8

Occlusions in Face Recognition: a 3D Approach

Alessandro Colombo, Claudio Cusano and Raimondo Schettini
Università degli studi di Milano Bicocca, Italy

1. Introduction

The real challenge in face detection and recognition technologies is the ability to handle all those scenarios where subjects are non-cooperative and the acquisition phase is unconstrained. In the last few years a great deal of effort has been spent to improve the performances where cooperative subjects are acquired in controlled conditions. However, in those scenarios other biometrics, such as fingerprints, have already proved to be well suited; in fact, the performances obtained using them are good enough to implement effective commercial systems. In all those cases where the application requires no constraints during the acquisition phase, the face is one of the best candidates among biometrics. The face is a non-touch biometrics and is also the natural way people use to recognize each other: for this reasons it is more accepted by the final users.

The fundamental problem in recognizing people in unconstrained conditions is the great variability of the visual aspect of the face introduced by various sources. Given a single subject, the appearance of the face image is disturbed by the lighting conditions, the head pose and orientation of the subject, the facial expression, ageing and, last but not least, the image may be corrupted by the presence of occluding objects. This great variability is the reason that makes face detection and recognition two of the toughest problems in the fields of pattern recognition, computer vision and biometrics.

One of the less studied aspects seems to be the presence of occluding objects. In unconstrained real-world applications it is not an uncommon situation to acquire subjects wearing glasses, scarves, hats etc.; or subjects talking on the phone, or, for some reason, having their hands between their face and the camera. In all these kinds of situations most of the proposed algorithms are not able to grant acceptable performances or to produce any kind of response at all.

Some approaches (Lin, 2004; Hotta, 2004) propose the detection of partially occluded faces in two dimensional images using Support Vector Machines (SVM) or a cascade of classifiers trained to detect subparts of the face. For recognition, few approaches working on 2D data are able to recognize people using only the visible parts of the face. For example, (Park et al., 2005) have proposed a method for removing glasses from a frontal image of the human face. A more general solution is needed, however, when the occlusions are unforeseen and the characteristics of the occluding objects are unconstrained. The problem has been addressed using local approaches which divide the face into parts which are independently compared. The final outcome is determined by a voting step. For example, see (Martinez, 2000; Martinez, 2002; Kim et al. 2005). A different approach has been investigated by (Tarrés &

Source: State of the Art in Face Recognition, Book edited by: Dr. Mario I. Chacon M.,
ISBN -3-902613-42-4, pp. 250, January 2009, I-Tech, Vienna, Austria

Rama, 2005). Instead of searching for local non-occluded features, they try to eliminate some features which may hinder recognition accuracy in the presence of occlusions or changes in expression. In (De Smet et al., 2006) a morphable model based approach is proposed. The parameters of a 3D morphable model are estimated in order to approximate the appearance of a face in a 2D image. Simultaneously, a visibility map is computed which segments the image into visible and occluded regions.

In this chapter we present our approach: a full automatic recognition pipeline based on 3D imaging (Fig. 1). We have chosen to use 3D data because by having depth information available, it is easier to detect and isolate occluding objects. This makes it possible to detect and recognize partially occluded faces. There are also other well known advantages using 3D sensors: lighting independence, the possibility to normalize pose and orientation, and last but not least, 3D sensors are more difficult to circumvent compared to 2D cameras (Fig. 2).

The first three modules of the pipeline are the core of our method. The face detector, based on an improved version of a primal approach presented in (Colombo et al., 2006), is able to localize the presence of human-size faces in depth images. The algorithm is based on curvature analysis, ICP-based normalization and Gappy PCA classification. The main advantages of the algorithm are its independence from scale and orientation in the image plane and its ability to deal with occlusions if at least two eyes or one eye and the nose are visible. The output of the detector is the normalized position and orientation of the faces in the acquired scene. At this point, the image of each detected face is analyzed by the occlusion detection module. Each pixel is classified as occluded or free using a statistical model of the human face. Once the visibility map is computed, the restoration module reconstructs the occluded parts using gappy PCA reconstruction (Everson & Sirovich, 1996). Since the images are restored, it is possible to adopt any state of the art feature extraction and matching module. Alternatively, a partial matching approach may be adopted using the non-restored face image and the visibility map.

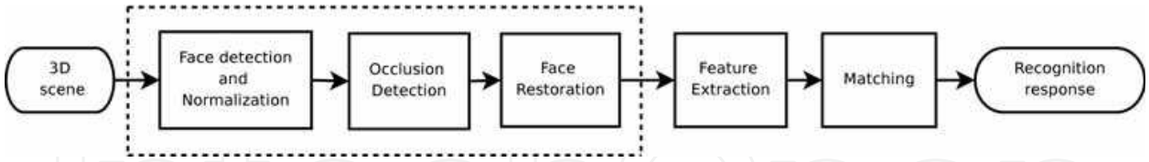

Fig. 1. The proposed approach. The modules enclosed in the dotted box are those described in the paper. Feature extraction and matching modules may be any state of the art approach.

2. Face detection and normalization

Face detection in 3D images is not a common task; a primal approach has been presented in (Colombo et al., 2006). Detecting faces in 3D images presents some benefits: first of all, 3D data is independent from scale. This is a big advantage because face detectors have to deal with one less degree of freedom. The second reason to adopt 3D images regards the nature of the 3D data itself. There are some properties of surfaces, like the curvatures, that are independent from the reference system. Using features like these, it is possible to recognize typical parts of the face independently from pose or viewpoint. For these reasons, 3D data facilitate the problem of face detection.

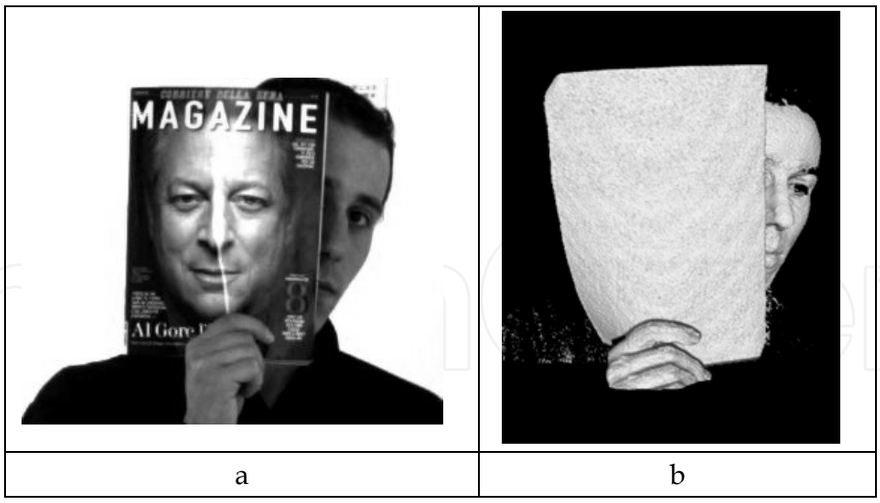

Fig. 2. (a) a subject trying to circumvent a 2D recognition system using a magazine. (b) the same scene acquired by a 3D sensor.

The approach that will be presented here reveals another important feature of 3D data. Considering the depth component of the image, the discrimination between face and occluding objects can be done using simple criteria. Intuitively an occluding object can be simply viewed, in terms of depth information, as something that is added to the original face surface, as can be seen in Figure 3. As will be shown, it is possible to handle this kind of noise in complex tasks like face detection and normalization.

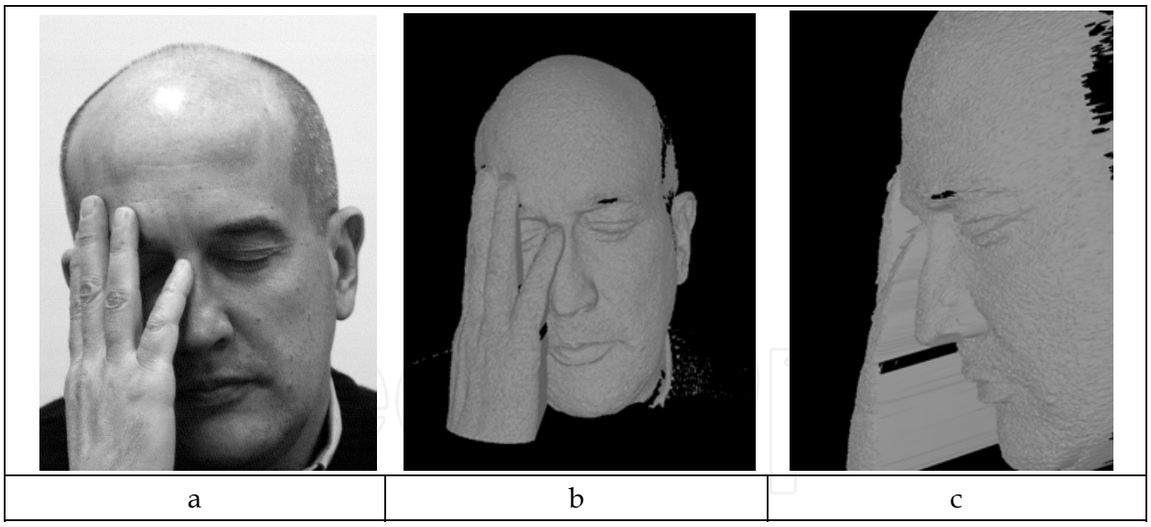

Fig. 3. (a) A two dimensional image of an occluded face. (b) The three dimensional image of the same scene represented in (a). (c) Profile view of (b).

2.1 Algorithm overview

The face detector is based on the work presented in (Colombo et al., 2006). The input of the algorithm is supposed to be a single 3D image of the scene. If other representations are available, a 3D image can be generated using simple and well known rendering techniques, like variants of the Z-Buffer algorithm (e.g. Watt, 1999). Figure 4 shows a concise diagram

representing the main steps of the algorithm. To render the problem less computationally intensive, single facial features, such as eyes and noses, are initially searched. Consequently, this first step results in an image segmentation in regions corresponding to candidate facial features. No relationships are established for the moment, between the segmented elements. In the next step potential faces are created from candidate noses and/or candidate eyes. The goal now is to discriminate between candidate faces that correspond to actual faces and those that do not. The method now registers each candidate face in a standard position and orientation, reducing intra-class face variability. The areas of the range image covered by each candidate face is further analyzed by applying a face versus non-face classifier based on a holistic approach. Knowledge about the structure of the face is used therefore only in the generation of a list of candidate face regions, while the actual classification of these regions as faces is purely holistic.

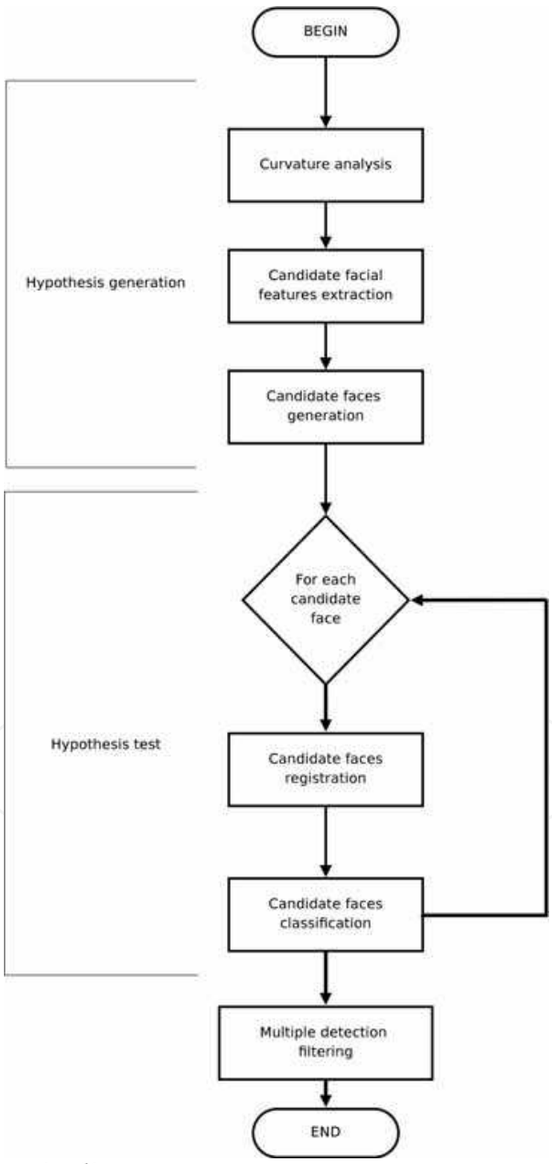

Fig. 4. The face detector main diagram

In Figure 5 a more detailed description of the processing steps is shown. Once the scene is acquired, surface curvature, which has the valuable characteristic of being viewpoint invariant, is exploited to segment candidate eyes and noses. In greater detail: (i) the mean (H) and Gaussian (K) curvature maps are first computed from a smoothed version of the original range image; (ii) a simple thresholding segments regions of high curvature which

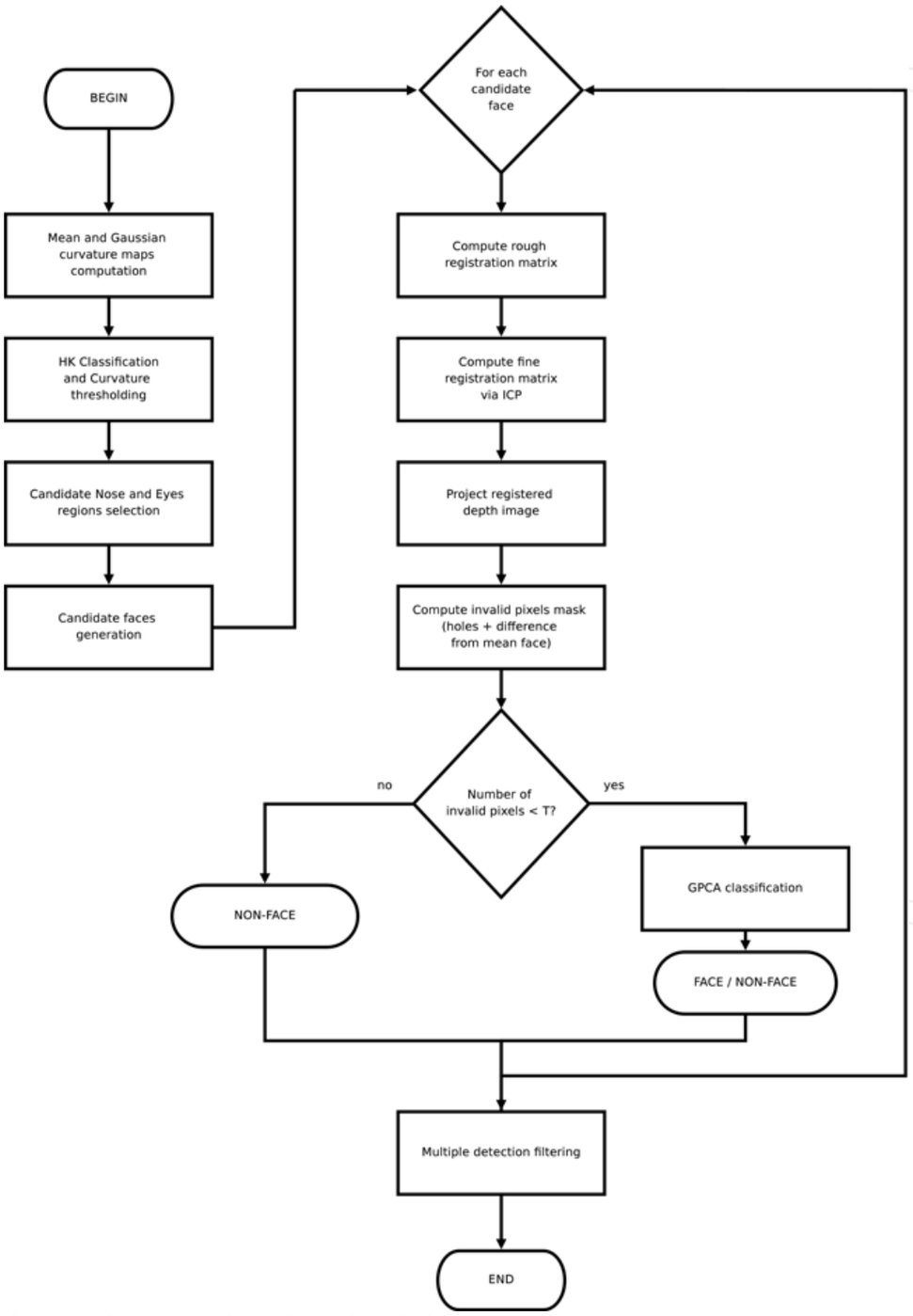

Fig. 5. The face detection algorithm: detailed diagram.

might correspond to eyes and noses; (iii) an HK classification, based on the signs of Gaussian and mean curvature, divides the segmented regions into four types: convex, concave, and two types of saddle regions. Regions that may represent a nose or an eye are then characterized by their type and some statistics of their curvature. The output of the processing step may contain any number of candidate facial features. If no nose or eyes are detected, the method assumes that no faces are present in the acquired scene; while there are no upper bounds on the number of features that can be detected and further processed. Combinations of candidate features are used to select corresponding 3D surface regions including the eyes and the nose, but excluding the mouth and part of the cheeks.
Each region is then rotated and translated into a standard position using a rough+fine registration approach based on an occlusions-tolerant version of the ICP algorithm (Besl & McKay, 1992), and a new depth image of the area containing the candidate facial features is computed. In order to select only the rigid part of the face, the image is cropped with a binary mask. Then, the image is analyzed in order to find occluding objects: if present, occluding objects are eliminated from the image invalidating the corresponding pixels. Finally, a face vs. non-face gappy PCA based classifier, which has been trained on several examples, processes the candidate depth image. The final output of the procedure is a list containing the location and orientation of each detected face.

2.2 Candidate faces generation

Given in input a 3D scene (in the form of a range image in our current implementation) a curvature analysis reveals regions of interest having similar characteristics to those of human eyes and noses. By combining all these regions, it is possible to generate hypotheses about the presence of faces in the scene. For the generation of a single hypothesis, at least two eyes or an eye and a nose must be present. The curvature analysis processing is based on the computation of gaussian and mean curvature maps. Using simple thresholding of low curvature values it is possible to isolate potential noses (convex regions) and eyes (elliptical concave regions). See (Colombo et al., 2006) for an in-depth description of the procedure adopted.
The generation of candidate facial features results in a set composed of two kinds of features: eyes and noses. A single candidate face is generated combining two eyes, or an eye and a nose, or two eyes and one nose. These cases can handle occlusions in some parts of the face; for example a hand on an eye or a scarf in front of the tip of the nose. The regions generating a candidate face must satisfy some constraints about distances between themselves (Gordon, 1991). Figure 6 shows an example of a candidate face generation.
From an actual face multiple candidates could be generated. Moreover, in the case of eye pairs or nose-eye pairs, double candidate faces are generated because of the ambiguities regarding the actual face orientation. For example, from a pair of eyes, either an upward or a downward face might be present. Multiple, spurious detections are eliminated as a final step by a simple filtering process.

2.3 Candidate face normalization

Once generated, candidate face images are normalized in pose and orientation in order to perform a final classification into faces or non-faces. The normalization is computed in two steps. First, a rough normalization is performed using the position of the candidate facial features. A reference position is defined aligning the eye positions with the X axis and the

plane passing through the eyes and the tip of the nose rotated by 45 degrees around the same axis. Finally, the tip of the nose is translated to the origin. In all those cases where one feature is missing (for example in case of occlusions) a degree of freedom is left undetermined.

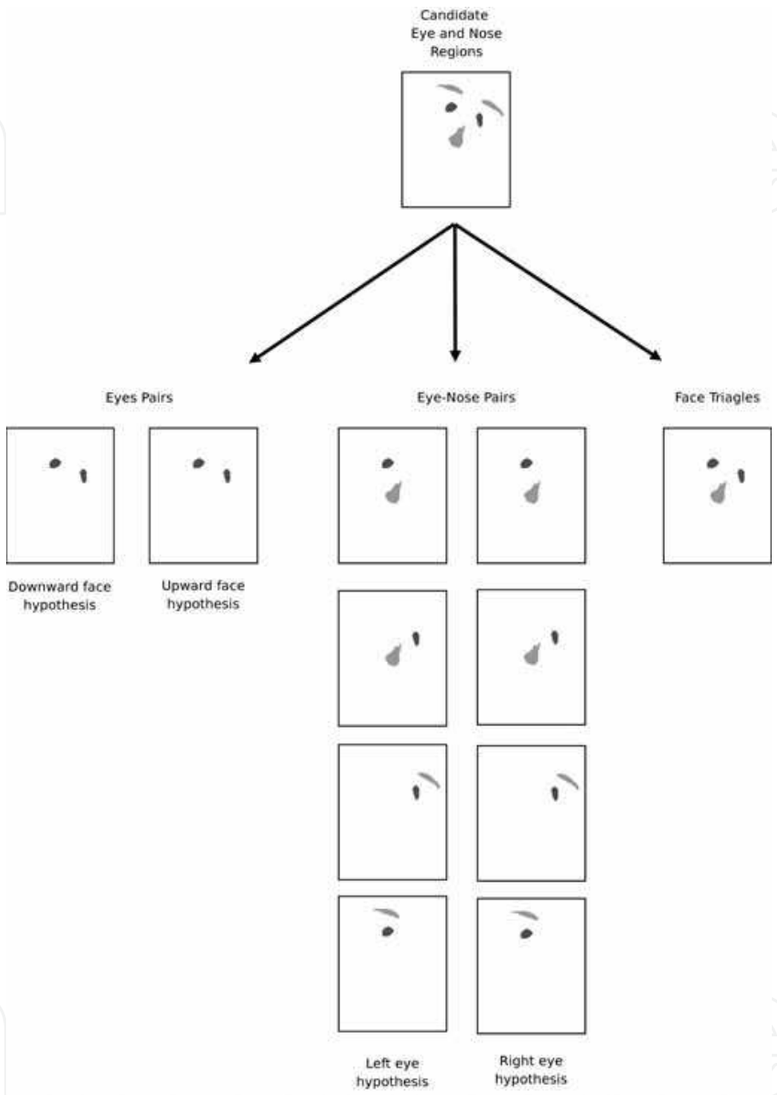

Fig. 6. An example of candidate face generation. The image on top shows the selected candidate features (light gray for noses, dark gray for eyes). Below, all the candidate faces generated by the algorithm are shown. In the case of eye pairs and eye-nose pairs, two candidate faces are generated for each pair because of the impossibility of determining the actual face orientation.

The rough normalization procedure generates a good starting position for the following refinement step based on the ICP algorithm (Besl & McKay, 1992). The idea here is to refine the position registering the candidate face with a mean face template (Fig. 7). We used a customized version of the ICP algorithm aimed to handle the presence of extraneous objects. The algorithm has been inspired by the variants presented in (Rusinkiewicz & Levoy, 2001).

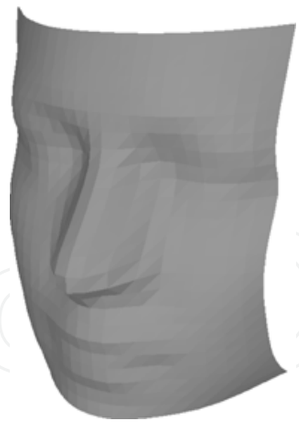

Fig. 7. The mean face template used for fine registration. The surface, composed of approximatively 600 vertexes, has been computed using a training set of manually normalized faces.

The ICP algorithm requires a matching criteria in order to find correspondences between the points of the surfaces to be registered. In our implementation we used a projective matcher . Based on the assumption that the rough registration computes a good registration (i.e. with a low error) between the mean face and the candidate face surface, the projective matcher tries to find correspondences using orthographic projections of each vertex. More precisely, at each iteration of the main ICP loop, the mean face template and the candidate face are orthographically projected using the same camera, resulting in a pair of 3D images, the data image and the model image. For each point located at coordinates *(i,j)* in the data image space, the correspondent point is searched in the model image space in locations *(i±r, j±r)*; where *r≥1* is an integer defining a square region around the current location. The correspondence criterion is the point at minimum distance using the 3D Euclidean metric.
In order to deal with the presence of occlusions, ICP has been customized by including a correspondence rejector which allows the registration process to avoid the use of those points which probably belong to the occluding objects.
Given a correspondence :

$$c=(\vec{p}_d, \vec{p}_\mu, \vec{n}_d, \vec{n}_\mu)$$

(where *p* indicates a surface point, *n* a surface normal; the subscript *d* indicates the data surface while *μ* indicates the mean surface) the rejector verifies that the following conditions are satisfied:

$$\arccos(\vec{n}_d \cdot \vec{n}_\mu) \leq T_\alpha, \tag{1}$$

$$\vec{p}_d - \vec{p}_{\mu 2} < T_d. \tag{2}$$

The first condition (Equation 1) checks if the angle between the two normals is inferior to a predefined threshold. We have chosen a value of 90 degrees; so the check filters out all those matches that are clearly wrong because the orientation of the surfaces is very dissimilar.
The second condition (Equation 2) assures that the distance between the two correspondent points must be below a predefined threshold. The value has been computed considering the variations between the normalized non-occluded faces from the training set and the mean

face template. A distance greater than 15mm is very improbable (see Fig.8 for the distribution of the differences) and thus we have chosen this value.
If at least one of the two conditions is not satisfied then the correspondence is rejected. Only the filtered correspondences are used to compute the registration.

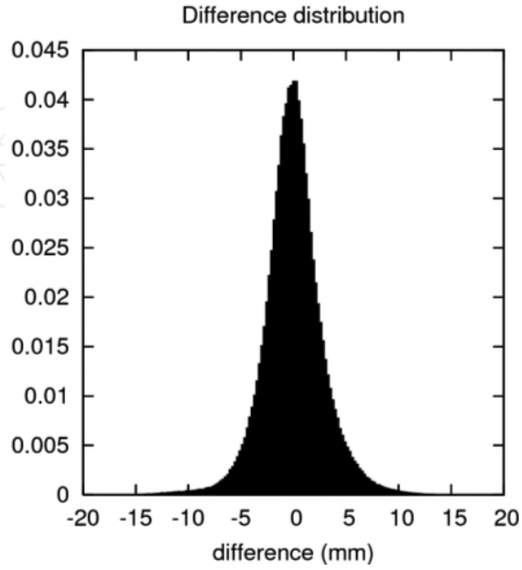

Fig. 8. The histogram of the differences in depth between the mean face template and the normalized faces from the training set.

2.4 Candidate face classification through GPCA

Once registration is computed for each candidate face, depth images are generated using orthographic projections of the original acquisition. After cropping, each image is compared with the mean face in order to detect occluding objects. For each pixel p the following condition is checked:

$$|Y(p) - \mu(p)| \leq T_d, \tag{3}$$

where Y is the candidate face depth image, while μ is the mean face depth image. T_d is the same threshold used in Equation 2. If the check fails, the pixel is invalidated. In this way, large parts of occluding objects can be eliminated. Those non-face pixels passing the check are assured to be limited in depth by T_d. Since large regions of the image may be invalid, a check on the fraction of valid pixels is performed:

$$\frac{n_v}{N} \geq T_v \tag{4}$$

where N is the number of total pixels in the image and T_v is the valid pixel threshold. Images with a low number of valid pixels are more difficult to classify because of the lack of information. The check is also used to avoid degenerative cases; i.e. images composed of only a few pixels. At this point, images present invalid regions of pixels due to occlusion detection or holes generated by acquisition artifacts. In order to classify them, a Gappy PCA classifier has been adopted. The classifier is based on Principal Component Analysis for gappy data (Everson & Sirovich, 1995).

GPCA extends PCA to data sets that are incomplete or gappy. When the intrinsic dimension is smaller than that of its representation, some of the information in the original representation is redundant, and it may be possible to fill in the missing information by exploiting this redundancy. The procedure requires knowledge of which parts of the data are available and which are missing.
A set of N patterns

$$\{\vec{x}_1, \cdots, \vec{x}_N\} \subset R^n,$$

extracted from a training set of normalized non-occluded faces is used to determine the PCA basis so that a generic pattern $\vec{x}$ can be approximated using a limited number, M, of eigenvectors:

$$\vec{x} \simeq \vec{\mu} + \sum_{i=1}^{M} \alpha_i \vec{v}_i, \tag{5}$$

where $\vec{\mu}$ is the mean vector, $\vec{v}_i$ is an eigenvector, and α_i is a coefficient obtained by the inner product between $\vec{x}$ and $\vec{v}_i$.
Suppose there is an incomplete version $\vec{y}$ of $\vec{x}$, and suppose that the location of missing components is encoded in the vector $\vec{m}$ ($\vec{m}_i = 0$ if the i-component is missing, otherwise $\vec{m}_i = 1$). GPCA searches for an expression similar to Equation 5 for the incomplete pattern $\vec{y}$:

$$\vec{y} = \vec{y}' \simeq \vec{\mu} + \sum_{i=1}^{M} \beta_i \vec{v}_i, \tag{6}$$

Note that $\vec{y}'$ has no gaps since the eigenvectors are complete. To compute the coefficients β_i the square reconstruction error *E* must be minimized:

$$E = \|\vec{e}\|^2 = \|\vec{y} - \vec{y}'\|^2. \tag{7}$$

However, this expression includes the missing components, while only the available information must be considered. To do so, it is useful to introduce the gappy inner product and the corresponding gappy norm:

$$(\vec{v}, \vec{u})_{\vec{m}} = \sum_{i=1}^{n} \vec{v}_i \vec{u}_i \vec{m}_i. \tag{8}$$

$$\|\vec{v}\|_{\vec{m}} = \sqrt{(\vec{v}, \vec{v})_{\vec{m}}} \tag{9}$$

Now the error E can be redefined in such a way that only the available components are considered:

$$E = \|\vec{e}\|_{\vec{m}}^2 = \|\vec{y} - \vec{y}'\|_{\vec{m}}^2. \tag{10}$$

The gappy pattern $\vec{y}$ can be reconstructed as $\vec{y}'$ using the Equation 6, where the coefficients β_i are found by minimizing E (for additional details see Everson & Sirovich, 1995). Figure 9 shows an example of a face range image reconstructed using GPCA.

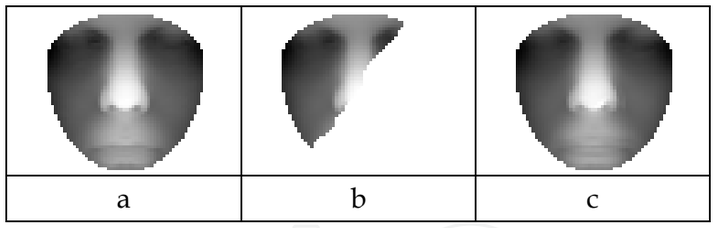

Fig. 9. Example of a reconstruction of a gappy image: (a) the original depth image; (b) the same image with a region of invalidated pixels; (c) the reconstruction of (b) through gpca.

The classifier used for the face detector constructs the vector $\vec{m}$ through lexicographical construction, considering all the invalidated or holes pixels. The error defined in Equation 10 is used as a measure of faceness. Thus, a candidate face is classified as a face if the following condition is satisfied:

$$\frac{E}{n_v} \leq T_f \tag{11}$$

where n_v is the number of valid pixels and T_f is a predefined threshold. The error E needs some kind of normalization because the number of valid components is not fixed. Here the most simple normalization, the division by the number of valid pixels n_v, has been adopted; but other kinds of normalization approaches could be experimented with as well.

3. Occlusion detection and face restoration

At this stage of the pipeline, faces are encoded as range images projected from the normalized position produced by the face detector. Any part of these images that does not look like part of a face and lies between the acquisition device and the acquired face is considered an occlusion (occluding objects may not touch the face). In other words, sets of points which do not fit the model of a non-occluded 3D face are classified as occlusions. The 3D face model is obtained by the popular eigenfaces approach (Sirovich & Kirby, 1990). Occluded faces cannot be well represented by the linear combination of the computed eigenfaces. Therefore, the distance from the face space (DFFS) of occluded faces is expected to be quite large and can be used to reveal the presence of occlusions, though not their location. Differences in the pixels of the reconstructed and the original image (vector $\vec{e}$) are likely to be more pronounced where the face is occluded.
A preliminary mask **M** of occlusions can be simply obtained by thresholding vector $\vec{e}$:

$$\mathbf{M}_i = \begin{cases} 1 & \text{if } \vec{e}_i > T_\tau \\ 0 & \text{if } \vec{e}_i \leq T_\tau \end{cases} \tag{12}$$

where T_τ is a threshold that must take into account the resolution of the imaging device, acquisition noise, and the accuracy achieved in face detection and pose normalization. Since occlusions must be located between the acquisition device and the acquired face, their z coordinates will be greater than those of the reconstructed face, resulting in positive components of $\vec{e}$. Unfortunately, the reconstruction error of the non-occluded regions is influenced by the occlusions, which may determine an inaccurate choice of the gappy

projection coefficients β_i. However, the effect is significant only for occlusions which are very far from the face, or very large in size. Occluding objects which are far from the face are detected on the basis of the difference between the (possibly) occluded face $\vec{y}$ and the mean (non-occluded) face $\vec{\mu}$. Components of $\vec{x}-\vec{\mu}$ which are positive and large enough can be considered part of an occlusion. A mask $\vec{B}$ of these occlusions is obtained as follows:

$$\mathrm{B}_i = \begin{cases} 1 & \text{if } \vec{y}_i - \vec{\mu}_i > T_\rho \\ 0 & \text{if } \vec{y}_i - \vec{\mu}_i \leq T_\rho \end{cases} \tag{13}$$

T_ρ threshold must be tolerant with respect to face variability in the data set to be processed. As previously discussed, it is very unlikely that any face differs more than 15 mm from the mean face. Consequently, the threshold T_ρ has been set at that value.

A more accurate estimate of the occlusion mask **M** can be obtained by excluding from the computation of the reconstruction error E the pixels selected in mask B. The key idea here is to use only non-occluded parts of the face to estimate the distance between the face space and the processed face. For this, we apply Gappy Principal Component Analysis (GPCA). The coefficients obtained by minimizing Equation 10 can now be substituted in Equation 6 to determine a more accurate reconstruction error.

The same technique may be applied to deal with cases in which the 3D face in input is incomplete (e.g. missing data from the scanner). Holes can be encoded in a mask **H**, so that the gappy products $(\cdot,\cdot)_B$ may be replaced by $(\cdot,\cdot)_{B\vee H}$, where $B \vee H$ is the component-wise logical OR of holes and occlusions. Analogously, color information could be exploited if available: points having a color which is not likely to be found on a face can be marked as occlusions.

On the basis of the new reconstruction error E obtained using the coefficients β_i, the occlusion mask **M** can now be determined more precisely. Here again, when the reconstruction error is high, the corresponding pixel is considered occluded. A variation of (12) is used:

$$\mathrm{M}_i = \begin{cases} 1 & \text{if } (\mathbf{e}_i > T_\tau) \vee \mathbf{B}_i = 1 \vee \mathbf{H}_1 = 1 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

As a final step, morphological filters can be used to clean the occlusion mask, enforce the locality of the occlusion, and discard tiny regions. If the resulting mask **M** is empty, the face is considered non-occluded, and can be directly processed for recognition. Otherwise **M** replaces **B** as the mask for the gappy projection. The solution yields the coefficients β_i, which employed as indicated in Equation 6 allow for the restoration of the input face.

At this stage, any recognition algorithm, whether holistic or feature-based, can then be applied to recognize the restored face. Holistic approaches may use the fully restored image while partial matching or feature-based approaches may exclude occlusions using the information contained in **M**. Fig. 10 summarizes the whole occlusion detection and restoration procedure. Figure 11 presents an example of restoration showing the steps in the computation of the occlusion mask.

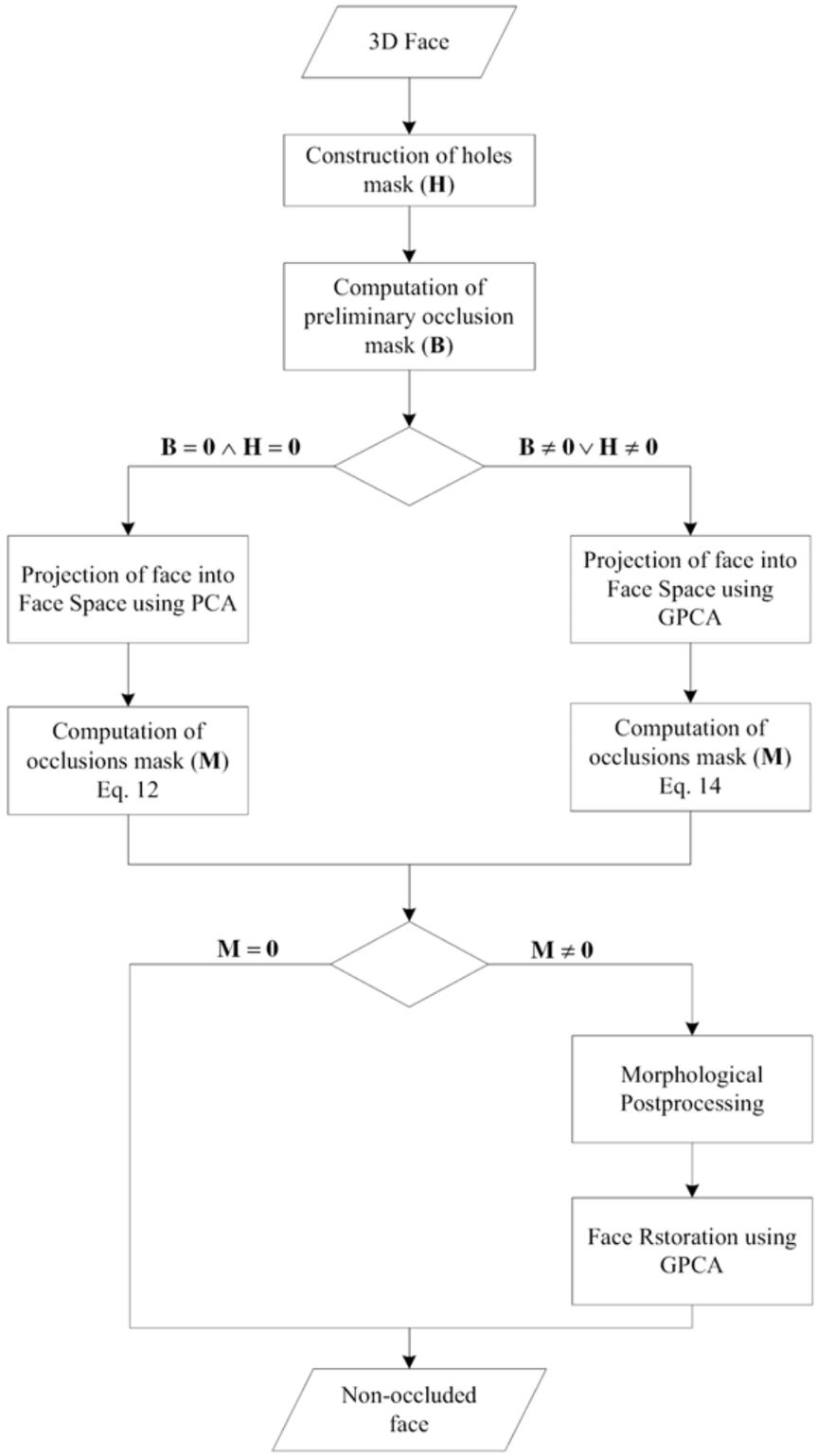

Fig. 10. The occlusion detection and face restoration diagram.

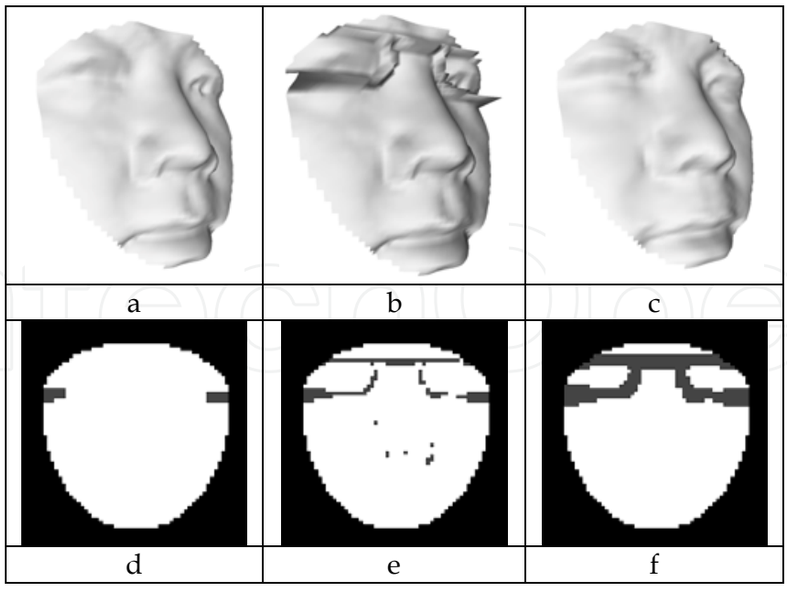

Fig. 11. Occlusion detection and restoration on a face artificially occluded by glasses (b). The restored face (c) is very similar to the original non-occluded face (a). A few points have been preliminarily detected (d) as occluded using Eq. 13; the second step correctly identifies most of the occlusion (e) using Eq.14. The final occlusion mask (f) is the result of the application of simple morphological operators.

4. Experimental results

4.1 Dataset and artificial occlusion generation

In order to perform an in-depth analysis of the algorithms (which needs the presence of a complete ground truth) we decided to adopt the artificial occlusion solution. It consists in taking an existing database of non-occluded faces and adding occluding 3D objects in each acquisition. Occluding objects are real world objects acquired at the IVL Laboratory at the University of Milano Bicocca. The entire set of objects includes: a scarf, a hat, a pair of scissors, two types of eyeglasses, a newspaper and hands in different configurations. We have used the UND Database (Chang et al., 2003), which consist of 951 3D+2D acquisition of 277 subjects.

Occluded acquisitions are generated inserting the objects in the acquisition space at plausible positions. This means, for example, that the eyeglasses are placed in the eyes region, the scarf in front of the mouth and so on. Thus, for each type of object, a starting position and orientation (T_o, R_o) is manually predefined considering the normalized mean face template as a reference. Then, for each acquisition an object is randomly chosen and placed on the face. The position and orientation is finally perturbed with random noise in order to increase the variability of the test set. Fig. 12 shows some examples.

4.2 Face detection results

The face detector and normalization module has been tested occluding a subset of 476 acquisitions from the 951 images of the UND Database. A training set of 150 non occluded faces, taken at the IVL Laboratory with the Minolta Vivid 900 range scanner, has been used as the training set for the GPCA eigenspace computation.

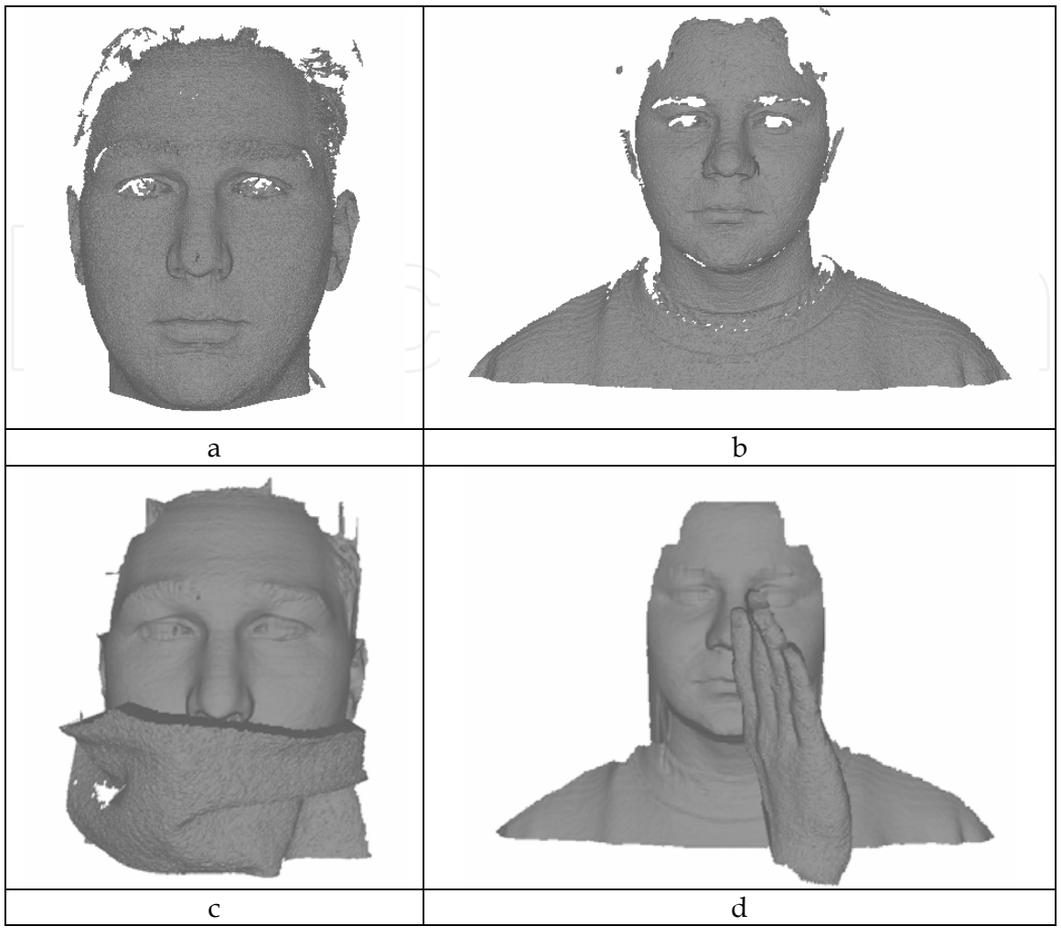

Fig. 12. Some examples of artificially occluded faces. (a,c) The original UND acquisitions; (b,d) the occluded faces. Note that automatic preprocessing is applied to the occluded faces, i.e. smoothing and simple hole closure via linear interpolation.

Figure 13 shows the receiver operator characteristic curves for the GPCA classifier at different values of the valid pixel threshold T_v. The curves have been computed considering only the candidate faces produced by the hypothesis generation phase of the algorithm. As can be seen, performance deteriorates as T_v decreases. This is an obvious consequence due to the lack of information. Depending on the application requirements, T_v must be chosen accordingly.

Table 1 reports the results obtained when choosing a value for the classifier threshold T_f aimed at reducing false positives and a value of T_v of 0.5 (i.e. at least half of the face image must be non-occluded). In this case, tests were also performed on the original non-occluded test set. A fraction of 83% of the total number of occluded faces have been successfully detected, generating 43 false alarms. The results are satisfactory considering the toughness of the problem and the fact that a large number of the acquisitions would be missed using conventional 3D approaches. The detector performs very well on non-occluded faces, reaching 100% of the detected faces and just one false alarm.

Figure 14 shows some examples of correctly detected faces while Fig. 15 shows a subset of missed faces.

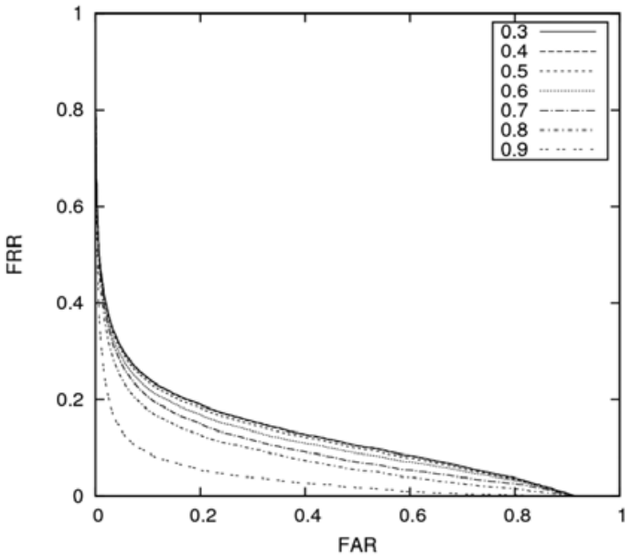

Fig. 13. The GPCA classifier ROC curves (FAR-False Acceptances Rate vs. FRR-False Rejections Rate) varying the value of the threshold T_v. and considering all the candidate faces from the 476 acquisitions of the test set.

Test Set	False Positives	False Negatives	Detected Faces
UND (non-occluded)	1	0	476/476 (100%)
UND (occluded)	43	77	399/476 (83.8%)

Table 1. Face detection results obtained on 476 acquisitions.

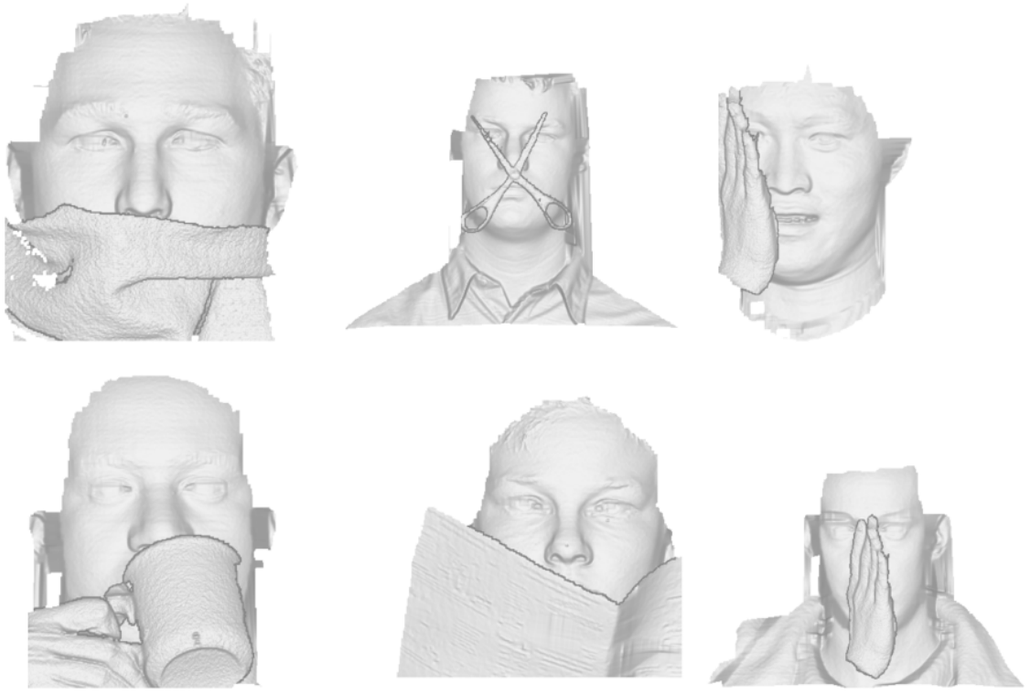

Fig. 14. Examples of detected faces taken from the artificially occluded UND dataset.

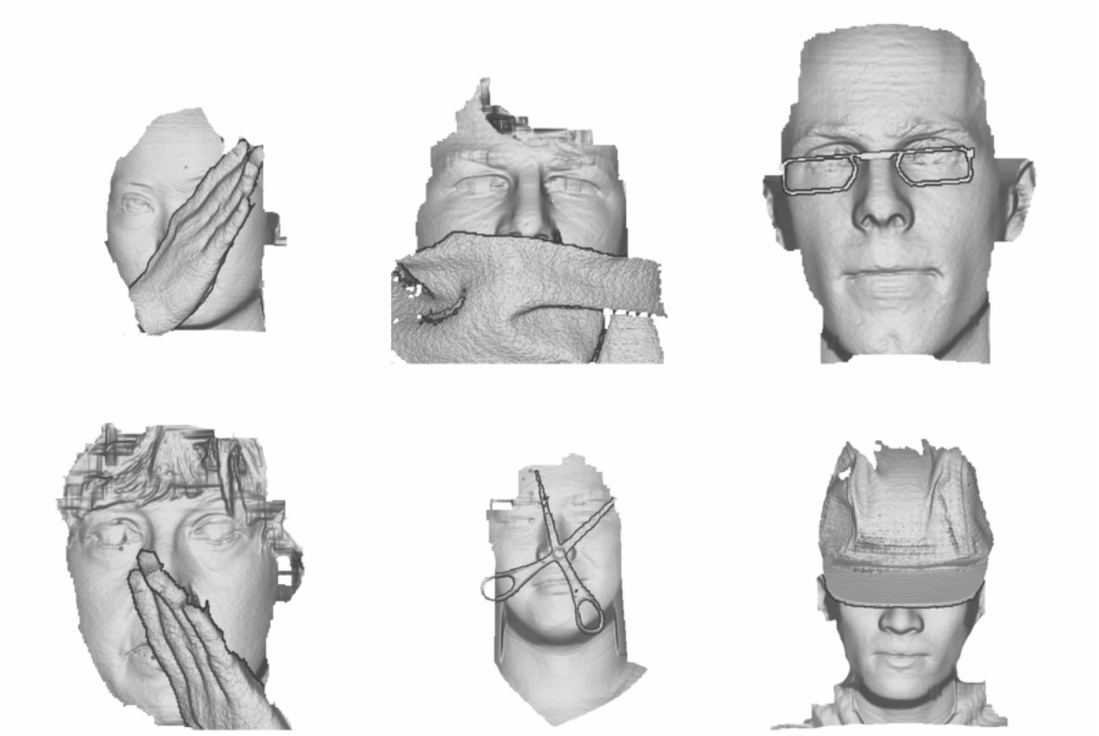

Fig. 15. Examples of missed faces taken from the artificially occluded UND dataset.

4.3 Face recognition results

The occlusion detection, face restoration and the following recognition modules have been tested considering the set of detected faces. For feature extraction and matching we used the popular Fisherfaces approach.

First, an analysis has been made of how the values of the thresholds T_ρ (Equation 13) and T_τ (Equation 14) influence the accuracy in the detection of occlusions. The training set has been used to estimate the mean face $\vec{\mu}$ and to evaluate the distribution of the differences between the pixels of non-occluded faces and those of the mean face. Taking into account such a distribution, the threshold T_ρ has been set to be 15 mm, which is more than five times the estimated standard deviation of the distribution (σ = 2.849 mm). Less than 0.01% of the pixels of the training images exceed the mean face by that value.

To evaluate the accuracy of the occlusion detection method the procedure has been run varying the threshold T_τ. Since it is known exactly which parts of the faces are occluded, it is possible to compute the precision (fraction of true positives among the pixels detected as occluded) and the recall (fraction of occluded pixels which have been detected). The results, obtained with both automatic as well as manual normalized dataset, are reported in Fig. 16. As can be seen, though the normalization error degenerates the occluded pixel detection performances, the approach is still effective in the case of automatic systems.

On the basis of these measures the authors considered the value of 1.9 mm for the threshold T_τ to be a good compromise which has been set to this value for the rest of the experiments.

As a measure of the accuracy of the restoration method, the pixel by pixel absolute difference between the original (non-occluded) and the restored faces has been considered.

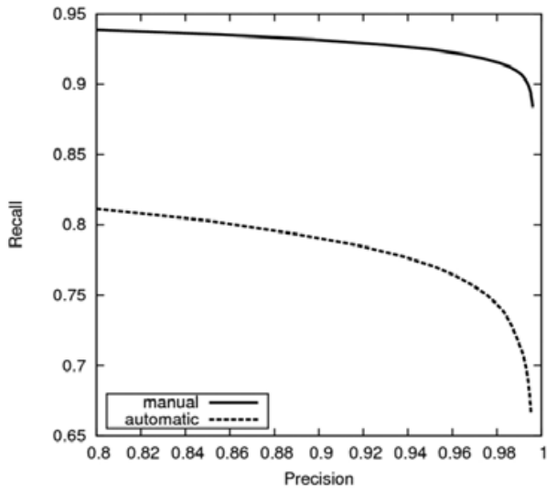

Fig. 16. Precision vs. recall corresponding to the pixels detected as occluded, obtained varying the value of the threshold T_τ. Manual (476 acquisitions) and automatic (399 acquisitions) normalization.

Since the restoration accuracy is expected to be highly correlated to the extent of the occlusions, the results obtained as a function of the area of occluded regions are reported (see Fig. 17).

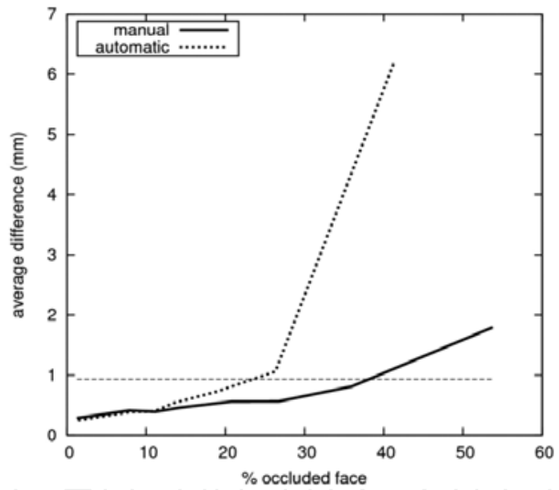

Fig. 17. Average absolute difference of the **z** coordinate between original and restored pixels, as a function of the fraction of the occluded face. Manual (476 images) and automatic (399 images) normalization. The horizontal line represents the average difference between two non-occluded acquisitions of the same subject, computed on manually normalized faces of the training set.

In Figure 18 shows two example of successful occlusion detection and restoration and an example of incorrect restoration.

However, since the error may be unevenly distributed over the faces, recognition performance could be affected more than expected. In order to ensure that a restored face may be reliably recognized, the same holistic recognition method has been applied to the original, the occluded, and the restored faces, and the performance has been compared in

the three cases. The Fisherfaces method (Belhumeur et al., 1997) has been adopted here, but any other method may be applied. Fisherfaces has been chosen for its popularity, and being a holistic approach, it is quite sensitive to normalization errors.

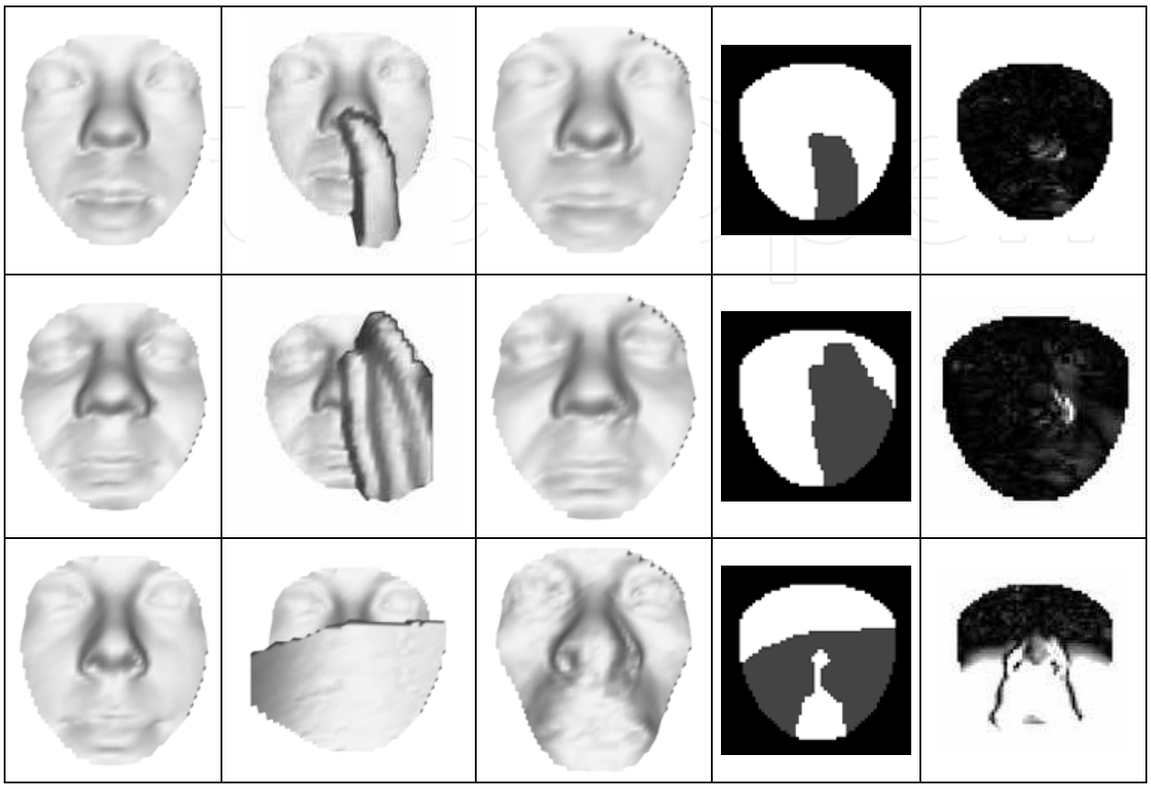

Fig. 18. First column: original faces. Second column: artificially occluded faces. Third column: restored faces. Fourth column: occlusion mask. Fifth columns: restoration error (brighter pixels indicate higher errors). The face in the third row represents a case of restoration failure.

We considered the identity verification scenario: the set of the remaining 475 non-occluded UND DB images not included in the test set have been used for building the known subject database and for training the Fisherfaces. Each test image has been proposed to the system, every time claiming a different identity (once for each subject included in the system DB).

Fig. 19 shows the ROC curves obtained on non-occluded, occluded, and restored faces in case of automatic detection and normalization. As expected, the recognition of occluded faces is very difficult: the Equal Error Rate (ERR) for occluded faces is 0.488, which is slightly better than random guessing. The application of the restoration strategy significantly improves the EER to a more acceptable 0.147, which is quite close to the 0.062 EER obtained on the original non-occluded faces.

In order to understand how much normalization errors degenerate results, we normalized the test set manually. Fig 20 shows the performances obtained with the manual procedure. As can be seen, normalization errors influence the system performances but the comparison of these results with those in Fig 19 shows the robustness of the automatic procedure.

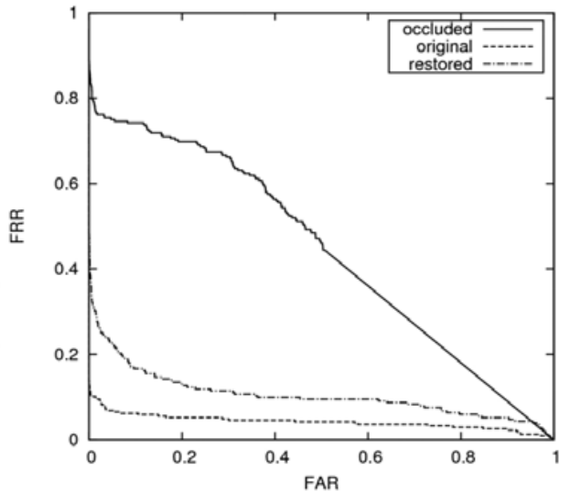

Fig. 19. ROC curves for the identity verification scenario. The curves refer to the performances obtained on the original (476 acquisitions), the artificially occluded and the restored test set (399 acquisitions). Images have been normalized automatically.

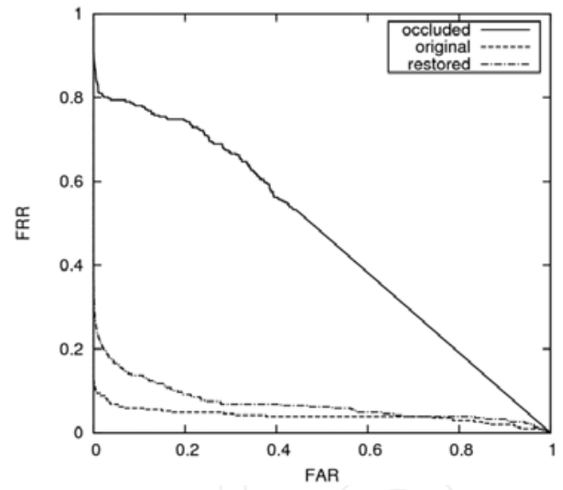

Fig. 20. ROC curves for the identity verification scenario. The curves refer to the performances obtained on the original (476 acquisitions), the artificially occluded and the restored test set (399 acquisitions). Images have been normalized manually.

To better understand the relationship between the occlusion extension and the recognition performance, the test set has been divided into three groups: 0-20%, 20-40% and over 40% (small, medium and large). Fig. 21 reports the ROC curves obtained using the three groups as test set. In the first case the EER is 0.07, which is comparable to the error rate measured on non-occluded faces. In the case of large occlusions, results are unacceptable (EER=0.5). For medium occlusions we obtained an EER of 0.27. In Fig 17 the restoration error begins to diverge approximatively around 30% of occluded face.

In conclusion the proposed face detection and normalization approach combined with the restoration module obtains promising results with occlusions smaller than approximatively 30% of the face.

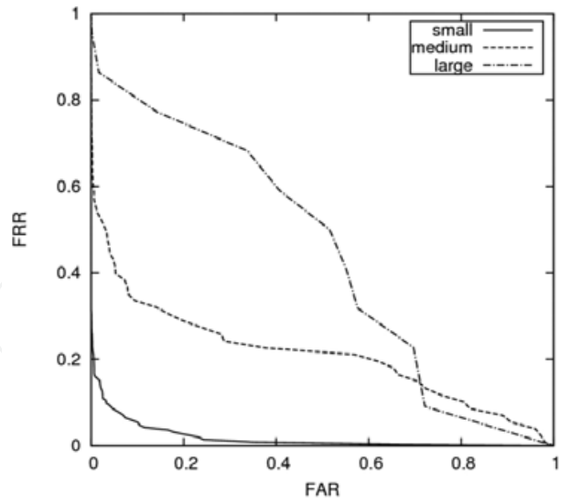

Fig. 21. ROC curves for the identity verification scenario obtained on the set of 399 detected faces, subdivided into "small", "medium" and "large" subsets.

5. Conclusions and future works

The presence of occluding objects in face recognition and more generally in object recognition tasks, is a "far from solved" problem. Here a solution has been presented which is composed of three core modules (detection, normalization and occlusion detection/face restoration) that could be employed in any 3D recognition system in order to improve its robustness. The results are quite promising and indicate that the use of 3D data may simplify the problem.

However, there are still a lot of open issues. First of all, is it possible to reduce the errors introduced by the automatic normalization of faces? Secondly, how does the proposed algorithm perform in the presence of emphasized facial expressions? How could the proposed algorithm be integrated with a facial expression tolerant solution?

Another important issue regards the comparison between face restoration approaches versus partial matching approaches. It is not clear at present if the missing or covered parts of the face should be reconstructed or simply detected and ignored, letting a partial matching strategy perform recognition.

Future studies must take into account all these aspects. A large database containing real occlusions and various types and degrees of facial expressions should be adopted in order to allow an in-depth study of the problem and the proposed solutions.

8. References

Belhumeur, P.N., Hespanha, J.P. & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on PAMI*, pp. 711-720, 1997.

Besl, P. J. & McKay N. D. (1992). A method for registration of 3-d shapes. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14:239-256, 1992.

Chang, K, Bowyer, K. W. & Flynn, P. J. (2003). Face recognition using 2D and 3D facial data. *ACM Workshop on Multimodal User Application*, pp. 25-32, 2003.

Colombo, A., Cusano, C. & Schettini, R. (2006). 3D Face Detection using Curvature Analysis. *Pattern Recognition*, vol. 39, no. 3, pp. 444-455, 2006.

Colombo, A., Cusano, C. & Schettini, R. (2008a). Recognizing Faces In 3D Images Even In Presence Of Occlusions. *IEEE Second International Conference on Biometrics (BTAS08)*, submitted, 2008.

Colombo, A., Cusano, C. & Schettini, R. (2008b). Gappy PCA Classification for Occlusion Tolerant 3D Face Detection. *IEEE Transactions on systems, man and cybernetics–Part B*, submitted, 2008.

Colombo, A., Cusano, C. & Schettini, R. (2008b). Three-dimensional Occlusion detection and Restoration Of Partially Occluded Faces. *IEEE Transactions on multimedia*, submitted, 2008.

De Smet, M., Fransens, R. & Van Gool, L. (2006). A Generalized EM Approach for 3D Model Based Face Recognition under Occlusions. IEEE *Computer Society Conference on Computer Vision and Pattern Recognition* vol.2, no., pp. 1423-1430, 2006.

Everson, R. & Sirovich, L. (1995). Karhunen-Loève Procedure for Gappy Data. *J. Optical Soc. of America A*, vol. 12, no. 8, pp. 657-1664, 1995.

Gordon, G. (1991). Face recognition based on depth maps and surface curvature. *In Proceedings of SPIE, Geometric Methods in Computer Vision,* volume 1570, pages 234-247, 1991.

Hotta, K (2004). A robust face detector under partial occlusion. *Proceedings of ICIP 2004,* pp. 597-600, 2004.

Kim, J, Choi, J.,Yi, J., & Turk, M. (2005). Effective representation using ICA for Face Recognition Robust to Local Distortion and Partial Occlusion. *IEEE Trans. PAMI*, vol 27, no. 12, pp. 1977-1981, 2005.

Kirby, M. & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on PAMI*, vol.12, no.1, pp. 103-108, 1990.

Lin, Yen-Yu, Liu, Tyng-Luh, & Fuh, Chiou-Shann (2004). Fast Object Detection with Occlusions. *The 8th European Conference on Computer Vision (ECCV-2004)*, Prague, May 2004.

Martinez, A. M. (2000). Recognition of Partially Occluded and/or Imprecisely Localized Faces using a Probabilistic approach. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 712-717, 2000.

Martinez, A. M. (2002). Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class. *IEEE Trans. PAMI*, vol. 24, on. 6, pp. 748-763, 2002.

Park, J. S., Oh, Y. H., Ahn, S. C., & Lee, S. W (2005). Glasses Removal from Facial Image Using Recursive Error Compensation. *IEEE Trans. PAMI*, vol. 27, no. 5, pp. 805-811, 2005.

Rusinkiewicz, S., Levoy, M. (2001). Efficient Variants of the ICP Algorithm. *Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, 2001.

State of the Art in Face Recognition

Edited by Julio Ponce and Adem Karahoca

ISBN 978-3-902613-42-4
Hard cover, 436 pages
Publisher I-Tech Education and Publishing
Published online 01, January, 2009
Published in print edition January, 2009

Notwithstanding the tremendous effort to solve the face recognition problem, it is not possible yet to design a face recognition system with a potential close to human performance. New computer vision and pattern recognition approaches need to be investigated. Even new knowledge and perspectives from different fields like, psychology and neuroscience must be incorporated into the current field of face recognition to design a robust face recognition system. Indeed, many more efforts are required to end up with a human like face recognition system. This book tries to make an effort to reduce the gap between the previous face recognition research state and the future state.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Alessandro Colombo, Claudio Cusano and Raimondo Schettini (2009). Occlusions in Face Recognition: a 3D Approach, State of the Art in Face Recognition, Julio Ponce and Adem Karahoca (Ed.), ISBN: 978-3-902613-42-4, InTech, Available from: http://www.intechopen.com/books/state_of_the_art_in_face_recognition/occlusions_in_face_recognition__a_3d_approach

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen