Noname manuscript No.
(will be inserted by the editor)

Gappy PCA Classification for Occlusion Tolerant 3D Face Detection

Alessandro Colombo · Claudio Cusano · Raimondo Schettini

the date of receipt and acceptance should be inserted later

Keywords Three dimensional face detection, three dimensional face recognition, face occlusions, curvatures, gappy principal component analysis, global registration.

Abstract This paper presents an innovative approach for the detection of faces in three dimensional scenes. The method is tolerant against partial occlusions produced by the presence of any kind of object. The detection algorithm uses invariant properties of the surfaces to segment salient facial features, namely the eyes and the nose. At least two facial features must be clearly visible in order to perform face detection. Candidate faces are then registered using an ICP (*Iterative Correspondent Point*) based approach aimed to avoid those samples which belong to the occluding objects. The final face versus non-face discrimination is computed by a Gappy PCA (*GPCA*) classifier which is able to classify candidate faces using only those regions of the surface which are considered to be non-occluded. The algorithm has been tested using the UND database obtaining 100% of correct detection and only one false alarm. The database has been then processed with an artificial occlusions generator producing realistic acquisitions that emulate unconstrained scenarios. A rate of 89.8% of correct detections shows that 3D data is particularly suited for handling occluding objects. The results have been also verified on a small test set containing real world occlusions obtaining 90.4% of correctly detected faces. The proposed approach can be used to improve the robustness of all those systems requiring a face detection stage in non-controlled scenarios.

DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione),
Università degli Studi di Milano–Bicocca,
Viale Sarca 336, 20126 Milano, Italy.
E-mail: {colomboal,cusano,schettini}@disco.unimib.it

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. 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, face is one of the best candidates among biometrics. Face is a non-touch biometrics; for this reason it is more accepted by the final users. Face is also the natural way people use to recognize each other. 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, aging and, last but not least, the image may be corrupted by the presence of occluding objects. This great variability is the reason that make face detection and recognition two of the toughest problems in the fields of pattern recognition, computer vision and biometrics.

Actually, commercial 3D scanners are not yet ready to be employed in completely unconstrained scenarios. Laser scanners, for example, require subjects to stand still for few seconds in front of the device. Although, 3D acquisition technology is rapidly converging to realtime scanners (e.g. see [29]) and in the next years 3D cameras

will be ready to be adopted for real unconstrained applications. The algorithm presented here can be adopted in scenarios where nowadays instruments are applicable and is ready to be employed on future acquisition devices.

Many works in the face related literature focus on the detection or recognition of faces even in presence of some source of variability. For instance, lighting has been approached using fisherfaces [1] or the relighting technique [2]. Facial expressions, instead, are the main focus of many works employing 3D data (see for example [3, 4, 27]). Pose and orientation normalization are considered an essential part and for this reason all detection and recognition systems include dedicated algorithms for this task.

One of the less studied problems 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 having for some reason, 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 (for example [5, 6]) 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. have proposed a method for removing glasses from a frontal image of the human face [7]. 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 an example, see [8], [9], [10]. A different approach has been investigated by Tarrés and Rama [11]. 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. De Smet et al. [23] proposed a morphable models based approach. 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. Gross et al. [28] proposed the use of an occlusion tolerant Active Appearance Models (AAM) for 2D face tracking. Alyüz et al. developed a system invariant from expressions and occlusions [31]. Their approach is based on Average Regional Models (ARMs) i.e. matching subparts of the face via ICP. Since their focus is on recognition, detection and coarse alignement is based on manually selected landmarks.

In this paper we try to address the occlusion problem for face detection in three dimensional images. Detecting faces in depth images is not a common task. One of the first algorithms has been proposed by the authors of this paper [12]. The known advantages in using 3D data for detection are the independency from the lighting condition and scale (if the target faces have human sizes). Here we will try to demonstrate a third advantage of 3D data: having depth information available, it is easier to detect and isolate occluding objects; this makes it possible to detect and recognize partially occluded faces.

Intuitively, an occluding object in a 2D image is something which has a different luminance/colour pattern from a typical face image. This is also true for the three-dimensional case if we consider the depth component. But here we have another important advantage because an occluding object is something between the camera and the face. This additional information is encoded in the 3D data itself. For example, in a range image a hat is far away from the facial surface. Considering the depth, this can be detected using simple geometric tests.

Our approach is based on the general idea behind the algorithm presented in [12], which was able to detect the presence of multiple faces in a single image and was also able to determine a normalized position for each face. The main advantages of the algorithm were its independence from scale, lighting condition and orientation on the image plane. Although, the algoritm was not able to detect faces in case of strong occlusion; i.e occlusions wich cover a large portion of the face or one of the most important features such as the eyes or the nose. There were also problems in dealing with missing data due to self occlusions or acquisition errors.

In this paper we propose a new algorithm which mantain only the main structure of our previous one, adding the capability to detect faces even in presence of self occlusions, occlusions generated by any kind of objects, and missing data. The algorithm is also able to determine a rough mask indicating the regions covered by the occluding objects and it is also able to give a better estimation of the face pose and orientation. Invariant properties of the surfaces are used to segment salient facial features, namely the eyes and the nose. At least two facial features must be clearly visible in order to perform face detection (this excludes, for instance the case of sunglasses). Candidate faces are then registered using an ICP [16] based approach aimed to avoid those

samples belonging to the occluding objects. The final face versus non-face discrimination in computed by a GPCA [13] classifier which is able to classify candidate faces using only those regions of the surface which are considered to be part of the face.

Pose and orientation normalization is a crucial step in a face recognition system. In particular, misaligned faces in 3D systems produce big performance degeneration [14]. There are two kinds of approaches in the literature: normalization to a canonical position or face-to-face registration. In the first case, a canonical position is defined a-priori; then each face is rotated and translated in order to assume the desired position. Sometimes the final position is represented by a mean model computed from a training set. In the second case, each face reaches the matching phase in its original pose and orientation. During matching, faces are registered one-to-one against the faces composing the database known subjects (e.g. [15]).

In our approach we have chosen to adopt the pre-matching strategy for face normalization. The face detector itself needs to compute surface registration in order to give the final face/non-face classification response. Moreover, it is in the authors opinion that pre-matching strategies are less time consuming and can easily implement more complex matching criteria during recognition, since matching does not have to cope with minimization processes which are typically involved in global registration algorithms. Our normalization algorithm uses global registration through ICP against a mean face model computed from the training set. Since we have to deal with the presence of occluding objects, the registration process has been customized in order to reject all those points which apparently do not belong to the facial surface.

Since no occluded 3D face database is publicly available, we built an artificial occlusion generator for our experiments. The generator is able to combine a set of occluding objects and an existing database. The output is an artificially occluded database presenting one occluding object for each face. We captured a set of scanned real world objects for building the dataset used in our experiments. The resulting occluded acquisitions have a realistic visual aspect and are quite undistinguishable from real occluded acquisitions.

The paper is organized as follows. Section 2 describes the detection algorithm in details. Section 3 describes the artificial occlusion generator and the database used for training and testing.Finally, Section 4 presents experimental results while Section 5 summarizes conclusions and future work.

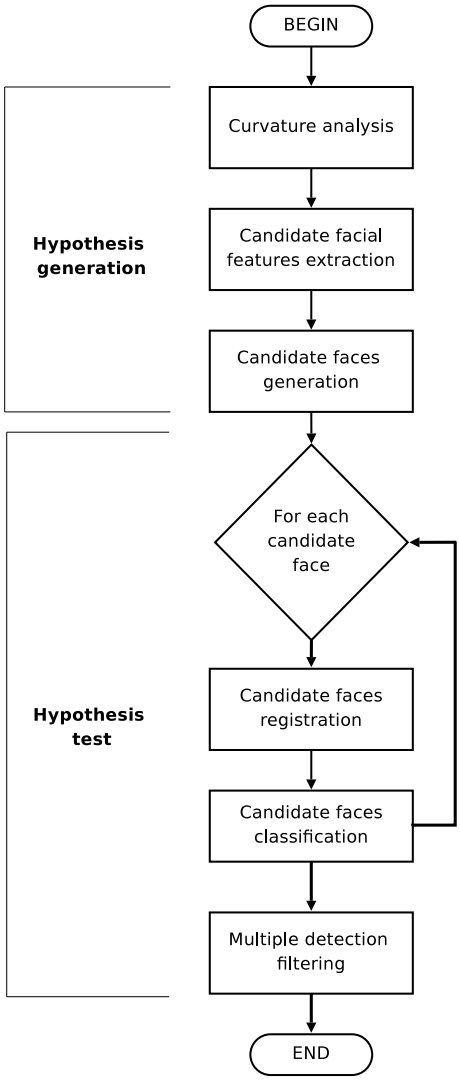


Fig. 1 The face detection algorithm: main diagram.

2 Detection algorithm

The face detection algorithm is based on the approach presented in [12]. The input of the algorithm is a single range image of the scene. If other representations are available, a range image can be generated using simple and well known rendering techniques, such as variants of the Z-Buffer algorithm. Figure 1 shows a diagram representing the main steps of the algorithm. The idea consists in generating hypothesis about the presence of faces. These hypothesis are generated starting from the position of potential facial features, namely the eyes and the nose. These regions are, among other facial features, the most stable and they can be well isolated using curvature analysis segmentation. For more details about the procedure used for the extraction of candidate eyes and nose regions see [12].

Given the set of all candidate facial features, hypothesis about the presence of faces are generated combining eyes and noses in a coherent way. Once the set of hypothesis is built, for each element of the set a registration followed by face vs. non-face discrimination is performed.

In Figure 2 a more detailed description of the processing steps is shown. The image is initially analyzed exploiting curvature characteristics, namely the gaussian and mean curvature. Candidate regions are generated through HK classification and curvature thresholding (see [12]).

Combinations of candidate features are used to select the corresponding 3D surface region 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 followed by a fine registration approach (rough+fine) based on an occlusions-tolerant version of the ICP algorithm [16], 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 GPCA 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.

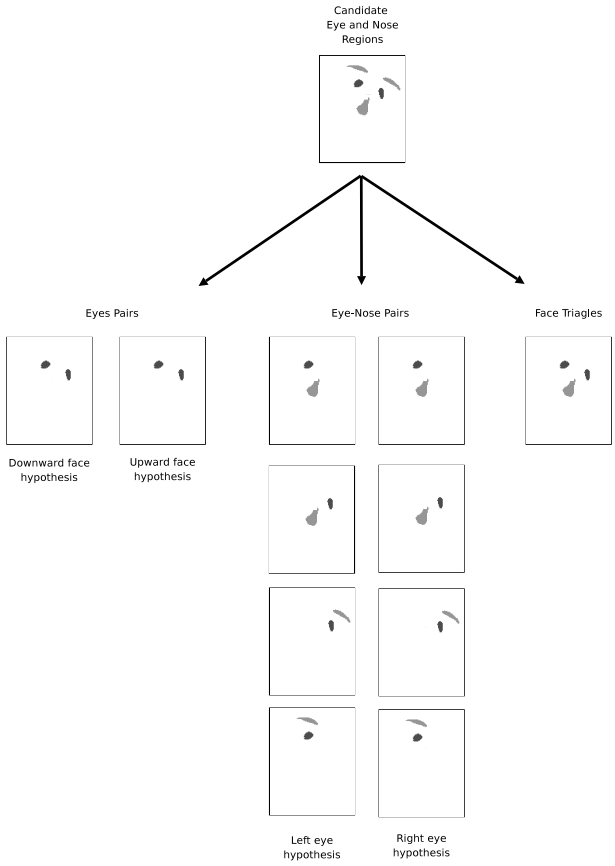


Fig. 3 An example of candidate face generation. The image on top shows the selected candidate features (light gray for noses, dark gray for eyes). Below, all the candidate faces generated by the algorithm are shown. In the case of eye pairs and eye-nose pairs, two candidate faces are generated for each pair because of the impossibility of determining the actual face orientation.

2.1 Candidate face generation

The generation of candidate facial features results in a set composed by two kinds of features: eyes and noses. A single candidate faces is generated combining two eyes, or an eye and a nose, or two eyes and one nose. This 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 ([12]). Figure 3 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 filtering process (see Section 2.5). When a candidate is composed of a nose and one eye, two hypothesis are generated. The wrong hypothesis will be registered in a wrong way and then it will be discarded by the GPCA classifier. This allows to determine if the eye is the left or the right one.

2.2 Candidate face rough registration

Candidate faces are freely oriented in 3D space. In order to obtain registered depth images ready for classification a rough+fine normalization approach is adopted.

Rough registration is computed starting from the reference points of the regions, which are the points of maximum curvature within each region (mean curvature for the nose, gaussian curvature for the eyes). The normalized reference position is defined starting from the three points case (left eye, right eye and nose) associated with each candidate face triangle. A reference system is built as follows (Figure 4):

- the x axis is oriented from the right eye to the left one;

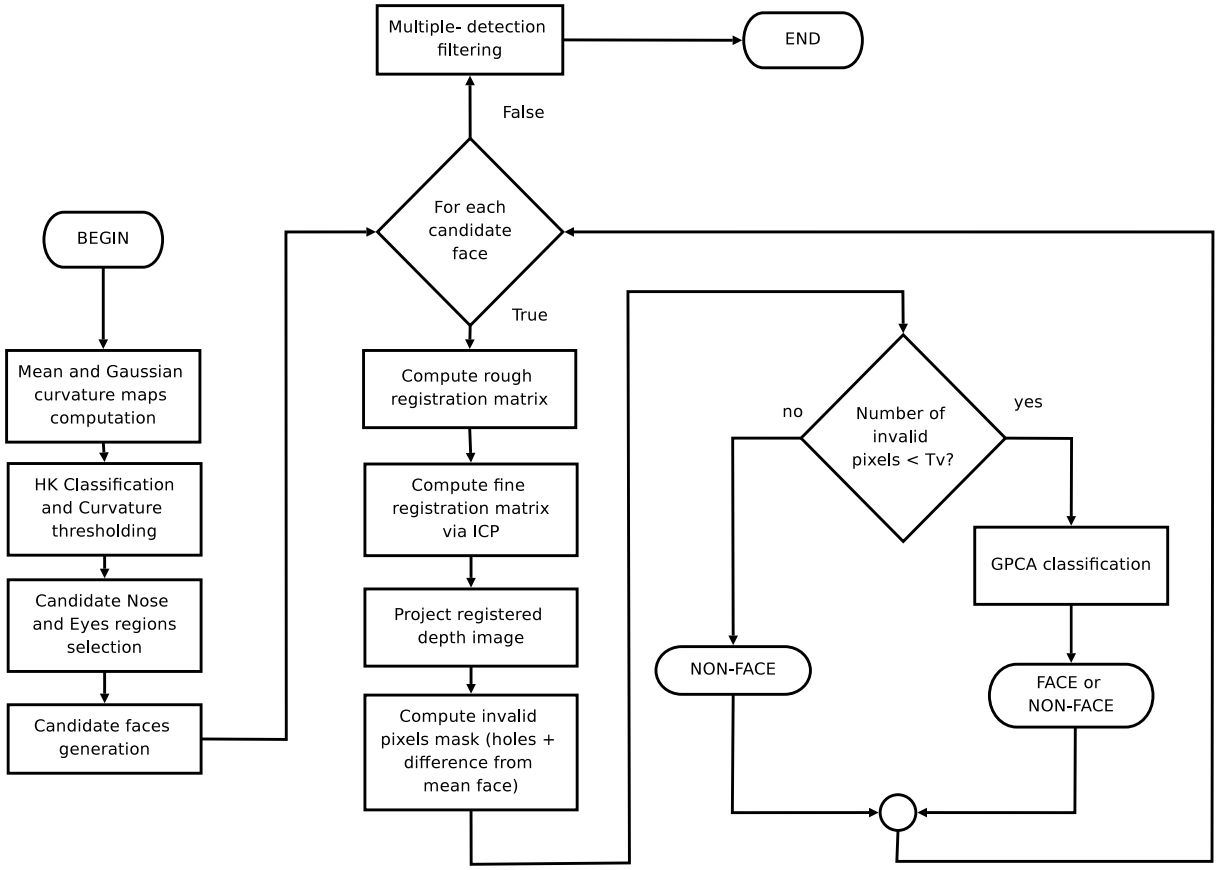
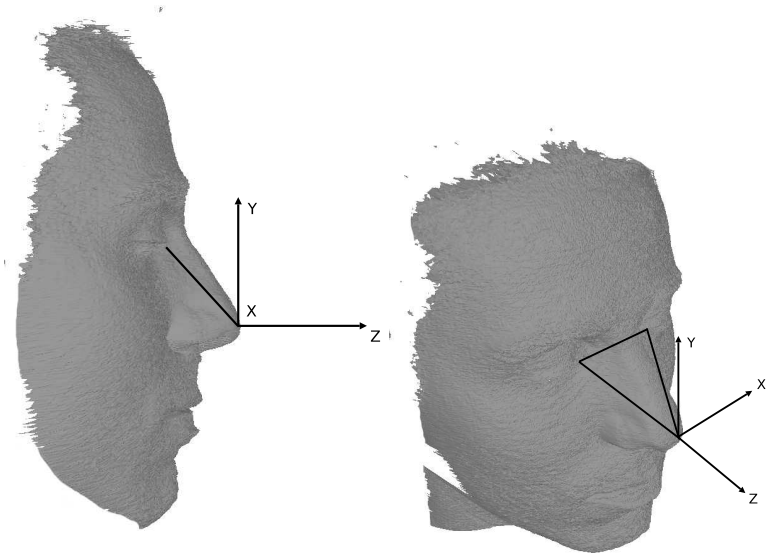


Fig. 2 The face detection algorithm: detailed diagram.



(a) (b)

Fig. 4 The reference face position used for rough normalization.

– the z axis is oriented in the same direction as the normal vector of the face triangle;
– the y axis is computed using the cross product between the other two axes;
– the origin of the axes is translated to the tip of the nose;
– finally the system is rotated about the x axis by 45 degrees.

Using this reference system, a transformation matrix S is built. The 3D model of the face is registered applying the transformation S on all the vertexes of the model. Denoting the axes of the new reference system as the vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ with origin $\mathbf{o}$, the matrix S can be seen as the product of the rotation matrix R and the translation matrix T:

$$S = R \cdot T, \tag{1}$$

$$R = \begin{bmatrix} \mathbf{u} & 0 \\ \mathbf{v} & 0 \\ \mathbf{w} & 0 \\ \mathbf{0} & 1 \end{bmatrix} \tag{2}$$

$$T = \begin{bmatrix} I_3 & \mathbf{o} \\ 0 & 1 \end{bmatrix} \tag{3}$$

Using column vector notation, a vertex $\mathbf{p}$ is transformed into a vertex $\mathbf{p}'$ using:

$$\mathbf{p}' = S\mathbf{p}. \tag{4}$$

When only two feature points are present, one degree of freedom is left undetermined and only a partial rough registration is computed. In application scenarios where the variability around the missed degree of freedom is small, this is not a problem because ICP is able to converge. In general, these are the cases when subjects are supposed to look approximatively toward the camera. However, in case of less constrained scenarios, the initial guess for the ICP could be not good enough and this could lead to registration errors. In these cases, the local surface information can be used to approximate the surface orientation. For example, it is possible to compute a mean normal vector using small regions around the detected points. This normal can be used to determine an orientation for the final degree of freedom. In our experiments, ICP produced good results without using local surface properties.

In the case of two-eyes, rough registration is simply computed bringing only the eyes to the reference position. Head pitch angle is left undetermined. The reference system is computed in this way:

– the x axis is oriented from the right eye to the left one;
– the z axis is oriented in the same direction as the scene z axis.
– the y axis is computed using the cross product between the other two axes;
– the origin of the axis is translated so that the middle point of the eyes corresponds to a reference point computed from a training set of normalized faces.

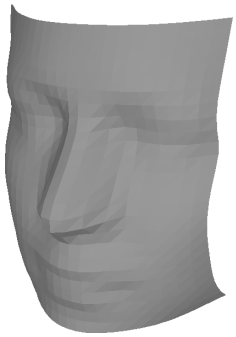


Fig. 5 The mean face template used for fine registration. The surface, composed of approximatively 600 vertexes, corresponds to the mean surface of the training set.

Finally, in the case of eye-nose pairs, the reference system is computed as follows:

– the origin of the axes is translated to the tip of the nose;
– the axes are rotated so that the eye-nose vector is in the same direction as the reference eye-nose vector computed on a training set of normalized 3D faces.

In this case, the correct rotation about the eye-nose vector is not determined.

2.3 Candidate face fine registration

Fine registration is accomplished using a variant of the popular ICP (Iterative Closest or Correspondent Point) algorithm [16] applied to the rough registered candidate faces through a mean face template. Figure 5 shows the template used for this scope. The algorithm has been inspired by the analysis of ICP variants conducted in [22].

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 [18]. 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 I_d (data image) for the mean face and I_m (model image) for the candidate face. 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 $(i \pm r, j \pm r)$; where $r \geq 1$ is an integer defining a square region around the current location. The correspondence criterion is the point at minimum 3D Euclidean distance. In our experiments we adopted a mean face model composed of a grid of 51x51 vertexes. All the vertexes has been used in the ICP matching phase.

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 $c = (\mathbf{p}_d, \mathbf{p}_m, \mathbf{n}_d, \mathbf{n}_m)$, where $\mathbf{p}_d$ and $\mathbf{p}_m$ are, respectively, the 3D model and data points while $\mathbf{n}_d$ and $\mathbf{n}_m$ are the surface normals to those points, the rejector verifies that the following conditions are satisfied:

$$\arccos(\mathbf{n}_d \cdot \mathbf{n}_m) \leq \alpha, \tag{5}$$

$$\|\mathbf{p}_d - \mathbf{p}_m\|_2 < T_{dist}. \tag{6}$$

The first condition checks if the angle between the two normals is inferior to a predefined threshold α. 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 conditions assures that the distance between the two correspondent points must be below a predefined threshold T_{dist}. The value of this threshold has been computed considering the variations between the normalized non-occluded faces from the training set and the mean face template. In Figure 6 is presented the histogram of the differences in depth computed at pixel level. As can be seen, a distance greater than 15mm is very improbable and thus we have chosen this value for T_{dist}. In Figure 7 is presented the cumulative histogram of the difference in depth between occluding objects and faces evaluated at pixel level. As can be seen, approximatively 10% of the occluded pixels are below the selected value of T_{dist}.

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. In each ICP iteration, the transformation can be calculated by different methods; we used the quaternion method of Horn [25]. For more details and for an in-depth description of the ICP algorithm see [16].

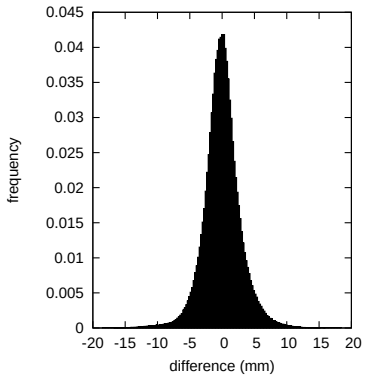


Fig. 6 The histogram of the differences in depth between the mean face template and the normalized faces from the training set.

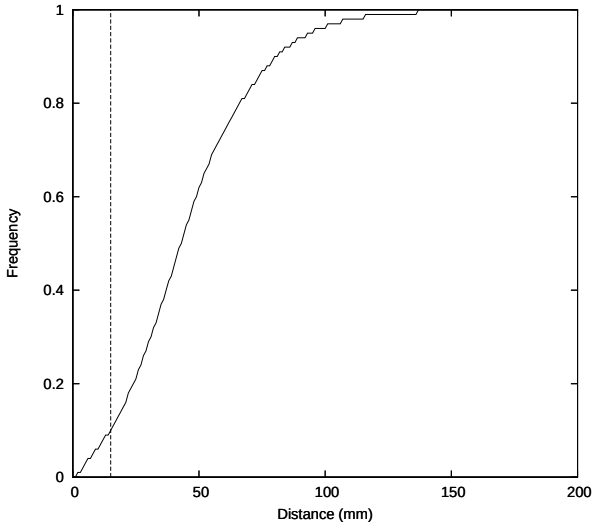


Fig. 7 The cumulative histogram of the difference in depth between occluding objects and faces. The distances are computed at pixel level. The dotted line shows the value of the threshold T_{dist} used to discriminate occluded pixels.

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. Only the depth information is retained. The images are then cropped; i.e. the borders of the candidate faces are invalidated using a fixed shaped mask in order to reduce irregularities and to make all the images equal in size and shape. At this point, each image is compared with the mean face in order to detect occluding objects. For each pixel (i, j) the following condition is checked:

$$|I(i, j) - M(i, j)| \leq T_{dist}, \tag{7}$$

where I is the candidate face depth image while M is the mean face depth image. T_{dist} is the same threshold used in 6. 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_{dist}. Since large regions of the image may be invalid, a check on the fraction of valid pixels is performed:

$$\frac{n_v}{N} \geq T_v \tag{8}$$

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 GPCA classifier has been adopted. The classifier is based on Principal Component Analysis for gappy data [13].

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.

Gappy Principal Component Analysis addresses two scenarios: the case where an incomplete pattern is restored with an existing PCA basis, constructed from complete data, and the case in where only incomplete data is available. The second scenario is more challenging and is not considered here, since it is assumed that at least a set of non-occluded faces is available (usually those acquired for the training step).

Assuming the first scenario, a set of N patterns $\{\mathbf{x}_1, \cdots, \mathbf{x}_N\} \subset \mathbb{R}^n$ is used to determine the PCA basis in such a way that a generic pattern $\mathbf{x}$ can be approximated using a limited number, M, of eigenvectors:

$$\mathbf{x} \simeq \boldsymbol{\mu} + \sum_{i=1}^{M} \alpha_i \mathbf{v}_i, \tag{9}$$

where $\boldsymbol{\mu}$ is the mean vector, $\mathbf{v}_i$ is an eigenvector, and α_i is a coefficient obtained by the inner product between $\mathbf{x}$ and $\mathbf{v}_i$.

Suppose there is an incomplete version $\mathbf{y}$ of $\mathbf{x}$ and suppose that the location of missing components encoded in the vector $\mathbf{m}$ ($\mathbf{m}_i = 0$ is the i-component, otherwise $\mathbf{m}_i = 1$) is missing. GPCA seeks for an expression similar to Equation 9 for the incomplete pattern $\mathbf{y}$:

$$\mathbf{y} = \mathbf{y}' \simeq \boldsymbol{\mu} + \sum_{i=1}^{M} \beta_i \mathbf{v}_i, \tag{10}$$

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

$$E = \|\mathbf{y} - \mathbf{y}'\|^2. \tag{11}$$

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 $(\mathbf{v}, \mathbf{u})_{\mathbf{m}}$ and the corresponding gappy norm $\|\mathbf{v}\|_{\mathbf{m}} = \sqrt{(\mathbf{v}, \mathbf{v})_{\mathbf{m}}}$:

$$(\mathbf{v}, \mathbf{u})_{\mathbf{m}} = \sum_{i=1}^{n} \mathbf{v}_i \mathbf{u}_i \mathbf{m}_i. \tag{12}$$

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

$$E = \|\mathbf{y} - \mathbf{y}'\|_{\mathbf{m}}^2. \tag{13}$$

Rewriting E in terms of gappy inner products, and using the equation 10:

$$E = \|\mathbf{y}\|_{\mathbf{m}}^2 - 2\sum_{i=1}^{M} \beta_i (\mathbf{y}, \mathbf{v}_i)_{\mathbf{m}} + \sum_{i=1}^{M}\sum_{j=1}^{M} \beta_i \beta_j (\mathbf{v}_i, \mathbf{v}_j)_{\mathbf{m}}. \tag{14}$$

Differentianting E with respect to each β_i yields a system of M linear equations:

$$\frac{\partial E}{\partial \beta_i} = -(\mathbf{y}, \mathbf{v}_i)_{\mathbf{m}} + \sum_{j=1}^{M} \beta_j (\mathbf{v}_i, \mathbf{v}_j)_{\mathbf{m}} = 0. \tag{15}$$

By defining $A_{ij} = (\mathbf{v}_i, \mathbf{v}_j)_{\mathbf{m}}$, for $i, j = 1, \cdots M$, and $\mathbf{z}_i = (\mathbf{y}, \mathbf{v}_i)_{\mathbf{m}}$, for $i = 1, \cdots M$, the system of linear equations can be rewritten as:

$$A\boldsymbol{\beta} = \mathbf{z}. \tag{16}$$

The gappy pattern $\mathbf{y}$ can be reconstructed as $\mathbf{y}'$ using the expansion 10, where the coefficients $\boldsymbol{\beta}$ are found by solving the system 16. Figure 8 shows an example of a reconstructed gappy face image.

The classifier used for the face detector constructs the vector $\mathbf{m}$ considering all the invalidated or holes pixels. The error defined in equation 13 is used as measure of faceness (i.e. how much a test is similar to a face). Thus, a candidate face is classified as a face if the following condition is satisfied:

$$\frac{E}{n_v} \leq T_f. \tag{17}$$

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.

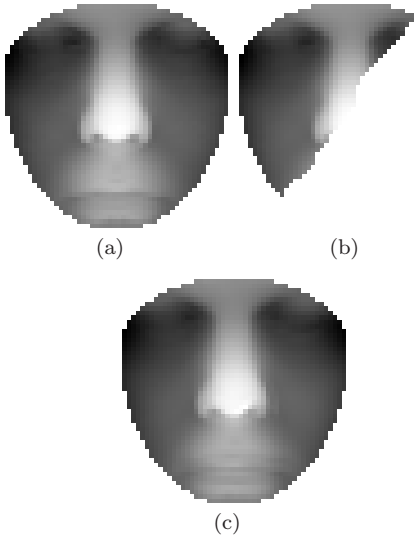


Fig. 8 Example of a reconstruction of a gappy image: (a) the original depth image; (b) the same image with a region of invalidated pixels; (c) the reconstruction of (b) through GPCA.

2.5 Multiple detection filtering

The patterns classified as faces are finally filtered in order to eliminate multiple detections for the same face. Multiple detections derive from the generation of multiple candidate faces starting from the same candidate facial features, as already explained in previous sections. As an example, in Figure 3 there are four correct candidates for the actual face. Moreover, sometimes false candidate facial features generated from curvature analysis are located close to real facial features; this also leads to the generation of unnecessary candidate faces.

Multiple detection filtering is performed using the following criteria:

- if a subset of detected faces have at least one facial feature in common, the face having minimum residual reconstruction error E is selected as the representative while the others are invalidated;
- if two faces overlap, the face with minimum residual reconstruction error E is selected as the representative while the other one is invalidated. The overlapping check is approximated using bounding spheres of radius d centered on the face tip of the nose. So, if the bounding spheres of two detected faces overlap, then the faces overlap.

3 Artificial Occlusion Generation

The first big problem in dealing with occlusions is how to produce them in order to test algorithms. Acquiring faces with real occlusions such as eyeglasses, scarves, hats, hands, cellular phones etc. presents some disadvantages. First of all, face related algorithms must be tested on a large number of subjects. For each subject, a considerable number of acquisitions with different kinds of occlusions in different positions is needed. Acquiring such a big dataset requires a great effort. Second, real occlusions do not allow accurate tests because the ground truth cannot always be determined easily. For example, feature points in occluded parts cannot be determined. Accurate pose normalization cannot be computed with precision. Moreover, the manual selection of the occluding objects requires a great deal of work because this must be done at the pixel level. Considering a dataset composed by hundreds or even thousands of acquisitions, real occlusions results in an unpractical solution.

In this paper the artificial occlusion generation solution has been adopted. It consists in taking an existing database of non-occluded faces and adding occluding 3D objects in each acquisition. This is a low-cost and effective solution because no effort is required to acquire new images, and the ground truth can be determined easily with automatic procedures.

Occluding objects are real world objects captured at the IVL Laboratory at the University of Milano Bicocca. Figure 9 shows the entire set of objects, which includes plausible objects such as a scarf, a hat, two types of eyeglasses, a newspaper and hands in different configurations. There is also an unusual object; i.e. a pair of scissors. Its complex shape has been considered a good test for the proposed detection algorithm. The eyeglasses are composed only of the frame; no lenses are considered. 3D acquisition devices usually introduce some kind of artifacts in the area of the lenses. We have not included lenses in our models because the artifacts are not easily reproducible. To test the algorithm with real glasses we have captured a small dataset of real occlusions, as described in Section 4.

The set of occluding objects has a good variability in shape and the set of target regions that will be possibly occluded covers all the extent of the face. For these reasons, we consider our choice a good test for detection algorithms.

3.1 Occluded acquisition generation

Occluded acquisitions are generated inserting the objects in the acquisition space in 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. So, 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.

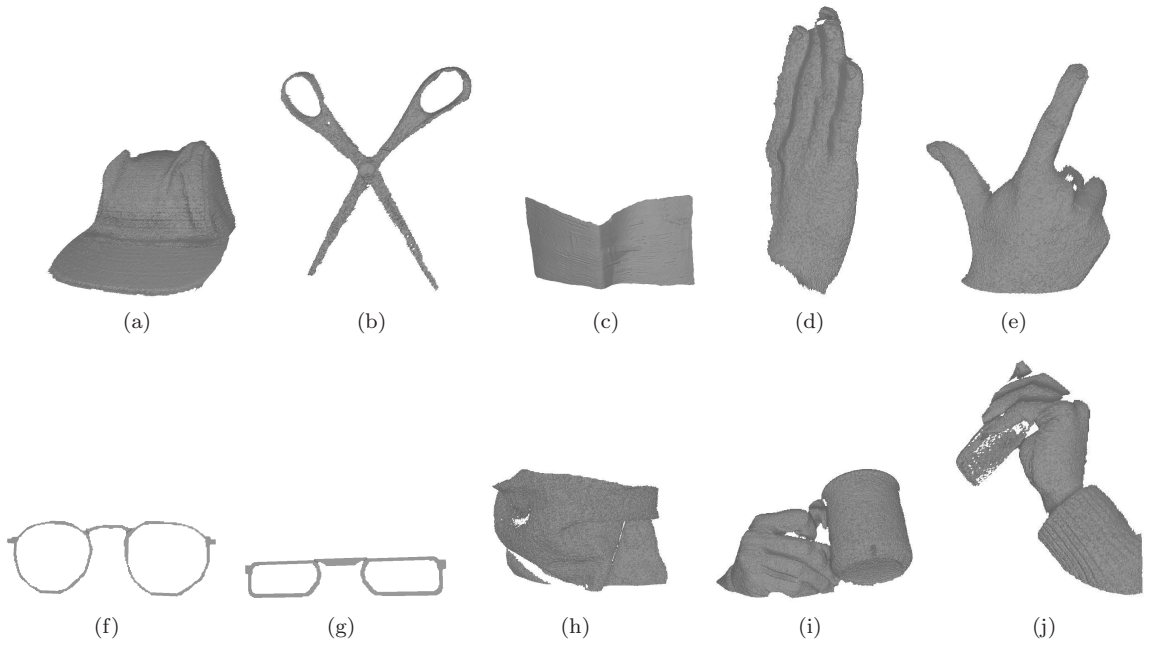


Fig. 9 The objects acquired at the IVL laboratory and used as artificial occlusions:(a) hat, (b) scissors, (c) newspaper, (d,e) hands with different gestures, (f,g) eyeglasses, (h) scarf, (i) hand with a teacup, (j) hand with a phone.

Then, for each acquisition i the following steps are applied:

- Given the feature points placed on the nose and on the corners of the eyes (which are part of the dataset ground truth), compute a rough registration (T_r, R_r) of the face, as described in more detail in section 2.2.
- Compute the fine registration (T_f, R_f) starting from the rough registration (T_r, R_r) using the ICP algorithm, as described in more detail in section 2.3.
- Choose a random occluding object o.
- Add a random noise $(N_t(\boldsymbol{\lambda}_{\mathbf{o_t}}), N_r(\boldsymbol{\lambda}_{\mathbf{o_r}}))$ to the object starting position and orientation (T_o, R_o). N_t and N_r are functions generating noise in ranges specified by the object o range vectors $\boldsymbol{\lambda}_{\mathbf{o_t}}$ and $\boldsymbol{\lambda}_{\mathbf{o_r}}$.
- Compute the final position of the object in the original acquisition space:

$$(T, R) = (T_f, R_f)^{-1} \times (T_o + N_t(\boldsymbol{\lambda}_{\mathbf{o_t}}), R_o + N_r(\boldsymbol{\lambda}_{\mathbf{o_r}})). \tag{18}$$

- Generate a range image using an orthographic projection of the acquisition i in its original position and of the object o transformed by (T, R). Orthographic projection is computed using variants of the Z-buffer algorithm [24].

In Figure 10 some examples of automatically occluded acquisitions are shown. As can be seen, the 3D component of the acquisitions have a realistic aspect.

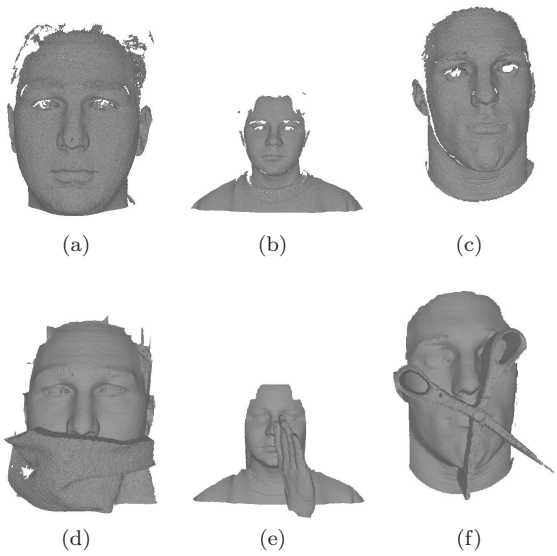


Fig. 10 Some examples of artificially occluded faces. (a,b,c) The original UND acquisitions; (d,e,f) the occluded faces. Note that automatic preprocessing is applied to the occluded faces, i.e. smoothing and simple holes closure via linear interpolation.

In Figure 11 the histogram of the fraction of the face (in percent) covered by artificial occlusions is reported.

4 Experimental results

The proposed method has been tested on the artificially occluded UND database [17], a publicly available 3D

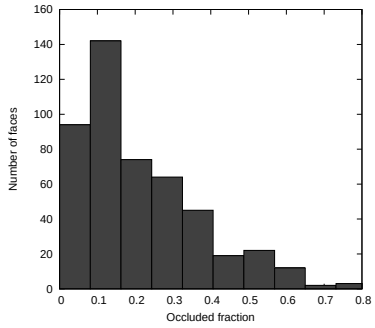


Fig. 11 Histogram of the fraction of face (in percent) covered by artificial occlusions.

face database from the University of Notre Dame. It is composed of 951 multimodal 2D + 3D acquisitions from 275 subjects. The ground truth included in the database distribution is composed of manually selected feature points; in particular we used the corners of the eyes, the tip of the nose and the pogonion. For the experiments conducted in this paper we used all the 951 acquisitions occluded using the technique explained in section 3.

For the training set we used the 100 neutral acquisitions (i.e no expressions) from the BU3DFE database [30]. The training set, once normalized, it has been used to build the face space for the GPCA classifier and for computing the mean face template. The normalization procedure is the same rough+fine approach described in Sections 2.2 and 2.3 using the manually selected feature points given with the DB ground truth. For the GPCA classifier we used only 16 eigenvectors, wich corresponds to 90% of the retained variance. Table 1 summarizes the thresholds values used during tests.

Table 1 Thresholds used in the experiments

Name	Meaning	Value
α	maximum normal angle difference between two correspondant points during ICP	90 degrees
T_{dist}	maximum distance between two correspondant points during ICP and occlusion detection	15mm
T_v	fraction of valid pixels accepted by the GPCA classifier	50%

Figure 13 shows the precision-recall curves for the GPCA classifier at different values of the valid pixel threshold T_v. The curves has been computed considering only the candidate faces produced by the hypothesis generation phase of the algorithm. As can be seen, accepting larger occlusions (i.e. decreasing the threshold T_v) the performances of the classifier decrease.

In order to understand the quality of the registration process, we computed the registration error produced by the ICP; Table 2 shows the results. The ground truth has been computed in this way: we used the feature points included in the UND DB distribution to compute the rough registration. Then, ICP ha been applied on non-occluded acquisition. The mean errors has been computed measuring the angles between the ground truth reference system axis and the axis of the computed reference system. The distance between the reference systems origins is also reported. We also compared the rough registration and the registration ground truth. This gives an idea of the initial guess performed by rough registration before ICP is computed.

Since the GPCA approach is based on reconstruction we evaluated the quality of the reconstructed patterns in the case of the artificially occluded UND database. In particularly, we computed the mean reconstruction error for each face evaluated on the occluded pixels. We obtained a mean error over all the faces of $2.59mm$. As a term of comparison, we computed on manually normalized faces the average difference between two non-occluded acquisitions. We obtained a mean difference of $1mm$.

In Table 3 are reported the results obtained choosing a value for the classifier threshold T_f aimed to reduce 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 89.8% of the total number of occluded faces has been successfully detected, generating 135 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 by most of the conventional 3D systems. The detector performs very well on non-occluded faces, reaching 100% of detected faces and 18 false alarms. In order to understand the impact of the occlusion type on detection performances, we reported in Table 4 the detection results for each object type. Errors are more frequent with objects wich usually cover a large portion of the face (i.e. newspaper, scarf and free hands). Figure 14 presents some examples of successfully detected faces. Figure 12 shows some examples of errors produced by the algorithm. Both figures refer to the occluded dataset. Usually, false negatives are due to the absence of the fundamental features needed to generate the candidate. False positives, instead, are usually generated by the GPCA classifier when large portion of the face is occluded. The first example in figure 12 is a frequent error in case of a scarf. The eyes are usu-

ally detected and two candidates are generated; i.e one candidate for the upright face and one candidate for the downright face. Since the mouth region of the face is occluded, sometime the GPCA classifier generate a lower reconstruction error for the wrong candidate (the downright one). In the second and third case, wrong facial features are detected. The corresponding candidate range image is reconstructed by the GPCA with a low error because large parts of the image is considered occluded. For example, the reconstructed image (l) appears very similar to the input pattern (j). In the last example, the reconstructed image (r) has an error very close to the rejection threshold T_f.

In order to verify the results obtained with artificial occlusions we have captured a small set of acquisitions containing real occlusions (see Figure 15 for some examples). This test set (IVL test set) is composed of 102 acquisitions containing 104 faces (in two acquisitions there were two subjects at the same time). We used the same device adopted in the UND database, the Minolta Vivid 900 laser range scanner. The occlusions are of the same kind as the artificial occlusions used with the UND Database and we used the same parameters as those selected in the previous experiments. As Table 3 shows, 90.4% of faces have been correctly detected producing 16 false alarms. Figure 16 shows two examples of false alarms. These results are in line with those obtained with the artificial dataset.

Since the eyeglasses case is very frequent in real-world applications, we collected an additional test set (IVL-EG) containing real occlusions. This test set is composed by 50 acquisition of 50 different subjects wearing eyeglasses. We obtained a detection accuracy of 100% and 6 false alarms. This shows the robustness of the method against the artifacts introduced by the laser scanner in the lenses regions.

To understand the advantages of the GPCA over a typical PCA approach, we have substituted the GPCA classification stage with a PCA classifier. In Table 5 we reported comparative results obtained with the UND database. As expected, in the non occluded case the performance of the two classifier are similar. However, in the occluded case the PCA classifier performances drop down to 49.2% of correctly detected faces against 89.9% of the GPCA classifier. The presence of occluding objects does not allow a correct reconstruction of the face pattern using standard PCA techniques.

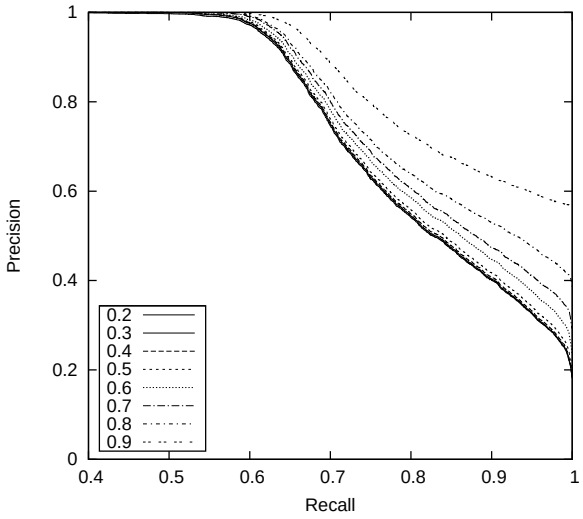


Fig. 13 The GPCA classifier Precision-Recall curves varying the value of the threshold T_v.

5 Conclusions and Future work

The presented algorithm is a first attempt to solve the occluding objects problem in 3D face detection and normalization. The algorithm is primarily aimed to be adopted in biometric recognition systems. The experiments were conducted on an artificially occluded database presenting very difficult cases were the objects occlude fundamental facial features. These occluded parts are often considered necessary by a great number of detection algorithms. Our approach is based on simple geometrical considerations about the depth nature of the problem. This clearly shows the advantage in adopting 3D data to deal with occlusions.

The detection and normalization solution presented here has been developed as the first step of a complete recognition system. Recognition algorithms may take advantage of our solution adopting partial matching strategies or using local features, like those described in [26], for example.

We are considering to approach the problem of detection in presence of emphasized facial expressions using similar strategies. Part of the face presenting large differences from the neutral face may be detected and GPCA could be used to neutralize facial expression.

We plan to extend the generator of artificial test cases. For example it should be useful to add cluttered backgrounds, to generate different head pose variations and to add a simulator of eyeglasses artifacts. We are also working on the acquisition of a larger dataset with real occlusions that could be make available to the research community.

References

1. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", *IEEE Transactions on Pattern Analisys and Machine Intelligence*, pp. 711–720, 1997.

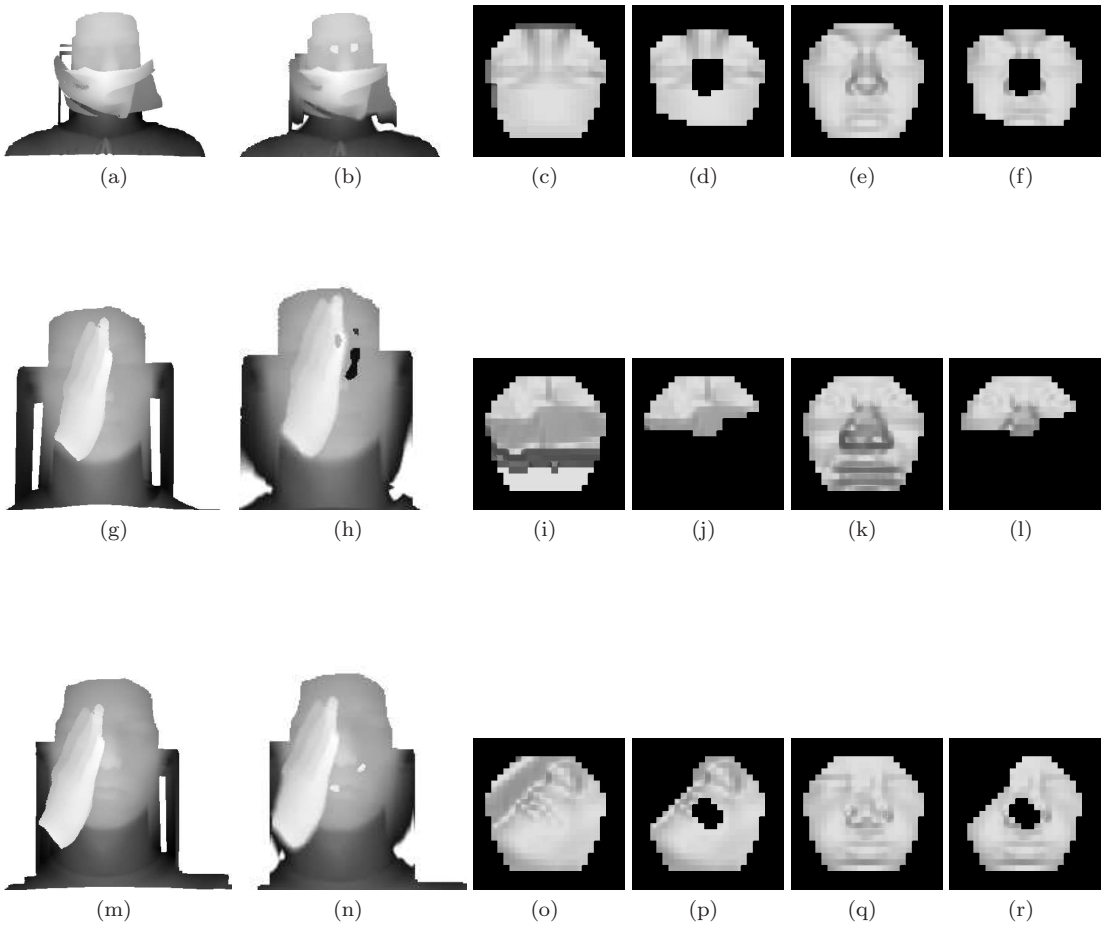


Fig. 12 Examples of errors from three acquisitions (one per row) from the artificially occluded UND dataset. The first column shows the original scan. The second column shows the detected features: (b) a pair of eyes; (h) a face triangle; (n) a pair of eyes. The third column represents the projected range image of the selected candidate. In (c), although the features were correct, the selected candidate is the downright face. In column four, the same image is masked in order to eliminate the detected occluding pixels. Columns 5 and 6 show the reconstructed image and its masked version.

Table 2 Mean registration errors produced by the rough and fine registration process. Angles are in degrees (X, Y, Z) while origins euclidean distance is in millimeters.

Test Set	Registration Type	X	Y	Z	Origin
UND	fine	1.07	1.07	1.08	4.68
(non-occluded)	rough	7.55	7.98	10.90	21.3
UND	fine	1.84	1.97	1.75	6.41
(artificially occluded)	rough	7.60	8.35	10.06	21.52

2. L. Qing; S. Shan; and X. Chen, "Face relighting for face recognition under generic illumination", *IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. (ICASSP)*, pp V-733-6 vol.5, 2004.

3. Bronstein, A. M. and Bronstein, M. M. and Kimmel, R, "Expression-invariant 3D face recognition", *"Proc. Audio and Video-Based Person Authentication"*, vol. LCNS 2688, pp 62–70, 2003.

4. Lu, Xiaoguang and Jain, Anil K. "Deformation Analysis for 3D Face Matching", *Application of Computer Vision. WACV/MOTIONS*, pp.99–104, Volume 1, 2005.

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

6. K. Hotta, "A robust face detector under partial occlusion". *Proceedings of ICIP* pp. 597–600, 2004.

Table 3 Face detection results obtained on the test sets used in the experiments.

Test Set	Total Faces	False Positives	False Negatives	Detected Faces
UND (non-occluded)	951	18	0	951 (100%)
UND (artificially occluded)	951	135	104	847 (89.8%)
IVL (occluded)	104	16	10	94 (90.4%)
IVL EG (occluded)	50	6	0	50 (100%)

Table 4 Face detection results obtained using the artificially occluded UND Database. Results are reported for each type of occluding object.

Object Type	Total Faces	False Positives	False Negatives	Detected Faces
hand with cup	91	5	3	88 (96.7%)
eyeglasses	184	4	0	184 (100.0%)
free hands	181	46	41	140 (77.3%)
newspaper	77	20	18	59 (76.6%)
hand with phone	97	1	0	97 (100.0%)
scarf	123	42	29	94 (76.4%)
scissors	109	9	7	102 (93.6%)
hat	89	8	6	83 (93.6%)
all	951	135	104	847 (89.8%)

Table 5 PCA vs. GPCA comparative detection results.

Test Set	Classifier	Total Faces	False Positives	False Negatives	Detected Faces
UND (non-occluded)	PCA	951	3	1	950 (99.9%)
	GPCA	951	18	0	951 (100%)
UND (artificially occluded)	PCA	951	165	483	468 (49.2%)
	GPCA	951	135	104	847 (89.8%)

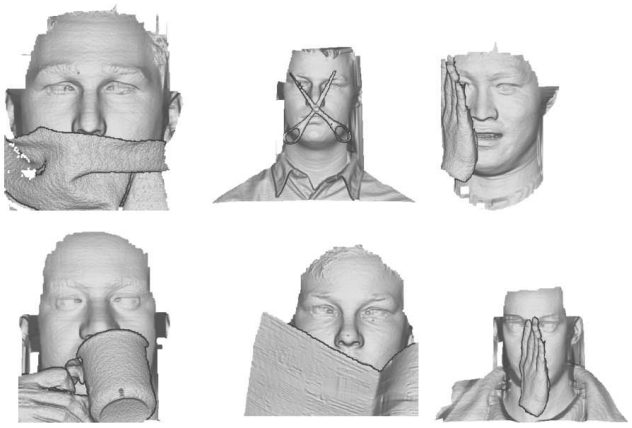


Fig. 14 Examples of detected faces taken from the artificially occluded UND dataset.

Fig. 15 Examples of real occluded faces taken from the IVL dataset.

7. J. S. Park, Y. H. Oh, S. C. Ahn, and S. W. Lee, "Glasses Removal from Facial Image Using Recursive Error Compensation", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 5, pp. 805–811, 2005.
8. A. M. Martinez, "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.
9. A. M. Martinez, "Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single

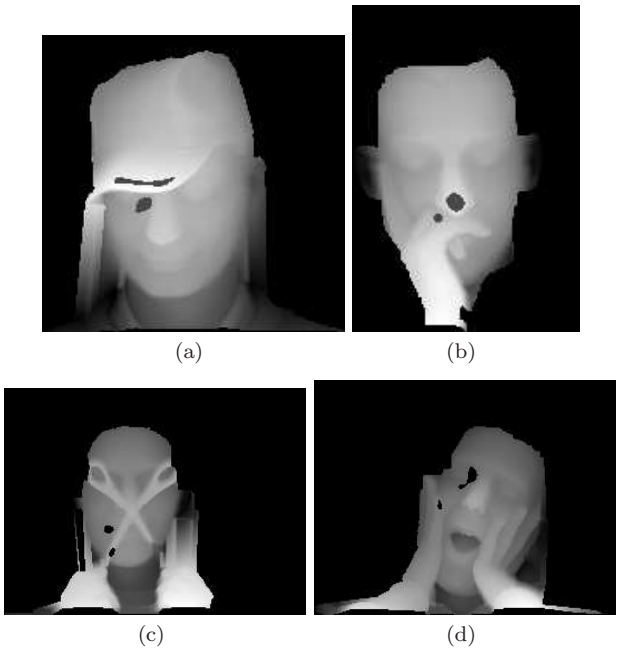


Fig. 16 Some examples of false acceptances from the IVL Dataset. Dark regions indicate the facial features of the false faces.

Sample per Class", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, on. 6, pp. 748–763, 2002.

10. J. Kim, J. Choi, J. Yi, and M. Turk, "Effective representation using ICA for Face Recognition Robust to Local Distortion and Partial Occlusion", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol 27, no. 12, pp. 1977–1981, 2005.
11. F. Tarrés, and A. Rama, "A Novel Method for Face Recognition under Partial Occlusion or Facial Expression Variations", *Proc. 47th Int'l Symp. ELMAR,* pp. 163–166, 2005.
12. A. Colombo, C. Cusano, and R. Schettini, "3D Face Detection using Curvature Analysis", *Pattern Recognition,* vol. 39, no. 3, pp. 444–455, 2006.
13. R. Everson and L. Sirovich, "Karhunen-Loève Procedure for Gappy Data", *J. Optical Soc. of America A,* vol. 12, no. 8, pp. 657–1664, 1995.
14. A. Colombo, C. Cusano, and R. Schettini, "A 3D Face Recognition System using Curvature-based Detection and Holistic Multimodal Classification", *Proc. 4th Int'l Symp. on Image and Signal Processing and Analysis,* pp. 179–184, 2005.
15. X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5D Face Scans to 3D Models", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 1, pp. 31–43, 2006.
16. Paul J. Besl and Neil D. McKay, "A Method for Registration of 3-D Shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol 14, pp. 239–256, 1992.
17. K. Chang, K. W. Bowyer, P. J. Flynn, "Face recognition using 2D and 3D facial data", *ACM Workshop on Multimodal User Application,* pp. 25–32, 2003.
18. K. Pulli, "Multiview Registration for Large Data Sets" *Proc 3DIM*, 1999.
19. Do Carmo, M. P., "Differential geometry of curves and surfaces",*Prentice Hall*, 1976.
20. Gordon, G., "Face recognition based on depth maps and surface curvature", *Proceedings of SPIE, Geometric Methods in Computer Vision*, pp. 234–247, vol. 1570, 1991.
21. Moreno, A. B. and Sánchez, A. and Vélez, J. F. and Díaz. F. J., "Face recognition using 3D surface-extracted descriptors", *Proceedings of the Irish Machine Vision and Image Processing Conference*, 2003.
22. Rusinkiewicz, S., Levoy, M., "Efficient Variants of the ICP Algorithm", *Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, 2001.
23. De Smet, M., Fransens, R. and Van Gool, L., "A Generalized EM Approach for 3D Model Based Face Recognition under Occlusions," *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* , vol.2, no., pp. 1423-1430, 2006.
24. A. H. Watt. *3D Computer Graphics*. Addison Wesley, 1999.
25. Horn, B. K. P., Closed-form solution of absolute orientation using quaternions. J. Opt. Soc. Am. A, 4(4), pp. 629-642, 1987.
26. Yingjie Wang, Chin-Seng Chua, Face recognition from 2D and 3D images using 3D Gabor filters, *Image and Vision ComputingVolume 23, Issue 11*, Pages 1018-1028, 2005.
27. Faltemier, T.C., Bowyer, K.W., Flynn, P.J. Using multi-instance enrollment representation to improve 3D face recognition, *Int. conference on biometrics, theory, Applications, and Systems, BTAS* pp.1-6, 2007.
28. Gross, Matthews, Baker, Active appearance models with occlusion, *Image and Vision Computing*, Vol. 24, No. 6, pp. 593-604, 2006.
29. Huang ,P. S., Zhang, S., "Fast Three-Step Phase Shifting Algorithm" *Applied Optics*, Vol. 45, pp 5086-5091, 2006.
30. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M. J. , A 3D Facial Expression Database For Facial Behavior Research, *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pp. 211–216, 2006.
31. Alyuz, N. and Gokberk, B. and Akarun, L.,A 3D Face Recognition System for Expression and Occlusion Invariance,*Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, pp 1–7, 2008.