Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition 39 (2006) 444–455
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

3D face detection using curvature analysis

Alessandro Colombo, Claudio Cusano, Raimondo Schettini*

DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione), Università degli Studi di Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy

Received 19 November 2004; received in revised form 27 September 2005; accepted 27 September 2005

Abstract

Face detection is a crucial preliminary in many applications. Most of the approaches to face detection have focused on the use of two-dimensional images. We present an innovative method that combines a feature-based approach with a holistic one for three-dimensional (3D) face detection. Salient face features, such as the eyes and nose, are detected through an analysis of the curvature of the surface. Each triplet consisting of a candidate nose and two candidate eyes is processed by a PCA-based classifier trained to discriminate between faces and non-faces. The method has been tested, with good results, on some 150 3D faces acquired by a laser range scanner.

Keywords: Three-dimensional face detection; Face curvatures; *HK* classification; Eigenfaces; Face localization

1. Introduction

Face detection and recognition have attracted the attention of many research groups. Recent key applications in fields such as human–computer interface, identity verification, criminal face recognition, and surveillance systems require the detection, and sometimes the recognition, of human faces.

In this paper, we propose an innovative approach to face detection which, as stated in Ref. [1], can be defined as follows: "Given an arbitrary image, the goal of face detection is to determine whether or not there are any faces in the image and, if present, return the image location and extent of each face". Face detection is therefore more challenging than face localization in which we assume that the image contains only one face.

Most of the literature concerned face detection investigates face detection in two-dimensional (2D) images [1]. An example is the popular work of Turk and Pentland [2]. The so-called "eigenfaces approach" projects face images onto the feature space ("face space") that best encodes the variation among faces. The difference between an image and its projection in that space is used as a measure of "faceness": high values of faceness indicate the presence of a face. The effectiveness of this and many other methods exploiting only 2D images is curtailed by the variability frequently found in imaging conditions (lighting, scale, occlusion. . .) as well as in the subject's expression, pose and orientation.

The decreasing cost of three-dimensional (3D) acquisition systems and their increasing quality, together with the greater computational power available nowadays, will make real-time 3D systems for face recognition a commonplace in the near future. 3D data is already widely used for this purpose. Apart from being less sensitive to viewpoint and lighting conditions, 3D data exploit information which is complementary to gray-level based approaches, enabling the fusion with those techniques. Beumier and Acheroy [3] present a surface analysis approach based on face profile extraction, registration and comparison. Similar approaches, working only with the central face profile, are proposed by Cartoux et al. [4] and Pan et al. [5]. Hesher et al. [6] investigate the use of eigenfaces with automatically registered range images. Similarly, Chang et al. [7] propose a 2D+3D multimodal approach applying eigenfaces to range images and 2D images. Lee and co-workers [8] extract depth areas of 3D face images using the contour line of the same depth value.

* Corresponding author. Tel.: +39 02 6448 7480; fax: +39 02 6448 7839.
E-mail addresses: colomboal@disco.unimib.it (A. Colombo), cusano@disco.unimib.it (C. Cusano), schettini@disco.unimib.it (R. Schettini).

0031-3203/$30.00
doi:10.1016/j.patcog.2005.09.009

These are resampled and stored in consecutive locations in feature vectors using a statistical multiple feature method. The comparison between two faces is based on their Euclidean distance in the feature space. Gordon [9] presents an accurate analysis of the face based on curvatures. She extracts features (nose tip, nose bridge, eyes corners etc...) for alignment purposes; the faces are then compared using their volume difference as a similarity measure. Moreno et al. [10] also segment the faces using surface curvatures. Feature vectors are formed using the extracted regions, and then compared. Wang et al. [11] use "point signature" representation [12] to code 3D surface information in a pose-invariant way. These signatures are combined with 2D information extracted using Gabor filters in order to construct a feature vector. The results are classified by support vector machines. In the great majority of publications concerning 3D face recognition, faces are manually detected and registered in a standard position (sometimes referred to as "normal position") using some ad hoc interactive tools that allow the selection of predefined landmark points, such as the position of the eyes (see for example Ref. [7]). In some cases face localization is automated on the basis of strong assumptions about position and orientation with respect to the imaging device, such as assuming that the subject's nose is the point nearest to the camera [6,8]. We must conclude that the problem of detecting faces in 3D acquisitions has yet to be thoroughly explored.

In this paper we propose an innovative 3D approach: our method assumes that the faces, if any, may be freely oriented with respect to the camera plane, the only restriction being that no self-occlusions and/or face camouflage hide the central part of the face containing the eyes and the nose of the subjects in the acquired scene. In Section 2, we present an overview of our face detection method which combines curvature surface analysis for face feature detection with eigenfaces classification for non-face discrimination. We then explain each processing steps in detail. The method has been tested on some 140 3D scenes acquired by a laser range scanner and containing 150 faces as well as several confusing, i.e. non-face, elements. The results, reported in Section 3, demonstrate the robustness of this approach even in the presence of a great variability of facial expressions and pronounced occlusions due to scarves, hats, etc. Our conclusions and plans for future work are discussed in Section 4.

2. Method description

Our method assumes that a 3D shot of a real world scene is available as a range image, i.e. an image where for each location (i, j) the coordinates (x, y, z) of the 3D scene are expressed with respect to the camera reference system. Some acquisition devices return data in the form of a polygonal model, usually a triangular mesh. In this case, the range image can be obtained using the well known z-buffer algorithm [13].

Fig. 1 shows the major steps used. We assume we have no knowledge about the contents of the scene that is acquired; the resulting image may therefore contain any number of faces, and these may be freely located and oriented with respect to the imaging device. To render the problem less computationally intensive we search initially for single facial features such as eyes and noses. 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, a "face triangle", a potential face, is created from a candidate nose and two candidate eyes. Our goal now is to discriminate between face triangles 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. We further analyze the areas of the range image covered by each face triangle 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. For these reasons, we define our method a combination of a feature-based approach and a holistic one.

Fig. 2 presents a more detailed description of the processing steps. 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 might correspond to eyes and noses; (iii) a 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 contain a nose and eyes are then characterized by their type and by some statistics of their curvature. The output of the processing step may contain any number of candidate facial features. If no nose or less than two eyes are detected we assume 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. Each triplet of a candidate nose and two candidate eyes, forming a face triangle, is used to select the

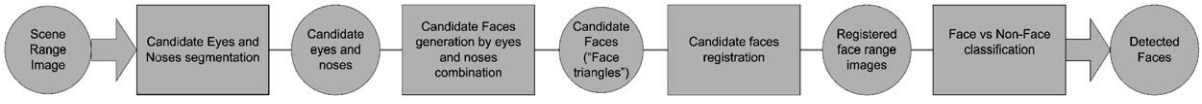

Fig. 1. Schematic diagram of our face detection method. Only the major steps are shown.

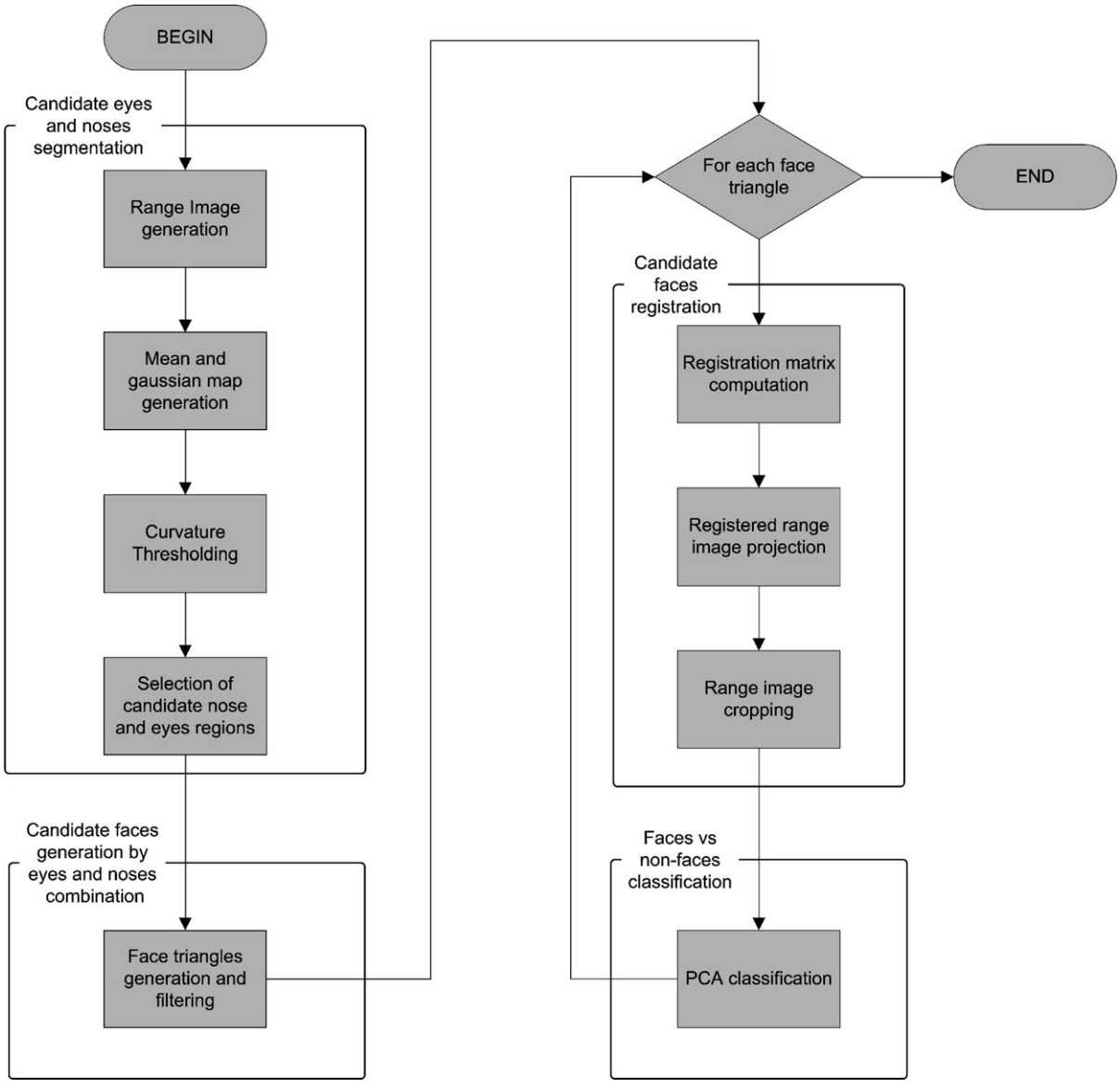

Fig. 2. Detailed diagram of the face detection method.

corresponding 3D surface region including the eyes and the nose, but excluding the mouth and part of the cheeks. This region is rotated and translated into a standard position, 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. Finally, a face vs. non-face 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 extension of each detected face.

2.1. Face curvatures

To analyze the curvature of 3D faces we let S be the surface defined by a twice differentiable real valued function $f : U \to R$, defined on an open set $U \subseteq R^2$:

$$S = \{(x, y, z) \mid (x, y) \in U; z \in R; f(x, y) = z\}.$$

For every point $(x, y, f(x, y)) \in S$ we consider two curvature measures, the mean (H) and the Gaussian (K) curvature [14]:

$$H(x, y) = \frac{\left(1 + f_y^2\right) f_{xx} - 2 f_x f_y f_{xy} + \left(1 + f_x^2\right) f_{yy}}{2\left(1 + f_x^2 + f_y^2\right)^{3/2}},$$

$$K(x, y) = \frac{f_{xx} f_{yy} - f_{xy}^2}{\left(1 + f_x^2 + f_y^2\right)^2},$$

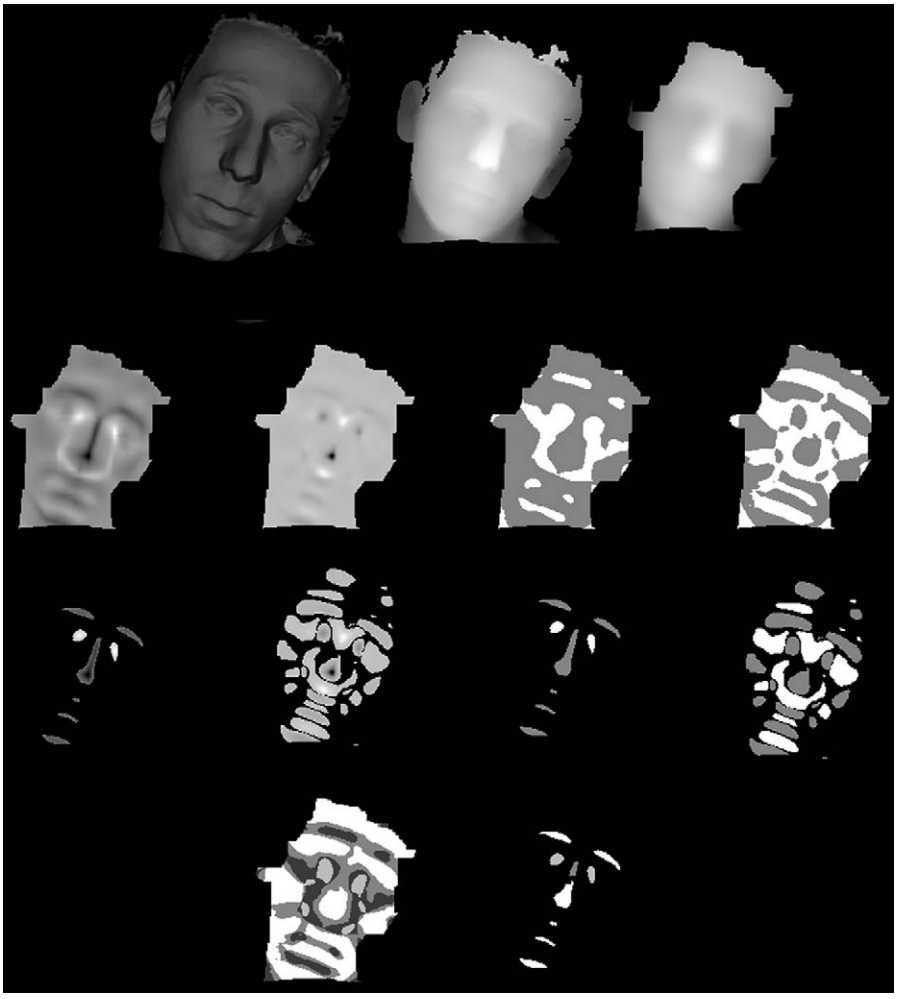

Fig. 3. Top row, left to right: polygonal model, projected range image and its smoothed version. Second row: mean and Gaussian curvature map (the darker zones are high curvature regions, and lighter low curvature regions); sign of mean and Gaussian curvature map (white zones are negative curvature regions; gray zones are positive curvature regions). Third row: thresholded mean and Gaussian curvature map; sign of thresholded mean and Gaussian curvature map. Last row: the *HK*-classification map and its thresholded version. From darker to lighter shades: hyperbolic concave, hyperbolic convex, elliptical concave and elliptical convex regions.

where f_x, f_y, f_{xy}, f_{xx}, f_{yy} are the first and second derivatives of f in (x, y). In our approach, a face is initially represented by a range image of $N \times M$ points. Since we have only a discrete representation of S, we must estimate the partial derivatives. For each point (x_i, y_j) on the grid we considered a biquadratic polynomial approximation of the surface:

$$g_{ij}(x, y) = a_{ij} + b_{ij}(x - x_i) + c_{ij}(y - y_j) + d_{ij}(x - x_i)(y - y_j) + e_{ij}(x - x_i)^2 + f_{ij}(y - y_j)^2, \quad i = 1 \ldots N, \quad j = 1 \ldots M,$$

where the coefficients a_{ij}, $b_{ij}, c_{ij}, d_{ij}, e_{ij}, f_{ij}$ are obtained by least squares fitting of the points in a neighborhood of (x_i, y_j) . The derivatives of f in (x_i, y_j) are then estimated by the derivatives of g_{ij}:

$$f_x(x_i, y_j) = b_{ij}, \quad f_y(x_i, y_j) = c_{ij}, \quad f_{xy}(x_i, y_j) = d_{ij},$$
$$f_{xx}(x_i, y_j) = 2e_{ij}, \quad f_{yy}(x_i, y_j) = 2f_{ij}.$$

Since the second derivative is very sensitive to noise, a smoothing filter must be applied to the surface. Before computing the curvature, we apply a Gaussian filter to the depth image, discarding high-frequency fluctuations of the surface, while the salient facial features, such as the eyes and nose, are still clearly distinguishable.

By analyzing the signs of the mean and the Gaussian curvature, we perform what is called an *HK* classification of the points of the surface to obtain a concise description of the local behavior of the surface (Fig. 3). *HK* classification was introduced by Besl in 1986 [15]. Image points can be labeled as belonging to a viewpoint-independent surface shape class

Table 1
HK classification

	$K<0$	$K=0$	$K>0$
$H<0$	Hyperbolic concave	Cylindrical concave	Elliptical concave
$H=0$	Hyperbolic symmetric	Planar	*Impossible*
$H>0$	Hyperbolic convex	Cylindrical convex	Elliptical convex

type based on the combination of the signs from the Gaussian and mean curvatures as shown in Table 1.

2.2. Eyes, noses and "face triangles"

As suggested by Gordon [9] we use a thresholding process to isolate regions of high curvature. Points with low curvature values are discarded:

$$|H(u,v)| \geqslant T_h, \quad |K(u,v)| \geqslant T_k,$$

where T_h and T_k are predefined thresholds. These thresholds were experimentally tested, before choosing values similar to those used by Moreno et al. [10] ($T_h = 0.04$; $T_k = 0.0005$). If we consider a smoothed face surface, those regions with the highest curvature values are the nose and the inside corners of the eyes. In all our experiments these regions were always well-isolated by using the proposed thresholding process. As other regions, like the mouth, or cheeks, or forehead, do not present particular or simple curvature characteristics that allow robust automatic detection, we decided to consider only the nose and the two inside corners of the eyes. More in detail, we search for the nose region in the thresholded mean curvature map; only positive (convex) regions are considered. We look for the eyes, instead, in the thresholded *HK*-classified map considering only the elliptical concave regions (Fig. 4). Observing the curvature inside the regions of interest we found that possible eyes and noses were well described by statistical descriptors such as the mean. Consequently, we could reduce the number of candidates by filtering each candidate region i, considering average mean and Gaussian curvature ($\overline{H_i}$ and $\overline{K_i}$, respectively):

$$\overline{H}_i \geqslant \overline{H}_{\min}$$

for noses and

$$\overline{K}_i \geqslant \overline{K}_{\min}$$

for eyes. The thresholds $\overline{H}_{\min}$ and $\overline{K}_{\min}$ have been experimentally tuned to consider a smaller number of cases and reduce the system pipeline overhead. The output of this process is then a list of potential noses and eyes (Fig. 4). We now take the face triangles, that is, the triplets (left eye, right eye, nose), generated by combining all the candidate regions. For each nose we compute its principal direction and use that to cut the image plane in half; we then combine the left eye region with the right eye region to build the face triangle, checking the direction in which the triangle is turned: only face triangles having normals pointing towards the camera are considered.

Further filtering again reduces the number of candidates: each face triangle (el, er, n) is described by the distances between the three regions composing it; face triangles with abnormal distances are rejected:

$$LR_{\min} \leqslant d(el, er) \leqslant LR_{\max},$$

$$LN_{\min} \leqslant d(el, n) \leqslant LN_{\max},$$

$$RN_{\min} \leqslant d(er, n) \leqslant RN_{\max},$$

where $d(x, y)$ is the relative distance between regions x and y, and $LR_{\min}, LR_{\max}, LN_{\min}, LN_{\max}, RN_{\min}, RN_{\max}$ are predefined thresholds. Since such distances refer to the actual distances between the features in the acquired scene, it is possible to tune the thresholds in order to retain only triangles compatible with the size and proportions of the human face. This significantly reduces the computational burden of the following steps.

Fig. 4. From left to right: a range image generated by an orthographic projection of a polygonal model; the image used to localize nose candidates; the image used to localize eye candidates; the regions selected for generation of the face triangle.

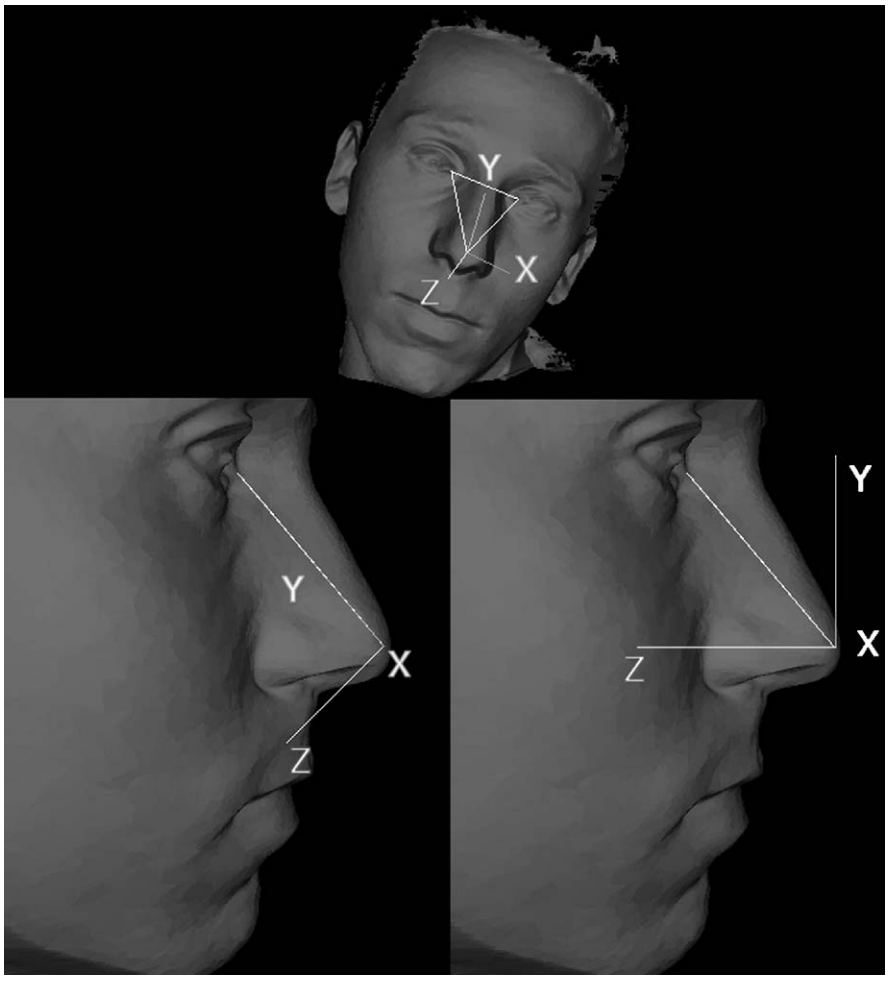

Fig. 5. The reference system used to register a face triangle and its corresponding face surface.

2.3. Registration

Face triangles are freely oriented on the image plane, with some degrees of freedom on the other two rotation planes. We decided to use them as a reference to register the surface of potential faces (Fig. 5). The reference system is built as follows:

- The x-axis is oriented from the right eye to the left one.
- 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°.

Using this reference system, a transformation matrix, $\mathbf{T}$, is built and the 3D model of the face is registered, applying the transformation $\mathbf{T}$ on all the vertices 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 $\mathbf{T}$ can be seen as the product of the translation matrix $\mathbf{S}$ and the rotation matrix $\mathbf{R}$:

$$\mathbf{T} = \mathbf{S} \cdot \mathbf{R},$$

$$\mathbf{R} = \begin{bmatrix} \mathbf{u}_x & \mathbf{u}_y & \mathbf{u}_z & \mathbf{0} \\ \mathbf{v}_x & \mathbf{v}_y & \mathbf{v}_z & \mathbf{0} \\ \mathbf{w}_x & \mathbf{w}_y & \mathbf{w}_z & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{o}_x \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{o}_y \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{o}_z \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix},$$

where the subscripts x, y, and z indicate the components of the vectors.

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

$$\mathbf{p}' = \mathbf{T} \cdot \mathbf{p}.$$

An orthographic projection makes it possible to generate a registered range image corresponding to the surface of the face triangle. Finally, the range image is cropped with a mask to eliminate errors on face borders, so that all the images are equal in shape and size, and only the rigid portion of the face, including eyes and nose (Fig. 6) is considered.

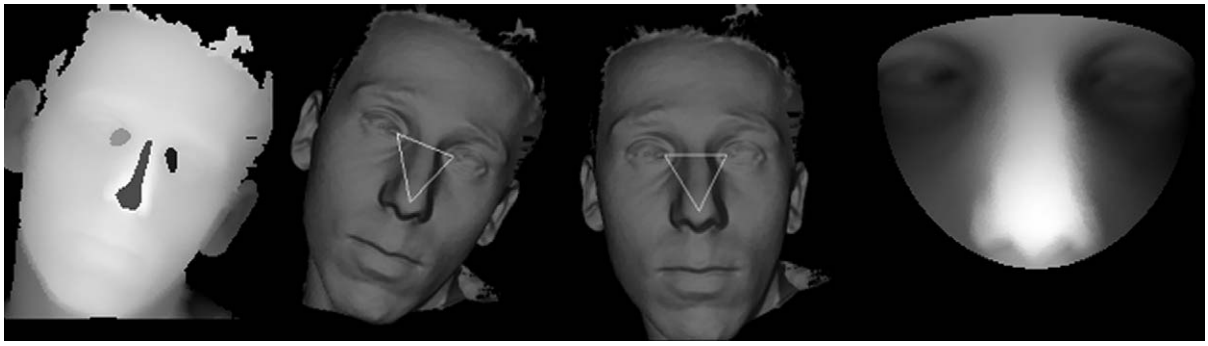

Fig. 6. From left to right: a face range image; the detected face triangle; the same face triangle drawn on the polygonal mesh; the polygonal model registered using the face triangle as a reference; the registered model projected on a range image and cropped with a mask.

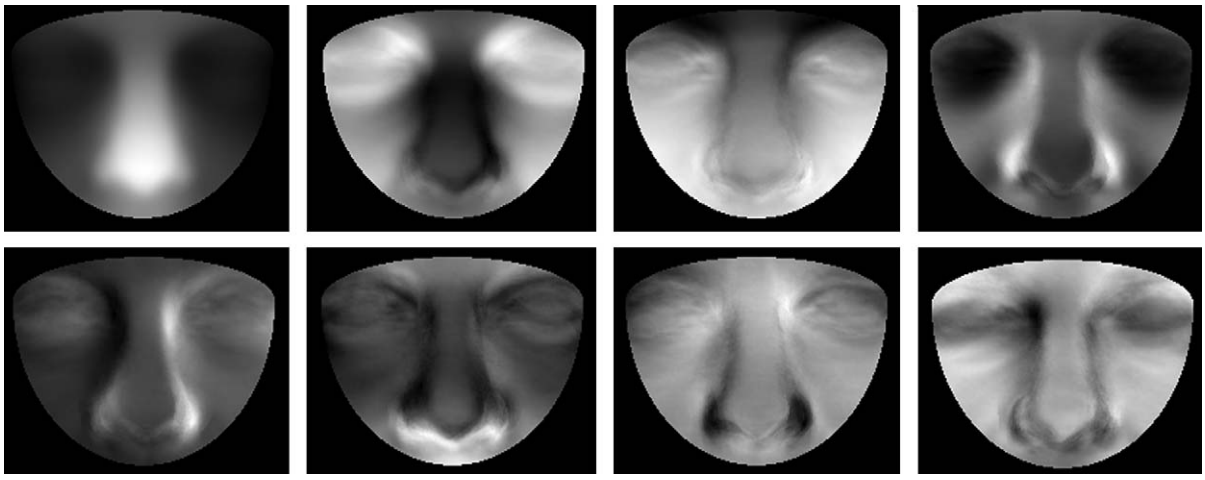

Fig. 7. An example of an eigenface base. Top left, the mean vector; the remaining images are the first seven eigenvectors ordered by variance from left to right, and from top to bottom.

2.4. Face validation

The input of the classifier is the registered and cropped range image of the surface behind each detected face triangle. We have used the eigenfaces technique proposed by Turk and Pentland [2] to classify input patterns as faces or non-faces. In this approach, each registered depth image is used to construct an n-dimensional vector, built by linking together each row of the image. An initial set of M training vectors ($\mathbf{\Gamma_1}, \ldots, \mathbf{\Gamma_M}$) is used to construct the "face space"; first, the mean vector is computed:

$$\mathbf{\Psi} = \frac{1}{M}\sum_{n=1}^{M}\mathbf{\Gamma}_n.$$

Then, the covariance matrix $\mathbf{C}$ is built:

$$\mathbf{C} = \frac{1}{M}\sum_{n=1}^{M}(\mathbf{\Gamma_n} - \mathbf{\Psi})(\mathbf{\Gamma_n} - \mathbf{\Psi})^{\mathrm{T}}.$$

The eigenvectors $\mathbf{u_i}$ of C form an orthogonal base for the n-dimensional space where the axes are oriented toward the directions of maximum variance (Fig. 7). Each eigenvalue λ_i represents the variance associated with the direction indicated by the corresponding eigenvector. The dimension is reduced by selecting the first M' eigenvectors with greater variance, that is, we selected the smallest number of vectors for which the retained variance is greater, or equal to a predefined threshold T_{rv}:

$$r(M') \geqslant T_{rv}, \quad r(k) = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{M}\lambda_i}.$$

The retained variance represents the percentage, not less than 90% in our experiments, of the original information preserved in the reduced space.

We call $\mathbf{U} = (\mathbf{u_1}, \ldots, \mathbf{u_{M'}})$ the base of the reduced space thus computed. A new test face vector can be projected in the reduced space as follows:

$$\boldsymbol{\omega} = \mathbf{U}^{\mathrm{T}}(\mathbf{\Gamma} - \mathbf{\Psi}).$$

Similarly, with some loss of information, the projected vector can be re-projected in the original space (Fig. 8):

$$\mathbf{\Gamma} \cong \mathbf{\Psi} + \mathbf{U}\boldsymbol{\omega}.$$

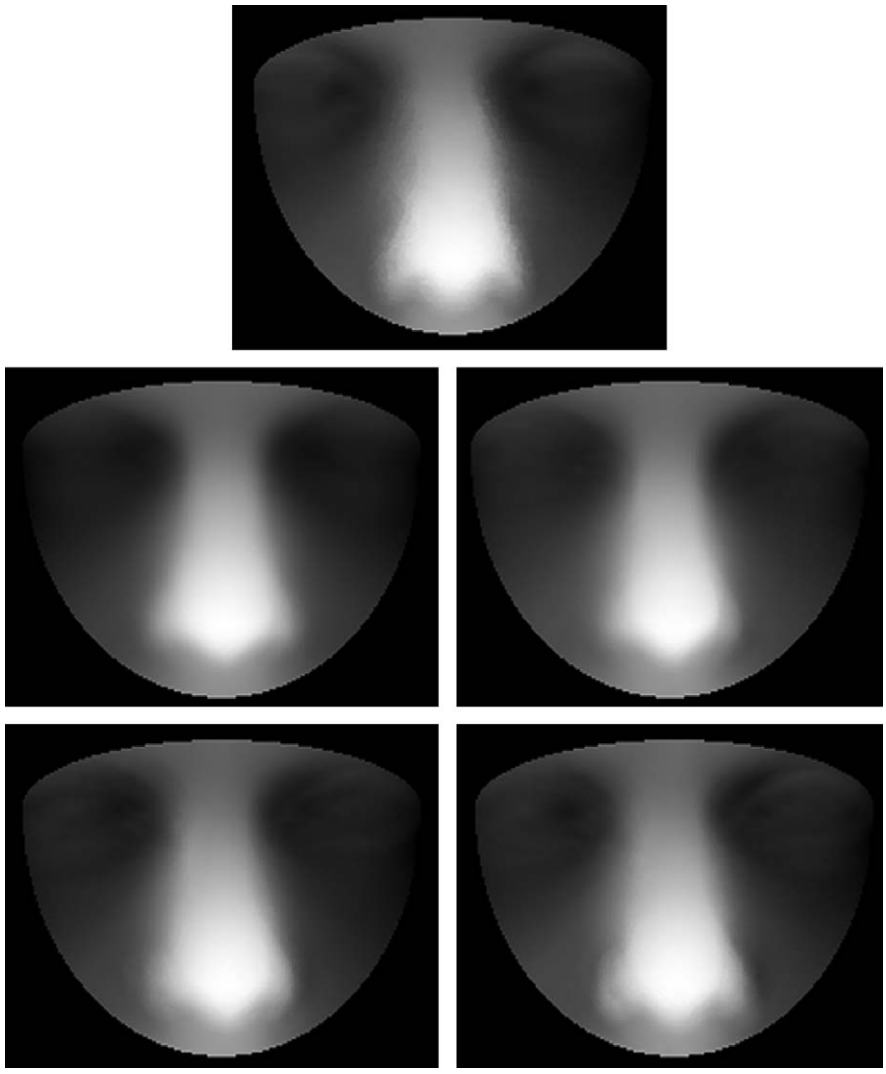

Fig. 8. Some examples of compressed range images using the eigenfaces technique. Top: the original image; and left to right, top to bottom, the same image compressed using 1, 5, 15 and 21 eigenvectors.

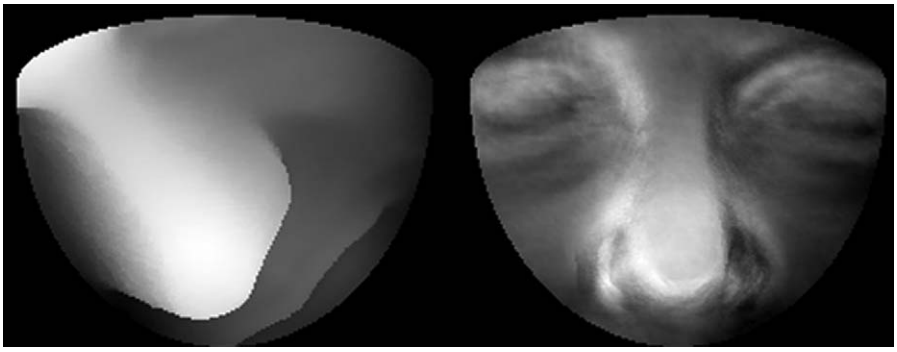

Fig. 9. Left: example of a non-face range image. Right: the same image compressed using the eigenface approach with 21 eigenvectors.

The amount of information lost is used as a measure of "faceness" because it represents the distance of the image from the face space. The idea behind this technique is that a non-face image is not well represented using the face space base as a decomposition base. An example of this is shown in Fig. 9. The reconstruction error is the distance between the original vector and the reconstructed vector:

$$\varepsilon(\mathbf{\Gamma}) = \|\mathbf{\Gamma} - (\mathbf{\Psi} + \mathbf{U}\boldsymbol{\omega})\|_2.$$

Table 2
The threshold $Thresh_f$ used for PCA classification, showing the different threshold values, the resulting false positives and false negatives, and the total error (EER = equal error rate, OPT = optimal value)

$Thresh_f (x10^5)$	False positives (%)	False negatives (%)	Error sum (%)
1.1752	0	5.56	5.56
1.2149	0	4.44	4.44
1.2963	1.11	4.44	5.56
1.3148	1.11	3.33	4.44
1.4013	2.22	3.33	5.56
1.4752	2.22	2.22	4.44 (EER)
1.6909	2.22	1.11	3.33
2.0661	2.22	0	2.22 (OPT)
4.4262	3.33	0	3.33

Vectors with a high reconstruction error are classified as non-faces; otherwise they are recognized as faces:

$$\varepsilon(\mathbf{\Gamma}) \leqslant Thresh_f.$$

The threshold $Thresh_f$ must be determined experimentally; in Table 2 we report the false positives and false negatives obtained by varying the threshold. The equal error rate (EER) is reached when the false positives and false negatives are both 2.22%. The optimal overall error (OPT) obtained is 2.22%.

3. Results

We tested our system on 140 acquisitions of 55 subjects. In 10 of these acquisitions two subjects were present at the same time, while in 24 acquisitions some distracting, i.e. non-face, objects were present. In 20 acquisitions, the subjects were wearing hats, scarves and so on, or assumed a non-neutral facial expression. All the acquisitions were obtained using a Minolta Vivid 900 laser range scanner. Subjects were positioned 3 m away from the device and the range images resolution used was 640 × 480. Our method is reasonably independent from image resolution: low resolution images can also be used if the facial features (i.e. the eyes and the nose) are clearly distinguishable.

Fig. 10 presents an example of a scene containing two people. As can be seen, the localization of eyes and noses has generated several false hits. Three face triangles have been generated; and one of these is a non-face. The classifier has rejected the registered range image associated with the

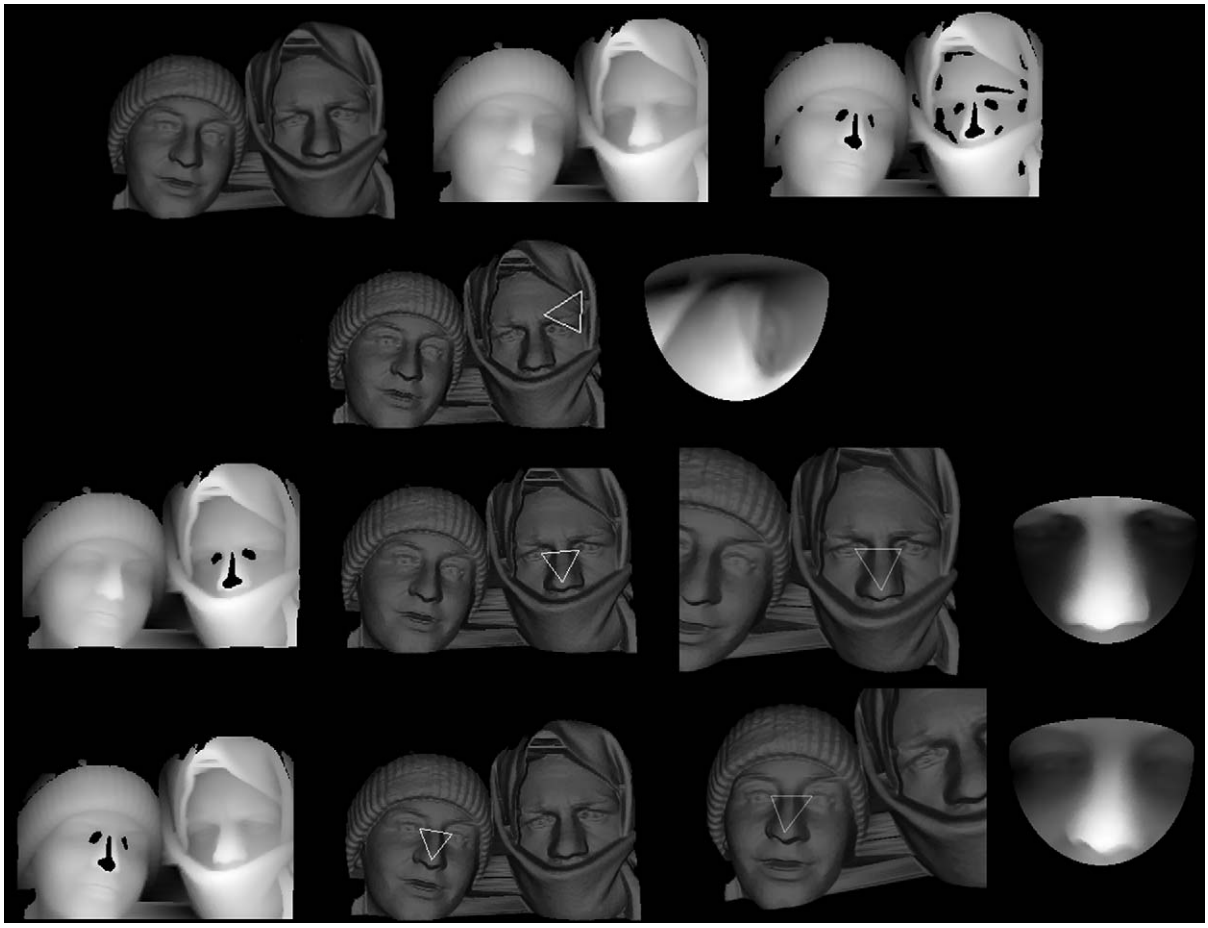

Fig. 10. An example of the system in action. Top row: input polygonal mesh, projected range image and eyes and noses candidates. Second row: a candidate face triangle and the associated range image that was classified as non-face. Third and fourth rows: the two face triangles correctly classified as faces.

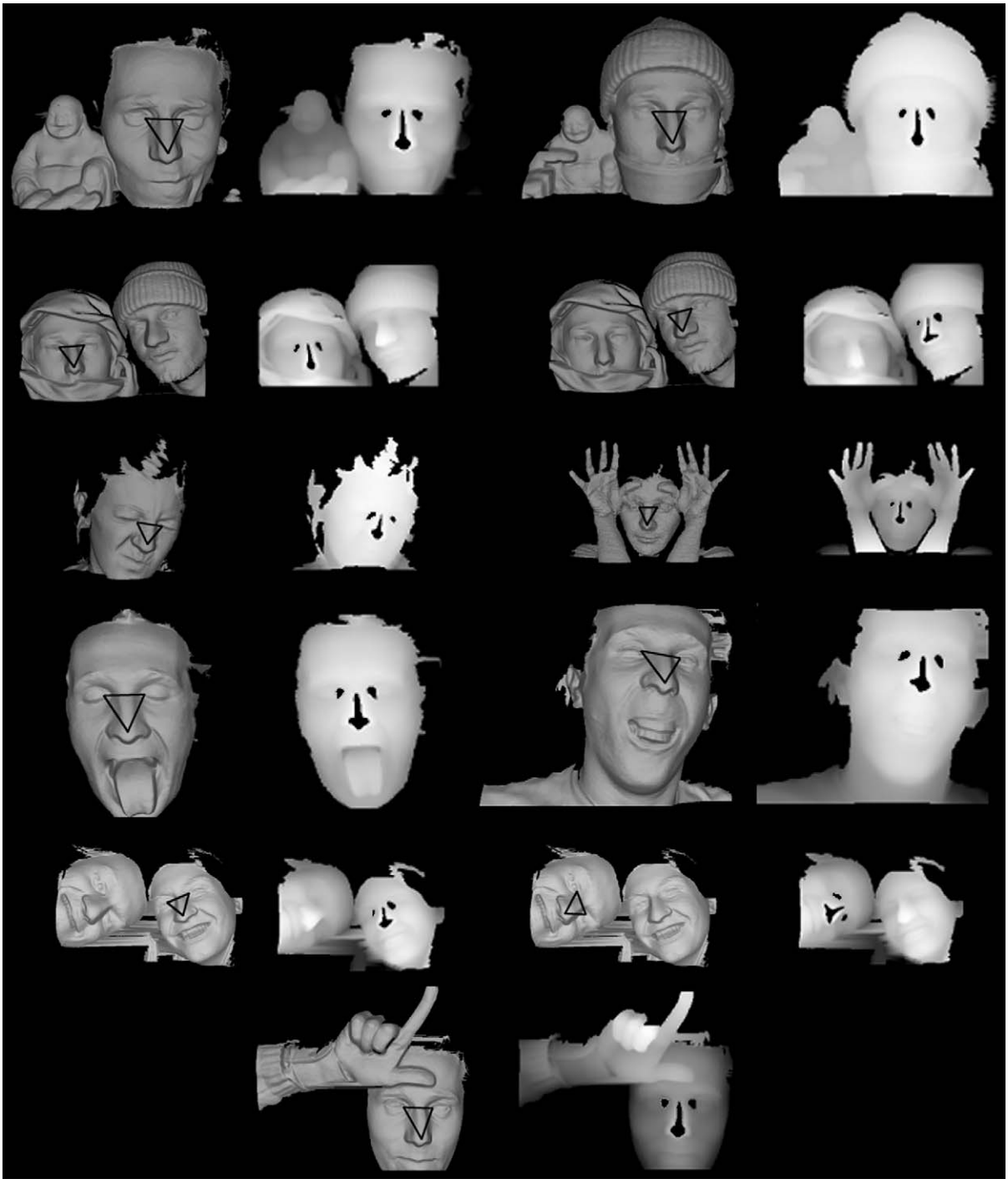

Fig. 11. Some significant examples of detection successes, showing the robustness of the system in the presence of marked occlusions, rotations, and non-neutral facial expressions.

false face triangle and accepted the other two. Fig. 11 shows some challenging examples when the method has correctly classified: in two of them, the subject has a little statue near his face; the system has not identified the statue's face as a candidate because the distances between its features are smaller than those of humans, and do not satisfy the constraints described in Section 2.2. This is one of the advantages in using real world 3D reference system and data. In another acquisition the subject is showing his tongue with the mouth open wide. The variation from a neutral expression in the 3D sense is very great, but the system shows the robustness which comes from using only the rigid part of the face for classification. In two other acquisitions, the subjects have fingers, hands, or arms near the face, sometime occluding part of it. In these cases, the facial feature detector has generated some false noses, but the filtering process

Table 3
Experimental results

Number of acquisitions	140
Number of subjects	55
Total faces	150
Errors due to holes	23
Other errors	4
Total errors	27
Successes (calculated on all images)	82%
Successes (calculated only on images without artifacts)	96.85%

described in Section 2.2 has prevented the generation of any face triangles. In some acquisitions the subjects are wearing hats and scarves, occluding the mouth and forehead; again, the choice of using only the rigid part of the face has allowed the system to do its work correctly. Finally, in some cases the subjects are not in an upright position, but as our facial feature detector is rotation-invariant, and the face vs. non-face classifier is applied after face registration, this did not present a problem.

The experimental results are summarized in Table 3. Most of the failures were caused by holes introduced by the scanner (Fig. 12 shows an example). These holes correspond to

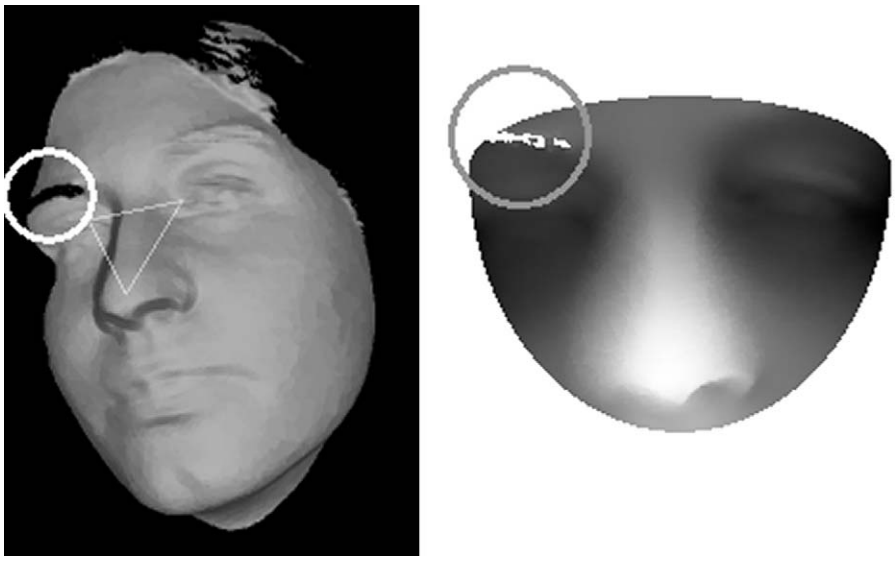

Fig. 12. An example of a hole in the eyebrows region. The projected range image contains points of null information that the classifier is not able to handle.

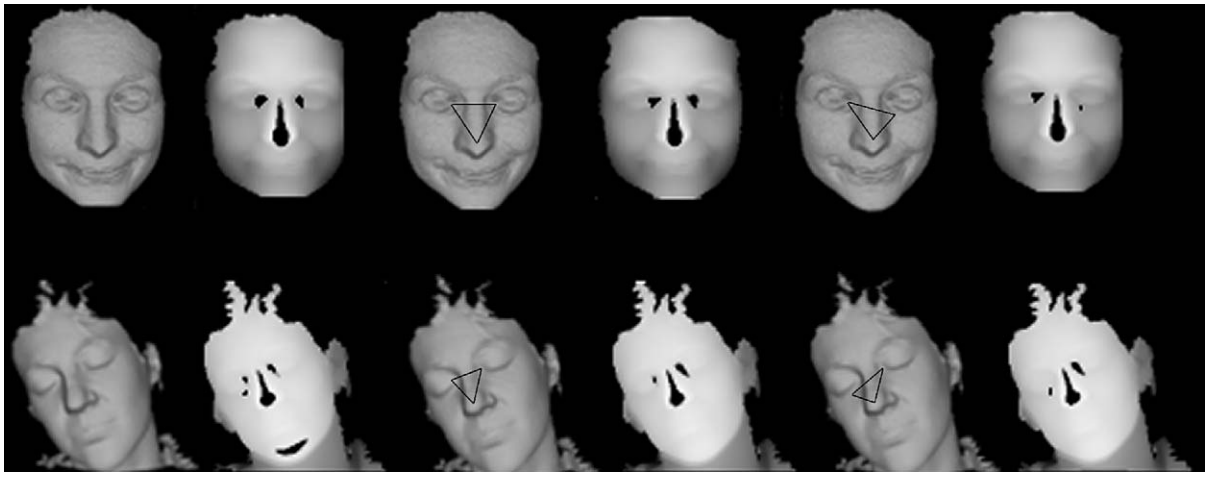

Fig. 13. Two examples of the overlapping face triangle problem. As can be seen, in each case a small candidate eye is generated near the true eyes; as a consequence more than one candidate face is generated for the same real face.

points of null information within range images, which the classifier has not been able to deal within its present version. These holes were usually located on the eyebrows or similar regions because laser technology has some difficulty in capturing that type of surface. In two cases we encountered the problem of overlapping face triangles: the segmentation step produced small regions near the eyes with curvature characteristics similar to those of the eyes. This generated more than one face triangle which was then classified as a face (see Fig. 13). The problem can be handled by introducing an overlapping check and selecting the face triangle with the highest faceness. In only one case was a non-face classified as a face, and in only one other case did a face go undetected, because the nose region was not well segmented by thresholding.

4. Conclusions and future work

We present an innovative approach to the problem of recognizing faces in frontal 3D acquisitions. The results are quite satisfactory: only a small number of faces were missed in a test set of 150 faces, and no false positives occurred. Moreover, since our method is based only on geometrical information, it has proved very reliable even in the presence of variations in the lighting and pose of subjects. Invariance with respect to facial expressions is achieved by considering only an almost rigid part of the face, a feature of the method confirmed by tests on several examples of highly emphasized expressions. Most failures are due to artifacts introduced in acquisition. We plan to reduce the possibility of such errors by adding a pre-processing module that will reduce the noise in raw data and eliminate artifacts (e.g. replacing the holes in the model with an interpolated surface). To further improve classification accuracy, we plan to extend the method integrating the geometrical features with pictorial information as well. Finally, in order to provide a better estimate of the performance of the method, we intend to amplify the dataset used to date.

References

[1] Yang Ming-Hsuan, J. Kriegman David, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (1) (2002) 34–58.
[2] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1991) 71–86, 24.
[3] C. Beumier, M. Acheroy, Face verification from 3D and grey level clues, Pattern Recognition Lett. 22 (12) (2001) 1321–1329.
[4] J.Y. Cartoux, J.T. Lapreste, M. Richetin, Face authentication or recognition by profile extraction from range images, IEEE Computer Society Workshop on Interpretation of 3D Scenes, 1989, pp. 194–199.
[5] G. Pan, Y. Wu, Z. Wu, Investigating profile extracted from range data for 3d face recognition, IEEE Int. Conf. Syst. Man Cybern. 2 (2003) 1396–1399.
[6] C. Hehser, A. Srivastava, G. Erlebacher, Principal component analysis of range images for facial recognition, Proceedings of the Multiconference in Computer Science, 2002.
[7] K.I. Chang, K.W. Bowyer, P.J. Flynn, Multi-modal 2d and 3d biometrics for face recognition, IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003, p. 187
[8] Y. Lee, K. Park, J. Shim, T. Yi, 3D face recognition using statistical multiple features for the local depth information, Proceedings of the 16th International Conference on Vision Interface, 2003
[9] G.G. Gordon, Face recognition based on depth maps and surface curvature, SPIE Geom. Methods Comput. Vision 1570 (1991) 234–274.
[10] A. Moreno, A. Sanchez, J. Velez, F. Diaz, Face recognition using 3d surface extracted descriptors, Proceedings of the Irish Machine Vision and Image Processing 2003, 2004 (ISBN 1-85923-177-2).
[11] Y. Wang, C.S. Chua, Y.K. Ho, Facial feature detection and face recognition from 2d and 3d images, Pattern Recognition Lett. 23 (2002) 1191–1202.
[12] C.S. Chua, R. Jarvis, Point signatures: a new representation for 3d object recognition, Int. J. Comput. Vision 17 (1997) 77–79.
[13] A. Watt, M. Watt, Advanced Animation and Rendering Techniques: Theory and Practice, Addison-Wesley, Reading, MA, 1992.
[14] M.P. Do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, Englewood Cliffs, NJ, 1976.
[15] P.J. Besl, R.C. Jain, Invariant surface characteristics for 3-d object recognition in range images, Comput. Vision, Graphics Image Process. 33 (1986) 33–80.

About the Author—ALESSANDRO COLOMBO is a Ph.D. student at DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione) of the University of Milano-Bicocca, where he took his degree in Computer Science in 2004. His current activity focuses on 3D imaging, face detection and classification.

About the Author—CLAUDIO CUSANO is a Ph.D. student at DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione) of the University of Milano-Bicocca, where he took his degree in Computer Science. Since April 2001 he has been a fellow of the ITC Institute of the Italian National Research Council. His current activity focuses on image and video analysis and classification, and on 3D imaging.

About the Author—RAIMONDO SCHETTINI is associate professor at DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione) of the University of Milano Bicocca, where he is in charge of the Imaging and Vision Lab. He has been associated with Italian National Research Council (CNR) since 1987, has been team leader in several research projects, and published over 160 refereed papers on image processing, analysis and reproduction, and on image content-based indexing and retrieval. He is an associate editor of the Pattern Recognition Journal, and has been co-guest editor of three special issues on Internet Imaging (Journal of Electronic Imaging, 2002), Color Image Processing and Analysis (Pattern Recognition Letters, 2003), and Color for Image Indexing and Retrieval (Computer Vision and Image Understanding, 2004). He was General Co-Chairman of the first Workshop on Image and Video Content-based Retrieval (1998), of the First European Conference on Color in Graphics, Imaging and Vision (2002), and of the EI Internet Imaging Conferences (2000–2006).