On Finding Differences Between Faces

Manuele Bicego[1], Enrico Grosso[1], and Massimo Tistarelli[2]

[1] DEIR - University of Sassari, via Sardegna 58 - 07100 Sassari - Italy
[2] DAP - University of Sassari, piazza Duomo 6 - 07041 Alghero (SS) - Italy

Abstract. This paper presents a novel approach for extracting characteristic parts of a face. Rather than finding a priori specified features such as nose, eyes, mouth or others, the proposed approach is aimed at extracting from a face the most distinguishing or dissimilar parts with respect to another given face, *i.e.* at "finding differences" between faces. This is accomplished by feeding a binary classifier by a set of image patches, randomly sampled from the two face images, and scoring the patches (or features) by their mutual distances. In order to deal with the multi-scale nature of natural facial features, a local space-variant sampling has been adopted.

1 Introduction

Automatic face analysis is an active research area, whose interest has grown in the last years, for both scientific and practical reasons: on one side, the problem is still open, and surely represents a challenge for Pattern Recognition and Computer Vision scientists; on the other, the stringent security requirements derived from terroristic attacks have driven the research to the study and development of working systems, able to increase the total security level in industrial and social environments.

One of the most challenging and interesting issue in automatic facial analysis is the detection of the "facial features", intended as characteristic parts of the face. As suggested by psychological studies, many face recognition systems are based on the analysis of facial features, often added to an holistic image analysis. The facial features can be either extracted from the image and explicitly used to form a face representation, or implicitly recovered and used such as in the PCA/LDA decomposition or by applying a specific classifier.

Several approaches have been proposed for the extraction of the facial features ([1–5], to cite a few). In general terms, all feature extraction methods are devoted to the detection of a priori specified features or gray level patterns such as the nose, eyes, mouth, eyebrows or other, non anatomically referenced, fiducial points. Nevertheless, for face recognition and authentication, it is necessary to also consider additional features, in particular those features that really characterize a given face. In other words, in order to distinguish the face of subject "A" from the face of subject "B", it is necessary to extract from the face image of subject "A" all features that are significantly different or even not present in face "B", rather than extract standard patterns.

T. Kanade, A. Jain, and N.K. Ratha (Eds.): AVBPA 2005, LNCS 3546, pp. 329–338, 2005.

This paper presents a novel approach towards this direction, aiming at "finding differences" between faces. This is accomplished by extracting from one face image the most distinguishing or dissimilar areas with respect to another face image, or to a population of faces.

2 Finding Distinguishing Patterns

The amount of distinctive information in a subject's face is not uniformly distributed within its face image. Consider, as an example, the amount of information conveyed by the image of an eye or a chin (both sampled at the same resolution). For this reason, the performance of any classifier is greatly influenced by the uniqueness or degree of similarity of the features used, within the given population of samples. On one side, by selecting non-distinctive image areas increases the required processing resources, on the other side, non-distinctive features may drift or bias the 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 [6–10].

In principle, this feature selection process can be performed by extracting the areas, within a given subject's face image, which are most dissimilar from the same areas in a "general" face. In practice, it is very difficult to define the appearance of a "general face". This is an abstract concept, definitely present in the human visual system, but very difficult to replicate in a computer system. A more viable and practical solution is to determine the face image areas which mostly differ from any other face image. This can be performed by feeding a binary classifier with a set of image patches, randomly sampled from two face images, and scoring the patches (or features) by their mutual distances, computed by the classifier. The resulting most distant features, in the "face space", have the highest probability of being the most distinctive face areas for the given subjects.

In more detail, the proposed algorithm extracts, from two face images, a set of sub-images centered at random points within the face image. The sampling process is driven to cover most of the face area[1]. The extracted image patches constitute two data sets of location-independent features, each one characterizing one of the two faces. A binary Support Vector Machine (SVM) [16, 17] is trained to distinguish between patches of the two faces: the computed support vectors define a hyperplane separating the patches belonging to the two faces. Based on the distribution of the image patches projected on the classifier's space, it is possible to draw several conclusions. If the patch projection "lies" very close to the computed hyperplane (or on the opposite side of the hyperplane), it means

[1] A similar image sampling model has been already used in other applications such as image classification (the so called patch-based classification [11–14]) or image characterization (the epitomic analysis proposed by Joijc and Frey in [15])

that the classifier is not able to use the feature for classification purposes (or it may lead to a misclassification). On the other hand, if the patch projection is well located on the subject's side of the hyperplane and is very far from the separating hyperplane, then the patch clearly belongs to the given set (*i.e.* to that face) and it is quite different from the patches extracted from the second face.

According to this intuition, the degree of distinctiveness of each face patch can be weighted according to the distance from the trained hyperplane. Since the classifier has been trained to separate patches of the first face from patches of the second face, it is straightforward to observe that the most important differences between the two faces are encoded in the patches far apart from the separating hyperplane (i.e. the patches with the highest weights).

In this framework the scale of the analysis is obviously driven by the size of the extracted image patches. By extracting large patches only macro differences are determined, loosing details, while by reducing the size of the patches only very local features are extracted, loosing contextual information. Both kinds of information are important for face recognition. A possible solution is to perform a multi scale analysis, by repeating the classification procedure with patches at different sizes, and then fusing the determined differences. The drawback is that each analysis is blind, because no information derived from other scales could be used. Moreover, repeating this process for several scales is computationally very expensive.

A possible, and more economic, alternative to a multi-scale classification, is to extract "multi-scale" patches, *i.e.* image patches which encode information at different resolution levels. This solution can be implemented by sampling the image patches with a log-polar mapping [18]. This mapping resembles the distribution of the ganglion cells in the human retina, where the sampling resolution is higner in the center (fovea) and decreases toward the periphery. By this resampling of the face image, each patch contains both low scale (high resolution) and contextual (low resolution) information.

The proposed approach for the selection of facial features consists of three steps:

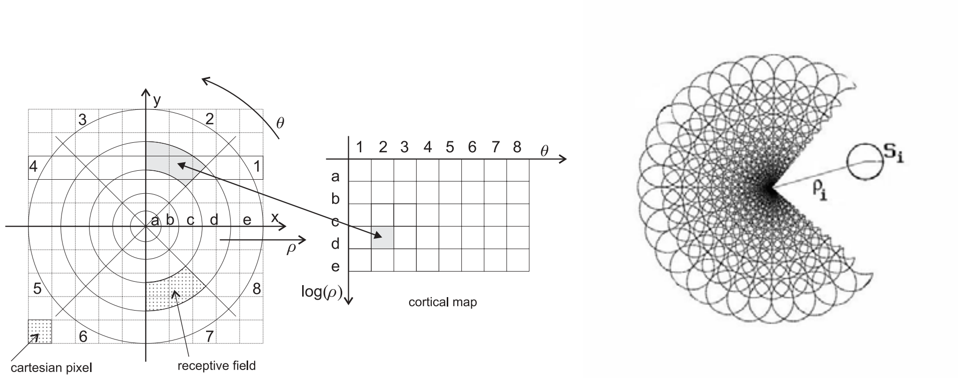
1. two distinct and geometrically disjoint sets of patches are extracted, at random positions, from the two face images;
2. a SVM classifier is trained to define an hyperplane separating the two sets of patches;
3. for each of the two faces, the face patches are ranked according to the distances from the computed hyperplane.

The processes involved by each step are detailed in the remainder of the paper.

2.1 Multi-scale Face Sampling

A number of patches are sampled from the original face image, centered at random points. The randomness in the selection of the patch center assures that the entire face is analyzed, without any preferred side or direction. Moreover, a random sampling enforces a blind analysis without the need for a priori alignment between the faces.

The face image is re-sampled at each selected random point following a log-polar law [18]. The resulting patches represent a local space-variant remapping of the face image, centered at the selected point. The analytical formulation of the log-polar mapping describes the mapping that occurs between the retina (retinal plane (x, y)) and the visual cortex (log-polar or cortical plane (log(ρ), θ)). The derived logarithmic-polar law, taking into account the linear increment in size of the receptive fields, from the central region (fovea) towards the periphery, is described by the diagram in figure 1(a).



(a) (b)

Fig. 1. (a) Retino-cortical log-polar transformation. (b) Arrangement of the receptive fields in the retinal model

The log-polar transformation applied is the same described in [18] which differs from the models proposed in [19, 20]. The parameters required to define the log-polar sampling are: the number of receptive fields per eccentricity (N_a) and the radial and angular overlap of neighboring receptive fields (O_r and O_a).

For each receptive field, located at eccentricity ρ_i and with radius S_i, the angular overlap factor is defined by $K_0 = \frac{S_i}{\rho_i}$. The amount of overlapping is strictly related to the number of receptive fields per eccentricity N_a. In particular if $K_0 = \frac{\pi}{N_a}$ all receptive fields are disjoint. The radial overlap is determined by:

$$K_1 = \frac{S_i}{S_{i-1}} = \frac{\rho_i}{\rho_{i-1}}.$$

The two overlap parameters K_0 and K_1 are not independent, in fact:

$$K_1 = \frac{\rho_i}{\rho_{i-1}} = \frac{1 + K_0}{1 - K_0}.$$

As for the angular overlap, the radial overlap is not null only if:

$$K_1 < \frac{1 + K_0}{1 - K_0}.$$

Given the log-polar parameters N_a, O_r, O_a, K_0 and K_1 are computed as:

$$K_0 = \pi \frac{O_a}{N_a}, \qquad K_1 = \frac{O_r + K_0}{O_r - K_0}.$$

The image resolution determines the physical limit in the size of the smallest receptive fields in the fovea. This, in turn, determines the smallest eccentricity:

$$\rho_0 = \frac{S_0}{K_0}$$

Defining $\rho_0 \in [0.5 - 5]$, the original image resolution is preserved.

2.2 The SVM Classifier

In the literature Support Vector Machines have been extensively employed as binary classifiers in face recognition and authentication [21, 22], object classification [23], textile defects classification [24] and other applications as well.

The SVM classifier holds several interesting properties: quick training process [25], accurate classification, and, at the same time, a high generalization power [17]. Moreover, only two parameters need to be set: the regularization constant C and the size of the kernel for the regularization function.

In the proposed approach the *Radial Basis Function (RBF)* regularization kernel has been adopted, because it allows the best compromise between classification accuracy and generalization power. In order to obtain an acceptable generalization from the input data, the value of sigma has been carefully determined.

The set of log-polar image patches, sampled from each face image, are firstly vectorized and subsequently fed to a Support Vector Machine [16, 17]. As the SVM is a binary classifier, the data from the two subjects are used to build a set of support vectors able to distinguish them. Therefore, according to the procedure adopted to build a classifier for authentication purposes, the patches from one subject are used to represent the "client" class, while the patches from the second subject represent the "impostor" class.

2.3 Determining Face Differences

The SVM classifier, obtained from the input patches, defines an hyperplane separating the features belonging to the two subjects. The differences between the two subjects could be determined, for each correctly classified patch, from the absolute distance from the hyperplane: higher distances identify more characteristic facial features.

More formally, let $\mathcal{C}(\mathbf{x})$ be the class assigned by the trained SVM to an unknown patch $\mathbf{x}$, then:

$$\mathcal{C}(\mathbf{x}) = sign(f(\mathbf{x})) \tag{1}$$

where $f(\mathbf{x})$ represents the distance between the point $\mathbf{x}$ and the hyperplane represented by the SVM. When using a kernel $K(\mathbf{x}_i, \mathbf{x}_j)$, the distance $f(\mathbf{x})$ is computed as

$$f(\mathbf{x}) = b + \sum_{i=1}^{D} \alpha_i \mathcal{C}(\mathbf{x}_i) K(\mathbf{x}, \mathbf{x}_i) \tag{2}$$

where b and α_i are parameters determined in the training phase, and $\mathbf{x}_i$ are the points of the training set.

Given the trained SVM, the weight ω of the patch P_i belonging to the face k is computed as follows:

$$\omega(P_i) = \begin{cases} |f(P_i)| & \text{if } \mathcal{C}(P_i) = k \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

This analysis is repeated for both faces. It is important to note that the patches which are in the uncorrect side of the hyperplane are discarded (weight equal to 0), since the classifier could not provide any useful information about them (it is not able to correctly classify those patches).

3 Experimental Results

In order to verify the real applicability of the proposed method, two experiments were performed. In the first experiment a synthetic artifact (a black dot) is added to a face image and this is compared against the original image (see Fig. 2). In the second experiment two face images from two different subjects are compared (see Fig. 3). In both experiments gray level images were used, with a resolution of 310x200 pixels. The images have been re-sampled, at random positions, with 1000 log-polar patches. Each log-polar patch has a resolution of 23 eccentricities and 35 receptive fields for each eccentricity, with an overlap equal to 10% along the two directions. The *Radial Basis Function (RBF)* regularization kernel has been adopted for the SVM, with parameters $\sigma = 400$ and $C = 10$.

The results of the synthetic experiment is displayed in Fig. 2. To facilitate the understanding of the computed image differences, only the first ten patches with higher weights (distances from the computed hyperplane) are displayed. From the sequence of patches resulting in figure 2 the black dot is clearly identified.

In the experiment performed on two real face images, the 52 patches with higher distances for each face have been considered. The computed results are shown in Fig. 4 and 5.

In order to facilitate the visualization, similar patches have been grouped together, using the K-means method [26]. For the first face, six semantically different regions have been found, whereas in the second face nine different regions were considered. For each patch retained in the figure, the number of similar patches in the group is displayed. From these pictures some relevant differences between the two faces are detected. In the first face, for example, the forehead (both right and left part), the nose and the eyes are clearly identified. It is worth noting that also the fold of the skin on the right cheek is detected. As for the

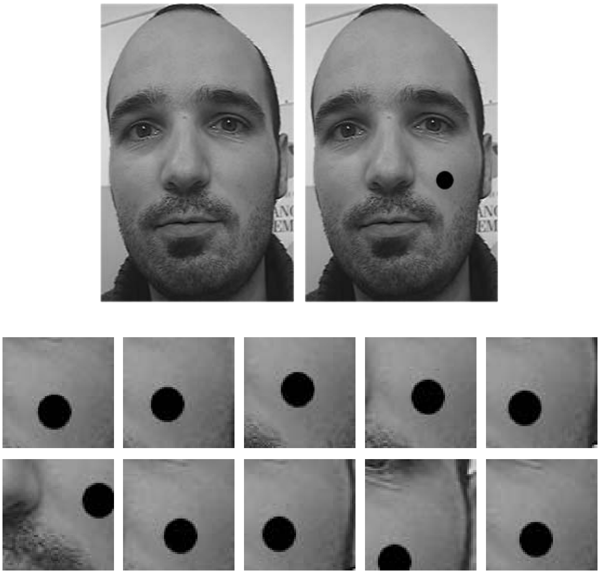
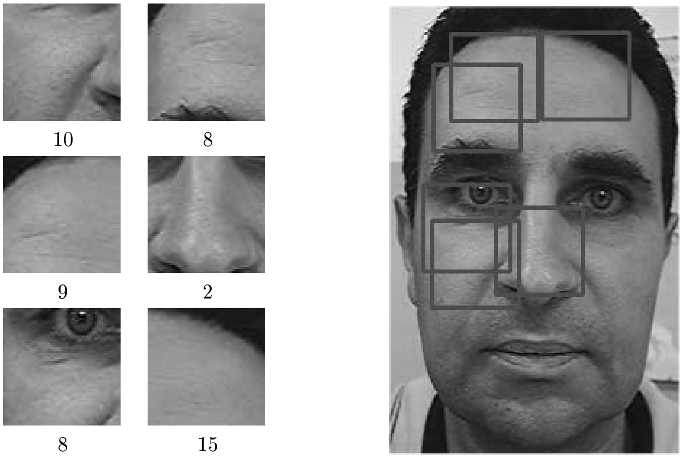


Fig. 2. Synthetic experiment. (top) The two images used in the experiment. (bottom) The 10 most weighted patches extracted when comparing the two faces. Only the patches related to the modified face are displayed

Fig. 3. (left) Original images used in the comparison experiment. (right) Random image points used for sampling the space-variant patches

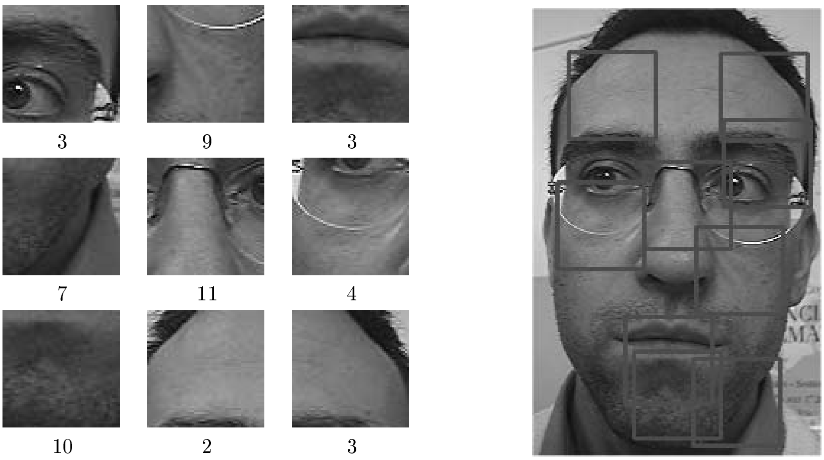
second person (Fig. 5) the eyeglass are clearly identified as distinctive features (both right, left, upper and central parts). In fact, 27 out of the first 52 most weighted patches are located on them. Another distinctive pattern is the shape of the mouth, together with the chin, and the shape of the forehead.

As it can be noted, the extracted patterns seem to have some complementarities for the two faces. In fact, some distinctive areas are still present in both faces (regions around the eyes and the nose) while other distinctive and subtle details are preserved.



(a) (b)

Fig. 4. Results of the detection of the most distinguishing features for the first face. Similar patches have been grouped together. (a) The representative patches (the number of components of each group is displayed below the patch) and (b) the location of the patches on the face



(a) (b)

Fig. 5. Results of the detection of the most distinguishing features for the second face. Similar patches have been grouped together. (a) The representative patches (the number of components of each group is displayed below the patch) and (b) the location of the patches on the face

4 Conclusions

In this paper a new approach for finding differences between faces has been proposed. A Support Vector Machines classifier is trained to distinguish between two sets of space-variant patches, randomly extracted from two different face images. The "distinctiveness" of each patch is computed as the distance from the separating hyperplane computed from the support vectors.

Even though the experiments performed are very preliminary, already demonstrate the potential of the algorithm in determining the most distinctive patterns in the analyzed faces. The proposed approach can be very effective to tailor the face representation according to the most distinctive features of a subject's face, for recognition purposes.

A future development of this research includes the combination of the extracted features, which could be performed by "back propagating" the patches weights to the face, to build a true "difference map".

Another interesting issue is the comparison of more than two faces, i.e. finding the differences between a given face and the rest of the world. In this way it may be possible to extract the general characteristic features of any given face. This can be achieved by choosing the negative examples in the SVM training as formed by all patches randomly sampled from several different faces. A further issue could be the investigation of different sampling techniques, *i.e.* methods that could reduce the number of samples needed to significantly cover the whole face.

Acknowledgments

This work has been supported by funds from the 6th Framework Programme European Network of Excellence "BioSecure".

References

1. Craw, I., Tock, D., Bennett, A.: Finding face features. In: Proc. of European Conf. Computer Vision. (1992) 92–96
2. Graf, H., Chen, T., Petajan, E., Cosatto, E.: Locating faces and facial parts. In: Proc. Int. Workshop Automatic Face and Gesture Recognition. (1995) 41–46
3. Campadelli, P., Lanzarotti, R.: Fiducial point localization in color images of face foregrounds. Image and Vision Computing **22** (2004) 863–872
4. Ming-Hsuan, Y., Kriegman, D., Ahuja, N.: Detecting faces in images: a survey. IEEE Trans. on Pattern Analysis and Machine Intelligence **24** (2002) 34–58
5. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys **35** (2003) 399 – 458
6. Goren, C., Sarty, M., Wu, P.: Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics **56** (1975) 544–549
7. Yarbus, A.: Eye movements and vision. Plenum Press, New York (1967)
8. Nahm, F., Perret, A., Amaral, D., Albright, T.: How do monkeys look at faces? Journal of Cognitive Neuroscience **9** (1997) 611–623
9. Haith, M., Bergman, T., Moore, M.: Eye contact and face scanning in early infancy. Science **198** (1979) 853–854
10. Klin, A.: 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.

11. Agarwal, S., Roth, D.: Learning a sparse representation for object detection. In: Proc. European Conf. on Computer Vision. Volume 4. (2002) 113–130
12. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: Proc. Int. Conf. on Computer Vision and Pattern Recognition. Volume 2. (2003) 264
13. Dorko, G., Schmid, C.: Selection of scale-invariant parts for object class recognition. In: Proc. Int. Conf. on Computer Vision. Volume 1. (2003) 634–640
14. Csurka, G., Dance, C., Bray, C., Fan, L., Willamowski, J.: Visual categorization with bags of keypoints. In: Proc. Workshop Pattern Recognition and Machine Learning in Computer Vision. (2004)
15. Jojic, N., Frey, B., A.Kannan: Epitomic analysis of appearance and shape. In: Proc. Int. Conf. on Computer Vision. Volume 1. (2003) 34–41
16. Vapnik, V.: The Nature of Statistical Learning Theory. Springer-Verlag (1995)
17. Burges, C.: A tutorial on support vector machine for pattern recognition. Data Mining and Knowledge Discovery **2** (1998) 121–167
18. Grosso, E., Tistarelli, M.: Log-polar stereo for anthropomorphic robots. In: Proc. European Conference on Computer Vision. Volume 1., Springer-Verlag (2000) 299–313
19. Braccini, C., Gambardella, G., Sandini, G., Tagliasco, V.: A model of the early stages of the human visual system: Functional and topological transformation performed in the peripheral visual field. Biol. Cybern. **44** (1982) 47–58
20. Tistarelli, M., Sandini, G.: On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. IEEE Trans. on Pattern Analysis and Machine Intellingence **15** (1993) 401–410
21. Jonsson, K., Kittler, J., Li, Y.P., Matas, J.: Support vector machines for face authentication. In: Proc. of Brit. Machine Vision Conf., Nottingham, UK. (1999) 543–553
22. Bicego, M., Iacono, G., Murino, V.: Face recognition with Multilevel B-Splines and Support Vector Machines. In: Proc. of ACM SIGMM Multimedia Biometrics Methods and Applications Workshop. (2003) 17–24
23. Pontil, M., Verri, A.: Support vector machines for 3-D object recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence **20** (1998) 637–646
24. Murino, V., Bicego, M., Rossi, I.: Statistical classification of raw textile defects. In: Proc. of IEEE Int. Conf. on Pattern Recognition. Volume 4. (2004) 311–314
25. Platt, J.C.: Fast training of support vector machines using sequential minimal optimization. Advances in Kernel Methods - Support Vector Learning (1998)
26. Theodoridis, S., Koutroumbas, K.: Pattern Recognition. Academic Press (1999)