The 2013 Face Recognition Evaluation in Mobile Environment

M. Günther,[*,1] A. Costa-Pazo,[2] C. Ding,[3] E. Boutellaa,[4] G. Chiachia,[5] H. Zhang,[6] M. de Assis Angeloni,[7]

V. Štruc,[8] E. Khoury,[1] E. Vazquez-Fernandez,[2] D. Tao,[3] M. Bengherabi,[4] D. Cox,[10] S. Kiranyaz,[6]

T. de Freitas Pereira,[7] J. Žganec-Gros,[9] E. Argones-Rúa,[2] N. Pinto,[10] M. Gabbouj,[6] F. Simões,[7] S. Dobrišek,[8]

D. González-Jiménez,[2] A. Rocha,[5] M. Uliani Neto,[7] N. Pavešić,[8] A. Falcão,[5] R. Violato,[7] S. Marcel[1]

[1]Idiap Research Institute (CH), [2]Galician Research and Development Center in Advanced Telecommunications (ES),

[3]University of Technology, Sydney (AUS), [4]Centre de Développement des Technologies Avancées (DZ),

[5]University of Campinas (BR), [6]Tampere University of Technology (FIN), [7]CPqD (BR),

[8]University of Ljubljana (SLO), [9]Alpineon Ltd. (SLO), [10]Harvard University (US)

Abstract

Automatic face recognition in unconstrained environments is a challenging task. To test current trends in face recognition algorithms, we organized an evaluation on face recognition in mobile environment. This paper presents the results of 8 different participants using two verification metrics. Most submitted algorithms rely on one or more of three types of features: local binary patterns, Gabor wavelet responses including Gabor phases, and color information. The best results are obtained from UNILJ-ALP, which fused several image representations and feature types, and UC-HU, which learns optimal features with a convolutional neural network. Additionally, we assess the usability of the algorithms in mobile devices with limited resources.

1. Introduction

Since it started in the late 1980s [1], automatic face recognition is an ongoing topic and enormous progress has been made [2]. On good quality facial images with frontal pose and moderate illumination, current automatic face recognition systems outperform humans [3] on this job. Still, recognizing facial images in low resolution images and under uncontrolled pose is a challenging task for automatic face recognition systems.

To evaluate different face recognition algorithms on these kind of data, the Biometric Group at the Idiap Research Institute organized a face recognition evaluation in mobile environment. Together with the speaker verification evaluation [4], it is the second in an ongoing series of face and speaker recognition evaluations [5]. The evaluation is carried out on the MOBIO database [6], which con-

tains videos of talking faces that were captured with mobile devices.

In total, 8 institutions participated in the face recognition evaluation. To assure a fair competition and comparable results, all participants had to strictly follow an unbiased evaluation protocol. In the first phase of the evaluation, the *training* and the *development* set was made available to the participants. In these sets, each image was labeled with the *client ID*, and the participants were allowed to optimize their system parameters according to these data. In the second phase, an *evaluation* set was given to the participants. In the evaluation set, the file names are anonymized, so that client ID could not be read out from them. Participants were asked to compute the scores according to the protocols and send the final score files to the organizers of the evaluation, which evaluated them.

The structure of this paper is as follows: first we introduce the database and the evaluation method, then we describe the face recognition systems of the participants. Afterward, we evaluate the verification performances of the submitted systems and discuss some further characteristics of them. Finally, we close this paper with a conclusion.

2. The MOBIO Database

The MOBIO database is a bi-modal (face/speaker) video database recorded from 152 people. The database has a female-male ratio of nearly 1:2 (100 males and 52 females) and was collected from August 2008 until July 2010 in 6 different sites from 5 different countries. In total 12 sessions were captured for each individual.

The database was recorded using two types of mobile devices: mobile phones (NOKIA N93i) and laptop computers (standard 2008 MacBook). In this evaluation we only use the mobile phone data. The MOBIO database is challenging

---

*corresponding author: manuel.guenther@idiap.ch

since the data are acquired with uncontrolled illumination, facial expression, and face pose, and sometimes only parts of the face are visible. For this face recognition evaluation, one facial image was extracted from each video by choosing the video frame after 10 seconds. The eye positions were labeled by hand and made available to the participants.

More technical details about the MOBIO database including some example images can be found in [6] and on its official web page[1], which also contains instructions on how to obtain the data.

Based on the gender of the clients, different evaluation protocols *female* and *male* exist. In order to have an unbiased evaluation, the clients of the database are split up into 3 different sets:

Training set. The images from training set can, e. g., be used to compute projection matrices for PCA. For some systems, these images also serve as the negative examples in a classification system. Furthermore, these images can be used as cohort images for score normalization. The training set consists of both male and female subjects, and participants are allowed to separate the clients in the training set according to the gender. If desired, participants are also allowed to use other data to train their algorithms, but none of the participants makes use of this option.

Development set. The development set can be used to choose the meta-parameters of the algorithms, e. g., the dimension of a PCA projection matrix and the distance function. For each client in this set, the images are subdivided into the *enrollment* and the *probe* group. For the enrollment of a client model, 5 images per client are provided, and **it is forbidden to use the information of other clients of the development set** for enrolling a client model. The remaining images of the clients serve as probe images, and similarity scores have to be computed between all probe images and all client models. In systems that required score calibration, these scores can be used to train the calibration system.

Evaluation set. The evaluation set is used to assess the final unbiased system performance. The subdivision into enrollment and probe groups is similar to the one in the development set, and also the model enrollment uses similar data. To disallow participants to optimize system parameters on the evaluation set, the probe image names are anonymized and shuffled.

Table 1 statistically details each of the sets described above. It specifies the number of files, clients, and probe images; and the number of scores that need to be computed.

3. Performance Measures

The metrics that are used to evaluate face verification performance are based on the *false acceptance rate* (FAR) and the *false rejection rate* (FRR). The definition of these rates is dependent on a certain *threshold θ*:

$$\text{FAR}(\theta) = \frac{|\{s_{\text{imp}} \mid s_{\text{imp}} \geq \theta\}|}{|\{s_{\text{imp}}\}|}$$
$$\text{FRR}(\theta) = \frac{|\{s_{\text{cli}} \mid s_{\text{cli}} < \theta\}|}{|\{s_{\text{cli}}\}|} \quad (1)$$

where s_{cli} are client scores, while s_{imp} are impostor scores. We evaluate the FAR and the FRR for both the development and the evaluation set independently.

The first evaluation metric is based on the *equal error rate* (EER) and the *half total error rate* (HTER). Particularly, it defines a score threshold θ^* based on the EER of the development set, and computes the final HTER using this threshold:

$$\theta^* = \arg\min_{\theta} |\text{FAR}_{\text{dev}}(\theta) - \text{FRR}_{\text{dev}}(\theta)|$$
$$\text{EER} = \frac{\text{FAR}_{\text{dev}}(\theta^*) + \text{FRR}_{\text{dev}}(\theta^*)}{2} \quad (2)$$
$$\text{HTER} = \frac{\text{FAR}_{\text{eval}}(\theta^*) + \text{FRR}_{\text{eval}}(\theta^*)}{2}$$

The second metric is the *receiver operating characteristics* (ROC). In this metric, the *correct acceptance rate* (CAR) with $\text{CAR} = 100\% - \text{FRR}$ is plotted against the FAR. Commonly, the focus is put on the low FAR values and, hence, we plot the FAR axis in logarithmic scale.

4. Participants

In total, 8 institutions participated in the evaluation. Please refer to table 2 for the names of the participants and their abbreviation. In this section we briefly review the face recognition systems of the participants, split into systems that use a single algorithm and systems that fuses scores from different algorithms.

As long as not stated otherwise, all face recognition systems use the provided hand-labeled eye positions to geometrically normalize (align) the faces.

4.1. Simple Systems

Baseline. The baseline system that we provide computes a PCA+LDA [7] projection matrix on the raw pixel values of aligned 64×80 pixel images that are enhanced using histogram equalization. The PCA and LDA matrices are computed on the training set, and the dimensions of the subspaces are limited to 200 and 199, respectively. The projected features of one model and a probe image are compared using the cosine similarity measure, and the highest

Table 1. PARTITIONING OF THE MOBIO DATABASE. *This table exemplifies the number of clients and images of the training set, as well as the number of clients and enrollment images, and the number of probe images and scores for the development and the evaluation set for the protocols Female and Male.*

Protocol	Training set		Development set				Evaluation set			
			Enrollment		Probe		Enrollment		Probe	
	Clients	*Images*	*Clients*	*Images*	*Probes*	*Scores*	*Clients*	*Images*	*Probes*	*Scores*
Female	13	2496	18	90	1890	34020	20	100	2100	42000
Male	37	7104	24	120	2520	60480	38	190	3990	151620

Table 2. PARTICIPANTS. *This table lists the participants and their institutions.*

Participant	Institution
CDTA	Centre de Développement des Technologies Avancées
UC-HU	University of Campinas; Harvard University
TUT	Tampere University of Technology
Idiap	Idiap Research Institute
UTS	University of Technology, Sydney
GRADIANT	Galician Research and Development Center in Advanced Telecommunications
CPqD	CPqD
UNILJ-ALP	University of Ljubljana; Alpineon Ltd.

similarity value for all model images of a client is chosen. Note that the baseline system solely uses open source software [8] that was developed at Idiap.[2]

CDTA. The CDTA system compares adapted histograms of *local binary patterns* (LBP) from aligned images of resolution 64×80, which are photometrically normalized [9]. Each image is divided into 80 blocks of size 8×8 pixels and histograms H^k of $\text{LBP}^{u2}_{8,2}$ features [10] are extracted from each block $k = 1, \ldots, 80$. A global histogram H^k_w is computed from the training images. Each model histogram is enrolled using (there are 5 model histograms per client):

$$\widehat{H}^k_c = \alpha\,H^k_w + (1-\alpha)\,H^k_c \quad : \quad k = 1, \ldots, 80 \qquad (3)$$

where the parameter α is optimized on the development set. The score between a model histogram $\widehat{H}^k_c$ and a probe histogram H^k_r is computed by:

$$S = \sum_k [-\chi^2(H^k_r, \widehat{H}^k_c) + \chi^2(H^k_r, H^k_w)] \qquad (4)$$

using the default χ^2 histogram similarity measure [10]. The second term in (4) aims for eliminating the cohort effect. For each client, the highest score of all 5 model histograms is taken.

---

[2] http://www.idiap.ch/software/bob

UC-HU. The UC-HU system employs a three layer hierarchical convolutional neural network similar to the one described in [11]. This neural network takes as input raw pixel values from gray scale images of 200×200 pixels, in which the faces are aligned to 56 pixels inter-eye-distance. Additionally, Fisher LDA is learnt on the training set and used in order to represent the samples in a more meaningful way, *i.e.*, adapted to the discriminant face aspects of the individuals, and the number of retained projection vectors is determined using the development set.

On top of this representation, person-specific linear models are learnt at enrollment time in a one-versus-all setting, considering the samples of the person being enrolled for the positive class and all samples in the training set for the negative class. Scores between the model and probe samples are computed by dot product.

TUT. The TUT system extracts *local Gabor binary pattern histogram sequences* (LGBPHS) [12] from the aligned images of resolution 140×154 pixels. First, a Gabor wavelet transform with the default 40 Gabor filters [13] is performed. Afterward, $\text{LBP}^{u2}_{8,2}$ [10] codes are extracted from the absolute values of the Gabor filter responses. The image is split into 49 blocks and 49×40 LBP histograms are extracted and concatenated. For dimensionality reduction, only the 8000 histogram bins with highest variance are kept (variances are estimated on 1000 images from the training set).

For face verification, the *partial least square* (PLS) classifier [14] is used. For each client, a PLS classifier is learnt using the 5 enrollment images as positive and random 1000 training images as negative class. All PLS classifiers are learnt with 64 latent components. The final score between a client model and a probe vector is given by the classification score of the according PLS classifier.

Idiap. The Idiap system extracts Gabor jets in a regular 20×25 node grid with 6 pixels inter-node distance and compares them using a Gabor phase based similarity measure. After aligning images to resolution 120×150 pixels, they are preprocessed by using non-uniform $\text{LBP}_{8,2}$ codes [15].

From the preprocessed images, responses of 40 Gabor filters in 5 scales and 8 orientations are computed using Ga-

bor wavelet parameters optimized on the development set. 500 normalized Gabor jets extracted at regular grid positions form the feature. The score between a client model, which incorporates all 5 model features, and a probe feature is computed as:

$$S = \sum_{i=1}^{500} \max_{m} S_{n+C}(\mathcal{J}_{mi}, \mathcal{J}_i) \qquad (5)$$

where $\mathcal{J}_{mi}$ is the Gabor jet of the ith position of the mth model feature, and $\mathcal{J}_i$ is the probe Gabor jet at the ith position. The Gabor jet similarity function S_{n+C} [16] exploits absolute and phase values of the Gabor jets.

4.2. Fusion Systems

UTS. The UTS system fuses the results of Gabor phases of 40 default Gabor wavelets [13] and *local phase quantization* (LPQ) features [17]. After performing a photometric normalization [9] on the aligned images, they are split into 7×7 local regions, in which histograms of LPQ patterns are extracted; and 8×8 regions for extracting histograms of Gabor phases.

For both subsystems, the local histograms are concatenated into one histogram sequence each. Scores between one model histogram and a probe histogram are computed using the histogram intersection measure and for each client the model histogram with the lowest score is taken. Scores for both subsystems are calibrated using mean and variance of the scores from the development set, and fused with weight 0.5.

GRADIANT. The GRADIANT system fuses 2 different face recognition systems: *patterns of oriented edge magnitudes* (POEM) [18] and *Gabor jets* (Gabor) [19] from a 10×10 regular grid using 40 Gabor wavelets with default parametrization [13]. After aligning the face images to 125×140 (POEM) or 85×100 (Gabor) pixels, an adaptation of the retina layer model [20] was used to preprocess them. To the Gabor feature vectors, a PCA without subspace truncation is performed, which is learnt on the training set.

For both subsystems, the scores between one model image and the probe are computed using a normalized correlation measure. For each client, the highest score of all 5 model images was taken. Finally, POEM and Gabor scores are fused after performing score calibration using logistic regression.

CPqD. The CPqD system fuses scores from 4 different subsystems: LBP, dLBP, MSLBP-108, and MSLBP-68. The LBP system extracts $\text{LBP}_{8,2}^{u2}$ histograms [10] from 64 blocks of aligned and preprocessed [9] 108×108 pixel images. The dLBP system extracts histograms of *direction-coded* LBP features [21] from 16 blocks of aligned and pre-
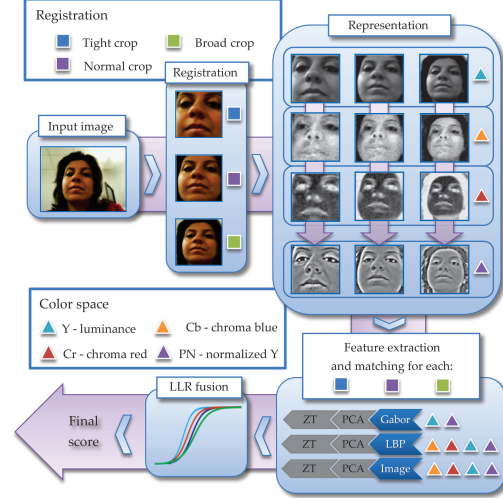

Figure 1. BLOCK DIAGRAM OF THE UNILJ-ALP FACE RECOGNITION SYSTEM.

processed [9] 68×68 pixel images. Both MSLBP systems extract *multispectral* $\text{LBP}_{8,2}^{u2}$ codes [22] from 16 blocks of 108×108 or 68×68 pixel color images preprocessed with histogram equalization of each color layer.

For the client model enrollment of the 3 subsystems LBP, dLBP, and MSLBP-68, linear *support vector machines* (SVM) are trained using the 5 samples from the model as positive and the images from the training set as negative class. The score between the client model and the probe feature is the SVM output of the according client. For MSLBP-108, an average histogram sequence is computed from the 5 model histogram sequences, and histogram intersection is used to compare client model and probe feature.

Finally, the scores of the 4 subsystems are fused using *linear logistic regression*.

UNILJ-ALP. The face recognition system submitted by UNILJ-ALP exploits multiple facial representations, PCA, and the whitened cosine similarity measure for recognition as shown in figure 1. In the first step, the system uses the manually marked eye-centers and an automatically detected mouth location to geometrically normalize the images. Afterward, it crops the facial area from the input image using 3 bounding boxes of different size. This procedure results in 3 distinct facial areas (a tightly, a normally, and a broadly cropped one), represented in *YCbCr* color space.

In the next step, the luminance components (*Y*) of cropped images are photometrically normalized [20] to *PN* (*cf.* figure 1). Together with the remaining color components *Y*, *Cb*, and *Cr* it forms the basis for feature extraction.

In the feature extraction step, various image descriptors, such as intensity values, Gabor features, and LBP features are computed from the image representations and subjected to PCA for dimensionality reduction. In the last step, all

feature vectors corresponding to the given probe image are matched against the corresponding enrollment feature vectors to produce 5 similarity scores per client, which are normalized using a special type of ZT normalization and ultimately combined using a variant of the recently proposed *temporal fusion* [23].

5. Results

5.1. Verification Performances

We evaluate the score files that have been sent by the participants using the 2 different verification metrics presented in section 3. Table 3 shows the equal error rates on the development set and the half total error rates on the evaluation set. Clearly, the error rates on the evaluation set are higher than on the development set. This is most probably caused by the fact that all algorithms are optimized to the development set and there is a slight mismatch between the conditions in both sets. In general, the male clients are slightly easier to recognize, maybe caused by the fact that women more often wore different hair styles or make-up in the different sessions of the database recording, or since the database (and also the training set) contains more men than women.

In this evaluation we discriminate the fusion systems that are marked with a * in table 3. The best overall rates are highlighted in bold font, while the best results of the non-fusion systems are in italics. Among the fusion systems, clearly UNILJ-ALP gets the best scores, but the performance differs drastically between development and evaluation set. Apparently, the fusion systems like UNILJ-ALP and GRADIANT seem to over-tune their parameters to the development set.

For the non-fusion systems, the clear winner in this evaluation is the UC-HU system since it gets best results in all 4 rates. In the evaluation set of the male protocol they could even outrival the fusion systems. With few exceptions, all other systems perform comparably well and all systems are able to outperform the baseline.

Inspecting the receiver operating characteristics of the evaluation set, which are given in figure 2 with a logarithmic FAR axis, one can see that different systems have different preferred working points. For example, the Idiap, TUT, and UTS systems perform comparably well in lower FAR regions while being average for higher FAR's. On the other hand, UC-HU and UTS have low performance on low FAR but improve with increasing FAR. Generally, the best system is again UNILJ-ALP.

5.2. System requirements

One important point, especially in mobile environments, is the requirements of the face recognition system in terms of speed and memory. Usually, the requirements can be split into an offline training and enrollment phase, and an online verification phase.

The participants were asked to report the size of their features and models; and also measure the time required for training, feature extraction and scoring. Since the time requirements are influenced by several factors, also the CPU and the tools that they use should be mentioned. Table 4 gives an overview of the requirements of the systems. The reported feature extraction time measures the average extraction time of one probe sample, and the scoring time is computed as the average time required to compare one model with one probe feature. Unfortunately, UNILJ-ALP could not report training and feature extraction times.

A special attention should be paid to the elements that are required online. In a verification scenario, usually a probe image is compared with an enrolled model. Hence, the device must possess enough memory to hold the model and the feature, but also the memory required for probe image preprocessing, feature extraction, and the comparison. In table 4, this includes the feature extraction memory, the feature size and the model size. Obviously, even though the best performing system UNILJ-ALP is able to compute the result rapidly, the required memory exceeds the capacity of mobile devices. The UC-HU system, which performs also very well, can run online since feature extraction and scoring do not require too much memory, but the model enrollment, which needs 600MB of memory, needs to be done offline.

The time required for feature extraction and scoring is also an important factor. Since TUT, UTS and CPqD need around 1.3 seconds[3] or 0.6 seconds to extract the features of one image, they might be considered as too slow. All other algorithms are fast enough to be useful in mobile devices.

Looking at the trade-off between verification performance and hardware requirements, the algorithms from UC-HU, GRADIANT, Idiap and UTS are most suitable for mobile applications.

Another interesting point is the reproducibility of the results and the algorithms. Therefore, in table 4 the institutions that provide the source code for their algorithms are highlighted as well. Unfortunately, additionally to the baseline[4] only one participant releases source code[4] for the algorithm.

6. Conclusion

In this paper we have shown the results of the face recognition evaluation in mobile environments. In total, 8 institutions participated, 4 of which submitted fusion algorithms. Most of the algorithms relied on one or more features of three different kinds: *local binary patterns*, *Gabor wavelet*

---

[3]TUT uses a slow Matlab implementation to extract the features.
[4]http://www.idiap.ch/resource/biometric

Table 3. EVALUATION RESULTS. *This table shows the equal error rates on the development set and the half total error rates on the evaluation set of the participants for the female and male protocol. Algorithms that perform score fusion are marked with *.*

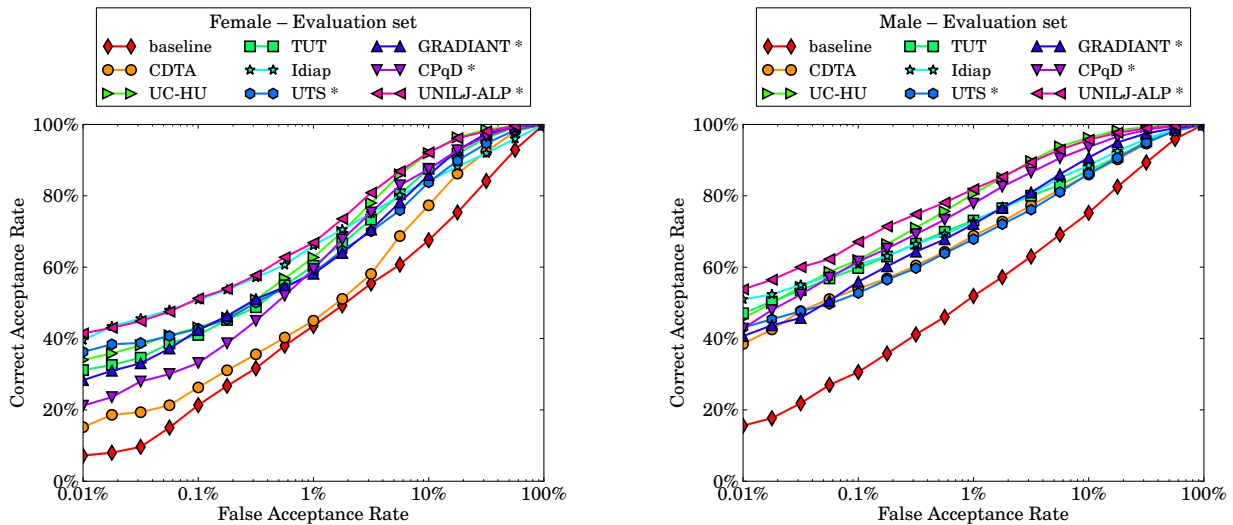|		Female		Male	
Participant	Algorithm	EER	HTER	EER	HTER
baseline	PCA+LDA	14.711%	20.942%	14.802%	17.115%
CDTA	ALBPH-χ^2	10.689%	28.482%	7.654%	11.929%
UC-HU	BIN	*4.709%*	*10.833%*	*3.492%*	**6.210%**
TUT	LGBP-PLSDA	8.571%	13.914%	7.341%	11.544%
Idiap	Disparity	6.243%	12.505%	6.627%	10.292%
UTS	Gabor+LPQ *	7.462%	13.569%	6.111%	11.955%
GRADIANT	POEM+Gabor *	5.384%	12.274%	3.144%	9.522%
CPqD	CPqD *	6.296%	11.209%	5.476%	7.668%
UNILJ-ALP	MR-PCA *	**2.751%**	**10.457%**	**1.707%**	7.450%

Figure 2. ROC CURVES OF EVALUATION SET. *This figure display the receiver operating characteristics with logarithmic FAR axis of the participants on the female and male protocol of the evaluation set. Algorithms that perform score fusion are marked with *.*

responses (especially *Gabor phases*) and *color information*. Notably, the best performing simple system does not rely on hand-designed features, but learns the extracted features with a convolutional neural network. The fusion systems perform comparably well on the development set, but these good results cannot generalize to the evaluation set. Unfortunately, the reproducibility of the results is limited since only one participant provides source code.

The usability of the algorithms in mobile devices is two-fold. Usually, algorithms need to have a good balance between requirements in terms of memory and execution time, and verification accuracy. Depending on the available hardware, different algorithms might be preferred. Also, all algorithms in this evaluation used the hand-labeled eye positions, while algorithms might work different if the face is

detected by an algorithm rather than annotated by a human.

Finally, future work will focus on a biometric verification system that fuses scores from face and speaker [4] verification systems.

Table 4. MEMORY AND TIME REQUIREMENTS. *This table shows the memory footprints and the time requirements of the algorithms. Entries marked with * are estimated based on theoretical values.* **Institutions that provide the source code are marked with** [+].

Participant	CPU	Tools	Training time	Training memory	Feature extraction time	Feature extraction memory	Feature extraction size	Model size	Scoring time
baseline[+]	Intel i7 3.4 GHz	Bob	5.5m	4 GB	4.3ms	1MB*	199	5×199	0.1ms
CDTA	Intel i7 2.8 GHz	Matlab	58s	69KB	39ms	10MB	4720	6×4270	0.9ms
UC-HU	Intel i7 2.6 GHz	Bob, Scipy, Scikit-learn	15m	6GB	180ms	10MB	21600	6400	0.3ms
TUT	Intel Core2 3.0 GHz	Matlab	3.9m	32MB*	1.3s	1.5MB*	115640	8001	0.5ms
Idiap[+]	Intel i7 3.4 GHz	Bob	*n/a*	*n/a*	62ms	7MB*	40000	5×40000	13ms
UTS	Intel i5 3.1 GHz	Matlab	*n/a*	*n/a*	0.6s*	6MB	53504	5×53504	1.2ms
GRADIANT	Intel Xeon 2.4 GHz	OpenCV	27m	120MB	44ms	25MB	55648	5×55648	1ms
CPqD	Intel Xeon 3.1 GHz	Bob, NumPy, OpenCV	52m	40GB*	0.6s	350MB	24864	90336	2.6ms
UNILJ-ALP	Intel i7 2.8 GHz	Matlab, PhD, INface	?	>1.8GB*	?	>2.3MB*	303360	5×303360 + cohort	1.4ms

sponsiveness.

References

[1] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3), 1987.

[2] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, and W. Worek. Preliminary face recognition grand challenge results. In *Proceedings of FG 2006*, 2006.

[3] A.J. O'Toole, P.J. Phillips, F. Jiang, J. Ayyad, N. Penard, and H. Abdi. Face recognition algorithms surpass humans matching faces over changes in illumination. *Transactions on PAMI*, 29, 2007.

[4] E. Khoury et al. The 2013 speaker recognition evaluation in mobile environment, 2013. accepted at: ICB 2013.

[5] S. Marcel et al. On the results of the first mobile biometry (MOBIO) face and speaker verification evaluation. In *Proceedings of the 20th International conference on Recognizing patterns in signals, speech, images, and videos*, ICPR, 2010.

[6] C. McCool et al. Bi-modal person recognition on a mobile phone: using mobile phone data. In *IEEE ICME Workshop on Hot Topics in Mobile Multimedia*, 2012.

[7] W. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. In *FG*. IEEE Computer Society, 1998.

[8] A. Anjos, L. El Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel. Bob: A free signal processing and machine learning toolbox for researchers. In *Proceedings of ACMMM*, 2012.

[9] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Transactions on Image Processing*, 19(6), 2010.

[10] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *European Conference on Computer Vision*. Proc. Workshop Dynamical Vision, 2004.

[11] D.D. Cox and N. Pinto. Beyond simple features: A large-scale feature search approach to unconstrained face recognition. In *FG*, 2011.

[12] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. *ICCV*, 1, 2005.

[13] L. Wiskott, J.-M. Fellous, N. Krüger, and C. v.d. Malsburg. Face recognition by elastic bunch graph matching. *Transactions on PAMI*, 19, 1997.

[14] R. Rosipal and N. Krämer. Overview and recent advances in partial least squares. In *Subspace, Latent Structure and Feature Selection*, number 3940 in Lecture Notes in Computer Science. Springer, 2006.

[15] G. Heusch, Y. Rodriguez, and S. Marcel. Local binary patterns as an image preprocessing for face authentication. In *Proceedings of FG*, 2006.

[16] M. Günther, D. Haufe, and R.P. Würtz. Face recognition with disparity corrected Gabor phase differences. In *ICANN*, volume 7552. Springer Berlin, 2012.

[17] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkil. Recognition of blurred faces using local phase quantization. In *ICPR*, 2008.

[18] N. Vu and A. Caplier. Face recognition with patterns of oriented edge magnitudes. In *Proceedings of ECCV: Part I*. Springer, 2010.

[19] E. Argones-Rúa, J.L. Alba-Castro, and C. García-Mateo. Quality-based score normalization and frame selection for video-based person authentication. In *BIOID*, 2008.

[20] N. Vu and A. Caplier. Illumination-robust face recognition using retina modeling. In *Proceedings of ICIP*. IEEE Press, 2009.

[21] J. Trefnỳ and J. Matas. Extended set of local binary patterns for rapid object detection. In *Proceedings of CVWW*, 2010.

[22] C. Chan, J. Kittler, and K. Messer. Multispectral local binary pattern histogram for component-based color face verification. In *Proceedings of BTAS*. IEEE, 2007.

[23] N. Poh, J. Kittler, and F.M. Alkoot. A discriminative parametric approach to video-based score-level fusion for biometric authentication. In *Proceedings of ICPR*, 2012.