Face Recognition in Unconstrained Environments: A Comparative Study

Rodrigo Verschae, Javier Ruiz-del-Solar and Mauricio Correa

Department of Electrical Engineering, Universidad de Chile
{rverscha,jruizd}@ing.uchile.cl

Abstract. The development of face recognition methods for unconstrained environments is a challenging problem. The aim of this work is to carry out a comparative study of existing face recognition methods that are suitable to work properly in these environments. The analyzed methods are selected by considering their performance in former comparative studies, in addition to be real-time, to require just one image per person, and to be fully online (no requirements of offline enrollment). The methods are compared using the LFW database, which was built to evaluate face recognition methods in real-world conditions. The results of this comparative study are intended to be a guide for developers of face recognition systems.

Keywords: Unconstrained Face Recognition, Face Recognition in Real-Life Images.

1 Introduction

Face information is by far the most used visual cue employed by humans. There is evidence of specialized processing units for face analysis in our visual system [1]. Face analysis allows localization and identification of other humans, as well as interaction and visual communication with them. Therefore, computational face analysis plays a key role in building these functionalities, and in allowing humans to interact with computational systems in a natural way. Currently, computational face analysis is a very lively and expanding research field. Face recognition, i.e. the specific process for determining the identity of an individual contained in an image area that has been already identified as containing a face (by a face detection system) and already aligned (by a face alignment process which usually includes eye detection), is a key functionality. Many different face recognition approaches have been developed in the last few years [2][3][4], ranging from classical Eigenspace-based methods (e.g. eigenfaces [5]), to sophisticated systems based on thermal information, high-resolution images or 3D models. Many of these methods are well suited to specific requirements of applications such as biometry, surveillance, or security. However, we are interested in demanding applications such as search of faces in non-annotated or partially annotated databases (i.e. news databases, the

Internet, etc) and HRI (Human-Robot Interaction), which require real-time operation, just one image per person, to be fully online (no requirements of offline enrollment), and that consider unconstrained environmental conditions.

In this general context, the aim of this paper is to carry out a comparative study of face-recognition methods under these requirements. The main motivation is the lack of direct and detailed comparisons of these kind of methods under the same conditions. The results of this comparative study are a guide for the developers of face recognition systems. As mentioned, we concentrate ourselves on methods that fulfill the following requirements: (i) *Full online operation*: No offline enrollment stages. All processes must be run online. The systems has to be able to build the face database from scratch incrementally; (ii) *Real-time operation* for achieving user interaction with low delays. The whole face analysis process, which includes detection, alignment and recognition, should run at least at 5fps; (iii) *One single image per person problem*: One face image of an individual should be enough for his/her later identification. Databases containing just one face image per person should be considered. The main reasons are savings in storage and computational costs, and the impossibility of obtaining more than one face image from a given individual in certain situations; and (iv) *Unconstrained environments*: No restrictions over environmental conditions such as scale, pose, lighting, focus, resolution, facial expression, accessories, makeup, occlusions, background and photographic quality are required.

This study analyzes four face recognition methods that are based on different representations of the image: (1) LBP Histograms, (2) Gabor Jet Descriptors (GJD), (3) SIFT Descriptors, and (4) ERCF (Extremely Randomized Clustering Forest) of SIFT Descriptors. These representations are used in different ways by the analyzed face recognition methods: (1) LBP Histograms are directly used as feature vectors together with distance metrics for comparing these histograms, (2) GJD are used together with Borda count, (3) SIFT Descriptors are used together with local and global matching methods, and (4) ERCF are used together with linear classifiers. These methods were selected by considering their fulfillment of the requirements mentioned above, and their performance in former comparative studies of face recognition methods [6][7][13][15].

The comparative study is carried out using the LFW database [14] (the restricted setting was considered, see below for details). Aspects such as scale, pose, lighting, focus, resolution, facial expression, accessories, makeup, occlusions, background and photographic quality are implicitly considered in this database. As part of this study we also consider the effect of an alignment algorithm, called funneling [12], which has improved the results of recognition systems, and which has reported results in the LFW database. In addition, for some of the best working methods, we also analyze the effect of variations in the size of the region containing the face (i.e. the amount of the background/context) on the recognition performance.

This paper is structured as follows. The methods under analysis are described in section 2. In sections 3 and 4 the comparative analysis of these methods is presented. Finally, conclusions of this work are given in section 4.

2 Methods Under Comparison

As mentioned above, the algorithms' selection criteria are their fulfillment of the defined requirements, and their performance in former comparative studies of face-recognition methods [3][6][7][13][15]. The first issue to be mentioned is that most of the holistic methods, which are normally based on eigenspace-decompositions, fail when just one image per person is available, mainly because they have difficulties building the required representation models. Although, this difficulty can be overcome if a generalized face representation is built, for instance using a generalized PCA model, for cases where the images are poorly aligned, where the face background is highly inhomogeneous, or the faces present very different expressions, as in the LFW database, the performance can be quite low [15]. For this reason holistic eigenspace-based methods are not considered in this study.

In general terms, local-matching methods behave well when just one image per person is available, and some of them have presented very good results in standard databases such as FERET. Taking into account the results of [7], and our requirements of high-speed operation, we selected two methods to be analyzed. The first one is based on the use of histograms of LBP features, and the second one is based on the use of Gabor filters and Borda count classifiers. Moreover, local interest points and descriptors (e.g. SIFT) have been used successfully for solving some similar wide baseline-matching problems as fingerprint verification [9], and as a first stage of a complex face-recognition system [10]. Therefore, we decided to test the suitability of a SIFT-based face recognition system. Finally, we also considered ERCF, a recently presented method [13] that was designed to classify pairs of images as corresponding to the same object or not. The results of this method on the LFW database were quite good [16], therefore we compare it with the previously mentioned methods. Thus, the methods under comparison are:

- *LBP Histograms*. A local-appearance-based approach with a single, spatially enhanced feature histogram for global information representation is described in [8]. In that approach, three different levels of locality are defined: pixel level, regional level and holistic level. The first two levels of locality are realized by dividing the face image into small regions from which LBP features are extracted for efficient texture information representation. The holistic level of locality, i.e. the global description of the face, is obtained by concatenating the regional LBP extracted features. The recognition is performed in the computed feature space, using one of the three following similarity measures: histogram intersection, log-likelihood statistic and Chi square. We implemented that recognition system without considering preprocessing (cropping, using an elliptical mask and histogram equalization are used in [8]). We have chosen the following parameters: (i) images divided in 10 (2x5), 40 (4x10) or 80 (4x20) regions, instead of using the original divisions which range from 16 (4x4) to 256 (16x16), and (ii) the Chi-square, and the mean square error as a similarity measure, instead of the log-likelihood statistic. Thus, considering the 3 different image divisions and the 2 different similarity measures, we get 6 flavors of this face-recognition method. In addition, we have considered several different windows sizes, which consider different amounts of background.

- *Gabor Jets Descriptors*. Different local-matching approaches for face recognition are compared in [7]. The study analyzes several local feature representations, classification methods, and combinations of classifier alternatives. Taking into account the results of that study, the authors implemented a system that integrates the best choice in each processing step. That system uses Gabor jets as local features, which are uniformly distributed over the images, one wave-length apart. In each grid position of the test and gallery image and at each scale (multiscale analysis) the Gabor jets are compared using normalized inner products, and these results are combined using the Borda Count method. Given that the LFW database only requires comparing pairs of faces, and that an important part of this method is the ranking done using Borda count, we had to adapt it to be able to work on pairs of images. This was done by means of a reference image set of faces, by ranking one of the input faces against these reference images plus the second input face using Borda count. The relative ranking obtained by the second input face, with respect to the reference faces, was considered as a measure of the similarity between the pair of input images. To obtain a symmetric dissimilarity measure, we repeated the same procedure by switching the roles of the first and the second face image, and then averaging the two obtained rankings. The average value was taken as the final similarity measure of the pair of images. The reference set of images was build by randomly selecting from the training set (e.g. 50 faces).We considered three sizes of reference image sets: 10, 50 and 100 faces. To show the importance of using Borda count, results using the Euclidean distance between the GJD descriptors are also given. In the Gabor feature representation, only the magnitude component of the filters is used, and 5 scales and 8 orientations of the filters are adopted. We implemented this system using all parameters described in [7] (filter frequencies and orientations, grid positions, face image size).

- *SIFT descriptors*. Wide baseline-matching approaches based on local interest points and descriptors have become increasingly popular and have experienced an impressive development in recent years. Typically, local interest points are extracted independently from both a test and a reference image, and then characterized by invariant descriptors, and finally the descriptors are matched until a given transformation between the two images is obtained. Lowe's system [11] using SIFT descriptors and a probabilistic hypothesis rejection-stage is a popular choice for implementing object-recognition systems, given its recognition capabilities, and near real-time operation. However, Lowe's system's main drawback is the large number of false positive detections. This drawback can be overcome by the use of several hypothesis rejection stages, as for example in the L&R system [9]. This system has already been used in the construction of robust fingerprint verification systems [9]. Here, we have used the same method for building a face-recognition system, with two different flavors. In the first one, *Full*, all verification stages defined in [9] are used, while in a second one, *Simple*, just the probabilistic hypothesis rejection stages are employed.

- *ERCF*. In [13] it is proposed to learn a similarity measure for comparing pairs of object images. The method is meant to be used in object recognition problems and makes use of Extremely Randomized Clustering Forest (ERCF) and SIFT. The authors propose to learn a similarity function to discriminate whether the pair of images corresponds to the same object or not. The learning is done for specific object classes, such as frontal faces or specific views of cars. The method basically consists of three stages. In the first stage, pairs of similar patches, measured in terms of a normalized cross-correlation, are selected. In the second stage each pair of patches is coded (quantize) by means of ERCF. In the third stage, the quantized pairs of patches are used to build a feature vector, which is finally used to evaluate the similarity of the image pair using a liner classifier.

We use the following notation to refer to the methods and their variations: A-B-C, where A describes the name of the face-recognition algorithm (H: Histogram of LBP features, GJD: Gabor Jets Descriptors, SD: L&R system with SIFT descriptors, ERCF: Extremely Randomized Clustering Forest); B denotes the similarity measure (MSE: Mean square error, XS: Chi square, BC: Borda Count, EU: Euclidian Distance); and C describes additional parameters of each algorithm (H: Number of image divisions, GJD: number of reference images: 10, 50 or 100; SD: verification procedure, *full* or *simple*).

In the first experiments images were cropped to 100x185 pixels, with the bounding box centered in the image (the input images are of 250x250 pixels). No scaling was applied. In the following experiments the faces were also aligned. In the last experiments, the faces were cropped considering larger and smaller bounding boxes, which include different amounts of background (see Figure 1).

(a) (b) (c) (d) (e) (f)

Figure 1: Example faces. (a) and (d): 100x185, unaligned; (b) and (e): 100x185, aligned; (c) and (f): 81x150, aligned.

3 Comparative Study using the LFW Database

The LFW database [14] consist of 13,233 images faces of 5,749 different persons, obtained from news images by means of a face detector. These images have a very large degree of variability in the face expression, age, race, background and illumination conditions. Also, unlike other databases, the recognition is only to be done comparing by pairs, instead of searching for the most similar face in the database. The idea is that the algorithm being evaluated is given a pair of images and it has to output whether the two images correspond to the same person or not. There are two evaluation settings already defined by the authors of the LFW: the image restricted setting and the image unrestricted setting. The image restricted setting is the most difficult one, and it is the one considered here. Under this setting the only information that the algorithm can use is the image pair, no information of the identity of the faces in the images can be used, i.e., the algorithm is restricted to work only using the image pair at hand. The systems are trained (if required) and evaluated using a 10-fold validation procedure, where the folds are symmetric in the sense that the number of matching pairs and non-matching pairs is the same. See [14] for details.

In the following sections different face region sizes are considered. In the experiments of sections 3.1 and 3.2 all methods are evaluated using a face region size of 100x185, which was selected taking into account the results of [15]. In these experiments we analyze and compare two cases, unaligned and aligned faces. Afterwards, in section 3.3, the LBP based method is evaluated, and several experiments are done considering different region sizes. In the following, all results referring to those of ERCF consider complete images (250x250) and correspond to those presented in [16]. We repeated the experiments presented it [16] and similar results were obtained.

3.1 Experiments using unaligned faces

Table 1 (second and third columns) shows the results for all methods under comparison in the unaligned LFW database. In all cases (except for ERCF), regions of 100x185 pixels containing the centered face in the 250x250 image were cropped. As it can be observed, the results obtained with our own implementation of the methods are consistent with those of other studies results (in terms of the relative order of the classification accuracy). However the accuracies are much lower, going form 60% to 72%, values that show the difficulty of the database at hand (e.g. in previous studies [7][8] using the FERET database, the performance was over 95%).

In the case of LBP based methods, best results are obtained with H-x-80, i.e. when using the largest number of divisions. The difference between using the Chi-Square and the Mean Square Error is not significant, although the Chi-Square measure gives slightly better results in all cases. For the method based on the GJD, best results are obtained when using the proposed Borda count methodology (increases the performance in circa 2% over the Euclidean distance); and 100 reference images gives slightly better results. Both methods based on SD present lower performance results (about 60-62%). In the Table 1, results for ERCF, taken from [16], are also included. Its performance is quite good, being 4% larger than the second best method (GJD-BC-100).

3.2 Experiments using aligned faces

Results for aligned faces are presented. As mentioned before, the faces were aligned using the funneling algorithm [12]. Funneling is an unsupervised algorithm for object alignment based on the concept of congealing. Congealing basically consists of searching a sequence of transformations (in this case affine transforms and translations) that are applied to a set of images in other minimize an entropy measure on the set of images. After having built the congealing model, the transformations can be applied to an unseen image (funneling it) to obtain an aligned image. The main advantage of this method is that it can work in complex objects and that it does not requires any labeling during training.

Table 1 (last two columns) shows the results for all methods under comparison using aligned faces. As in the case of unaligned faces, (except for ERCF), the face region was cropped considering a region of 100x185 pixels centered in the 250x250 image. Compared to the case of unaligned faces, all methods, but GJD, improved or maintained their performance. The method presenting the greater improvement is the one based on LBP Histograms, which improves between 2% and 3% depending on the variant being considered. Again, in the case of LBP based methods, best results are obtained with H-x-80, i.e. when using the largest number of divisions, and the Chi-Square distance measure, with a performance similar to GJD in the case of unaligned faces. For the method based on the GJD, best results are obtained when Borda Count is used (increases the performance in circa 3% over the Euclidean distance), and 100 reference images gives slightly better results. However, in this case, the results were slightly worst than the ones obtained for the case of unaligned faces. Again, best results are obtained by ERCF, but this time being about 5% over the second best method.

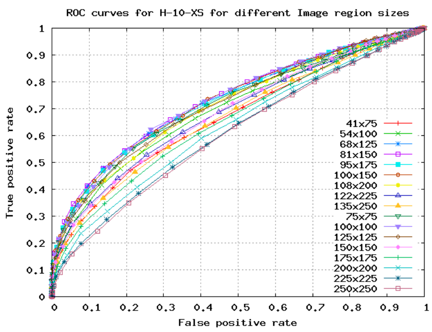
Table 1. Correct classification rates (LFW database, restricted setting). Experiments were performed on cropped regions of size 100x185, except for ERCF which considered the full image. MCA: *Mean classification accuracy.* SME: *Standard error of the mean.*

	Without alignment		**With alignment (funneling)**	
Method	**MCA**	**SME**	**MCA**	**SME**
H-MSE-10	0.6375	0.0049	**0.6585**	0.0046
H-XS-10	0.6500	0.0043	**0.6668**	0.0044
H-MSE-40	0.6217	0.0055	**0.6527**	0.0057
H-XS-40	0.6383	0.0064	**0.6650**	0.0059
H-MSE-80	0.6527	0.0047	**0.6725**	0.0032
H-XS-80	0.6532	0.0053	**0.6785**	0.0055
GJD-EU	**0.6410**	0.0084	0.6375	0.0071
GJD-BC-10	**0.6777**	0.0080	0.6753	0.0082
GJD-BC-50	**0.6770**	0.0075	0.6742	0.0061
GJD-BC-100	**0.6798**	0.0065	0.6762	0.0069
SD-SIMPLE	0.6015	0.0049	**0.6215**	0.0036
SD-FULL	**0.6295**	0.0071	0.6288	0.0051
ERCF [13] (results from [16])	0.7245	0.0040	**0.7333**	0.0060

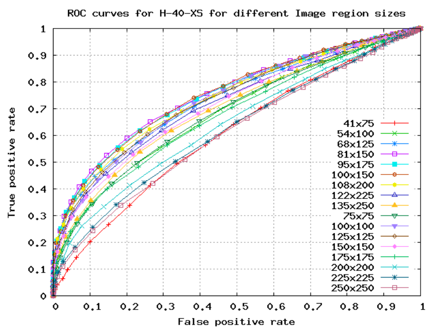
3.3 Experiments using different windows sizes

In this section we analyze the effect of using different region sizes in the performance of the LBP based method, for the case of aligned faces. We have chosen this method for these experiments, because its high processing speed allows to perform this kind of analysis and because it showed the best performance on aligned faces after ERCF. Note that increasing the size of the regions corresponds to adding or removing different amounts of background to the region being analyzed. The experiment were performed considering squared image regions, ranging from 50x50 to 250x250, with a step of 25 pixels and considering regions of ratio 1:1.85 (as in the previous section), ranging from 41x75 to 135x250, with a step of 25 pixels. Results are presented in figures 2 and 3.

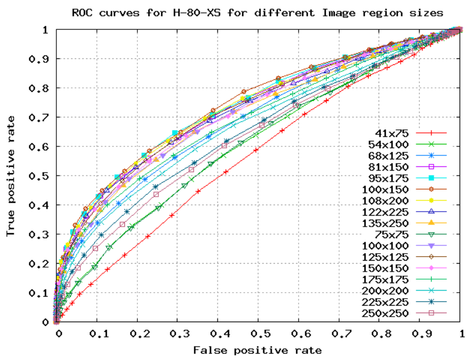
By observing figures 2(a), 2(b) and 2(c), the first thing we can see is the importance of the size of the region on the performance of the algorithm. The second thing is that in all cases (independently of the distance measure and the number of divisions), small region sizes present the worst results, followed by the largest region sizes. Best results are obtained using medium size regions. Figure 3 shows the results for the region size that presented the best overall results. This region size corresponds to 81x150, which contains some background, but not much (see figure 1 a for an example). For this region size, best results are obtained with 40 divisions and worst results with 80 divisions. As in previous experiments, for a fixed number of divisions, the Chi-Square measure works better than the mean square error.



(a)



(b)



(c)

Figure 2. Effect of the image region size on the performance of the method based on LBP Histogram and the Chi square distance. ROCs were obtained using (a) 10, (b) 40, and (c) 80 image divisions. Experiments were performed on faces aligned using funneling.

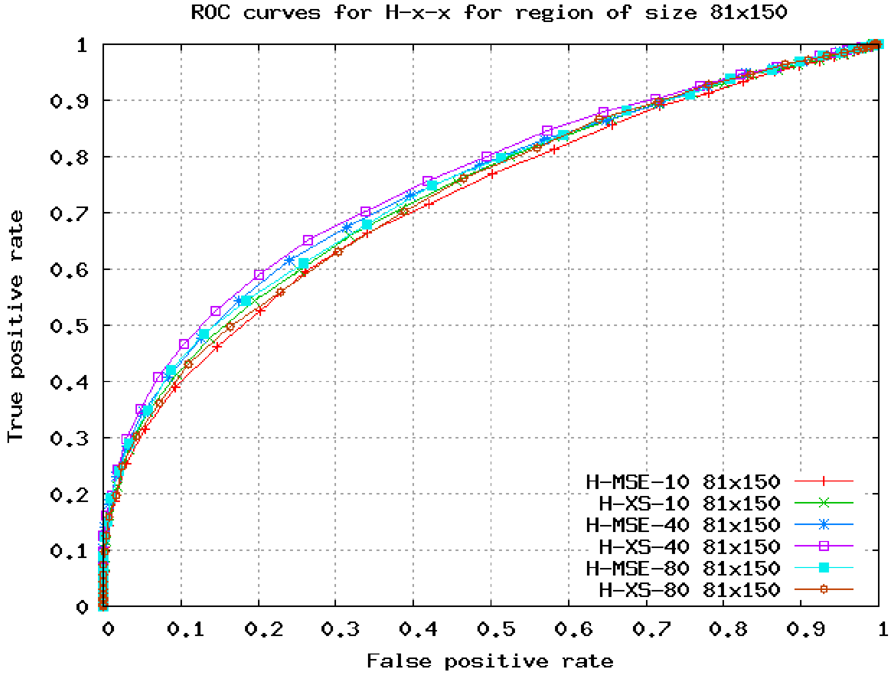


Figure 3. Comparison of the method based on LBP Histogram for a regions sizes of 81x150 when using different distance measures and image divisions. Experiments were performed on faces aligned using funneling.

4 Discussion

If one compares the best LBP method, to the best GJD method and to ERCF only based on their performance, ERCF is clearly better, with LBP being almost 4% below ERCF (see Table 3 and Figure 4) and with LBP 1.7% over GJD. If one now evaluates the processing speed, the LBP method is at least 65 times faster that ERCF (see Table 2), and 10 times faster that GJD with Borda count (for the case of 50 reference images). The slow processing time of ERCF and GJD can be too restrictive for some applications, in particular in the ones that require real-time operation (e.g. HRI), as well in applications where very large amounts of data are being analyzed (e.g. search operation in a very large multimedia database). The slowest part of GJD is to perform the Borda count ranking of each of the features, while the slowest part of the method based on ERCF is the selection of pairs using normalized cross correlations. Any improvement in these methods, reducing these bottlenecks, would be of great help.
We have also performed additional experiments, but for space reasons we are not reporting them in detail, thus we are just giving a brief overview. The first experiment was to use, as [15], local grids for GJD, i.e. grids that include less background. In [15] this improved the results in some cases, but here in all cases the results were worst, by 2% to 5%. The second of these experiments consist of applying ERCF using smaller

region sizes, not just the complete (250x250) image. Preliminary results indicate that the method's performance does not change much, which tells us that its processing time could be partially reduced by considering just the optimal region size. In the third experiment we used LBP histograms and ERCF methods in a cascade setting, using LBP histogram in an operation point with high true detection rate (90% detection rate and 70% false positive rate) in order to use it to filter out "clearly" different faces. The basic idea was that if the errors of ERCF and LBP histograms were correlated, then, by using both methods in cascade, a higher processing speed could be achieved, in particular if it is considered that most of the time, in a typical application, the faces to be compared correspond to different persons. Nevertheless, this idea did not work very well, because the cascade system presented error rates almost equal to the case of the LBP histograms working alone at a operation point with a lower false positive rate.

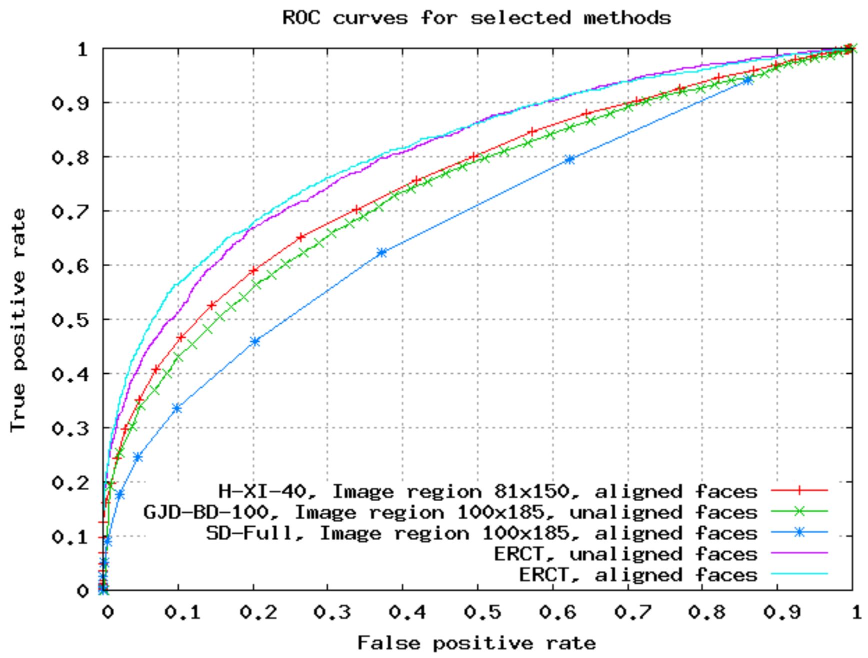


Figure 4. ROC curves of the best working variant of each method. Experiments were performed on faces aligned using funneling. Result for ERCF [13] were taken from [16].

Table 2. Processing Time. Time measures are in milliseconds. An image size of 100x185 pixels is considered. The experiments we carried out on a computer running Linux with an AMD Athlon (tm) XP 2100+ cpu (MHz 1738)

Method	**H**	**H**	**H**	**GJD**	**GJD**	**GJD**	**GJD**	**SD**	**ERCF**
Parameters	**x-10**	**x-40**	**x-80**	**EU**	**BC- 10**	**BC- 50**	**BC-100**	**X**	**From [13]**
Time (milliseconds)	26	30	32	130	210	290	400	150	2000

Table 3. Correct classification rates (LFW database, restricted setting). MCA: *Mean classification accuracy.* SME: *standard error of the mean.*

Method	**Region Size**	**MCA**	**SME**
H-XS-40, aligned faces	100x150	0.6905	0.0047
H-XS-40, aligned faces	81x150	0.6945	0.0048
H-XS-40, aligned faces	95x175	0.6913	0.0044
GJD-40, unaligned faces	100x185	0.6798	0.0065
ERCF , aligned faces [13] (results traken from [16])	250x250	0.7333	0.0060

5 Conclusions

In this article, a comparative study among face-recognition methods was presented. The analyzed methods were selected by considering their suitability for the defined requirements - real-time operation, just one image per person, fully online (no training), and robust behavior in unconstrained environments -, and their performance in former studies. The comparative study was carried out using the LFW database, which includes aspects such as scale, pose, lighting, focus, resolution, facial expression, accessories, makeup, occlusions, background and photographic quality, and the test protocol defined by this database.

The methods under comparison are LBP histograms, Gabor Jets descriptors, SIFT descriptors and ERCF (see descriptions in [7][8][13][15]). ERCF outperforms all methods using aligned and non-aligned images. However, ERCF is about 65 times slower than a LBP variant (Chi square distance and 40 regions image partition), which achieves a mean classification accuracy only ~4% smaller than ERCF (69.5% vs. 73%). The best GJD variant also shows a slightly lower performance, with a 67.8 % correct classification rate in the case of unaligned face, but it is more than 10 times slower than then LBP method. Thus, in applications that require real-time operation (e.g. HRI), as well in applications where very large amounts of data are being analyzed (e.g. search operation in very-large multimedia databases), LBP-based methods can be a very interesting alternative.

As future work we would like to study the behavior of the Gabor Jet descriptors method when using image regions of different sizes, and to analyze in detail the possibility of combining different methods for achieving, at the same time, high classification accuracy and high processing speed.

References

1. P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, Face Recognition by Humans: 19 Results All Computer Vision Researchers Should Know About, *Proc. of the IEEE*, Vol. 94, No. 11, Nov. 2006, pp. 1948-1962
2. W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face Recognition: A Literature Survey, *ACM Computing Surveys*, 2003, pp. 399-458.
3. X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, Face recognition from a single image per person: A survey, *Pattern Recognition*, Vol. 39, pp. 1725–1745, 2006.
4. R. Chellappa, C.L. Wilson, S. Sirohey, Human and Machine Recognition of Faces: A Survey, Proceedings of the IEEE, Vol. 83, Issue 5, May 1995, pp. 705-740.
5. M. Turk, A. Pentland, Eigenfaces for Recognition, *Journal of Cognitive Neurosicence*, Vol. 3, No. 1, 1991, pp. 71-86.
6. J. Ruiz-del-Solar, P. Navarrete, Eigenspace-based face recognition: a comparative study of different approaches, *IEEE Transactions on Systems, Man and Cybernetics, Part C*, Vol. 35, Issue 3, August 2005, pp. 315-325.
7. J. Zou, Q. Ji, G. Nagy, A Comparative Study of Local Matching Approach for Face Recognition, *IEEE Transactions on Image Processing*, Vol. 16, Issue 10, Oct. 2007, pp. 2617-2628.
8. T. Ahonen, A. Hadid, and M. Pietikainen, Face recognition with local binary patterns, *European Conference on Computer Vision – ECCV 2004*, pp. 469–481, 2004.
9. J. Ruiz-del-Solar, P. Loncomilla, and Ch. Devia, A New Approach for Fingerprint Verification based on Wide Baseline Matching using Local Interest Points and Descriptors, *Lecture Notes in Computer Science* 4872 (PSIVT 2007), pp. 586-599, 2007.
10. A.S. Mian, M. Bennamoun, and R. Owens, An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition, IEEE Trans. on Patt. Analysis and Machine Intell., Vol. 29, No. 11, pp. 1927-1943, Nov. 2007.
11. D. Lowe, Distinctive Image Features from Scale-Invariant Keypoints. *Int. Journal of Computer Vision*, 60 (2): 91-110, Nov. 2004.
12. G.B. Huang, V. Jain, E. Learned-Miller, Unsupervised Joint Alignment of Complex Images, IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, vol., no., pp.1-8, 14-21 Oct. 2007.
13. E. Nowak, F. Jurie, Learning Visual Similarity Measures for Comparing Never Seen Objects, IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07. , vol., no., pp.1-8, 17-22 June 2007.
14. Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49, October, 2007.
15. M. Correa, J. Ruiz-del-Solar, F. Bernuy. Face Recognition for Human-Robot Interaction Applications: A Comparative Study. *Lecture Notes in Computer Science* (*RoboCup Symposium 2008*) (in press).
16. Labeled Faces in the Wild database, Results webpage. Available in http://vis-www.cs.umass.edu/lfw/results.html