Face Authentication Competition on the BANCA Database

Kieron Messer[1], Josef Kittler[1], Mohammad Sadeghi[1], Miroslav Hamouz[1], Alexey Kostyn[1], Sebastien Marcel[2], Samy Bengio[2], Fabien Cardinaux[2], Conrad Sanderson[2], Norman Poh[2], Yann Rodriguez[2], Krzysztof Kryszczuk[2], Jacek Czyz[3], L. Vandendorpe[3], Johnny Ng[4], Humphrey Cheung[4], and Billy Tang[4]

[1] University of Surrey, Guildford, Surrey, GU2 7XH, UK
[2] Dalle Molle Institute for Perceptual Artificial Intelligence, CP 592, rue du Simplon 4, 1920 Martigny, Switzerland
[3] Université Catholique de Louvain, Batiment Stevin, Place du Levant 2, 1348 Louvain-la-Neuve, Belgium
[4] Titanium Technology Research Centre, 10/F, Tianjin Building, 167 Connaught Road West, Hong Kong

Abstract. This paper details the results of a face verification competition [2] held in conjunction with the First International Conference on Biometric Authentication. The contest was held on the publically available BANCA database [1] according to a defined protocol [6]. Six different verification algorithms from 4 academic and commercial institutions submitted results. Also, a standard set of face recognition software from the internet [3] was used to provide a baseline performance measure.

1 Introduction

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. The use of common datasets along with evaluation protocols can help alleviate this problem.

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 [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 [15] was organised [14]. As part of AVBPA 2003 a second competition on exactly the same data and testing protocol was organised [10]. All the data from the Xm2vts database is available from [5]. 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 [6]. 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. Again this 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 competition rules and performance criterion are described. Section 3 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 Competition

All the experiments were carried according to the Matched Controlled (MC) configuration of the BANCA database. In this configuration a high quality camera was used to capture all the data and the lighting was controlled.

There were two 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 (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 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.

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 this 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:

$R = 0.1 \rightarrow$ FA is an order of magnitude less harmful than FR,
$R = 1 \rightarrow$ FA and FR are equally harmful,
$R = 10 \rightarrow$ 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 390 true client claims and 520 impostor attacks.

3 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 [3] to provide a baseline performance measure on this database. Due to space limitations we have published the results from these experiments at [2].

3.1 Dalle Molle Institute for Perceptual Artificial Intelligence (IDIAP)

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 [23], followed by post-processing using a Multi-Layer Perceptron [20] to provide the final localized face. The probabilistic classifier uses DCTmod2 features [22] and models faces using pseudo-2D Hidden Markov Models (HMMs) [7]. 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 modeled 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 Maximization 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. As the generic face HMM is deemed to be a good representation of the general population, it is used as the impostor model.

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 [23], followed by post-processing using a Multi-Layer Perceptron (MLP) [20]. The first two classification subsystems are based on local features and generative models (namely, DCTmod2 features, Gaussian Mixture Models & pseudo-2D Hidden Markov Models [7]), while the third subsystem uses Linear Discriminant Analysis based feature extraction (i.e. holistic features) and a MLP for classification [13]. Finally, the opinions of the three subsystems are fused using an MLP based approach [19].; a global threshold (i.e. the same for all clients) is used in making the final verification decision.

3.2 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 $I_s = (I(x, y) + I(-x, y))/2$ [9].

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 [8].

3.3 Titanium Technology Research Centre

Dynamic Local Feature Analysis (DLFA) is used in this experiment. DLFA is developed from the concept of LFA. Texture and shape information is combined using the Local Feature Analysis (LFA) technique to develop a robust face recognition algorithm. The shape information is obtained by using an adaptive edge detecting method that can reduce the effect of different lighting conditions, while the texture information provides normalized facial features conveying more details about the image.

The approach can be divided into two main steps. The first step is preprocessing. The goal of this step is to reduce noise, transform the input face image into a binary one by dynamic edge detection and then extract the texture of the face. The second step employs the local feature analysis to combine both edge of face shape and the texture [16].

3.4 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 [21] 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 part II 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 [11]. 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 Results and Discussion

Table 1 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 2. The last column shows the average WER across all six test conditions. The best performing algorithm was UCL-Fusion which achieved an average WER of 1.95%. Second was IDIAP-Fusion with 2.70%. It is interesting to note that these two best performing algorithms used intra-modal fusion and are approximately 50% better then the two best single expert systems (i.e. UniS 2.99% and IDIAP-HMM 3.53%). This seems consistent with other published results [12]. However, there is a grey area over what constitutes a system using intramodal fusion. For example, a neural network based system uses hundreds of simple experts to arrive at a decision.

Table 2 shows the results using automatic registration for at least the verification stage. This time the best result is IDIAP-HMM with a performance of 3.78%. What is interesting about this result is that there is very little degradation in performance from the manual case (i.e. 3.53%). The fusion systems which performed well in part I of the test degrade in performance by a factor 2. The experts used in these fusion systems have not been so robust to the localisation errors. 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.

5 Conclusions

This paper presents a comparison of face verification algorithms on a new publically available and challenging face database. It was organised in conjunction with the First International Conference on Biometric Authentication. Six different verification algorithms from a variety of academic and commercial institutions entered the competition.

Table 1: The Weighted Erro 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	7.52	4.90	5.45	0.64	2.56	0.12	3.53
IDIAP-Fusion	6.99	2.42	3.85	1.76	0.70	0.47	2.70
UCL-LDA	6.53	1.17	7.05	2.88	1.28	2.10	3.50
UCL-Fusion	3.90	0.26	4.32	1.44	1.28	0.47	1.95
UniS	7.12	0.89	5.58	1.98	1.47	0.92	2.99
Titanium	4.12	3.90	3.04	3.10	1.97	2.12	3.04

Table 2: The Weighted Erro Rates on the two groups at the three different operating points using automatic localisation.

	R=0.1 (WER)		R=1 (WER)		R=10 (WER)		
	G1	G2	G1	G2	G1	G2	Av
IDIAP-HMM	7.78	3.76	5.13	2.08	1.17	2.74	3.78
IDIAP-Fusion	6.53	8.68	7.53	6.73	2.10	1.40	5.50
UCL-LDA	8.13	7.11	10.58	9.46	5.89	6.12	7.88
UCL-Fusion	4.28	3.64	9.13	5.12	2.44	1.80	4.40
UniS	5.75	3.00	6.38	4.50	1.95	1.97	3.93
Titanium	5.84	5.12	4.42	3.94	3.01	2.76	4.18

References

1. *BANCA; http://www.ee.surrey.ac.uk/banca/.*
2. *BANCA; http://www.ee.surrey.ac.uk/banca/icba2004.*
3. *The CSU Face Identification Evaluation System; http://www.cs.colostate.edu/evalfacerec.*
4. *Face Recognition Vendor Tests; http://www.frvt.org.*
5. *The XM2VTSDB; http://www.ee.surrey.ac.uk/ Research/VSSP/xm2vtsdb/.*
6. 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.
7. 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.
8. J Czyz, M Sadeghi, J Kittler, and L Vandendorpe. Decision fusion for face authentication. In *First International Conference on Biometric Authentication*, 2004.

9. Jacek Czyz. *Decision fusion for identity verification using facial images*. PhD thesis, Universite Catholique de Louvain, 2003.
10. 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.
11. 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.
12. Josef Kittler, Marco Ballette, Jacek Czyz, Fabio Roli, and Luc Vandendorpe. Enhancing the performance of personal identity authentication systems by fuion of face verification experts. In *International Conference on Multimedia and Expo*, 2002.
13. S. Marcel. A symmetric transformation for lda-based face verification. In *Proc. Int. Conf. Automatic Face and Gesture Recognition (AFGR), Seoul, Korea*, 2004.
14. 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.
15. 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.
16. Johnny Ng and Humphrey Cheung. Dynamic local feature analysis for face recognition. In *First International Conference on Biometric Authentication*, 2004.
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. 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.
20. 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.
21. 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.
22. C. Sanderson and K.K. Paliwal. Fast features for face authentication under illumination direction changes. *Pattern Recognition Letters*, 24(14):2409–2419, 2003.
23. 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.