Face Recognition by Local and Global Analysis

Massimo Tistarelli, Andrea Lagorio and Enrico Grosso

Computer Vision Laboratory, DAP,
University of Sassari, Alghero (SS), 07041, Italy

Abstract

Faces are highly deformable objects which may easily change their appearance over time. Not all face areas are subject to the same variability and they do not have the same relevance for recognition. Therefore, selecting and decoupling the information from independent areas of the face is of paramount importance to improve the robustness of any face recognition technique. In forensic applications it is rather important to identify an individual by peculiar, subjective features, which uniquely characterize his/her face. This paper discusses how to select relevant local features on the face and use these features to uniquely identify a subject. For identification purposes, both a global and local (as recognition from parts) matching strategy is proposed. The local strategy is based on matching individual salient facial SIFT features as connected to selected facial landmarks. As for the global matching strategy, relevant SIFT features are combined together to form a single feature.

1. Introduction

Face recognition is one of most challenging research areas in biometrics as well as computer vision [1,2]. The variability in the appearance of face images, either due to intrinsic and extrinsic factors, makes the identification problem ill-posed and difficult to solve. Moreover, additional complexities like the data dimensionality and the motion of face parts causes major changes in appearance. In order to make the problem well-posed, vision researchers have adapted and applied an abundance of algorithms for pattern classification, recognition and learning. To cope for the data dimensionality, several appearance-based techniques have been successfully used, such as the Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Fisher Discriminant Analysis (FDA), and Independent Component Analysis (ICA) [1]. Other methods have been studied based on the extraction of salient facial features by means of cascaded scale-space filtering [3-6].

A largely underestimated problem is the relevance of physical details in the characterization of human faces. While the geometry of facial components (such as the eyes and mouth) allow to provide a general characterization of the face, the identification of a subject requires to analyze fine details and combine them. In most of the approaches proposed in the literature, these details are related as "facial features" [7-12]. These approaches are different in nature, robustness, computational requirements, accuracy, but they all extract salient regions from a face by analyzing *only* the considered subject. In order to distinguish different faces it is required to define the facial features, which are mostly subject to change among different faces. In a recent paper we proposed a computational model to find the most distinguishing facial features to be used to characterize a human face [13]. This methodology can be beneficially applied to extract salient points on the face and build a face template of each subject.

Most of the times, one missing part is the link between the features extracted from the face images and the geometry of the face itself. Both information are relevant in humans to identify a subject [14]. An approach recently proposed employs a two face-matching technique, based on the fusion of local and global information [15]. In the local matching strategy, SIFT keypoint features are extracted from face images in the areas corresponding to relevant facial landmarks. Matching of a pair of feature vectors is performed by a minimum Euclidean distance metric. Matching scores produced from each pair of salient features are fused together using the sum rule. In the global matching strategy, the SIFT features extracted from the facial landmarks are fused together by concatenation. Also in this case, matching is performed by means of a minimum Euclidean distance metric.

2. Detection of distinguishing patterns

The amount of distinctive information in a human face is not uniformly distributed within its image. An image of an eye is often considered to convey more information that of say the chin, both being sampled at the same resolution. Indeed, the performance of any classifier is likely to be influenced by the uniqueness, or degree of similarity, of the features being used, within the given population of samples. Selecting non-distinctive image areas will tend to increase the required processing resources, and non-distinctive features may also drift, or bias, a classifier's response. This assert is also in accordance with the mechanisms found in the human visual system. Neurophysiological studies from impaired people demonstrated that the face recognition process is heavily supported by a series of ocular saccades, performed to locate and process the most distinctive areas within a face [16-20].

In detail, the algorithm extracts, from the two face-images, a set of patches centered upon specific points. These points are randomly distributed across the face-image in order to cover most part of the face. This process samples most of the face, in a way similar to that adopted in patch-based image classification and image characterization [21]. Each patch maps on to a coordinate in a multi-dimensional feature space by virtue of a feature extraction process. The patches from one face-image will tend to form their own cluster in this space: the other face-image ought to form a different cluster. The extracted patches thus constitute two data-clusters of location-independent features, each of which characterize one of the two faces. Based on the distribution of those patches within feature space, it is possible to derive different conclusions: if a patch that lies within, or close to, the other cluster, then it is very similar to patches of the other face, thus may lead to misclassification and so ought to be avoided by any classifier. Conversely a patch that lies on the limb of its own cluster that is most distant from the other might be well employed by a classifier, since it is very different from patches of the other set. We thus formalize the degree of distinctiveness of each face patch by weighting it according to its distance from the projection of the other data-cluster. Patches with the highest weights are then interpreted as encoding the most important differences between the two face-images. This idea is sketched in Fig. 1, using an oversimplified 2-D feature space.

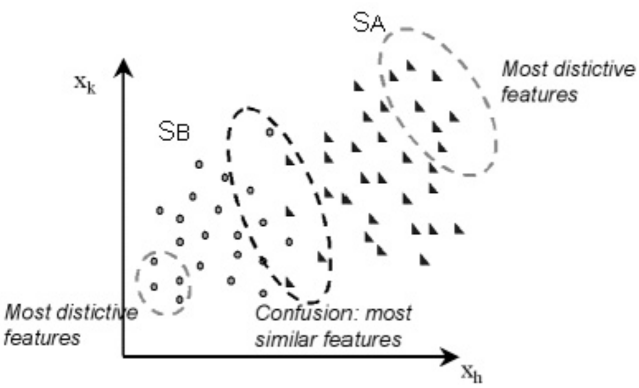

Figure 1: Sketch of the facial features distribution for two subjects S_A and S_B, in the case of a simple 2D feature space

Figure 2: (Top) Comparison of two faces from the BANCA database [22] and (Bottom) extraction of the most distinguishing patterns.

In order to preserve scale and orientation invariance, the patches are sampled according to a logarithmic-polar law [13].

Experiments performed on the BANCA database [22] demonstrated that:

i. The extracted face areas convey most of the discriminant information for identification
ii. The amount of data required to perform the identification is reduced with respect to the use of all available information in the face

As a direct consequence, this approach allows to automatically extract the most distinguishing features from the face image. These areas or points can be used to assist a human operator in a forensic application scenario. Moreover, this approach is also well sited to define the face areas to be analyzed at high resolution to perform a texture-based identification. Regardless of the resolution of the face image, the extracted patches will still retain a fraction of the pixels forming the original image.

3. Local and global matching

In this section two matching strategies are described namely the local, based on parts, and the global face matching. In addition, a classifier fusion technique is applied, where the scores obtained from the local strategy are fused together in terms of matching scores obtained from individual classifiers.

While raw gray levels have been used to extract the most relevant features from the face image, other methods can be employed to characterize local patterns. The scale invariant feature transform (SIFT) has been proposed by Lowe [8,9] and proved to be invariant to image rotation, scaling, translation, and partly illumination changes. The basic idea of the SIFT descriptor is detecting feature points efficiently through a staged filtering approach that identifies stable points in the scale-space. Local feature points are extracted by searching peaks in the scale-space from a difference of Gaussian (DoG) function. The feature points are localized using the measurement of their stability and orientations are assigned based on local image properties. Finally, the feature descriptors, which represent local shape distortions and illumination changes, are determined.

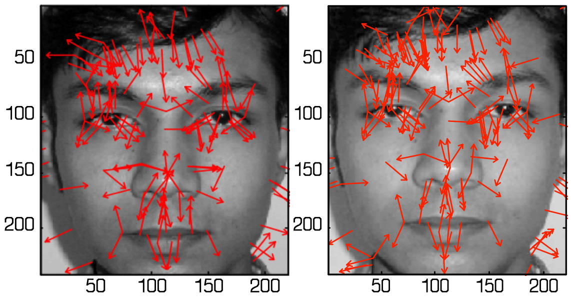

Figure 3: Invariant SIFT feature extraction are shown on a pair of face images.

3.1. Local face matching strategy

Faces are deformable objects which are generally difficult to characterize with a rigid representation. Different facial regions, not only convey different information on the subject's identity, but also suffer from different time variability either due to motion or illumination changes [14]. A typical example is the case of a talking face. While the eyes can be almost still and invariant over time, the mouth moves changing its appearance over time. As a consequence, the features extracted from the mouth area cannot be directly matched with the corresponding features from a static template. Moreover, single facial features may be occluded making the corresponding image area not usable for identification. For these reasons to improve the robustness of the identification process it is mandatory to decouple the image information corresponding to different face areas. The aim of the proposed local matching technique is to correlate the extracted SIFT features with independent facial landmarks.

In Figure 2 and 3 an example showing the concept of independent matching facial features from local areas is presented.

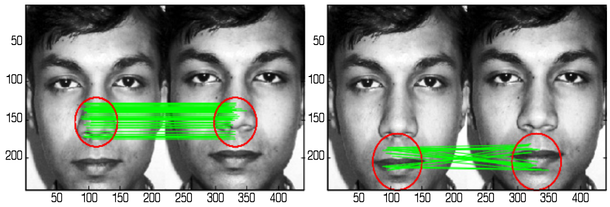

Figure 4: Example of independent matching of static and dynamic facial features.

Given a face image *I, N* independent ROIs are extracted. The SIFT feature points are then extracted from these regions and gathered together. From these groups pair-wise salient feature matching is performed. Finally, the matching scores obtained are fused together by the sum fusion rule [13] and the fused score are compared against a threshold. More formally, if $D^i(I^{gallery}, I^{gallery})$ is the distance between a pair of groups, then the distance can be defined as follows:

$$D^i(I^{test}, I^{gallery}) = \sqrt{\sum_{i \in m, j \in n} (I_j^{test}(k_i) - I_i^{gallery}(k_i))^2} \;. \le \Psi_i^k$$

where, m and n are the dimensions of concatenated feature points for a pair of gallery and test samples and k is the keypoint descriptor. Ψ^k is the threshold, which is computed a priori from a training set of face images. This face set must be disjoint from the image sets used for testing and validation. Finally, the fused matching score $FD(I^{test}, I^{gallery})$ is computed by combining the N individual matching scores together using the sum rule [13]:

3.2. Global face matching strategy

While for the local matching each face area is handled independently, in the global matching all SIFT features are grouped together. In particular, the SIFT features extracted from the image areas corresponding to the located facial landmarks, are grouped together to form an augmented vector by concatenation. The actual matching is performed by comparing the global feature vectors for a pair of face images. Before performing the face matching a one to one correspondence is established for each pair of facial landmarks, as discussed in Section 3.1.

Two concatenated keypoint sets can be computed as:

$$I_{gallery}^{Left-eye}(k_i) = \{I_{gallery}^i(k_1), I_{gallery}^i(k_2), \ldots, I_{gallery}^i(k_{m_i})\} \quad \because i \in N;$$

where, m_i is the dimension of the extracted keypoint feature sets computed from the i^{th} facial landmark. In order to obtain a fused feature set for a gallery sample face, the keypoints of the N components are joined together:

$$I^{gallery}(k) = \{I^{i}_{gallery}(k_{m_i}) \bigcup I^{j}_{gallery}(k_{m_j}) \bigcup .... \ I^{N}_{gallery}(k_{m_N})\};$$

Similarly, the concatenated feature set for a probe sample is built. The final matching score $FD_{GLOBAL}(I^{probe}, I^{gallery})$ is computed by first determining all the minimum pair distances and then computing a mean score of all the minimum pair distances as:

$$FD_{GLOBAL}(I^{probe}, I^{gallery}) = \sqrt{\sum \min_{i \in M}\{\min_{j \in N}\{I^{gallery}(k_i), I^{probe}(k_j)\}\}}$$

the final distance is determined by the Hausdorff distance metric and the distance score is compared against a threshold computed heuristically from a training set of face images. As for the local matching threshold, this face set is disjoint from the image sets used for testing and validation.

3.3. Fusion of local and global matching scores

Local and global matching can be fused together with the same fusion strategy as adopted for each individual sets or by means of a more complex strategy such as the Dempster-Shafer belief theory [23,24]. In this framework the evidences obtained from different sources are combined to compute the probability of an event. This is obtained combining three elements: the basic probability assignment function (*bpa*), the belief function (*bf*) and the plausibility function (*pf*) [15]. By applying this fusion method an Equal Error Rate of 2% was achieved on the Yale database. In figure 5 the comparative results obtained by applying the fusion of global and local matching scores are presented for three databases: the ORL (50 subjects, 10 samples per subject), the IITK (200 subjects, 4 samples per subject) and the Yale (15 subjects, 11 samples per subject with varying facial expression, illumination and occlusions) databases. More details on the experimental procedure are reported in [15].

4. Conclusion

Humans naturally perform identification of others from the face appearance. This process involves a variety of cues which can be both classifies as “local” and “global”. Characteristic features, as well as the general face shape and hair style, are all useful cues. Depending on several factors including viewing distance, familiarity, movement and facial occlusions we may use a variable number of facial features. There are also specific applications, such as forensics, where face images can be of great importance to establish the identity of a subject. In this specific case, it is crucial to strongly limit the probability of false acceptance to avoid convicting an innocent. For this reason, forensic professionals are trained to identify subtle marks, which uniquely characterize a face, thus reducing the probability of errors. Still global face features are also employed.

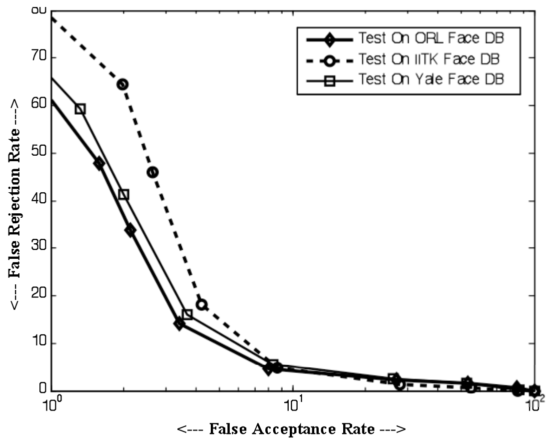

Figure 6: ROC curves determined from three face databases: IITK, ORL and Yale. The error rates were computed by fusing the local and global matching scores.

While global features are easier to capture they are generally less discriminative than localized features, but are less sensitive to localized changes in the face due to the partial deformability of the facial structure. On the other hand, local features on the face can be highly discriminative, but may suffer for local changes in the facial appearance or partial face occlusion. The optimal face representation should then allow matching localized facial features, but also determining a global similarity measurement for the face [14,15].

In this paper a method to locate subject-specific features, or landmarks, on the face has been presented. These landmarks can be used to build a subject-specific representation, which may reduce the probability of false acceptance. A robust, integrated classification paradigm for face recognition has been also discussed, comparing a local and a global face representation.

At the time being most face recognition algorithms rely on low resolution imaging, which do not allow to capture as many peculiar features as the human visual system does. By employing the distinctive feature detection process proposed in [13] it will be possible to process high resolution face images retaining only the relevant information which guarantees a specific representation, thus maximizing the probability of correct identification.

References

[1] G. Shakhnarovich and B. Moghaddam. Face recognition in subspaces. In S. Li and A. Jain, editors, Handbook of Face Recognition, pp. 141–168. Springer Verlag, 2004.

[2] G. Shakhnarovich, J. W. Fisher, T. Darrell, "Face recognition from long-term observations", Proc. IEEE European Conference on Computer Vision, 2002, pp. 851-865.

[3] L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg. Face recognition by elastic bunch graph matching. IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-19:775–779, 1997.

[4] J. Bigun. Retinal vision applied to facial features detection and face authentication. Pattern Recognition Letters, 23(4):463–475, 1997.

[5] G. Zhang, X. Huang, S. Li, Y. Wang, and X. Wu. Boosting local binary pattern (lbp)-based face recognition. In L. 3338, editor, SINOBIOMETRICS 2004, pp. 179–186. Springer Verlag, 2004.

[6] G. Heusch, Y. Rodriguez, and S. Marcel. Local binary patterns as an image preprocessing for face authentication. IDIAP-RR 76, IDIAP, 2005.

[7] D.R.Kisku, A.Rattani, E.Grosso and M.Tistarelli, "Face identification by SIFT-based complete graph topology", IEEE workshop AutoId, 2007, pp. 63-68.

[8] D. Lowe. Object recognition from local scale-invariant features. In Int. Conf. on Computer Vision, pp. 1150–1157, 1999.

[9] D. Lowe. Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision, 60(2):91–110, 2004.

[10] U. Park, S. Pankanti, A. K. Jain, "Fingerprint verification using SIFT features", Biometric Technology for Human Identification V. Edited by Vijaya Kumar, B. V. K.; Prabhakar, Salil; Ross, Arun A. Proceedings of the SPIE, Volume 6944, pp. 69440K-69440K-9 (2008).

[11] F. Smeraldi and J. Bigun, "Facial feature detection by Saccadic Exploration of the Gabor Decomposition", Proceedings of the 1998 International Conference on Image Processing, Chicago (USA), vol. 3, pp. 163 - 167, October 4 - 7, 199

[12] N. Gourier, D. H. James, and L. Crowley, "Estimating Face orientation from Robust Detection of Salient Facial Structures", FG Net Workshop on Visual Observation of Deictic Gestures (POINTING) – 2004.

[13] M. Bicego, G. Brelstaff, L. Brodo, E. Grosso S. Lagorio and M. Tistarelli, "Distinctiveness of faces: a computational approach", ACM Transactions on Applied Perception, 5(2):1-18, 2008.

[14] B. Heisele, P. Ho, J. Wu, T. Poggio, "Face recognition: component-based versus global approaches", Computer Vision and Image Understanding, Special Issue on Face Recognition, Volume 91, Issues 1-2, July-August 2003, pp. 6-21.

[15] D.R. Kisku J.K. Sing, M. Tistarelli, P. Gupta, "Face Recognition by Fusion of Local and Global Matching Scores using DS Theory: An Evaluation with Uni-classifier and Multi-classifier Paradigm", Proc. IEEE Int.l Conference on Computer Vision and Pattern Recognition (CVPR 2009) Biometrics Workshop, pp. 60-5, June 25th 2009, Miami, USA.

[16] C. Goren, M. Sarty, and P. Wu, "Visual following and pattern discrimination of face-like stimuli by newborn infants," *Pediatrics*, vol. 56, pp. 544–549, 1975.

[17] A.L. Yarbus, *Eye movements and vision*, Plenum Press, New York, 1967.

[18] F.K.D. Nahm, A. Perret, D. Amaral, and T.D. Albright, "How do monkeys look at faces?," *Journal of Cognitive Neuroscience*, vol. 9, pp. 611–623, 1997.

[19] M.M. Haith, T. Bergman, and M.J. Moore, "Eye contact and face scanning in early infancy," *Science*, vol. 198, pp. 853–854, 1979.

[20] A. Klin, "Eye-tracking of social stimuli in adults with autism," 2001, Paper presented at the meeting of the NICHD Collaborative Program of Excellence in Autism. Yale University, New Haven, CT.

[21] N. Jojic, B. Frey, and A.Kannan, "Epitomic analysis of appearance and shape," in *Proc. Int. Conf. on Computer Vision*, 2003, vol. 1, pp. 34–41.

[22] E. Bailly-Baillire, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Marithoz, J. Matas, K. Messer, V. Popovici, F. Pore, B. Ruiz, and J.-P. Thiran, "The BANCA database and evaluation protocol," in *Proc. Int. Conf. on Audio- and Video-Based* Biometric Person Authentication (AVBPA03). 2003, pp. 625–638, Springer-Verlag.

[23] J. A. Barnett, "Computational methods for a mathematical theory of evidence", Proceedings IJCAI, pp. 868-875, 1981.

[24] M. Bauer, "Approximation Algorithms and Decision-Making in the Dempster-Shafer Theory of Evidence—An Empirical Study, International Journal of Approximate Reasoning, vol. 17, pp. 217–237.