Image Preprocessing for Appearance-based Face Recognition

César Fernández, M. Asunción Vicente, Ramón P. Ñeco, Rafael Puerto

Miguel Hernández University, Av. de la Universidad s/n,
03202 Elche, Spain
{c.fernandez, suni, ramon.neco, r.puerto}@umh.es

Abstract. The effect of different image preprocessing algorithms on face recognition is evaluated through exhaustive experiments. Principal component analysis is used as a baseline for the comparison; and the preprocessing algorithms considered include image registration, lighting correction and image cropping. The experiments are carried out over commonly used face databases. Two main conclusions are drawn: first, excessively controlled or excessively biased image databases are not valid for evaluating preprocessing algorithms; second, a mere histogram equalization should be preferred over more complex lighting correction techniques which produce a higher loss of image information. The Matlab code used for the experiments, as well as usage information, is available for download.

Keywords: Face recognition, preprocessing, principal component analysis.

1 Introduction

Research on face recognition started around 30 years ago with the initial papers of Turk and Pentland [1]. Even though nowadays face recognition is a mature research topic, there is not a consensus on the best performing techniques. The present paper is focused on the preprocessing step, where different transformations are applied to the original images in order to increase the homogeneity and to minimize the effects of lighting, pose, etc.

Face recognition techniques can be roughly classified in two different categories: appearance-based and interest point-based. Algorithms belonging to the first category use a global description of the image to measure similarity (i.e. to determine whether two images belong to the same person or not). On the other side, interest point-based algorithms describe images considering only their most discriminative areas. Such algorithms measure similarity between two images by evaluating the number and quality of the matches between interest points.

Classical face recognition techniques belong to the first category. By using linear transformations, the information present in the original images is compressed and represented in a different basis (namely, a feature extraction is performed). The most commonly used feature extraction technique is principal component analysis or PCA

[1][2]. Other options include independent component analysis or ICA [3][4][5] and linear discriminant analysis or LDA [6].

Among the techniques belonging to the second category, SIFT (Scale Invariant Feature Transform, [7]) and SURF (Speeded Up Robust Features, [8]) are the most widely used descriptors. Both of them are highly independent of scale, lighting and orientation, and partially independent of pose (or relative orientation between subject and camera). Thus, image preprocessing is not strictly necessary when using such techniques.

Since most interest point based face recognition techniques do not rely on a previous preprocessing step, the present paper will focus on appearance-based face recognition, the goal being the evaluation of the most commonly used image preprocessing algorithms.

2 Baseline used for the experiments

Among the most widely used feature extraction techniques for appearance-based face recognition, we have decided to use PCA for our experiments. Recently, many authors have chosen other techniques, like ICA and LDA; but their advantages have not been fully contrasted and there is not a consensus in the research community concerning which feature extraction method should be preferred. A brief analysis is carried out in the next sections.

2.1 PCA vs. ICA

Concerning ICA and PCA, there are seemingly contradictory results in the literature: the authors of [9], [10], [11], [12], [13], [14], [15] and [16] claim that ICA outperforms PCA in classification systems; but [17] and [18] state that both approaches perform equally; while [19] and [20] claim that PCA is superior to ICA. Finally, [21] suggests that the performance of ICA is very dependent on the data set.

PCA seeks principal components which offer the maximum data variance, under the constraint of orthogonality; while ICA tries to represent the original data over a basis of statistically independent random vectors. A theoretical analysis of PCA and ICA foundations reveals that, under certain circumstances, both methods should perform equally. Most ICA algorithms seek for a matrix where the rows have maximally non-Gaussian distributions and are mutually uncorrelated (to the maximum extent). A simple way to do this is to first whiten the data (using a PCA or singular value decomposition approach) and then to seek for orthogonal non-normal projections: namely, a whitening operation followed by a rotation.

The most widely used classifiers (e.g. those relying in Euclidean distances or angles) are rotation-invariant, so ICA should not offer advantages over a much simpler feature extraction method such as whitening (which in turn is PCA with a further scale normalization step).

This analysis is carried out in higher detail in one of our previous papers [22], including experiments with synthetic and real data. The conclusion drawn is that ICA should not be preferred over PCA.

2.2 PCA vs. LDA

Several authors prefer a supervised feature extraction technique like LDA over an unsupervised one like PCA. While PCA seeks for the directions that better explain the data, LDA seeks for the directions that best discriminate between classes.

However, face recognition is a particular field where the number of training examples is several orders of magnitude lower than the dimension of the data (number of pixels of the images). In such a scenario, LDA is not guaranteed to outperform PCA.

Basically, LDA creates a linear combination of the original features (pixel values) which yields the largest mean differences between the classes (subjects of the face database). Two different matrixes are defined: the within-class scatter matrix and the between-class scatter matrix, the goal being to maximize the between class variance while minimizing the within-class variance. Such a strategy obviously outperforms PCA in classification scenarios where each instance is defined by a low dimension vector.

In order to determine whether LDA also outperforms PCA when images are used (e.g. in a face recognition application), we can rely in a previous paper from Martínez and Kak [23]. In such paper, an exhaustive comparison is carried out using the AR face database [24]. The results show that LDA does not consistently outperform PCA, except for specific cases where the number of training examples is large enough (which is not the case in a common face recognition scenario).

2.3 Conclusion

As a conclusion, we will use PCA for the comparison: ICA or LDA have not proved to perform better and they introduce extra computational load. Whitening (PCA plus a scaling) could have been used instead with similar results. However, we have to keep in mind that the goal of the present paper is not to obtain the best performing feature extraction algorithm (PCA, ICA, LDA or whitening), but to use one of them as a baseline for comparing the performances of different image preprocessing algorithms.

4 Face image databases considered

Five different face databases have been used for the experiments:

- AT&T database [25]: contains 400 gray level images (92x112 pixels) belonging to 40 different subjects (10 images per subject). All images where captured under controlled lighting and fixed distance to the camera. The main variations between images of the same subject are small pose changes and different facial expressions.

- FERET database [26]: contains 14126 images of 1199 subjects where at least two images are available per subject (some subjects have a higher number of images taken in different days, in some cases more than one year apart). Images are 24-bit color, 512x768 pixels. Over images taken in the same day, the main variation is facial expression; over images taken in different days there are much more variations, including pose, hairstyle and even aging. We use a subset of FERET containing only the first 100 subjects.

- YALE database or *extended Yale database B cropped* [27]: contains 16128 gray level images of 28 subjects, with differences in lighting (64 different lightings) and pose (9 different poses), which makes 64x9 = 576 images per subject. Images have been manually aligned, cropped, and then re-sized to 168x192 pixels. It must be stated that the 64 different images per subject and pose are taken almost instantly, so that facial expression can be considered invariant. For our experiments, we use a subset where only 10 images per subject are kept (in our subset, all images show the same pose and only lighting differs; among the 64 possible lighting scenarios we have manually selected 10 clearly different ones).

- LFW database [28]: contains 13233 color images of 5749 subjects, obtained from the internet. The number of images per subject varies from 1 (for the less popular subjects) to 530 (for the most popular subject). There is a high level of variability between images, which makes this database realistic in terms of similarity to a real face recognition scenario. Images have been automatically registered using the Viola-Jones algorithm, and resized to 250x250 pixels. For our experiments, we use a subset of 158 subjects (those subjects with 10 or more images). Besides, we use only the first 10 images of each subject (in order to avoid biasing the classifiers towards the most populated classes) so our subset contains a total of 1580 images.

- UMH database [29]: our own database contains 510 color images of 17 subjects (30 images per subject). Lighting is uncontrolled, and variations between images of the same subject include distance to the camera, pose, background, facial expression and focus (different cameras were used). Apart from that, dark glasses and caps introduce occlusions in some of the images. Image size is 121x151 pixels.

In order to show the kind of images present in our own database (UMH database), figure 1 shows the first image for all 17 subjects; and figure 2 shows the complete set of 30 images for one of the subjects.

As figure 2 shows, some images are captured against a white background while some other images are captured against a non-uniform background; some of them even include occlusions such as dark glasses and caps. It is clear that, even in the first image of each subject (see figure 1), where there are no occlusions and the background is uniform, the capture conditions are less controlled than those of other databases like AT&T, FERET or YALE (in terms of distance to the camera, lighting, focus, etc.).

Fig. 1. First image for each of the 17 subjects of UMH database.

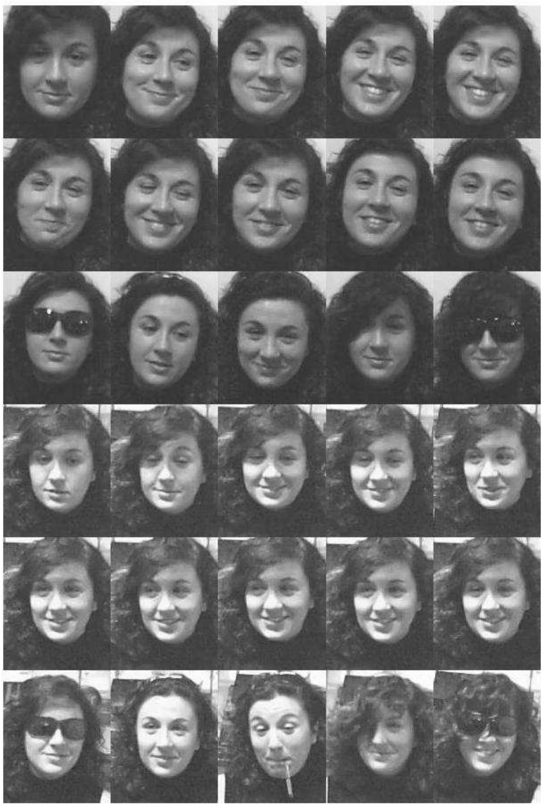

Fig. 2. Set of 30 images for one of the subjects of UMH database.

4 Image preprocessing techniques evaluated

Image preprocessing techniques evaluated fall in three categories:

- Image registration, namely face detection and alignment. We have selected the well-known Viola-Jones algorithm [30] for our experiments, due to its widespread usage and proved performance.

- Lighting correction. We have selected both the basic histogram equalization technique and a more complex filter proposed by Tan and Triggs [31].

- Background suppression (by image cropping). Background suppression is carried out by cropping the images with a mask defined by two elliptical borders, as suggested in [32].

The Viola-Jones algorithm, as well as both lighting correction techniques selected for our experiments are detailed in next sections.

4.1 Viola-Jones algorithm

Viola-Jones algorithm was initially designed to detect any kind of objects (cars, faces, etc); where the kind of object detected depends just on the training images used. Nowadays, it is widely used for frontal face detection, which performs reliably and fast with independence of scale and lighting.

Viola-Jones works by evaluating a huge number of extremely simple features (160000 features per sub-window of 24x24 pixels) over the image. Features are computed as additions and subtractions of pixel values corresponding to different areas; and the concept of integral image is used to speed up the process.

A simple classifier (a perceptron with only one neuron) is adjusted for every feature, and Adaboost [33] is used in order to avoid overfitting.

Training is carried out with a large dataset of correct images (containing the object of interest) and wrong images (not containing it). The training process is extremely slow (it may take days for a standard computer); but once the system is trained, the on-line computing time is below 50ms for a 300x300 image. In order to achieve such processing speed, a cascade of classifiers is used.

4.2 Histogram equalization

Histogram equalization is one of the most common preprocessing techniques. It is used both for improving the visual appearance of images (by adjusting the contrast) and for making images more homogeneous in terms of lighting.

Basically, the cumulative histogram of the image is linearized, thus making contrast more homogeneous in all the gray level range. The goal when applied as preprocessing in face recognition is to compensate for differences in lighting between images.

4.3 Tan-Triggs filter

The objective of the Tan-Triggs filter is to make images as independent as possible from lighting, even in the presence of shadows. It is a three step process:

- First, a gamma correction is performed by applying a power law to the pixel intensity values (by default, the exponent is set to 0.2). The goal is to make the resulting image independent of the overall illumination intensity.
- Then, a difference of Gaussians (DoG) filter is applied. The goal is to reduce the effect of shadows in the images by band-pass filtering them. By default, the standard deviations of the inner and outer Gaussians are 1.0 and 2.0 pixels.
- Finally, image intensities are rescaled by contrast equalization.

5 Experimental results

The preprocessing algorithms will be compared in terms of percentage of correct classifications. First, the image database is split in a training subset and a testing subset. Then, PCA is applied to the training subset in order to determine the directions (features) that will be used to compress the data. Then, all images (training and testing) are represented in the basis defined by such directions, and a nearest neighbor algorithm is used to classify the test images. The percentage of correct classifications (test images classified as corresponding to the correct subject) is obtained. The Matlab code used for the experiments, as well as usage information, are available at [34].

The same process is carried out for original images and for images where different preprocessing algorithms have been applied; and all the experiments are run for all image databases considered (AT&T, FERET, YALE, LFW and UMH).

All the possible preprocessing combinations are tested, which results in 12 different scenarios, which are be coded according to the following order:

- 2 options for image registration: no preprocessing (0) or Viola-Jones algorithm (1).
- 3 options for lighting correction: no preprocessing (0), histogram equalization (1) or Tan-Triggs algorithm (2).
- 2 options for image cropping: no preprocessing (0) or cropping by elliptical borders (1).

A complete preprocessing for an image is coded as *nmp*, where n represents the image registration performed; *m* represents the lighting correction applied and *p* represents the image cropping carried out (e.g. code *110* corresponds to first applying the Viola-Jones algorithm, then performing an histogram equalization and, finally, not cropping the image).

Figures 3 and 4 show the results obtained when all possible preprocessing combinations are applied to one of the images belonging to the UMH database. In figure 3 the processing codes (from left to right) are: 000, 001, 010, 011, 020, 021 (note: all possible combinations without image registration); in figure 4 the processing codes are: 100, 101, 110, 111, 120, 121 (all possible combinations with an initial image registration).

Fig. 3. Example of preprocessing combinations without image registration.

Fig. 4. Example of preprocessing combinations with an initial image registration.

5.1 First set of experiments

In the first set of experiments, one image is used for training and one image is used for testing. The number of PCA components kept (a measure of the level of data compression) varies from 1 to the maximum possible value. The results are shown in figures 5 to 9.

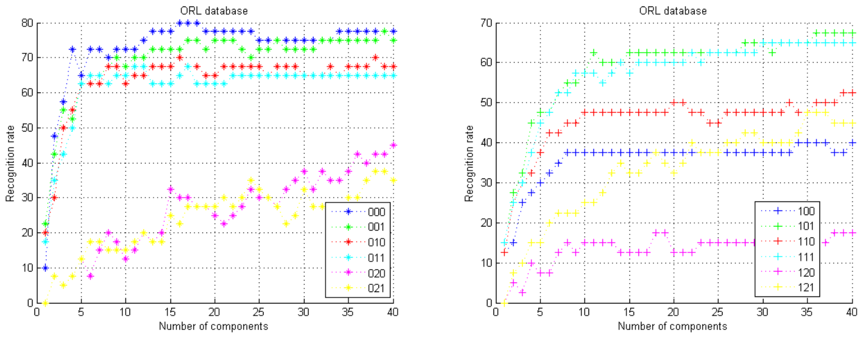

Fig. 5. Recognition rates for AT&T database against number of PCA components.

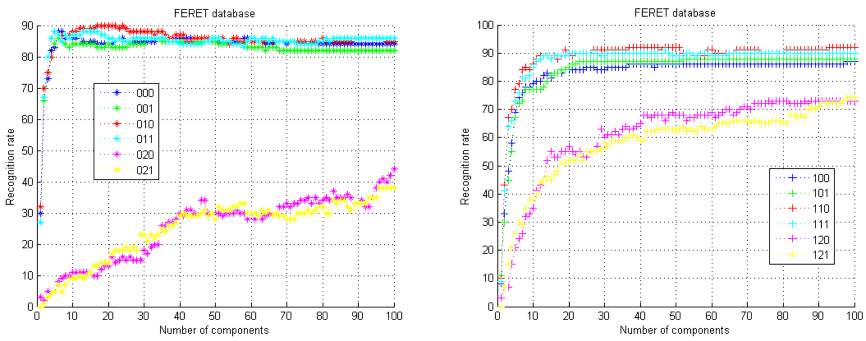

Fig. 6. Recognition rates for FERET database against number of PCA components.

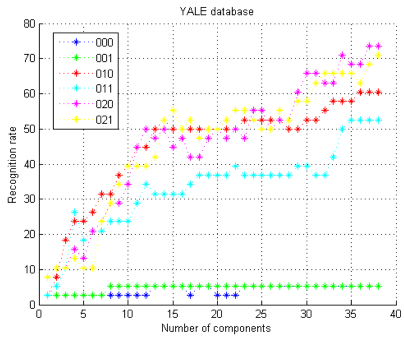

Fig. 7. Recognition rates for YALE database against number of PCA components. No registration is carried out because the database images are already cropped.

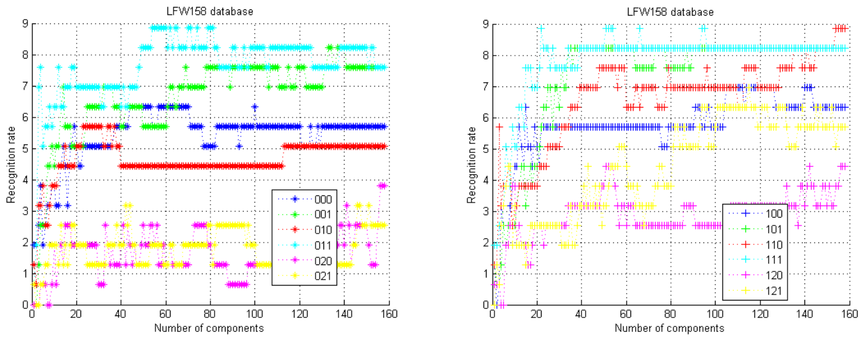

Fig. 8. Recognition rates for LFW database against number of PCA components.

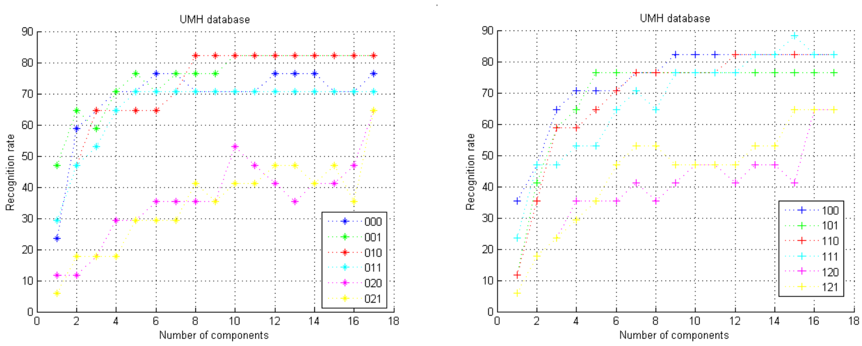

Fig. 9. Recognition rates for UMH database against number of PCA components.

Different conclusions can be drawn from the results:

- First: the behavior depends to a great extent on the database. AT&T database represents an example of highly controlled capture environment; under such circumstances image preprocessing is not needed and may worsen the results. In fact, the combination *000* is the best option (see [35] for a related analysis of AT&T and other databases). FERET database shows a similar behavior. On the other hand, YALE database, though controlled, shows extreme variations in lighting and thus requires lighting correction. Finally, LFW and UMH are examples of uncontrolled databases, so results in such databases may be closer to those expected in the application of face recognition to a real environment.

- Second: focusing on LFW and UMH databases, the best option seems to be image registration followed by histogram equalization, though the improvement in the results as compared to no preprocessing at all is not as high as expected.
- Third: Tan-Triggs algorithm performs extremely well with the YALE database and offers poor results with the other databases. The reason may be found in the amount of image information that is lost during the process, which can only be compensated when lighting variations are extreme. In most cases, a mere histogram equalization should be preferred.
- Fourth: image cropping by elliptical borders seems to improve results in most cases.

5.2 Second set of experiments

A second set of experiments has been carried out, the main difference being the number of test images: one image per subject is used for training, as before, but now all remaining images of the same subject are used for testing.

In the previous set of experiments, the best performing results were usually obtained when the maximum number of PCA components were used. In order to improve the readability of the figures, in this second set of experiments the results are only shown for this maximum number of components. Figures 10 to 14 show the results obtained.

In the figures, experiments are numbered from 1 (preprocessing code 000) to 12 (preprocessing code 121), except for the YALE database, were images are already cropped and we have decided not to perform experiments with image registration. The results show both the average recognition rate and the standard deviation of the results.

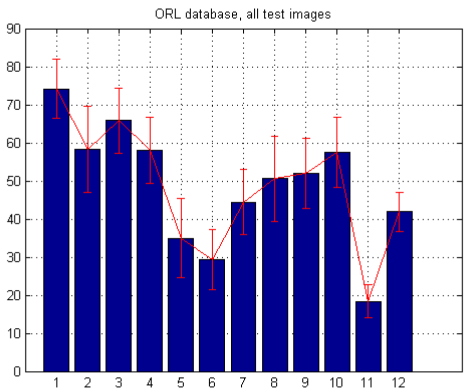

Fig. 10. Recognition rates for AT&T database, 1 training image and 9 test images.

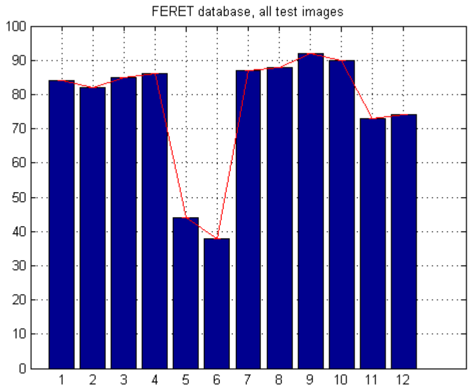

Fig. 11. Recognition rates for FERET database, 1 training image and 1 test image.

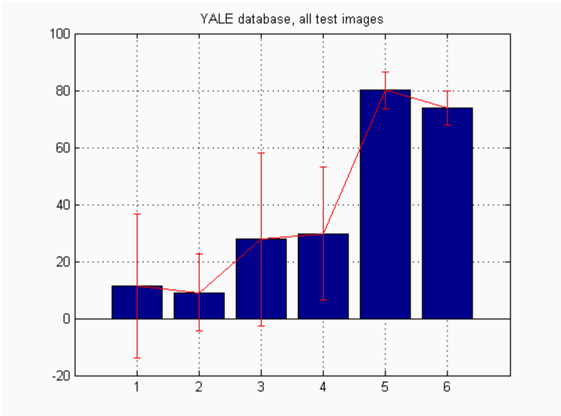

Fig. 12. Recognition rates for YALE database, 1 training image and 9 test images (no registration is performed because the images are already cropped).

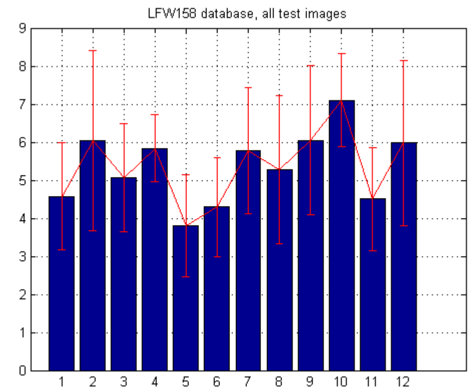

Fig. 13. Recognition rates for LFW database, 1 training image and 9 test images.

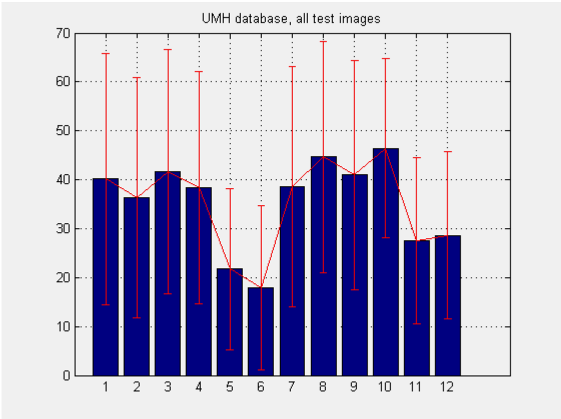

Fig. 14. Recognition rates for UMH database, 1 training image and 29 test images.

According to the results, similar conclusions as those of section 5.1 can be drawn. AT&T database is not a good test-bench for preprocessing algorithms, as images were captured under extremely controlled conditions. YALE cannot be considered valid either, as results on this database are always biased towards algorithms performing a high level of lighting correction, even if this supposes a considerable loss of image information. LFW, UMH and FERET (in this second set of experiments) seem to be more valid test-benches.

On the abovementioned three databases, the best preprocessing combination includes image registration through the Viola-Jones algorithm followed by lighting correction through histogram equalization and, finally, image cropping using elliptical borders.

6 Conclusions

As opposed to algorithms based on interest points, appearance based face recognition algorithms are highly dependent on image preprocessing.

There is not a consensus on which image preprocessing algorithms perform best, so an exhaustive experimental analysis can help in identifying the best techniques.

Not all the commonly used face databases can be considered valid for evaluating image preprocessing algorithms. In particular, AT&T database images were captured under excessively controlled conditions and thus image preprocessing is not needed. On the other side, YALE database is biased towards algorithms that perform a high level of lighting corrections (e.g. Tan-Triggs algorithm) even at the expense of a high loss of image information. Such algorithms do not perform well on other databases.

Using the results obtained on the remaining databases, our recommendation for image preprocessing includes a first registration step (e.g. Viola-Jones algorithm) followed by histogram equalization (which shows a good compromise between lighting correction and information loss) and finally, image cropping trough predefined elliptical borders.

Acknowledgments. The authors wish to thank *Ministerio de Ciencia e Innovación*, for the funding provided for this research, in the frame of project TIN2010-17513.

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. J.Cognitive Neuroscience. 3, 1, 71—86 (1991)
2. Kirby, M., Sirovich, L: Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans. Pattern Analysis and Machine Intelligence. 12, 1, 103--108 (1990)
3. Bartlett, M.S., Movellan, J.R., Sejnowski, J.: Face recognition by independent component analysis. IEEE. Trans. Neural Networks. 13, 6, 1450--1464 (2002)
4. Draper, B., Baek, K., Bartlett, M.S., Beveridge, R.: Recognizing faces with PCA and ICA. Computer Vision and Image Understanding. 91, 1, 115--137 (2003)
5. Liu, C.: Face Enhanced independent component analysis and its application to content based face image retrieval. IEEE. Trans. Systems, Man, and Cybernetics, Part B: Cybernetics. 34, 2, 1117--1127 (2004)
6. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Analysis and Machine Intelligence. 19, 7, 711--720 (2002)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 60, 2, 91--110 (2004)
8. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. Proceedings of the 9th International Conference on Computer Vision, pp. 404--417 (2006)
9. Bartlett, M.S., Movellan, J.R., Sejnowski, J.: Face Recognition by Independent Component Analysis. IEEE. Trans. Neural Networks. 13, 6, 1450--1464 (2002)
10. Draper, B., Baek, K., Bartlett, M.S., Beveridge, R.: Recognizing Faces with PCA and ICA. Computer Vision and Image Understanding. 91, 1-2, 115--137 (2003)
11. Ekenel H.K., Sankur, B.: Multiresolution Face Recognition. Image and Vision Computing. 23, 469--477 (2005)
12. Fortuna, J., Shuurman, D., Capson, D.: A Comparison of PCA and ICA for Object Recognition under Varying Illumination. Proc. 16th Int'l Conf. Pattern Recognition. 3, 11--15 (2002)
13. Martiriggiano, T., Leo, M., D'Orazio, T., Distante, A.: Face Recognition by Kernel Independent Component Analysis. LNAI. 3533, 55--58 (2005)
14. Socolinsky, D.A., Selinger, A.: A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery. Proc. 16th Int'l Conf. Pattern Recognition (2002)
15. Vicente, M.A., Fernández, C., Reinoso, O., Payá, L.: 3D Object Recognition from Appearance: PCA versus ICA Approaches. LNCS. 3211, 547--555 (2004)
16. Yuen, P.C., Lai, J.H.: Independent Component Analysis of Face Images. Proc. IEEE Workshop Biologically Motivated Computer Vision, 545--553 (2000)
17. Moghaddam, B.: Principal Manifolds and Probabilistic Subspaces for Visual Recognition. IEEE Trans. Pattern Analysis and Machine Intelligence. 24, 6, 1--10 (2002)
18. Yang, J., Zhang, D., Yang, J.: Is ICA Significantly Better than PCA for Face Recognition. Proc. IEEE Int'l Conf. Computer Vision. 198--203 (2005)
19. Baek, K., Draper, B.A., Beveridge, J.R., She, K.: PCA vs. ICA: A Comparison on the FERET Data Set. Proc. Sixth Joint Conf. Information Sciences. 824--827 (2002)

20. Fortuna, J., Quick, P., Capson, D.: A Comparison of Subspace Methods for Accurate Position Measurement. Proc. Sixth IEEE Southwest Symp. Image Analysis and Interpretation (2004)
21. Sahambi, H.S., Khorasani, K.: A Neural-Network Appearance-Based 3-D Object Recognition Using Independent Component Analysis. IEEE Trans. Neural Networks. 14, 1, 138--149 (2003)
22. Vicente, M.A., Hoyer, P.O., Hyvarinen, A.: Equivalence of Some Common Linear Feature Extraction Techniques for Appearance-Based Object Recognition Tasks. IEEE Trans. Pattern Analysis and Machine Intelligence. 29, 5, 896--900 (2007)
23. Martinez, A.M., Kak, A.C.: PCA versus LDA. IEEE Trans. Pattern Analysis and Machine Intelligence. 23, 2, 228--233 (2001)
24. Martinez, A.M., Benavente, R.: The AR Face Database. CVC Technical Report #24 (1998)
25. AT&T, The Database of Faces (formerly known as “The ORL Database of Faces”). http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.
26. Phillips, P.J., Moon, H., Rauss, P.J., Rizvi, S.: The FERET evaluation methodology for face recognition algorithms. IEEE Trans. Pattern Analysis and Machine Intelligence. 22, 10, 1090--1104 (2000)
27. Lee, K.C., Ho, J., Kriegman, D.: Acquiring Linear Subspaces for Face Recognition under Variable Lighting. IEEE Trans. Pattern Anal. Mach. Intelligence. 27, 5, 68--698 (2005)
28. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49 (2007)
29. UMH face database. http://lcsi.umh.es/investigacion/facerecognition/face-database-umh.
30. Viola, P., Jones, M.J.: Robust Real-Tine Face Detection. International Journal of Computer Vision, 57, 2, 137--154 (2004)
31. Tan, X., Triggs, B.: Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. IEEE Trans. Image Processing. 19, 6, 1635--1650 (2010)
32. Chen, L.F., Liao, H.Y., Lin, J.C., Han, C.C.: Why recognition in a statistic-based face recognition system should be based on the pure face portion: a probabilistic decision-based proof. Pattern Recognition. 34, 1393--1403 (2001)
33. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55, 1, 119--139 (1997)
34. Matlab code for PCA experiments. http://coolab.umh.es/ica_pca/index_en.html
35. Shamir, L.: Evaluation of Face Datasets as Tools for Assessing the Performance of Face Recognition Methods. Int. J. Comput. Vis. 79, 22--230 (2008)