the use of HMMs etc. An interesting study by Plamondon [12] found that the shape of the velocity profile for rapid aimed strokes as in a signature is approximately Gaussian but asymmetric, and that it is almost preserved for strokes that vary in duration, spatial extent or peak velocity. At QUT, we have implemented a signature verification system based on higher order spectral features and an “ordered part” approach to signal segmentation. It is an on-line signature verification system and uses spatial coordinates and pressure and their derivatives. It can also use pen angles if available. The system is robust to scale variations, rotation, slant, etc. It depends primarily on the handwriting characteristics of the person, is independent of the language and can work on any signature that the person trains to reproduce fairly well. A real-time prototype has been implemented.

IV. SECURE INTERNET ACCESS

A number of scenarios are possible when we use biometric information for secure Internet access. One method may be the storage of secret keys for encryption on smart cards [13-15] or local disks, accessible only with biometric verification of user identity. Multiple biometrics can be used to improve reliability. One could take this a step further to enforce privacy by providing third parties with only encrypted biometric information and reserving one biometric (this could be a learned signature) or one password. Thus biometrics could be made “cancellable”. Face and voice and even the

static signature are freely publicly available, but other biometrics need not be. The encryption algorithm could also make use of a timestamp to ensure that decryption is only possible within a reasonable timeframe. Biometric information could also be passed on to the presentation layer of every packet and continued availability be made a requirement for decryption, as for example, when the user is taking an examination or playing an on-line chess game at a remote terminal (with perhaps face as the continuously or frequently available biometric).

V. MULTI-MODAL SYSTEMS

In order to achieve better performance, a system that combines more than one biometric to make an identification or verification decision may be used. Such systems are called multi-modal systems because they use more than one modality in signal processing terminology. However, there are a number of issues that arise with the use of such systems. One has to decide the architecture of the system, whether it combines data before extracting features, combines features before classification, combines classifier scores before making a decision or combines decisions before making a final decision, for example. There are also many possible rules of combination [17]. Not all methods of such combination, or fusion, lead to better performance. Some can even lead to catastrophic fusion and poorer performance. Optimal combinations and weights are difficult to find without considerable experimentation and may not generalize well. A great deal of research has gone into the problem of fused classification and some general results are available.

Data fusion usually makes sense only for inputs that are of the same type or at least similar. When data refers to segments obtained from a sequence of samples, as in the case of speech or handwriting, they are often fused before feature extraction and the algorithm is designed to group optimal numbers of samples together. When data comes from different modalities such as fingerprint and face, it is not practical to fuse them because the type of processing used to extract features from the modalities (methods that work well) may be radically different. Feature extraction methods are necessary because raw data is usually too high in dimension and contains a lot of redundant information.

Features can be fused provided they share similar range of values or have been normalized to do so, provided again that the dimensionality is not increased beyond the capabilities of the classification subsystem. For example energy features may be positive, phase features may lie between $\pm\pi$, and topological features such as the Euler number may be integers. Feature fusion will result in improved performance only if the classifier can learn the joint probability density function well. If, for example, a diagonal covariance matrix is assumed in a Gaussian mixture model classifier, the dimensions are treated as being uncorrelated and no real advantage is gained from the fusion. In fact, the default computation of posterior probabilities would use the product of contributions from each

dimension and may be heavily influenced by the dimension that has the lowest contribution. Each individual contribution is a likelihood score.

Classifiers can be combined using their scores (soft fusion) or their decisions (hard fusion). When the available discriminatory information in the feature is highly ambiguous owing to high levels of noise (as may be the case with features from blocks of speech), a sum rule combination of scores will often outperform the product rule, as explained in [17]. It was also observed in [17] that correlations between features within the same modality are much stronger than those between features from different modalities.

Some modalities such as speech or handwriting are amenable to implementation in prompted systems where the client is asked to speak (or write as the case may be) a series of prompts. For example, he/she may speak out the digits in a randomly generated sequence, or be required to write a random sequence of words from a predefined vocabulary. In such systems text-dependent classifiers can be used and classifier decisions can be fused to achieve better performance. In the most typical scenario, the identity of the user is accepted only if it is accepted by every classifier in the sequence. This will lead to a lowering of false acceptance rates. Although not a multimodal system, strictly speaking, there is some independence in the information available from the response to each prompt and fusion of classifiers leads to improved performance in a similar manner. There is one important difference - a response to a prompt can be repeated if it is rejected. This leads to a trade-off of time rather than a trade-off of false rejection rates when false acceptance rates are brought down by classifier combination. Let us analyze this system assuming that the classifiers are making statistically independent decisions from independent information about the source. Further, it is assumed that each classifier is identical to simplify the algebra below without loss of generality.

Let the false acceptance rate of each classifier be α and the false rejection rate be ρ . Let there be N classifiers corresponding to N prompts. Assume that each prompt and response event takes $\frac{\tau}{N}$ seconds to complete. Then the false

acceptance rate of the combination is $\alpha_{1,N} = \alpha^N$ where 1 refers to the number of attempts allowed for each prompt. The probability of true acceptance up to the k -th classifier is $(1 - \rho)^{k-1}$. The probability that there will be a false rejection from any one classifiers is therefore,

$$\begin{aligned} \rho_{1,N} &= \rho + (1 - \rho)\rho + (1 - \rho)^2\rho + \dots + (1 - \rho)^{N-1}\rho \\ &= N\rho - \rho - \rho^2 - \dots - \rho^{N-1} \\ &\approx N\rho \text{ when } \rho \ll 1 \end{aligned}$$

This seems to suggest that the lower false acceptance is at the expense of a higher false rejection rate. However, if multiple attempts are allowed at each prompt, say M times, the

effective false rejection rate of each classifier is reduced and

$$\rho_{M,N} \approx N\rho^M$$

The consequent change in false acceptance rate is

$$\alpha_{M,N} = (M\alpha)^N = M^N \alpha_{1,N}$$

The trade-off here becomes the increase in total time for a verification attempt to an upper limit of $M\tau$. Note that the multiple attempts need not be evenly distributed over all the prompts for the above result to hold.

Assume that the false acceptance rate of each of 10 classifiers is 0.1 at the equal error rate and the verification time is 10 seconds without multiple prompts. With 2 attempts per prompt (or a total of 20 attempts allowed), the false acceptance rate of the combination is ideally $2^{10}10^{-10}$ or approximately 10^{-7} at a false rejection rate of 0.1. The total transaction time will increase to 20 seconds at most under these assumptions. In practice, the number of attempts required by a true claimant to get verified will be far less because the user will adapt. A false claimant on the other hand will not be very good at adapting his/her voice or handwriting to that of the true identity. It thus seems as if the false acceptance rate can be reduced arbitrarily without trading off the false rejection rate, at the expense of some increased time for a verification transaction. This increased time, in an application scenario such as an online transaction from a home or office computer, will be borne by the claimant. It need not hold up other users or be added on as a cost to the service provider. This is especially so, if the processing is done locally and only results are communicated. The models required may need to be downloaded but that is a one-off operation.

Is there a caveat in the above example? Of course, the assumption of statistically independent decisions and independent information for classification is an ideal one and not always met in practice. However, in a text-dependent speaker recognition system using speaker-dependent HMM classifiers for each word, an assumption of independence is good when the phonemes involved in the word are different and will hold reasonably well even when they share some phonemes but differ in the order in which they are put together. In a handwriting recognition system, the same will hold true if the alphabets involved are different and words are spelt differently.

There is also a limitation set by the need to have models trained for each prompt for each claimant. Training requires several repetitions of each prompt. Hence, vocabulary sizes will need to be manageable. A vocabulary in the range of 10 to 100 words is probably quite reasonable from a storage and processing perspective. It must be noted that the vocabulary need not be identical for each user. If a false claimant is unaware of the vocabulary used by the true identity in training models, it will be added protection against fraudulent access.

Such systems could be considered pseudo-multi-modal. They may not provide truly independent information as for example a fingerprint and voice. However, they have several

advantages over true multi-modal systems as well. The processing of data from different prompts can be identical because they are after all the same modality. This makes it possible to use identical classification strategies and soft fusion of classifier scores easier. Further, the number of classifiers can be quite large and not constrained by the availability of newer biometrics. It can also be noted that text-dependent systems can capture more than just the voice or the general handwriting of the identity. Speech processing systems of this nature can capture the accent and idiosyncrasy in uttering particular words. Handwriting systems can capture idiosyncrasy in writing particular alphabets and words, if properly designed.

One might ask whether multiple presentations of the same modality such as a face or a fingerprint from the same finger would achieve similar reduction in error rates. The answer is no. Even though the multiple acquisitions are independent trials, they contain a lot of common information. Hence classifier fusion such as by product rule would lead to lower false acceptances only at the expense of consequent increase in false rejection in these cases. Many face recognition systems utilize multiple frames but do so mainly to select the better ones for which pre-processing operations such as eye location and normalization work well. They may also screen the frames to locate the best frontal views for comparison. The analog of text in face recognition would come from variations in expression, pose, lighting etc. and in order to build text-dependent systems one would have to compare a smiling face with a smiling face model, a gloomy face with a gloomy face model, etc. Further, prompting one to reproduce particular facial expressions is not in general going to produce a workable, user-friendly system. A limited use of the principle is possible by capturing several poses simultaneously with an array of cameras.

VI. CLIENT-CENTERED APPROACH

Biometrics based secure access for online transactions over the Internet or over the counter using other networks such as ATM will be implemented as a distributed system. Most such systems at present use passwords or magnetic stripe cards. Some smart card systems with biometric information in non-volatile memory have also been introduced. The remote system or client could be (a) passive, acquiring biometric data and passing it on to the server for verification, or (b) active, acquiring data, processing it locally using locally stored models and passing on only results to the server. In addition, the system could be active during both the training and testing phases – that is, the models are even created at the client. Such an approach transfers the responsibility of acquiring data for training and creating good models to the user. The server may require the user to register a trained client and obtain some form of certification. Such an approach to biometrics could be called a client-centered approach. It is not conventional from the point of view of applications such as border control or law enforcement. However, it is natural from a commercial

transaction point of view where passwords and magnetic stripe cards are the responsibility of the user to maintain. It also offers many advantages in the development and deployment of biometrics solutions for large populations.

All the processing takes place in an embedded system possessed by the client (or user). Only a decision is communicated to the central system, which will keep a log of the transaction details and the verification status. The decision can be encrypted and time-stamped. A high end technology implementation would communicate the decision to the server via an infrared or radio link. However, we can even scale down in technology and communicate the decision through an LCD display if necessary. The low-tech version is not fool proof but good for human verification at a counter. In order to prevent circumvention attacks, one must ensure that the software in the client cannot be modified by the user and physical look-alike copies with tampered software cannot be made. This is similar to forging a card and can be prevented by watermarked identification of the hardware.

A client centered system can easily integrate all the factors for secure access - what you know (password), what you have (a smart card or pen or phone like embedded system) and what you are (biometric information) in one system.

A client centered system will make it unnecessary for every remote access point to have every type of biometric sensor. They need only have the communication interface which could be standardized.

When the smart card is issued, there will be a period over which the user trains the system and then registers it and gets certification. If a card is lost before training, the user can notify the issuing authority by some quick means and prevent certification through identification of the hardware. This would of course imply that the system must self-destruct on any attempt to open with the intention of changing the hardware settings.

Client centered systems will be better from the point of view of development – there is no need to collect very large databases and maintain those centrally. Verification can be done using a true identity model and a background model. These can be quite easily stored on a small embedded system even if they are hundreds of kilobytes. The true identity model can be generated by the user and a background model can be either supplied with the hardware or trained by the user with data acquired from a small local population such as family members and friends. A good near-set background model is obtained from other persons with similar voice, similar handwriting etc. as the case may be and they are more likely to be close family members. The collection of such data and its processing in a centralized manner can be a massive organization effort and prone to errors.

A client-centered approach allows easy customization. It can adapt to the language of the user, use one fingerprint or both that grasp the pen, allow prompted operation with user chosen vocabulary etc.

The client-centered approach can be implemented in a progressive manner with respect to factors such as coverage of

population, sophistication of the technology used, etc.

The cost of the embedded system would eventually be not much more than the annual fee for a credit card. It may be subsidized by financial institutions and insurance companies who are likely to benefit from lower incidence of fraudulent transactions.

The client-centered approach would allow protection of privacy because all biometric information need not be made available by the user to third parties. Encrypted and cancelable forms may be transferred upon request. For recognition and watch-list tasks, such information may be required by the government and law enforcement agencies.

A client centered approach will also reduce the risk of catching a virus or bacterial infection through having to touch a fingerprint sensor or keypad that is touched by many others. Each user carries the sensors with him or her and only data is exchanged.

Finally, with a client centered approach it will be unnecessary for every bank or credit card company to develop its own custom biometric solutions and maintain its own biometric databases. They can develop products using a standardized communication interface.

Are all biometric modalities suitable for client-centered implementation? It is unlikely that an iris sensor would go on a small embedded system, and face verification would pose computational and storage challenges to these systems with current technology. Fingerprints and handwriting would be natural to a pen-like device and voice would be to a mobile phone or wired microphone type device. However, the use of speech may be difficult in noisy public environments like an airport or shopping mall. Used from the privacy of ones home or office, it is an attractive modality for such implementation and less likely to degrade in performance owing to channel variations.

VII. CONCLUSION

There is enormous potential for biometric verification in secure internet transactions that will benefit the economy through lower-overhead and faster commercial transactions at lower risk to both consumer and supplier. A new perspective on multi-modal and prompted text-dependent or pseudo-multi-modal systems is presented showing that high performance can be achieved without the need to resort to one highly reliable biometric. A client-centered approach to biometric verification offers many advantages for application to large populations such as scalability in technology and coverage of population, customizability for the individual user, better protection of privacy, and elimination of the need to collect and maintain large biometric databases centrally.

ACKNOWLEDGMENT

Dr. Chandran is grateful to QUT for travel and conference support. Support from Prof. Sridharan and many postgraduate students in the Speech and Image laboratories at QUT are also acknowledged.

Facial images used in the illustrations in Figs. 1 and 2 were obtained from [19].

REFERENCES

- [1] J. L. Wayman, "Digital Signal Processing in biometric identification: A review," Proc. of Intl. Conf. Image Processing (ICIP), vol. 1, pp. 22-25, September 2002.
- [2] K. Delac and M. Grgit, "A survey of biometric recognition methods," Proc. of 46th Intl. Symposium on Electronics in Marine Elmar 2004, pp. 184-193, June 2004.
- [3] <http://www.biometrics.org>
- [4] <http://www.nist.gov>
- [5] P.J. Phillips, P. Grother, R.J. Michaels, D.M. Blackburn, E. Tabbassi and M. Bone, "Face Recognition Vendor Test 2002: Overview and Summary," NIST Technical Report, March 2003.
- [6] M. Yang, D.J. Kriedman and N. Ahuja, "Detecting faces in images: a survey," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, Jan. 2002.
- [7] R. Chellappa, C.L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," Proc. of the IEEE, vol. 83, no. 5, pp. 705-740, May 1995.
- [8] K.W. Bowyer, K. Chang and P. Flynn, "A survey of approaches to three-dimensional face recognition," Proc. of 17th Intl. Conf. on Pattern Recognition (ICPR), vol. 1, pp. 358-361, Aug. 2004.
- [9] D. A. Reynolds and R. C. Rose, "Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models," IEEE Trans. On Speech and Audio Processing, vol. 3. no. 1, pp. 72-83, Jan. 1995.
- [10] V. Chandran, D. Ning and S. Sridharan, "Speaker Identification using Higher Order Spectral Phase Features and their Effectiveness vis-à-vis Mel -Cepstral Features," Proc. of Intl. Conf. on Biometric Authentication (ICBA), 2004.
- [11] G. Gupta and A. McCabe, "A Review of Dynamic Handwritten Signature Verification," Technical Report, James Cook University, Australia, 1997.
- [12] R. Plamondon, "Looking at Handwriting Generation from a Velocity Control Perspective," Acta Psychologica, vol. 82, pp. 89-101, 1993.
- [13] M.S. Hwang and L.H. Li, "A new remote user authentication scheme using smart cards," IEEE Trans. on Consumer Electronics, vol. 46, pp. 28-30, 2000.
- [14] J.K. Lee, S.R. Ryu and K.Y. Yoo, "Fingerprint based remote user authentication scheme using smart cards," IEE Electronics Letters, vol. 38, no. 12, pp. 554-555, 2002.
- [15] U. Uludag and A.K. Jain, "Multimedia content protection via biometrics-based encryption," Proc. of Intl. Conf. on Multimedia and Expo (ICME), vol. 3, pp. 237-240, July 2003.
- [16] B.T. Tsieh, H.T. Yeh, H.M. Sun and C.T. Lin, "Cryptanalysis of a Fingerprint-based Remote User Authentication Scheme Using Smart Cards," Proc. of 37th Annual Intl. Carnahan Conf. on Security Technology, pp. 349-350, Oct. 2003.
- [17] J. Kittler, M. Hatef, R.P.W. Duin and J. Matias, "On Combining Classifiers," IEEE Trans. on Pattern Anal. And Mach. Intelligence, vol. 20, no. 3, pp. 226-239, March 1998.
- [18] Tony Mansfield, Gavin Kelly, David Chandler and Jan Kane, "Biometric Product Testing Final Report", CESG contract X92A/4009309, Issue 1.0, 19 March 2001. www.eatesam.com/product/pdf/uk_report.pdf
- [19] T. Sim, S. Baker and M. Bsat, "The CMU Pose, Illumination and Expression (PIE) database," Proc. of the 5-th International Conference on Automatic Face and Gesture Recognition, 2002

Vinod Chandran (S'85–M'90–SM'01) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1982, the M.S. degree in electrical engineering from Texas Tech University, Lubbock, in 1985, and the Ph.D. degree in electrical and computer engineering and the M.S. degree in computer science from Washington State University, Pullman, WA, in 1990 and 1991, respectively.

He is currently an Associate Professor at the Queensland University of Technology, Brisbane, Australia, in the School of Engineering Systems. His research interests include pattern recognition, higher order spectral analysis, speech processing, and image processing.

Associate Professor Chandran is a Senior member of the Institute of Electrical and Electronic Engineers (IEEE), USA, and a member of the Association for Computing Machinery.

Anthony Nguyen (S'00–M'05) received the B.Eng. (Aerospace Avionics) degree with first class honours in 1999, and a Ph.D. degree in the area of image processing in 2005 from Queensland University of Technology (QUT), Brisbane, Australia.

He is currently appointed as a Research Fellow within the Image and Video Research Laboratory within the School of Engineering Systems at QUT. He is also a tutor and a telecoms laboratory development coordinator for the digital communications and wireless communications units offered at Queensland University of Technology. His research interests include image processing, image compression, and pattern recognition.

Dr. Nguyen is also a member of the Institute of Electrical and Electronic Engineers (IEEE), USA, and the Australian Pattern Recognition Society (APRS).