Thursday May 11

9:20 - 10:20 Invited conference #1 Chair: J.-C. Junqua

Multi-Modal Biometrics: Orthogonal, Independent, and Collaborative
Kevin Bowyer

The topic of multi-modal biometrics has attracted great interest in recent years. This talk will categorize different approaches to multi-modal biometrics based on the biometric sources, the type(s) of sensing used, and the depth of collaborative interaction in the processing. By "biometric source" we mean the property of the person that is used for identification, such as fingerprint, voice, face appearance or iris texture. By type of sensing we mean different sensor modalities, such as 2D, 3D, or infra-red. By collaboration we mean the degree to which the processing of one biometric is influenced by the results of processing other biometrics. One common category of multi-modal biometrics might be called orthogonal. In this category, the biometric sources are different, such as face plus fingerprint used as a multi-modal biometric or a multi-biometric. In this category, there appears to be little or no opportunity for interaction between the processing of the individual biometrics. Another common category of multi-modal biometrics might be called independent. This type of processing is common with different modalities of sensing the face. For example, the 2D image of the face and the 3D shape of the face might be processed independently as biometrics, and then two results combined at a score or rank level. A less common category of multi-modal biometrics might be called collaborative. In this category, the processing of each individual biometric may be influenced by the other biometrics. For example, if specular highlights are found in the 2D face image, this might inform the processing of the 3D shape of the face, since specular highlights in the 2D often result in artifacts in the 3D shape. It is argued that the area of collaborative processing among multi-modal biometrics, although relatively less explored, holds the potential for important gains in accuracy.

10:45 - 12:15 Oral session #1 Chair: K. Bowyer

10:45 Human identification using heart sound
K. Phua, T. H. Dat, J. Chen, L. Shue

In this paper, we investigate the possibility of using heart sound as a biometric for human identification. The most significant contribution of using heart sound as a biometric is that it cannot be easily simulated or replicated as compared to other conventional biometrics. Our approach consists of a robust feature extraction scheme which is based on cepstral analysis with a specified configuration, combined with Gaussian mixture modeling. Various experiments have been conducted to determine the relationship between various parameters in our proposed scheme. The results suggest that parameter values appropriate for heart sounds should be significantly different for equivalent parameters used in conventional cepstral analysis for speech processing. In particular, heart sounds should be processed within segments of 0.5 second and using the full resolution in frequency domain. Secondly, higher order cepstral coefficients, carrying information on the excitation, are also useful. Preliminary results indicate that with well-chosen parameters, an identification rate of up to 96% is achievable for a database consisting of 7 individuals, with heart sounds collected over a period of 2 months.

11:15 2D-3D Hybrid Face Recognition Based on PCA Feature Modelling
C. McCool, V. Chandran, S. Sridharan

Hybrid face recognition, using image (2D) and structural (3D) information, has explored the fusion of Nearest Neighbour classifiers. This paper examines the

effectiveness of feature modelling for each individual modality, 2D and 3D. Furthermore, it is demonstrated that the fusion of feature modelling techniques for the 2D and 3D modalities yields performance improvements over the individual classifiers. By fusing the feature modelling classifiers for each modality with equal weights the average Equal Error Rate improves from 12.60% for the 2D classifier and 12.10% for the 3D classifier to 7.38% for the Hybrid 2D+3D classifier.

11:45 Gender and Identity Classification for a Naive and Evolving System
M. Castrillon-Santana, O. Déniz-Suarez, J. Lorenzo-Navarro,M. Hernandez-Tejera

This paper does not propose a new technique for face representation or classification. Instead the work described here investigates the evolution of an automatic system which, based on a currently common framework, and starting from an empty memory, modifies its classifiers according to experience. In the experiments we reproduce up to a certain extent the process of successive meetings. The results achieved, even when the number of different individuals is still reduced compared to off-line classifiers, are promising.

13:45 – 14:45 Invited conference #2 Chair: C.Sénac

Government initiatives concerning biometric authentication and their impact on the industry and international standards
Bill Perry

The international community has now been working on delivering the first ePassports for some three years. During that time, standards have been developed in parallel to countries delivering a product, causing some consternation and difficulties. However, the industry and governments have stood strong, developing ISO 19794 standards for biometrics, ICAO standards for ePassports including Passive and Basic Access Control structures. This has been the long awaited boost biometrics has been waiting for to a) raise its profile internationally and b) generate investment interest opportunities that can be realized by universities, research labs and the industry as a whole. This talk will discuss the above in depth with significant reference to large scale projects akin ePassports, EU SIS II, EU BMS, Identity Cards, etc., and their impact on our livelihoods.

14:45 – 15:45 Oral session #2 Chair: B. Perry

14:45 Revisiting Doddington's Zoo: A Systematic Method to Assess User-dependent Variabilities
N. Poh, S. Bengio, A. Ross

A systematic analysis of user-dependent performance variability in the context of automatic speaker verification was first studied by Doddington et al (1998). Different categories of users were identified and labeled as sheep, goats, lambs and wolves. While this categorization is significant, it does not provide a criterion to rank the users in a database based on their variability in performance. In this work we design and evaluate a user-dependent performance criterion that requires only a limited number of client (i.e., genuine) training scores. We then extend such a study to formulate a user-specific score normalization scheme (a variant of the classical F-norm) and show that user-dependent variabilities can be reduced by employing such a scheme. The results of 13 experiments confirm the efficacy of the proposed scheme.

15:15 Multimodal Speaker Authentication – Evaluation of Recognition Performance of Watermarked References
C. Vielhauer, T. Scheidat, A. Lang, M. Schott, J. Dittmann, T.K. Basu, P.K. Dutta

In this paper a new approach for combining biometric authentication and digital watermarking is presented. A digital audio watermark method is used to embed metadata into the reference data of biometric speaker recognition. Metadata in our context may consist of feature template representations complementary to the speech modality, such as iris codes or biometric hashes, ancillary information about the social, cultural or biological context of the originator of the biometric data or technical details of the sensor. We suggest a well-known watermark embedding technique based on LSB (least significant bit) modulation for this purpose; perform experiments based on a database taken from 33 subjects and 5 different utterances and a known cepstrum based speaker recognition algorithm in verification mode. The goal is to perform a first evaluation of the recognition precision for our selected algorithm. The first tests show that the recognition precision is not significantly deteriorated by the embedding of the information, as in three out of five cases, no degradation was observed at all and in the worst case the relative increase in false recognition was limited to 1%.

16:00 – 17:30 Poster session #1

The Multi-biometric, Multi-device and Multilingual (M3) Corpus
H. Meng, P.C. Ching, T. Lee, M. Wai. Mak, B. Mak, Y.S. Moon, M.H. Siu, X. Tang, H. P.S. Hui, A. Lee, W.-K. Lo, B. Ma, E. K.T. Sio

This paper presents an overview of the M3 (multi-biometric ,multi-device and multilingual) Corpus. M3 aims to support research in multi-biometric technologies for pervasive computing using mobile devices. The corpus includes three biometrics – facial images, speech and fingerprints; three devices – the desktop PC with plug-in microphone and webcam, Pocket PC and 3G phone; as well as three languages of geographical relevance in Hong Kong –Cantonese,Putonghua and English. The multimodal user interface can readily extend from desktop computers to mobile handhelds and smart phones which have small form factors. Multimodal biometric authentication can also leverage the mutual complementarity among modalities,which is particularly useful in dynamic environmental conditions encountered in pervasive computing. For example, we should emphasize facial images over speech when verification is performed in noisy acoustic environments. M3 is designed to include variable environmental factors indoors and outdoors, simultaneous recordings across multiple devices to support comparative and contrastive investigations, bilingual text prompts to elicit both application-oriented and cognitive speech data, as well as multi-session data from a fairly large set of subjects.

Discriminant approaches for GMM speaker recognition
A. Preti, N. Scheffer, J.-F. Bonastre

This paper presents some experiments on discriminative training for GMM/UBM speaker recognition. We propose two MMIE adaptation methods for GMM component weights suitable for speaker recognition. The impact on performance of this training method is compared to the standard weight estimation/adaptation criterion, MLE and MAP on standard GMM systems and on SVM systems. The results enforce the difficulty to introduce discriminative behaviour in GMM where as it is inherent in SVM systems.

A Max Kernel For Text-Independent Speaker Verification Systems
J. Mariéthoz, S. Bengio

In this paper, we present a principled SVM based speaker verification system. A general approach to compute two sequences of frames is developed that enables the use of any kernel at the frame level. An extension of this approach using the Max operator is then proposed. The new system is compared to state-of-the-art GMM and

other SVM based systems found in the literature on the Polyvar database. The new system outperforms, most of the time, the other systems, statistically significantly.

Biometric Fusion: Robust Approach
O. Ushmaev, S. Novikov

We have developed biometric fusion technique based on stochastic theory. The suggested method is a robust adaptation of the Neyman-Pearson technique to the specifics of biometrics. The method makes possible to achieve almost optimal performance as measured by the ROC curve.

The BioSecure Talking-Face Reference System
H. Bredin, G. Aversano, C. Mokbel, G. Chollet

In the framework of the BioSecure Network of Excellence (http://www.biosecure.info), a talking-face identity verification reference system was developed: it is open-source and made of replaceable modules. This is an extension of the BECARS speaker verification toolkit, implementing the GMM approach. In this paper, the audio and visual features extraction front-ends are presented. The performances of the system on the Pooled protocol of the BANCA database are described.

17h45 SOCIAL EVENT: Garonne cruise and dinner

Friday May 12

9:00 – 10:00 Invited conference #3 Chair: J.L. Dugelay

Fusing Noncontact Biometrics: Face, Gait and Ear
Mark Nixon

This talk will describe progress in multimodal biometric fusion and in the newer vision-based biometrics, ear and gait. The current interest in multimodal biometric fusion arises not only from ability to improve enrolment, but also from the potential performance advantage achieved by fusing complementary technologies. Currently there are two portal systems in development aiming to provide fusion based access control, one using face and iris and the other using face, gait and ear. Gait is attractive as a biometric since it concerns motion, is available at a distance and can cue acquisition of other biometrics. Ears are attractive as a biometric since their structure changes little with age and their covariate structure is considerably less complex than other approaches. As such, this talk will encompass two of the newer vision-based biometrics and whilst describing progress, also assess suitability for deployment in a multimodal biometric system.

10: 00 – 10: 15 Introduction to the demo session I. Ferrané

10:30 – 12:15 Poster session #2

Audio/Video Fusion: a Preprocessing Step for Multimodal Person Identification
G. Jaffre, J. Pinquier

In the audiovisual indexing context, we propose a system for automatic association of voices and images. This association can be used as a preprocessing step for existing applications like person identification systems. We use a fusion of audio and video indexes (without any prior knowledge) in order to make the information brought by each of them more robust. If both audio and video indexes are correctly segmented, this automatic association yields excellent results. In order to deal with over segmentation, we propose an approach which uses one index to improve the segmentation of the other. We show that the use of the audio index improves an over segmented video index on a corpus composed of French TV broadcasts.

On face image quality measures
K. Kryszczuk, A. Drygajlo

The classification reliability is an essential problem in biometric verification systems. In the presence of the possible mismatch between system's training and testing conditions, a measure of that mismatch is necessary in order to estimate the degree of trust one can have verification from plain image quality, and thus create a condition mismatch. In this paper we identify the most commonly encountered causes of face image degradation due to the recording conditions, and propose corresponding quality measure methods. We show on publicly available databases that proposed methods can be effectively used to estimate the verification reliability using a probabilistic framework.

Enabling EBGM Face Authentication on mobile devices
Y.S. Moon, K. H. Pun, K.C.Chan, M.F. Wong, T.H.Yu

This paper presents a systematic optimization strategy to implement the complicated EBGM face recognition on low processing power mobile devices We propose a tailor-made fixed point arithmetic and various memory access optimization techniques to speed up an EBGM authentication process from ~550s down to ~1s in a typical Intel PXA255 400 MHz mobile platform. The result shows that real time face recognition can be completed without any noticeable accuracy loss. This finding not only provides a guideline for porting various biometric applications in the mobile systems, but also exploits new opportunities for different mobile e-commerce applications with biometric identity checking.

Effectiveness of LP Based Features for Identification of Professional Mimics in Indian Languages
H. A. Patil, P. K. Dutta, T. K. Basu

Automatic Speaker Recognition (ASR) is an economic tool for voice biometrics because of availability of low cost and powerful processors. For an ASR system to be successful in practical environments, it must have high mimic resistance, i.e., the system should not be defeated by determined mimics which may be either identical twins or professional mimics. In this paper, we demonstrate the effectiveness of Linear Prediction (LP) based features viz. Linear Prediction Coefficients (LPC) and Linear Prediction Cepstral Coefficients (LPCC) over filterbank based features such as Mel-Frequency Cepstral Coefficients (MFCC) and newly proposed Teager energy based MFCC (T-MFCC) for the identification of professional mimics in Marathi and Hindi languages.

Multi-modal biometric authentication on the SecurePhone PDA
J. Koreman, A.C. Morris, D. Wu, S. Jassim, H. Sellahewa, J. Ehlers, G. Chollet, G. Aversano, H. Bredin, S. Garcia-Salicetti, L. Allano, B. Ly Van, B. Dorizzi

We present an overview of the development of the Secure Phone mobile communication system in which multimodal biometric authentication gives access to the system's built-in e-signing facilities, enabling users to deal m-contracts using a mobile call in an easy yet secure and dependable way. Authentication uses an neutral biometrics: the user reads a prompt into a camera and microphone, and signs on a touch screen. The state of the art techniques used for each biometric modality were initially developed using the benchmark databases BANCA (audio-visual) and BIOMET (signature). A suitable PDA was then selected and a multimodal database was recorded on the device itself. Several fusion techniques were tested for biometric evidence combination. Best performance achieved for voice, face, signature and fused modalities was 2.3, 17.3, 4.3 and 0.6% EER for BANCA/BIOMET and 3.2, 27.6, 8.0 and 0.8% EER for the PDA database.

Frame purification for cluster comparison in speaker diarization
X. Anguera, C. Wooters, J. Hernando

Speaker diarization is often performed as a first step to speaker or speech recognition systems, which work better when the input signal is split into its speakers. When performing speaker diarization, it is common to use an agglomerative clustering approach in which the acoustic data is first split in small pieces and then pairs are merged until reaching a stopping point. The speaker clusters often contain non-speech frames that jeopardize discrimination between speakers, creating problems when deciding which two clusters to merge and when to stop the clustering. In this paper, we present one algorithm that aims to purify the clusters, eliminating the non-discriminant frames –selected using a likelihood-based metric– when comparing two clusters. We show improvements of over 15.5% relative using three datasets from the most current Rich Transcription (RT) evaluations.

Bi-Modal Face and Speech Authentication: a BioLogin Demonstration System (Demo)
S. Marcel, J. Mariethoz,Y. Rodriguez, F. Cardinaux

This paper presents a bi-modal (face and speech) authentication demonstration system that simulates the loginof a user using its face and its voice. This demonstration is called BioLogin. It runs both on Linux and Windows and the Windows version is freely available for download. Bio-Login is implemented using an open source machine learning library and its machine vision package.

13:45 – 14:45 Discussion "Where are the European projects in Biometric authentication heading?"
Invited participants: Aladdin Ariyaeeinia and Bernadette Dorizzi
Chair: J.F. Bonastre

14:45 – 15:45 Oral session #3 Chair: M. Nixon

14:45 An Examination of Audio-Visual Fused HMMs for Speaker Recognition
D. Dean, T. Wark, S. Sridharan

Fused hidden Markov models (FHMMs) have been shown to work well for the task of audio-visual speaker recognition, but only in an output decision-fusion configuration of both the audio- and video-biased versions of the FHMM structure. This paper looks at the performance of the audio and video-biased versions independently, and shows that the audio-biased version is considerably more capable for speaker recognition. Additionally, this paper shows that by taking advantage of the temporal relationship between the acoustic and visual data, the audio-biased FHMM provides better performance at less processing cost than best-performing output decision-fusion of regular HMMs.

15:15 On the relation between biometric quality and user-dependent score distributions in fingerprint verification
F. Alonso-Fernandez, R. N. J. Veldhuis, A. M. Bazen, J. Fierrez-Aguilar, J. Ortega-Garcia

The lack of robustness against image quality degradation is a open issue in fingerprint verification. It has been found in previous studies that the behavior of a fingerprint verification system may vary depending on the quality of the fingerprints. In this paper, we study the performance for individual users under varying image conditions using a multisensor database acquired with three different fingerprint sensors. We propose a user-dependent score normalization scheme that exploits quality information, reaching an EER improvement of s 15% in one particular sensor. We have also included the proposed score normalization scheme in a multisensor fingerprint verification system that combines the three sensors, obtaining an EER improvement of s 13% in the best case.

16:00 – 17:00 Oral session #4 Chair: B.Dorizzi

16:00 Continuous Biometric Authentication Using Dynamic Bayesian Networks
J. Muncaster, M. Turk

In this paper we address the issue of post-login user verification using biometrics. We

propose a general framework for continuously monitoring a user and his or her characteristics throughout the session in order to provide continuous verification of identity. We present a multimodal approach using dynamic Bayesian networks to account for classification uncertainty and to encode the system's dynamic model. The system supports an extendable number of modalities that can be integrated at both the score level and the decision-level. The system can also be extended to incorporate important contextual information. We provide a simple way for an operator to tailor the dynamic model using known domain knowledge. Initial tests with the system using face recognition and keystroke dynamics are promising, and we expect performance to improve as more modalities are incorporated.

16:30 Fusing Generative and discriminative UBM-BASED systems for speaker verification
N. Scheffer, J.-F. Bonastre

In the past few years, discriminative approaches to perform speaker detection have shown good results and an increasing interest. Among these methods, SVM based systems have lots of advantages, especially their ability to deal with a high dimension feature space. Generative systems such as UBMGMM systems show the greatest performance among other systems in speaker verification tasks. Combination of generative and discriminative approaches is not a new idea and has been studied several times by mapping a whole speech utterance onto a fixed length vector. This paper presents a straight-forward, cost friendly method to combine the two approaches with the use of a UBM model only to drive the experiment. We show that the use of the TFLLR kernel, while closely related to a reduced form of the Fisher mapping, implies a performance that is close to a standard GMM/UBM based speaker detection system. Moreover, we show that a combination of both outperforms the systems taken independently.