Face Verification Competition on the XM2VTS Database

Kieron Messer[1], Josef Kittler[1], Mohammad Sadeghi[1], Sebastien Marcel[2], Christine Marcel[2], Samy Bengio[2], F.Cardinaux[2], C.Sanderson[2], J. Czyz[3], L. Vandendorpe[3], Sanun Srisuk[4], Maria Petrou[1], Werasak Kurutach[4], Alexander Kadyrov[1], Roberto Paredes[5], B. Kepenekci[6], F. B. Tek[6], G. B. Akar[6], Farzin Deravi[7], and Nick Mavity[7]

[1] University of Surrey, Guildford, Surrey, GU2 7XH, UK
[2] Dalle Molle Institute for Perceptual Aritificial Intelligence, CP 592, rue du Simplon 4, 1920 Martigny, Switzerland
[3] Universite Catholique de Louvain, Batiment Stevin, Place du Levant 2, 1348 Louvain-la-Neuve, Belgium
[4] Mahanakorn University of Technology, 51 Cheum-Sampan Rd, Nong Chok, Bangkok, 10530 Thailand
[5] DSIC, Universidad Politcnica de Valencia, Camino de Vera, s/n. 46022, Valencia, Spain
[6] Tübitak Bilten, ODTÜ Campus, 0653, Ankara, Turkey
[7] Electronic Engineering Laboratory, University of Kent, Canterbury, CT2 7NT, UK

Abstract. In the year 2000 a competition was organised to collect face verification results on an identical, publicly available data set using a standard evaluation protocol. The database used was the Xm2vts database along with the Lausanne protocol [14]. Four different institutions submitted results on the database which were subsequently published in [13]. Three years later, a second contest using the same dataset and protocol was organised as part of AVBPA 2003. This time round seven seperate institutions submitted results to the competition. This paper presents the results of the competition and shows that verification results on this protocol have increased in performance by a factor of 3.

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.

However, 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 [16]. 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 [24], Harvard [21], Olivetti [23], M2VTS [22], ([1] 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 [18]. 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 [17].

More recently, two Face Recognition Vendor Tests [2] 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 indepently 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 along with the Lausanne protocol [14] was carried out. Four different institutions submitted results on the database which were subsequently published in [13]. This paper presents the results of a second contest using the same dataset and protocol, that has been organised as part of AVBPA 2003. This time round seven seperate institutions submitted results to the competition.

The results published are based completely on self-assessment of the submitted methods by the participating research groups. All the data from the Xm2vts database to perform the tests is available from [3]. We believe that this open approach will increase, in the long term, the number of algorithms that will be tested on the XM2VTS database as each research institution is able to assess their algorithms performance at any time. To date over 100 institutions have obtained copies of the XM2VTS database.

The rest of this paper is organised as follows. In the next section the database and evaluation protocol are described. In section 3 an overview of each algorithm which entered the competition is given. In section 4 the results according to the

protocol are presented along with a discussion. Finally, some conclusions are made.

2 The XM2VTS database

The XM2VTS database [14] is a multi-modal database consisting of face images, video sequences and speech recordings taken of 295 subjects at one month intervals. This database is available at the cost of distribution from the University of Surrey (see [3] for details). The database is primarily intended for research and development of personal identity verification systems where it is reasonable to assume that the client will be cooperative. Since the data acquisition was distributed over a long period of time, significant variability of appearance of clients, e.g. changes of hair style, facial hair, shape and presence or absence of glasses, is present in the recordings - see figure 1.

Fig. 1. Sample images from XM2VTS database

The subjects were volunteers, mainly employees and PhD students at the University of Surrey of both sexes and many ethnical origins. The XM2VTS database contains 4 sessions. During each session two head rotation and "speaking" shots were taken. From the "speaking" shot, where subjects are looking just below the camera while reading a phonetically balanced sentence, a single image

with a closed mouth was chosen. Two shots at each session, with and without glasses, were acquired for people regularly wearing glasses.

For the task of personal verification, a standard protocol for performance assessment has been defined. The so called Lausanne protocol splits randomly all subjects into a client and impostor groups. The client group contains 200 subjects, the impostor group is divided into 25 evaluation impostors and 70 test impostors. Eight images from 4 sessions are used.

From these sets consisting of face images, training set, evaluation set and test set are built. There exist two configurations that differ by a selection of particular shots of people into the training, evaluation and test sets. The training set is used to construct client models. The evaluation set is selected to produce client and impostor access scores, which are used to find a threshold that determines if a person is accepted or not (it can be a client-specific threshold or global threshold). According to the Lausanne protocol the threshold is set to satisfy certain performance levels (error rates) on the evaluation set. Finally the test set is selected to simulate realistic authentication tests where impostor's identity is unknown to the system. The evaluation set is also used in fusion experiments (classifier combination) for training, but this is not relevant in the context of this paper.

The performance measures of a verification system are the False Acceptance rate (FA) and the False Rejection rate (FR). False acceptance is the case where an impostor, claiming the identity of a client, is accepted. False rejection is the case where a client, claiming his true identity, is rejected. FA and FR are given by:

$$FA = EI/I * 100\% \qquad FR = EC/C * 100\% \tag{1}$$

where EC is the number of impostor acceptances, I is the number of impostor claims, EC the number of client rejections, and C the number of client claims. Both FA and an FR are influenced by an acceptance threshold. To simulate real application the threshold is set on the data from the evaluation set to obtain certain false acceptance (FAE) and false rejection error (FRE). The same threshold is afterwards applied to the test data and FA and FR on the test set are computed. Three thresholds are defined on the evaluation set:

$$\begin{aligned} T_{FAE=0} &= \arg\min_T(FRE|FAE = 0) \\ T_{FAE=FRE} &= (T|FAE = FRE) \\ T_{FRE=0} &= \arg\min_T(FAE|FRE = 0) \end{aligned} \tag{2}$$

Consequently, performance on the test set is characterised by six error rates:

$$\begin{array}{ll} FA_{FAE=0} & FR_{FAE=0} \\ FA_{FAE=FRE} & FR_{FAE=FRE} \\ FA_{FRE=0} & FR_{FRE=0} \end{array} \tag{3}$$

3 Overview of the Algorithms and the Scope of their Evaluation

This section describes the face verification methods that participated in the contest. For this competition, it was decided just to report the results at the equal error rate, i.e. $FAE = FRE$.

Both configurations of the protocol are considered under two face image registration conditions: manual registration and fully automatic registration. Manual registration is self-explanatory. Fully automatic registration requires that the face has to be localised automatically for the test phase.

3.1 Best Results From ICPR2000 (Unis-ICPR2000)

In the ICPR 2000 competition, [13], the best verification results for both semi-automatic and fully automatic registration techniques were performed by a method developed at the University of Surrey. It was based on a technique reported in [10] which performs face verification based on linear discriminant analysis. A novel way of measuring the distance between probe image and the client template was used. We have included the results of this technique in this paper to give a baseline comparison and indicate how the algorithms have improved over the past three years.

3.2 Dalle Molle Institute for Perceptual Artificial Intelligence (IDIAP)

IDIAP entered two seperate face verification algorithms into the competition. A brief description of each technique is given below.

IDIAP - Cardinaux The proposed face verification method is based on Gaussian Mixture Models (GMMs), [19] and [4]. The face images are analyzed on a block by block basis. Each block is decomposed in terms of an extension of the 2D Discrete Cosine Transform (DCT), namely DCT-mod2. The GMM approach uses a combination of Maximum Likelihood (ML) and Maximum a Posteriori (MAP) criteria.

IDIAP - Marcel We use skin color information in addition to the gray-level face image in order to train face verification systems using artificial neural networks, [12] and [11].

The representation used to code input images is based on gray-scale face image. The face bounding box is computed using manually located eyes coordinates. The face is cropped and the extracted sub-image is down-sized to a 30x40 image. After enhancement and smoothing, the face image becomes a feature vector of dimension 1200. The skin color feature is chosen to be simply the RGB color distribution of filtered skin pixels inside the face bounding box. For each

color channel, an histogram is built using 32 discrete bins. Hence, the feature vector produced by the concatenation of the 3 histograms (R, G and B) has 96 components.

Our face verification method is based on Multi-Layer Perceptrons (MLPs). For each client, an MLP is trained to classify an input to be either the given client or not. The input of the MLP is a feature vector corresponding to the face image with its skin color. The output of the MLP is either 1 (if the input corresponds to a client) or -1 (if the input corresponds to an impostor). The MLP is trained using both client images and impostor images, often taken to be the images corresponding to other available clients.

3.3 Universidad Politécnica de Valencia (UPV)

The local feature representation approach is used in this face verification contest, [15] [7]. Using this local feature representation scheme each image is represented by several smaller images. To classify each test image a nearest neighbor classifier is used by taking a suitable voting scheme. Given a test image, the k-nearest neighbors of its local feature images are found among the feature vectors computed for the training images. Each neighbor votes for its own class and a vector of votes (per class) is obtained by simply counting all votes. Following a direct voting scheme, the test image is classified into the most voted class. This sum rule of the votes of each local feature image is similar to the sum rule used in the Combining Classifiers theory.

3.4 Tübitak Bylten (TB)

The method uses a full Gabor wavelet transform for both finding feature points and extracting feature vectors [6]. The feature extraction algorithm of the proposed method has two steps: (1) Feature point localization, (2). Feature vector generation. Feature vectors are extracted at points with high information content on the face image. The features are not limited to eyes, nose, etc., i.e. special facial features such as dimples are also extracted. The face image is then convolved with Gabor filters, and R_j is found to be the response of the face image to the jth Gabor filter. Feature localization is done by searching local maximums of R_j which are also having the value above the mean of all pixel values of R_j. Feature vectors are generated at the feature points as a composition of Gabor wavelet transform coefficients. To measure the similarity of two complex valued feature vectors, a normalized cross-correlation function is used which ignores the phase.

Face comparison is done in two steps. In the first step, the feature vectors of reference images those are not close enough to the feature vectors of the test image in means of both location and similarity, are eliminated. In the second step, the similarity of two faces is calculated as the mean of similarities of matched features.

3.5 Universite Catholique de Louvain (UCL)

This fuses results from three different face verification experts. It combines the 3 scores given by the algorithm using a weighted averaging. The first algorithm uses Gradient Direction Metric in the LDA subspace to compute the score (developed in UniS). The second algorithm uses the Probabilistic Matching to compute the score (developed in UCL). The third method computes the score by taking the L1 norm between the colour histogram of the face image (developed in UCL). The images are registered using manually located eye coordinates. More details can be found in [5].

3.6 Commercial System

The University of Kent used a well known commercial system to perform face verification using fully automatic registration according to the Lausanne protocol. The package was used with the default settings. In enrollment some images were rejected by the system. This meant that some client templates were built with only one or two examples. The package recommends a minimum of four suitable training images.

3.7 University of Surrey (UniS)

UniS entered three seperate face verification algorithms into the competition. A brief description of each is given below. The third method based on the the Shape Trace Transform was done in conjunction with a visiting researcher from the Mahanakorn University of Technology (MUT).

Normalised Correlation in LDA Space (UniS-NC) Linear Discriminant Analysis (LDA) projects the input image data into fisher faces which maximise the class separability. In [9], it has been demonstrated that in the context of face verification, a matching score based on Normalised Correlation (NC) works effectively in the LDA space. Histogram equalisation was used to normalise the registered face photometrically. The thresholds in the decision making system have been determined using the Client-Specific Thresholding technique.

Error Correcting Codes (UniS-ECOC) In [8] a novel approach to face verification based on the Error Correcting Output Coding (ECOC) classifier was presented. In the training phase the client set is repeatedly divided into two ECOC specified subsets to train a set of binary classifiers. The output of the classifiers defines the ECOC feature space, in which it is easier to separate transformed patterns representing clients and impostors. The faces were first transformed in LDA space and the binary classifiers used to generate the binary codes were neural networks.

Shape Trace Transform (MUT-UniS-STT) A new face representation, the Shape Trace Transform (STT), for recognizing faces in an authentication system [20] has been developed. The STT offers an alternative representation for faces that has a very high discriminatory power. We estimate the dissimilarity between two shapes by a new measure we propose, the Hausdorff context. The reinforcement learning is used to search the optimal parameters of the algorithm, for which the within-class variance of the STT is minimized. This research demonstrates that the proposed method provides a new way for face representation. Our system is verified with experiments on the XM2VTS database.

4 Results and Discussion

Tables 1 and 2 shows the results using manual registration for both configurations I and II. The results on configuration I show that the best performing algorithm, the Shape Trace Transform, achieves an error rate of 1.47%. In fact three different methods have achieved a very similar low error rate, i.e. MUT-UniS-STT, UniS-NC and UniS-ECOC. This is an increase in performance by a factor of 3 over the best performing semi-automatic technique in the year 2000 competition where the best TER obtained was 4.8%.

Tables 3 and 4 shows the results using fully automatic registration for both configurations I and II. In the year 2000 competition the best performance for configuration I was 13.1%, in this competition it was 3.86%. An increase in performance of factor 3.5. Again, three different methods have achieved a similar level of performance, i.e. UPV, IDIAP-Cardinaux and UniS-NC.

	Evaluation Set			Test Set		
Method	FA	FR	TER	FA	FR	TER
UniS-ICPR2000	-	-	5.00	2.30	2.50	4. 80
IDIAP-Marcel	1.67	1.67	3.34	1.748	2.000	3.75
IDIAP-Cardinaux	0.75	2.00	2.75	1.84	1.50	3.34
MUT-UniS-STT	1.16	1.05	2.21	0.97	0.50	1.47
UCL	1.17	1.17	2.34	1.71	1.50	3.21
TB	2.34	1.00	3.34	5.61	5.75	11.36
UniS-ECOC	0.0	0.0	0.0	0.86	0.75	1.61
UniS-NC	0.33	1.33	1.36	0.48	1.00	1.48

Table 1. Error rates according to Lausanne protocol for configuration I with manual registration

5 Conclusions

This paper presents a comparison of face verification algorithms that was organised in conjunction with the Audio Visual Biometric Person Authentication

	Evaluation Set			Test Set		
Method	FA	FR	TER	FA	FR	TER
IDIAP-Marcel	1.25	1.25	2.5	1.465	2.250	3.715
IDIAP-Cardinaux	0.75	0.75	1.50	1.04	0.25	1.29
TB	1.10	0.50	1.60	3.22	4.50	7.72
UniS-NC	0.33	0.75	1.08	0.25	0.50	0.75

Table 2. Error rates according to Lausanne protocol for configuration II with manual registration

	Evaluation Set			Test Set		
Method	FA	FR	TER	FA	FR	TER
UniS-ICPR2000	-	-	14.0	5.8	7.3	13.1
Commercial System	11.00	11.10	22.10	2.83	13.50	16.33
IDIAP-Cardinaux	1.21	2.00	3.21	1.95	2.75	4.70
UPV	1.33	1.33	2.66	1.23	2.75	3.98
UniS-NC	0.82	4.16	4.98	1.36	2.5	3.86

Table 3. Error rates according to Lausanne protocol for configuration I using full automatic registration

	Evaluation Set			Test Set		
Method	FA	FR	TER	FA	FR	TER
Commercial System	13.20	13.40	26.6	14.30	11.25	25.55
IDIAP-Cardinaux	1.25	1.20	2.45	1.35	0.75	2.10
UPV	1.75	1.75	3.50	1.55	0.75	2.30
UniS-NC	0.63	2.25	2.88	1.36	2.0	3.36

Table 4. Error rates according to Lausanne protocol for configuration II using full automatic registration

conference of 2003. Many different verification algorithms from 7 different institutions were tested using identical data from a large, publicly available multi-modal database, the XM2VTS. Training and evaluation was carried out according to an a priori known protocol. Results indicate that in the last three years the performance of the algorithms have increased by a factor of three.

References

1. *The Face Recognition Homepage; http://www.cs.rug.nl/ peterkr/FACE/face.html.*
2. *Face Recognition Vendor Tests; http://www.frvt.org.*
3. *The XM2VTSDB; http://www.ee.surrey.ac.uk/ Research/VSSP/xm2vtsdb/.*
4. Fabien Cardinaux, Conrad Sanderson, and Sébastien Marcel. Comparison of mlp and gmm classifiers for face verification on xm2vts. In *To appear in the Proceedings of the Audio Visual Biometric Person Authentification*, Guildford, Surrey, June 2003.
5. J. Czyz, J. Kittler, and L. Vandendorpe. Combining face verification algorithm. In R Kasturi, D Laurendeau, and C Suen, editors, *Proceedings 16th International Conference on Pattern Recognition III*, 2002.
6. B. Kepenekci, F. B. Tek, and G. B. Akar. Occluded face recognition based on gabor wavelets. In *Proc International Conference on Image Processing*, September 2002.
7. D. Keysers, R. Paredes, H. Ney, and E. Vidal. Combination of tangent vectors and local representations for handwritten digit recognition. In *International Workshop on Statistical Pattern Recognition*, 2002.
8. J Kittler, R Gadheri, T Windeatt, and J Matas. Face verification via ecoc. In *Proceedings of British Machine Vision Conference 2001*, pages 593–602, 2001.
9. J Kittler, Y P Li, and J Matas. On matching scores for lda-based face verification. In M Mirmehdi and B Thomas, editors, *Proceedings of British Machine Vision Conference 2000*, pages 42–51, 2000.
10. Y.P. Li, J. Kittler, and J. Matas. On Matching Scores of LDA-based Face Verification. In Tony Pridmore and Dave Elliman, editors, *Proc British Machine Vision Conference BMVC2000*, page submitted, London, UK, September 2000. University of Bristol, British Machine Vision Association.
11. Sébastien Marcel and Samy Bengio. Improving face verification using skin color information. In *Proceedings of the 16th International Conference on Pattern Recognition*. IEEE Computer Society Press, 2002.
12. Sébastien Marcel, Christine Marcel, and Samy Bengio. A state-of-the-art Neural Network for robust face verification. In *Proceedings of the COST275 Workshop on The Advent of Biometrics on the Internet*, Rome, Italy, 2002.
13. 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.
14. 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.

15. R. Paredes, J. C. Prez, A. Juan, and E. Vidal. Local representations and a direct voting scheme for face recognition. In *In Proc. of the Workshop on Pattern Recognition in Information Systems*, July 2001.
16. P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki. An introduction to evaluating biometric systems. *IEEE Computer*, pages 56–63, February 2000.
17. 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.
18. 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.
19. C Sanderson. *Automatic Person Verification Using Speech and Face Information.* PhD thesis, Griffith University, Brisbane, Australia., 2002.
20. Sanun Srisuk, Maria Petrou, Werasak Kurutach, and Alexander Kadyrov. Face authentication using the trace transform. In *To appear in CVPR2003.* IEEE Computer Society Press, 2003.
21. ftp://hrl.harvard.edu/pub/faces.
22. http://ns1.tele.ucl.ac.be/M2VTS/.
23. http://www.cam-orl.co.uk/facedatabase.html.
24. http://cvc.yale.edu/projects/yalefaces/yalefaces.html.