DETECTION AND RESTORATION OF OCCLUSIONS FOR 3D FACE RECOGNITION

Alessandro Colombo, Claudio Cusano, Raimondo Schettini

DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione)
Università degli Studi di Milano-Bicocca
Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy
colomboal@disco.unimib.it, cusano@disco.unimib.it, schettini@disco.unimib.it

ABSTRACT

This paper presents an innovative restoration strategy which allows for an effective recognition of 3D faces, even when they are partially occluded by unforeseen, extraneous objects such as scarves, hats, glasses, and so on. First, the occluded regions are detected by considering their effects on the projections of the faces in a suitable face space; the non-occluded regions are then used to restore the missing information. Any recognition algorithm can be applied as usual to restored faces. This restoration strategy led to very satisfactory results on a test set of 52 three-dimensional faces presenting various kinds of occlusions.

1. INTRODUCTION

Although current automatic recognition systems have reached a certain level of reliability, their success is limited from the conditions imposed by many practical applications. The main difficulties in recognizing faces come from the great variability of their appearance, caused by variability in the acquisition conditions. The problem of recognition under variable lighting conditions and under variable poses of the subjects may be solved by exploiting three-dimensional images [1-6].

This paper addresses the problem of recognition of 3D partially occluded faces. Occlusions occur when the face is covered by extraneous objects such as glasses, scarves, hats, and so on, a condition frequently encountered in many real applications.

This problem has not been frequently discussed in literature. The main strategy consists in the application of a local approach in which the face is subdivided in several regions which are independently analyzed. The final outcome is determined by a voting step [7-9]. For example, Martinez divided each face into local regions which are analyzed in isolation. He presented a probabilistic method that analyzes the correctness of local matches [10, 11]. This approach relies on the fact that the contribution of occluded regions is compensated by that of non-occluded regions.

The innovative strategy, here illustrated, approaches the problem by performing a normalization of the faces: first the occluded regions are detected, and then the non-occluded regions are used to restore the missing information. Any recognition algorithm, both holistic and feature-based, can then be applied to restored faces.

The paper is organized as follows: Section 2 describes the restoration strategy for occlusions. Section 3 reports the result obtained on a dataset of occluded faces. Section 4 discusses the results obtained and our plans for future improvements.

2. RESTORATION STRATEGY

The detection of occluded regions is based on the consideration that occlusions appear as local deformations of the face. These deformations are initially detected from a comparison of the input image with a generic model of face. Then, the detection of the occluded regions is refined by exploiting the locality of occlusions.

A general model of faces is provided by the eigenface approach [12] which represents faces as linear combinations of a small set of images (the eigenfaces):

$$\mathbf{x} + \mathbf{e} = \mathbf{m} + \sum_{i=1}^{N} y_i \mathbf{v}_i , \qquad (1)$$

where $\mathbf{x}$ is the input image (encoded as a d-dimensional vector), $\mathbf{e}$ is the approximation error, $\mathbf{m}$ is the mean face, $\mathbf{v}_i$ are the N eigenfaces considered, and y_i are the coefficients of the linear combination. The mean face and the eigenfaces are selected maximizing the amount of retained variability in a sample of face images. Each coefficient y_i is obtained projecting the vector ($\mathbf{x}$-$\mathbf{m}$) onto the corresponding eigenface $\mathbf{v}_i$.

The quantity $\|\mathbf{e}\|$ is called *Distance From Face Space* (DFFS) and can be considered a measure of "faceness". The DFFS of occluded faces is expected to be quite high and can be used to reveal the presence of occlusions. In fact, occluded faces cannot be well represented by a linear combination of eigenfaces computed on a sample of non-occluded faces.

The differences between the original image and the reconstruction are likely to be more pronounced where the face is occluded as in the case reported in Fig. 1-b which shows the error $\mathbf{e}$ for an image of a face partially occluded by glasses.

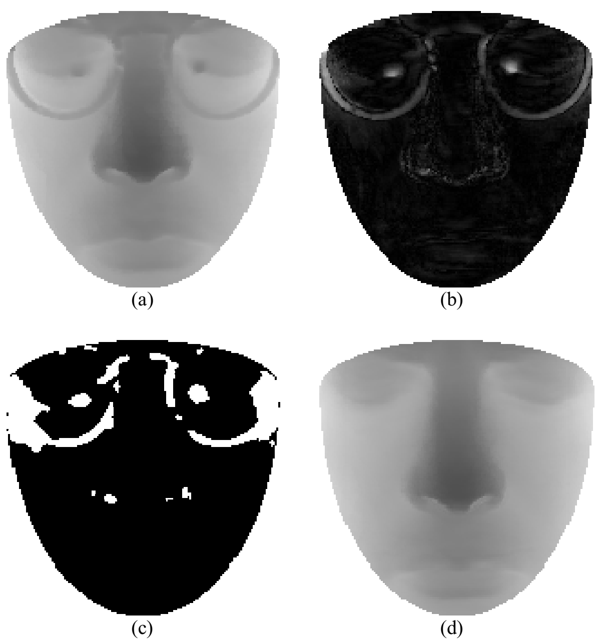


Fig. 1. Restoration of an image representing a subject wearing glasses. The input range image (a) is projected in the face space and the differences with its reconstruction (b) are computed. The occluded regions are selected (c), and excluded from the restoration process which produces a non occluded face (d).

The differences between the original image and the reconstruction are likely to be more pronounced where the face is occluded as in the case reported in Fig. 1-b which shows the error **e** for an image of a face partially occluded by glasses. The greater the difference between the original and the reconstructed pixel, the higher the probability that the original pixel will be occluded is expected to be. A thresholding on the difference image selects the pixels that are likely to be occluded. This selection can be rather inaccurate for an inadequate value of the threshold (which is determined empirically) or for the distortions of the reconstruction caused by occlusions (see Fig. 1-c for an example of the result of the occlusion detection process).

Morphological filters are used to clean the occlusion mask and to enforce the locality of the occlusion. The following steps are executed in more detail: i) a dilation operator fills the gaps of the thresholded mask; ii) the connected regions are labeled and only those with an area greater than a set threshold are retained. The first step is used to clean the occlusion mask filling its gaps, the second step discards small regions that are unlikely to correspond to occlusions.

2.1. Face restoration

Once the occlusions have been detected, the face must be restored. Due to the distortion caused by the occlusions, the image reconstructed with the eigenface approach cannot be considered good enough to be recognized. A more powerful restoration technique comes from an approach called Gappy Principal Component Analysis (GPCA) [13] on which is based our restoration strategy.

A training set of n complete patterns $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ (i.e., faces without occlusions) is used to determine the PCA basis formed by the mean face and the eigenfaces. An incomplete pattern $\mathbf{x}$ (i.e., an occluded face) to be restored may be expressed as the linear combination of the eigenfaces as in (1). To compute the coefficients y_i the DFFS $\|\mathbf{e}\|^2$ must be minimized. However, DFFS is computed including the missing components (i.e., the detected occlusions), while only the available information must be considered. To do so, it is useful to introduce the gappy inner product $(\mathbf{v}, \mathbf{u})_\mathbf{z}$, where the vector $\mathbf{z}$, obtained from the occlusion detection procedure previously described, indicates the missing components ($\mathbf{z}_i = 0$ is the i-th component of $\mathbf{x}$ is missing, otherwise $\mathbf{z}_i = 1$):

$$(\mathbf{u}, \mathbf{v})_\mathbf{z} = \sum_{i=1}^{d} \mathbf{u}_i \mathbf{v}_i \mathbf{z}_i \ . \tag{2}$$

The corresponding gappy norm $\|\mathbf{v}\|_\mathbf{z}$ is defined as $\sqrt{(\mathbf{v}, \mathbf{v})_\mathbf{z}}$. The coefficients y_i are then chosen to minimize the error $\|\mathbf{e}\|_\mathbf{z}^2$:

$$\begin{aligned} \|\mathbf{e}\|_\mathbf{z}^2 &= \left\| -\mathbf{x} + \mathbf{m} + \sum_{i=1}^{N} y_i \mathbf{v}_i \right\|_\mathbf{z}^2 = \|\mathbf{x} - \mathbf{m}\|_\mathbf{z}^2 - \\ &- 2\sum_{i=1}^{N} y_i (\mathbf{x} - \mathbf{m}, \mathbf{v}_i)_\mathbf{z} + \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j (\mathbf{v}_j, \mathbf{v}_i)_\mathbf{z}. \end{aligned} \tag{3}$$

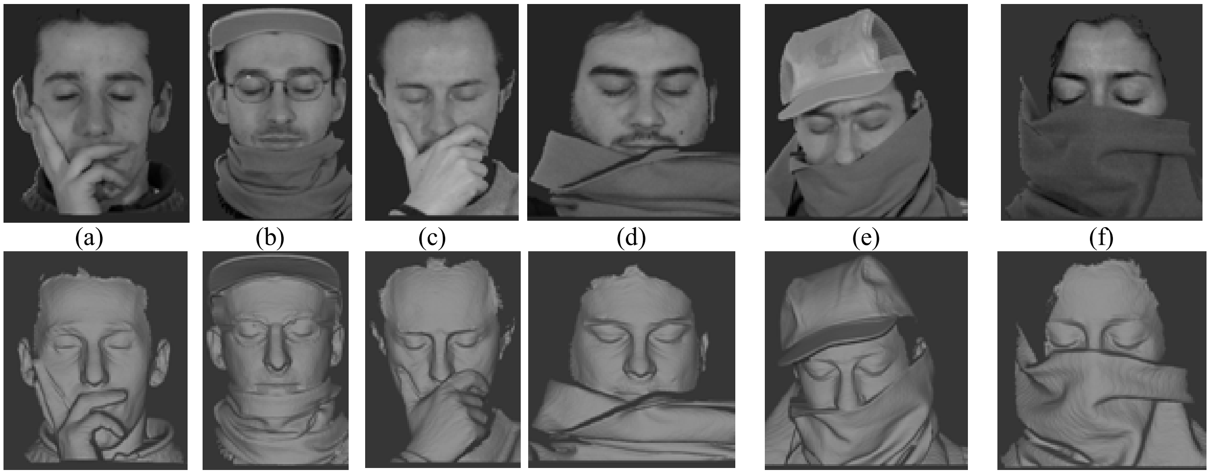
Differentiating $\|\mathbf{e}\|_\mathbf{z}^2$ with respect to each y_i, and requiring that the partial derivatives be null, yields a system of N linear equations:

$$\frac{\partial \|\mathbf{e}\|_\mathbf{z}^2}{\partial y_i} = -2(\mathbf{x} - \mathbf{m}, \mathbf{v}_i)_\mathbf{z} + 2\sum_{j=1}^{N} y_j (\mathbf{v}_j, \mathbf{v}_i)_\mathbf{z} = 0, \tag{4}$$

for i in $\{1,\ldots,N\}$. The gappy pattern $\mathbf{x}$ can be reconstructed using Equation (1), where the coefficients are found solving the system (4).

3. EXPERIMENTAL RESULTS

To test our method we collected a dataset of 208 three-dimensional faces acquired with the Minolta VIVID 900 range scanner. All the faces have been manually detected and normalized in a standard position and scale (see [14] for a description of an automatic procedure for these tasks); images have also been cropped to retain only the central portion of the faces. The dataset contains 22 different subjects, each of them acquired while posing different facial expressions (neutral, smiling, angry and bored). The training set includes 132 images which do not present occlusions. In the 76 acquisitions included in the test set, 52 faces are occluded by glasses, scarves, caps, by the hands of the subject, or by other extraneous objects. Some acquisitions include multiple occluding objects which are present. Fig. 2 shows a small sample of occluded faces.



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

(g) (h) (i) (j) (k) (l)

Fig. 2. Examples of weakly (a-d) and strongly (e,f) occluded faces, with the corresponding 3D acquisitions (g-l). Strong occlusions cover salient facial features such as the mouth or the eyes

Occluded faces has been subdivided into a set of 37 weakly occluded faces and a set of 15 strongly occluded faces on basis of the amount of details covered. The remaining 24 acquisitions do not present any occlusion and have been employed to evaluate the rate detection of false occlusions.

As a first test, we evaluated detection performance of the strategy. All the strong occlusions, which occur in 15 faces, have been correctly detected. Six weak occlusions have been missed, and the other 31 have been detected. The detection procedure correctly did not found any occlusion in the 24 non-occluded faces in the test set. Summing up, the procedure detected about 88.5% of the occlusions, without false positives. The missed occlusions are usually produced by thin objects, such as a scarf, which cover the face but without evident deformations.

In a second test we evaluated the improvement in recognition performances obtained when our strategy is applied. For this purpose we applied the fisherfaces method [15], but any other recognition method may be applied as well. The training set has been used to compute eigenface basis and to train fisherfaces. The images in the training set have also been used to build the gallery of known subjects. Each image of the test set has been restored using our restoration strategy, and the fisherfaces method has been evaluated on the test set before and after restoration. We considered two application scenarios: verification, in which the system must state if the identity claimed by the subject corresponds to his actual identity, and identification, in which the system must simply state which identity best matches with a given probe image. Fig. 3 shows the ROC curves (which report the False Acceptance Rates and the False Rejection Rates obtained varying an acceptation threshold) obtained in the identity verification scenario, and Fig. 4 the CMC curves (which report the Identification Rate within a variable rank) obtained the identification scenarios.

As expected, recognition on occluded faces is very difficult. The results obtained on weakly occluded faces are interesting: in the verification scenario, the equal error rates obtained on original (i.e., non-restored) and on restored faces are 22.5% and 13.5% respectively; in the identification scenario, the identification rates (within rank 1) obtained on original and on restored faces are 46.3% and 70.3%, respectively. The improvement obtained employing our restoration strategy is significant.

On strongly occluded faces, only a small improvement has been obtained (of about 7% for the identification rate, and less than 1% for the equal error rate). This fact is not surprising since strong occlusions hide most of the details of the faces.

Since no false occlusions have been detected, the recognition performances on the 24 non occluded faces are not affected by our restoration strategy (these results are not reported in the figures).

4. CONCLUSIONS AND FUTURE WORK

We presented an innovative strategy for the restoration of partially occluded 3D faces. Being independent on the recognition algorithm, our strategy can improve the reliability with respect to occlusions of any 3D recognition system.

Most of weakly occluded faces are detected, restored and correctly recognized. Strongly occluded faces are not recognized but the occlusions are correctly detected: in a concrete application these cases would be probably rejected.

Since our restoration strategy rarely find false occlusions (it never happened in our tests), non occluded faces can be regularly recognized without performances penalties.

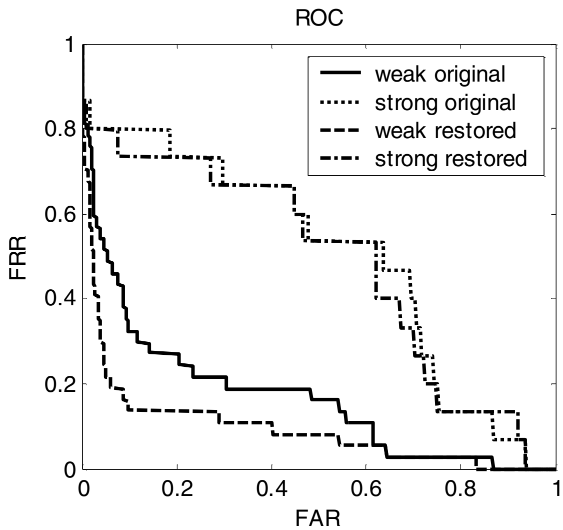


Fig. 3. Receiver Operator Characteristic (ROC) curves for the evaluation of the restoration strategy in the verification scenario. The curves refer to the performances obtained on the original and on the restored faces, for both weak and strong occlusions.

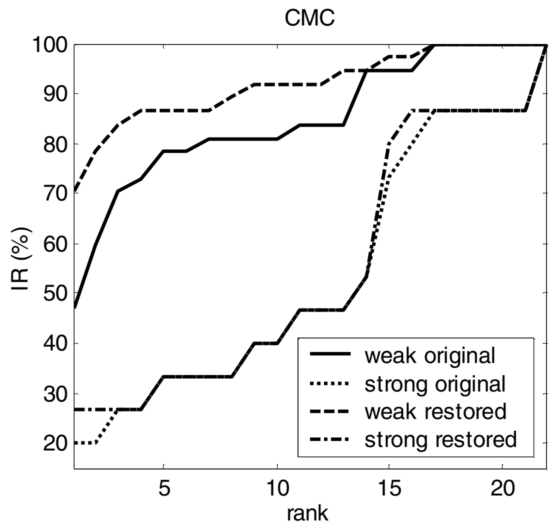


Fig. 4. Cumulative Matching Characteristic (CMC) curves for the evaluation of the restoration strategy in the identification scenario. The curves refer to the performances obtained on the original and on the restored faces, for both weak and strong occlusions.

To date, our strategy exploits three-dimensional data but we plan to investigate a similar approach on 2D and multimodal 2D+3D images as well. In order to provide a more accurate evaluation of our strategy, we also plan to extend the dataset used in the experimentation

5. REFERENCES

[1] B. Achermann and H. Bunke, “Classifying range images of human faces with Hausdorff distance,” *Proc. of 15-th Int. Conference on Pattern Recognition*, pp. 809–813, 2000.

[2] G. Medioni and R. R. Waupotitsch, “Face recognition and modeling in 3d,” *Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, pp. 232–233, 2003.

[3] J. Y. Cartoux, J. T. Lapreste, and M. Richetin, “Face authentification or recognition by profile extraction from range images,” *Proc. of IEEE Computer Society Workshop on Interpretation of 3D scenes*, pp. 194–199, 1989.

[4] T. Nagamine, T. Uemura, and I. Masuda, “3d facial image analysis for human identification,” *Proceedings of International Conference on Pattern Recognition*, pp. 324–327, 1992.

[5] C. Beumier and M. Acheroy, “Automatic 3d face authentication,” *Image and Vision Computing, Special Issue on Facial Image Analysis*, vol. 18(4), pp. 315–321, 2000.

[6] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, “Three-dimensional face recognition,” *International Journal of Computer Vision*, vol. 64(1), pp. 5–30, 2005.

[7] R. Brunelli and T. Poggio, “Face recognition: features versus templates”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15(10), pp. 1042–1052, 1993.

[8] A. Pentland, B. Moghaddam, and T. Starner, T, "View-based and modular eigenspaces for face recognition", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84–91, 1994.

[9] K. Ohba and K. Ikeuchi, "Detectability, uniqueness, and reliability of eigen windows for stable verification of partially occluded objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19(9), pp. 1043–1048, 1997.

[10] A. M. Martinez, Recognition of partially occluded and/or imprecisely localized faces using a probabilistic approach", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 712–717, 2000.

[11] A. M. Martinez, "Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24(6), pp. 748–763, 2002.

[12] M. Turk, and A. Pentland, "Eigenfaces for Recognition", *J. of Cognitive Neuroscience*, vol. 3(1), pp. 71–86,1991.

[13] R. Everson and L. Sirovich. “Karhunen-loève procedure for gappy data”. *Journal of Optical Society of America A*, vol. 12(8):, pp. 657–1664, 1995.

[14] A. Colombo, C. Cusano, R. Schettini, “3D face detection using curvature analysis”, *Pattern Recognition*, vol. 39(3), pp. 444–455, 2006.

[15] P. Belhumeur, J. Hespanha, and D. Kriegman. “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 19(7), pp. 711–720, 1997.