Face Authentication Test on the BANCA Database

Kieron Messer, Josef Kittler, Mohammad Sadeghi, Miroslav Hamouz and Alexey Kostin
University of Surrey, Guildford, Surrey, GU2 7XH, UK.

Fabien Cardinaux, Sebastien Marcel, Samy Bengio, Conrad Sanderson,
Norman Poh and Yann Rodriguez
IDIAP Research Institute, CP 592,
rue du Simplon 4, 1920 Martigny, Switzerland.

Jacek Czyz and L. Vandendorpe
Université Catholique de Louvain, Batiment Stevin, Place du Levant 2,
1348 Louvain-la-Neuve, Belgium.

Chris McCool, Scott Lowther, Sridha Sridharan and Vinod Chandran
Queensland University of Technology, Image and Video Research Lab.

Roberto Parades Palacios and Enrique Vidal
Departmento de Sistemas Informáticos y Computación,
Universidad Politécnica de Valencia.

Li Bai, LinLin Shen and Yan Wang
School of Computer Science and IT, University of Nottingham, UK.

Chiang Yueh-Hsuan, Liu Hsien-Chang and Hung Yi-Ping
Department of Computer Science and Information Engineering,
National Taiwan University.

Alexander Heinrichs, Marco Müller, Andreas Tewes,
Christoph von der Malsburg, Rolf Würtz
Institut für Neuroinformatik, Lehrstuhl Systembiophysik,
Ruhr-Universutät Bochum, Germany.

Zhenger Wang, Feng Xue, Yong Ma, Qiong Yang, Chi Fang and Xiaoqing Ding
Department of Electronic Engineering, Tsinghua University,
Beijing 100084, P.R. China.

Simon Lucey, Ralph Goss and Henry Schneiderman
Carnegie Mellon University.

Abstract

This paper details the results of a Face Authentication Test (FAT2004) [5] held in conjunction with the 17th International Conference on Pattern Recognition. The contest was held on the publicly available BANCA database [1] according to a defined protocol [7]. The competition also had a sequestered part in which institutions had to submit their algorithms for independent testing. 13 different verification algorithms from 10 institutions submitted results. Also, a standard set of face recognition software packages from the Internet [2] were used to provide a baseline performance measure.

IEEE COMPUTER SOCIETY

1. Introduction

In recent years the cost and size of biometric sensors and processing engines has fallen, a growing trend towards e-commerce, teleworking and e-banking has emerged and people's attitude to security since September 11th has shifted. For these reasons there has been a rapid increase in the use of biometric technology in a range of different applications. Many of these systems are based on the analysis of face images as they are non-intrusive and user-friendly. Moreover, personal identity can be ascertained without the client's assistance.

Face recognition technology is still developing and many papers on new face verification and recognition algorithms are being published almost daily. However, direct comparison of the reported methods can be difficult because tests are performed on different data with large variations in test and model database sizes, sensors, viewing conditions, illumination and background. Typically, it is unclear which methods are the best and for which scenarios they should be used. Evaluation protocols can help alleviate this problem.

Typically, an evaluation protocol defines a set of data, how it should be used by a system to perform a set of experiments and how the performance of the system should be quantified [21]. The protocol should be designed in such a manner that no bias in the performance is introduced, e.g. the training data is not used for testing. It should also represent a realistic operating scenario. As different scenarios normally require different protocols, no single protocol will be able to cover all scenarios.

Over the past few years standard datasets for testing face authentication systems have become available, e.g. Yale [37], Harvard [34], Olivetti [36], M2VTS [35], ([3] gives a more comprehensive list). However, for many of them no associated protocol has been defined. Experiments carried out by different organisations on these datasets will divide the data into different test and training sets and consequentially they measure performance differently.

The FERET database has defined a protocol for face identification and face verification [23]. However, only a development set of images from the database are released to researchers. The remaining are sequestered by the organisers to allow independent testing of the algorithms. To date three evaluations have taken place, the last one in the year 2000 [22].

More recently, two Face Recognition Vendor Tests [4] have been carried out, the first in 2000 and the second in 2002. The tests are done under supervision and have time restrictions placed on how quickly the algorithms should compute the results. They are aimed more at independently testing the performance of commercially available systems, however academic institutions are also able to take part. In the more recent test 10 commercial systems were evaluated.

In the year 2000 a competition on the XM2VTS database using the Lausanne protocol [19] was organised [18]. As part of AVBPA 2003 a second competition on exactly the same data and testing protocol was organised [11]. All the data from the Xm2vts database is available from [6]. We believe that this open approach increases, in the long term, the number of algorithms that will be tested on the XM2VTS database. Each research institution is able to assess their algorithmic performance at any time.

In this paper we detail a competition on a new database known as the BANCA database [7]. The database was captured under 3 different realistic and challenging operating scenarios. Several protocols have also been defined which specifies which data should be used for training and testing. Recently, on a competition on the Match Controlled protocol was held in conjunction with the International Conference on Biometric Authentication [12]. The BANCA database is being made available to the research community through [1].

The rest of this paper is organised as follows. In the next section the BANCA database is detailed. Next the competition rules and performance criterion are described. Section 4 gives an overview of each algorithm which entered the competition and in the following section the results are detailed. Finally, some conclusions are made.

2 The BANCA database

The BANCA database contains 52 subjects (26 males and 26 females). Each subject participated to 12 recording sessions in different conditions and with different cameras. Sessions 1-4 contain data under *Controlled* conditions while sessions 5-8 and 9-12 contain *Degraded* and *Adverse* scenarios respectively. Each session contains two recordings per subject, a true client access and an informed impostor attack. For the face image database, 5 frontal face images have been extracted from each video recording, which are supposed to be used as client images and 5 impostor ones. In order to create more independent experiments, images in each session have been divided into two groups ($G1$ and $G2$) of 26 subjects, 13 males and 13 females. Fig. 1 shows a few examples of the face data.

In the BANCA protocol, 7 different distinct experimental configurations have been specified, namely, Matched Controlled (MC), Matched Degraded (MD), Mat-

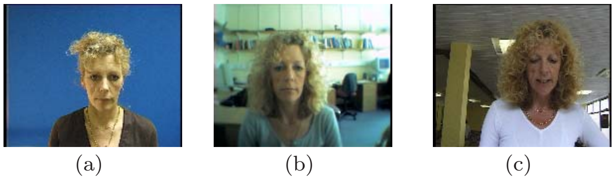

Figure 1. Examples of the database images. *a:* **Controlled,** *b:* **Degraded and** *c:* **Adverse scenarios.**

ched Adverse (MA), Unmatched Degraded (UD), Unmatched Adverse (UA), Pooled test (P) and Grand test (G). Table 1 describes the usage of the different sessions in each configuration. "T" refers to the client training while "C" and "I" depict client and impostor test sessions respectively. The decision function can be trained using only 5 client images per person from the same group and all client images from the other group. More details about the database and experimental protocols can be found in [7].

	MC	MD	MA	UD	UA	P	G
1	TI			T	T	TI	TI
2	CI					CI	CI
3	CI					CI	CI
4	CI					CI	CI
5		TI		I		I	TI
6		CI		CI		CI	CI
7		CI		CI		CI	CI
8		CI		CI		CI	CI
9			TI		I	I	TI
10			CI		CI	CI	CI
11			CI		CI	CI	CI
12			CI		CI	CI	CI

Table 1. The usage of the different sessions in the BANCA experimental configurations.

2.1 The Sequestered Data

The sequestered BANCA data set consists of 21 of the original 52 database subjects, 11 men and 10 women. It was captured over 24 months after the original recordings. Five images of each client were captured under two scenarios, controlled and degraded. Also, images of 22 non BANCA subjects were captured under the two conditions. These images were used to simulate impostor attacks. The cameras used to capture the controlled scenario used was high quality whilst the camera used for the degraded was a cheap web-cam. However, they were both different to those used in the original recordings.

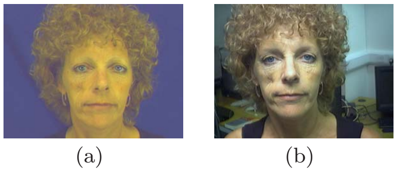

Figure 2. Examples of the sequestered database images. *a:* **Controlled,** *b:* **Degraded**

The images shown in figure 2 show some example images for subject 1008 from the sequestered data set.

3. The Competition

All the experiments were carried out according to the Pooled (P) configuration of the BANCA database. This configuration is the most challenging of the BANCA protocol. The training phase of every client model is done using only the 5 images from the Controlled Scenario, i.e. session 1 (en_video_sc1_1). For example, the training of the client 1001 use for training only the 5 images: 1001_f_g1_s01_1001_en_1, 1001_f_g1_s01_1001_en_2, 1001_f_g1_s01_1001_en_3, 1001_f_g1_s01_1001_en_4 and 1001_f_g1_s01_1001_en_5.

The BANCA protocol is an open-set protocol. It is forbidden to use any data from other clients from the same group when training a classifier or generating a PCA/LDA matrix. For example, the training of the client model 1001 from group g1 should not use any images from clients: 1002, 1003, 1004, ..., 1036, 1037 in any way.

The verification tests are then performed across all three scenarios; controlled, degraded and adverse. As no image data from the degraded and adverse conditions has been used in client enrolment this makes BANCA protocol P very challenging. The enviroment, subject pose and camera sensor have all changed.

There were three separate parts to the competition.

Part I: Pre-registered Images were supplied which had already been localised (by hand) and geometrically normalised. The resulting resolution of the images were 55x51 pixels, 8-bit grey-scale and contained just face pixels.

Part II: Automatic Full video resolution colour images as shown in figure 1(a) were supplied. All participants had to use an automatic method of localisation for the at least the test phase of the protocol. Manual localisation for the training and evaluation phases was allowed.

Part III: Sequestered In this part of the competition the algorithms were tested on the sequestered data. The entrants had to submit their trained face recognition system to the University of Surrey (UniS). The system had to input a claimed BANCA client id and test image. The system then output either a one for client acceptance or zero for impostor attack.

Part I of the competition allows us to assess the underlying performance of the core face verification technology as the images had all been localised and geometrically normalised by the same method. Part II of the competition was aimed at testing the complete verification system, including the detection and localisation stage. Part III allows an independent verification of the algorithms.

It was upto to the competing institution to specify which parts of the test they wished to enter.

To assess the algorithmic performance the *False Rejection Rate* P_{FR} and *False Acceptance Rate* P_{FA} are typically used. These two measures are directly related, i.e. decreasing the false rejection rate will increase the number of false acceptances. The point at which $P_{FR} = P_{FA}$ is known as the EER (Equal Error Rate).

For parts I and II of the competition we requested that the entrants submit their results for 3 specific operating conditions which corresponded to 3 different values of the Cost Ratio $R = C_{FA}/C_{FR}$, namely $R = 0.1, R = 1, R = 10$. Assuming equal *a priori* probabilities of genuine clients and impostor, these situations correspond to 3 quite distinct cases:

$R = 0.1$	→	FA is an order of magnitude less harmful than FR,
$R = 1$	→	FA and FR are equally harmful,
$R = 10$	→	FA is an order of magnitude more harmful than FR.

The entrants were asked to submit the Weighted Error Rate (WER) for the test data of groups $G1$ and $G2$ at the three different values of R. WER is defined as:

$$WER(R) = \frac{P_{FR} + R\, P_{FA}}{1 + R} \,. \tag{1}$$

For each group and at each operating point there are in total 1170 true client claims and 1560 impostor attacks.

For part III of the test the half-total error rate is reported (HTER) which is simply the average of the P_{FA} and P_{FR}. In these sequestered tests there were 210 true client claims (i.e. 21 clients x 5 images x 2 conditions). Also, the 22 non BANCA subjects were used to impost each of the 21 clients according to same sex, making a total there were 2320 impostor tests.

4. Overview of Algorithms

In this section the algorithms that participated in the contest are summarised. Also, we downloaded a standard set of face recognition software from the Internet [2] to provide a baseline performance measure on this database. Due to space limitations we have published the results from these experiments at [5].

4.1. IDIAP Research Institute

Pseudo-2D HMM (IDIAP - HMM)

The system is comprised of two main parts: an automatic face locator and a local feature probabilistic classifier. To locate faces, a fast cascade of boosted Haar-like features is applied to the integral image to detect potential faces [32], followed by post-processing using a Multi-Layer Perceptron [25] to provide the final localized face. The probabilistic classifier uses DCTmod2 features [27] and models faces using pseudo-2D Hidden Markov Models (HMMs) [8]. In DCTmod2 feature extraction, each given face is analyzed on a block by block basis; from each block a subset of Discrete Cosine Transform (DCT) coefficients is obtained; coefficients which are most affected by illumination direction changes are replaced with their respective horizontal and vertical deltas, computed as differences between coefficients from neighbouring blocks. For the pseudo-2D HMM topology, we use a top-to-bottom main HMM with each state being modelled by a left-to-right HMM. To circumvent the problem of small amount of client training data, parameters for each client model are obtained via Maximum *a Posteriori* (MAP) adaptation of a generic face HMM; the generic face HMM is trained using the Expectation Maximisation algorithm, using world model training data. A score for a given face is found by taking the difference between the log-likelihood of the face belonging to the true client and the log-likelihood of the face belonging to an impostor; a global threshold (i.e. the same for all clients) is used in making the final verification decision.

Fusion (IDIAP - Fusion)

The system is composed of an automatic face locator, three classification subsystems and a fusion stage. The face locator has two components: a fast cascade of boosted haar-like features is applied to the integral image to detect potential faces [32], followed by post-processing using a Multi-Layer Perceptron (MLP) [25]. The first two classification subsystems are based on local features and generative models (namely, DCTmod2 features, Gaussian Mixture Models & pseudo-2D Hidden Markov Models [8]), while the third subsystem uses Linear Discriminant Analysis based feature extraction (i.e. holistic features) and a MLP for classification [17]. Finally, the opinions of the three subsystems are fused using an MLP based approach [24].; a global threshold (i.e. the same for all clients) is used in making the final verification decision.

4.2. Queensland University of Technology (QUT)

The feature extraction technique uses Principal Component Analysis (PCA) on grayscale and chrominance images. Intra- and Inter-Class variation is then modelled by Gaussian Mixture Models (GMMs). Automatic localisation performed through a fusion of PCA reconstruction error and Intra-Class probabilities. This is optimised through use of a multi-scale coarse to fine search algorithm making the assumption of an upright face with limited rotation.

4.3. Universidad Politécnica de Valencia (UPV)

This method is based on Learning Prototypes and Distances, LPD [20]. LPD is a prototype reduction algorithm which simultaneous train both a *reduced* set of prototypes and a suitable *local metric* for *these* prototypes. Starting with an initial selection of a small number of prototypes, it iteratively adjusts both the position (features) of these prototypes and the corresponding local-metric weights. The resulting prototypes/metric combination minimises a suitable estimation of the 1-Nearest Neighbour classification error probability.

In the present work the prototypes are 55×51 feature vectors representing the image space. Each of these vectors have a suitable weighted distance associated. We use only one prototype for each class, this prototype is randomly selected from the training set. The LPD algorithm "moves" each prototype and "learns" a suitable local distances minimising an index that approximate the 1-Nearest Neighbour error.

The training set are the original pre-registered images and the mirror version. To compute the score of a new test image we use the inverse of the local distance from this new test vector to the prototype vector that represents each class. This distance is normalised between 0 and 1 using the distance to the other classes.

4.4. University of Nottingham

Images are processed using a sequence of Gabor filters. A set of discriminative features is extracted from each image. This enables an image to be represented as a Gabor feature vector and subjected to subspace projection for dimension reduction. Support Vector Machines (SVMs) are trained for each person in the BANCA image set as given in the protocol. A person is then verified against the trained SVMs associated with the claimed identification. A face detector capable of identifying eye features is used to perform the tasks specified in Part II and Part III of the competition.

4.5. National University of Taiwan

In this method Linear Discriminant Analysis forms the basis of this technique. The images are split into left images and right images in order to simplify the lighting condition. Different lighting conditions models are combined to accomplish higher performance.

4.6. University of Surrey (UniS)

The input image data is firstly projected into the fisher faces space using the Linear Discriminant Analysis. The Isotropic Gradient Direction metric [26] is then used as the scoring function which measures the degree of similarity between the test image and the claimed identity template. For the first part of the competition only the intensity image was used to comply with the competition rules. For parts II and III of the competition this process was performed in three different colour spaces namely intensity (I), chroma-g (G/I) and opponent chroma-rg ((R-G)/I) spaces [15]. The final score is then calculated by averaging the individual channel scores. The resulting score is finally compared to a pre-set threshold in order to decide whether the claim is genuine or impostor. We have used the XM2VTS database for the LDA training, the histogram equalisation for the photometric normalisation and client specific thresholding method for calculating the thresholds.

4.7. Université Catholique de Louvain (UCL)

Linear Discriminant Analysis (UCL-LDA)

The method is based on classical Linear Discriminant Analysis (LDA) or fisherfaces. The matching score is computed in the LDA subspace using normalised correlation. A large auxiliary dataset is used to compute the LDA basis. Note that instead of using original image $I(x, y)$, we take advantage of face symmetry and use the *symmetrised* image $I_s = (I(x, y) + I(-x, y))/2$ [10].

Fusion (UCL - Fusion)
The method is based on a combination of two different face experts. The first expert is based on classical Linear Discriminant Analysis (LDA) or fisherfaces. The matching score is computed in the LDA subspace using normalised correlation. The second expert uses a SVM classifier with linear kernel trained directly in the image space. The two expert scores are conciliated by a fusion module based on a Support Vector Classifier [9].

4.8. Institut für Neuroinformatik

A new method was applied, which is based on the elastic bunch graph matching algorithm described in [33].

The main difference in the new method is the calculation of the similarities of model graphs. In the "classical" approach, the normalized scalar-product between the jets of the nodes is calculated. Now, a statistical model based on a PCA of a given dataset is applied on the jets and the resulting vectors are compared. A publication of the details of this method and the improvements is underway.

The new method has been tested on several datasets and delivered better results than bunch graph matching. However, in this context only the first part of the contest is performed with the new method. Since the results of the classical approach were better on the second part, we decided to use the old approach in the second and third part of the contest.

4.9. Tsinghua University

The method can be broken down into the following steps.

1. Image data set: A large face-image database (including images from FERET, CMUPie, ARData etc.) and the world model of BANCA are used to generate face models, and the development set of BANCA are used to adjust parameters such as thresholds etc..

2. Face detection and eye localisation: Both use appearance-based methods.

3. Normalisation: The face is scaled and rotated such that eyes lie at predefined positions, and the illumination is also normalised by the histogram equalisation method.

4. Feature extraction: A family of Gabor kernels is used to extract features, and then the subspace LDA method is applied to improve the discriminant ability of these features.

5. Matching score: The nearest neighbour methodology is applied to calculate the similarity between the input image and the five training samples, and the normalised correlation is used as the measure of similarity.

6. Classifier combination: The holistic matching method and the component-based matching method are combined to improve the performance of our system.

4.10. Carnegie Mellon Institute (CMU)

The CMU technique [16] employs two representations of the face namely a monolithic and parts based representation. We verify the monolithic representation of the face using a linear discriminant analysis (LDA) based subspace representation and a probabilistic cohort distance measure. The parts based representation of the face was verified by first decomposing the image into N overlapping 2D patches. These patches are then compacted using a modified 2D DCT from which a Gaussian Mixture Model (GMM) is then estimated for each client during enrolment. The GMM estimation process employs MAP adaptation to make the most of the enrolment observations made available. During evaluation the likelihood scores from the GMM are also normalised based on world model of previously seen cohorts. Finally the a posteriori probability estimates from both the monolithic and parts representations are combined using a simple weighted sum rule.

Before verification we pre-process the images for illumination variation using the Gross and Brajovic [14] illumination normalisation module, based on a technique for estimation of the illumination field. For Part II of our submission we employed the Schneiderman et. al [28] [29] parts-based object detection module that detects the face area and eyes.

4.11. National University of Singapore (NUS)

NUS entered two alogorithms to part III of the competition. Unfortunately, the results for one of the algorithms (NUS-1NN) were not ready in time for publication of this paper. However, the results will appear on the competition web-page [5]. Both algorithms use image synthesis to verify face images. It is an extension of ARENA [31]. During the training stage, we synthesize images to augment our training set. These images can be synthesized with simple geometric transformations (i.e., translation, rotation and scaling). In order to cope with different illuminations, we subtract the best-fit brightness plane [30] then reduce the input image to 56 by 64 size.

NUS - 1NN

The first algorithm uses the 1-nearest neighbor and L_p distance measure. The norm is defined as $L_p(a)(|a_i|^p)^{1/p}$. In this competition we use $p = 0.5$. Given the augmented training set, we compute the distance between the input image and images in our training set. When the nearest neighbor is the claimed identity, our algorithm returns true (1), otherwise false (0).

NUS - FKT

The second algorithm uses Fukunaga Koontz Transform (FKT) [13] to separate one person from others. FKT decomposes the whole space into subspaces by choosing a subset of features. In the subspace, the dominant eigenvector of the person with claimed identity is the weakest eigenvector of others, and vice versa. Therefore by projecting the new input image into the subspace, we can verify whether the input image is with the claimed identity or not.

5. Results and Discussion

Table 2 shows the results for the pre-registered part of the competition. They show the WER for each group at the three operating points specified in Section 3. The last column shows the average WER across all six test conditions. The best performing algorithm was from Tsinghua University which achieved an average WER of 1.39%. Second was the University of Nottingham with 3.33%. It is interesting to note that both these methods rely upon Gabor filters for their feature extraction.

Table 3 shows the results using automatic registration for at least the verification stage. Again, the best two performing algorithms are from Tsinghua University (2.21%) and the University of Nottingham (4.58%). What is clear from these results is that accurate localisation is critical to verification performance and still needs to improve to match the performance provided by a manual operator.

The results to part III of the competition are reported in table 4. The best performing algorithm is again from Tsinghua University. However, the result is 5 times worse than achieved for part II of the competition, 13.47%. This trend is seen across all institutions who entered part III of the competition.

Several reasons can explain this drop in performance.

Over tuning The face models and algorithm parameters were over tuned to work on the data available to the institution to ensure the optimal performance on that data. Thus losing generality of the system.

Incorrect threshold The FA and FR rates are unequal. As the images were taken in a new environment and with different sensors the values of the matching scores shifted. The pre-selected thresholds were no longer valid.

Ageing The images of the subjects were taken over 24 months after the images used for the enrolment.

6. Conclusions

This paper presents a comparison of face verification algorithms on a new publicly available and challenging face database. The competition included a sequestered part where institutions submitted their algorithms for independent testing. It was organised in conjunction with the 17th International Conference on Pattern Recognition. 13 different verification algorithms from 10 institutions entered the competition.

Some of the results on part I and part II of the competition were very impressive on this challenging protocol. Eventhough the enviroment, subject pose and camera sensor had all changed the algorithms still reached a high level of performance. However, the results on part III still demonstrated that there is a level of overtuning algorithms to the data. More research is required to robustify these techniques to previously unseen situations.

Table 2: The Weighted Error Rates on the two groups at the three different operating points using the pre-registered images.

	$R = 0.1(WER)$		$R = 1(WER)$		$R = 10(WER)$		
	$G1$	$G2$	$G1$	$G2$	$G1$	$G2$	Av
IDIAP- HMM	8.69	8.15	25.43	20.25	8.84	6.24	12.93
IDIAP - FUSION	8.15	7.43	21.85	16.88	6.94	6.06	11.22
QUT	7.70	8.53	18.08	16.12	6.50	4.83	10.29
UPV	5.82	6.18	12.29	14.56	5.55	4.96	8.23
Univ Nottingham	1.55	1.77	6.67	7.11	1.32	1.58	3.33
National Taiwan Univ	7.56	8.22	21.44	27.13	7.42	11.33	13.85
UniS	4.67	7.22	12.46	13.66	4.82	5.10	7.99
UCL - LDA	8.24	9.49	14.96	16.51	4.80	6.45	10.08
UCL - Fusion	6.05	6.01	12.61	13.84	4.72	4.10	7.89
NeuroInformatik	6.40	6.50	12.10	10.80	6.50	4.30	7.77
Tsinghua Univ	1.13	0.73	2.61	1.85	1.17	0.84	1.39
CMU	5.79	4.75	12.44	11.61	6.61	7.45	8.11

Table 3: The Weighted Error Rates on the two groups at the three different operating points using automatic registration.

	$R = 0.1(WER)$		$R = 1(WER)$		$R = 10(WER)$		
	$G1$	$G2$	$G1$	$G2$	$G1$	$G2$	Av
IDIAP- HMM	8.16	8.57	22.97	18.54	5.91	5.34	11.58
IDIAP - FUSION	7.86	8.68	23.40	17.64	5.79	5.50	11.48
QUT	9.01	8.53	18.52	15.71	6.12	5.51	10.56
Univ Nottingham	3.34	3.20	8.51	7.59	2.51	2.30	4.58
UniS - Fusion	7.92	10.06	16.07	18.00	4.58	5.42	10.34
UCL - LDA	9.77	10.84	20.30	19.55	7.21	6.97	12.44
UCL - Fusion	6.66	8.62	14.00	17.68	5.78	5.21	9.66
NeuroInformatik	6.80	7.40	16.70	16.10	6.70	7.20	10.15
Tsinghua Univ	2.68	1.37	4.07	2.08	1.65	1.41	2.21
CMU	7.72	8.78	26.08	23.05	18.19	9.12	15.49

Table 4: Results of the face verification systems on the sequestered BANCA data.

	FAR	FRR	HTER
IDIAP	3.49	63.81	33.65
UniS	27.46	24.29	25.87
Tsinghua Univ	8.36	18.57	13.47
Univ Nottingham	66.98	35.23	51.11
NUS - FKT	15.00	60.00	37.50
NeuroInformatik	33.70	16.67	25.19

References

[1] *The BANCA Database; http://www.ee.surrey.ac.uk/banca.*

[2] *The CSU Face Identification Evaluation System; http://www.cs.colostate.edu/evalfacerec.*

[3] *The Face Recognition Homepage; http://www.cs.rug.nl/ peterkr/FACE/face.html.*

[4] *Face Recognition Vendor Tests; http://www.frvt.org.*

[5] *The ICPR FAT2004 Homepage; http://www.ee.surrey.ac.uk/banca/icpr2004.*

[6] *The XM2VTSDB Homepage; http://xm2vtsdb.ee.surrey.ac.uk.*

[7] E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, V. Popovici, F. Poree, B. Ruiz, and J. P. Thiran. The BANCA database and evaluation protocol. In *Audio- and Video-Based Biometric Person Authentication: Proceedings of the 4th International Conference, AVBPA 2003*, volume 2688 of *Lecture Notes in Computer Science*, pages 625–638, Berlin, Germany, June 2003. Springer-Verlag.

[8] F. Cardinaux, C. Sanderson, and S. Bengio. Face verification using adapted generative models. In *Proc. Int. Conf. Automatic Face and Gesture Recognition (AFGR), Seoul, Korea.*, 2004.

[9] J Czyz, M Sadeghi, J Kittler, and L Vandendorpe. Decision fusion for face authentication. In *First International Conference on Biometric Authentication*, 2004.

[10] Jacek Czyz. *Decision fusion for identity verification using facial images*. PhD thesis, Universite Catholique de Louvain, 2003.

[11] Kieron Messer et al. Face verification competition on the xm2vts database. In *4th International Conference on Audio and Video Based Biometric Person Authentication*, pages 964–974, June 2003.

[12] Kieron Messer et al. Face verification competition on the banca database. In *First International Conference on Biometric Authentication*, August 2004.

[13] F.Fukunaga and W.Koontz. Applications of the karhunen-loeve expansion to feature selection and ordering. *IEEE Trans. Computers*, 19(5):311–318, 1970.

[14] R Gross and V Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *4th International Conference on Audio and Video Based Biometric Person Authentication*, June 2003.

[15] J. Kittler and M. Sadeghi. Physics-based decorrelation of image data for decision level fusion in face verification. In *The Fifth Workshop on Multiple Classifier Systems (MCS 2004)*, Cagliari, Italy, June, 2004.

[16] Simon Lucey. The symbiotic relationship of parts and monolithic face representations in verification. In *Workshop on Face Processing in Video (FPIV)*, 2004.

[17] S. Marcel. A symmetric transformation for lda-based face verification. In *Proc. Int. Conf. Automatic Face and Gesture Recognition (AFGR), Seoul, Korea*, 2004.

[18] J Matas, M Hamouz, K Jonsson, J Kittler, Y P Li, C Kotropoulos, A Tefas, I Pitas, T Tan, H Yan, F Smeraldi, J Bigun, N Capdevielle, W Gerstner, S Ben-Yacoub, Y Abdeljaoued, and E Mayoraz. Comparison and face verification results on the xm2vts database. In A Sanfeliu, J J Villanueva, M Vanrell, R Alquezar, J Crowley, and Y Shirai, editors, *Proceedings of International Conference on Pattern Recognition, Volume 4*, pages 858–863, 2000.

[19] K Messer, J Matas, J Kittler, J Luettin, and G Maitre. XM2VTSDB: The Extended M2VTS Database. In *Second International Conference on Audio and Video-based Biometric Person Authentication*, March 1999.

[20] R. Paredes and E. Vidal. Learning prototypes and distances (lpd). a prototype reduction technique based on nearest neighbor error minimization. In *International Conference of Pattern Recognition*, August 2004.

[21] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki. An introduction to evaluating biometric systems. *IEEE Computer*, pages 56–63, February 2000.

[22] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi. The feret evaluation methodology for face-recognition algorithms. volume 22, pages 1090–1104, October 2000.

[23] P.J. Phillips, H. Wechsler, J.Huang, and P.J. Rauss. The FERET database and evaluation procedure for face-recognition algorithm. *Image and Vision Computing*, 16:295–306, 1998.

[24] N. Poh and S. Bengio. Non-linear variance reduction techniques in biometric authentication. In *Proc. Workshop on Multi-Modal User Authentication (MMUA),Santa Barabara*, pages 123–130, 2003.

[25] H.A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Tran. Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[26] M. Sadeghi and J. Kittler. Decision making in the lda space: Generalised gradient direction metric. In *The 6th Int. Conf. on Automatic Face and Gesture Recognition*, Seoul, Korea, May, 2004.

[27] C. Sanderson and K.K. Paliwal. Fast features for face authentication under illumination direction changes. *Pattern Recognition Letters*, 24(14):2409–2419, 2003.

[28] H Schneiderman. Feature-centric evaluation for cascaded object detection. In *International Conference on Computer Vision*, 2004.

[29] H Schneiderman. Learning an approximate bayesian network for object detection. In *International Conference on Computer Vision*, 2004.

[30] K Sung. and T. Poggio. Example-based learning for view-based human face detection. *IEEE PAMI*, 20(1):39–51, 1998.

[31] T.Sim, R.Sukthankar, M.Mullin, and S.Baluja. Memory-based face recognition for visitor identification. In *Proceedings of the IEE International Conference on Automatic Face and Gesture Recognition*, March 2000.

[32] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)Hawai*, volume 1, pages 511–518, 2001.

[33] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[34] *The Harvard Database; ftp://hrl.harvard.edu/pub/faces.*

[35] *The M2VTS Database; http://ns1.tele.ucl.ac.be/M2VTS/.*

[36] *The Olivetti Database; http://www.cam-orl.co.uk/facedatabase.html.*

[37] *The Yale Database; http://cvc.yale.edu/projects/yalefaces/yalefaces.html.*