Eurographics Workshop on 3D Object Retrieval (2011)
H. Laga, T. Schreck, A. Ferreira, A. Godil, I. Pratikakis, and R. Veltkamp (Editors)

Partial Match of 3D Faces using Facial Curves between SIFT Keypoints

Stefano Berretti, Alberto Del Bimbo and Pietro Pala

Dipartimento di Sistemi e Informatica, University of Firenze, Italy

Abstract

In this work, we propose and experiment an original solution to 3D face recognition which supports partial matching of facial scans as occurs in the case of missing parts and occlusions. In the proposed approach, distinguishing traits of the face are captured by first extracting SIFT keypoints on the face scan and then measuring how the face changes along facial curves *defined between pairs of keypoints. Facial curves are also associated with a measure of salience so as to distinguish curves that model characterizing traits of some subjects from curves that are frequently observed in the face of many different subjects. The recognition accuracy of the approach has been experimented on the Face Recognition Grand Challenge dataset.*

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications— I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Curve, surface, solid, and object representations

1. Introduction

Automatic recognition of human faces is a challenging computer vision task especially in presence of illumination variations or in the case parts of the face are missing. Recently, the availability of 3D facial data acquired with scanner devices has increased the interest in 3D face recognition solutions that are expected to feature less sensitivity to pose and illumination changes. Based on this, many 3D face recognition approaches have been proposed and experimented in the last years [BCF06], [BDP10a]. In summary, these approaches can be grouped in two broad categories: *global* (or *holistic*), that perform face matching based on representations extracted from the whole face; and *local* (or *region-based*), that partition the face surface into regions, and extract and match appropriate descriptors for each of them. Many of these approaches have been designed to support face recognition also in presence of expression variations reporting very high accuracy on benchmark databases like the *Face Recognition Grand Challenge* (FRGC version 2.0 dataset) [PFS*05]. However, a problem that can substantially affect the accuracy of recognition and that has not been extensively addressed is the effect induced by *missing parts* and *partial occlusions* of the face. More generally the problem of supporting the recognition of a subject when only a part of her/his facial scan is available. Missing parts can be determined by self-occlusions of the face due to pose variations, whereas face occlusions are likely to occur in real applications due to the hair, glasses, scarves or caps.

The effects of face occlusions have been first studied in 2D face recognition applications. In 3D, just few solutions have explicitly considered the problems posed by missing parts and occlusions in the design and experimentation of recognition methods. In general, global approaches cannot effectively manage these conditions, instead local approaches have the potential to address partial face matching. In [PPT*09], an automatic face landmarks detector is used to identify the pose of the facial scan so as to mark regions of missing data and to roughly register the facial scan with an Annotated Face Model (AFM). The AFM is fitted using a deformable model framework that exploits facial symmetry where data are missing. Wavelet coefficients extracted from a geometry image derived from the fitted AFM are used for the match. Experiments have been performed using the FRGC v2.0 gallery scans and side scans with 45° and 60° rotation angles as probes. In [DBDS10], the facial surface is represented as a collection of radial curves originating from the nose tip and face comparison is obtained by elastic matching of the curves. A quality control permits the exclusion of corrupted radial curves from the match, thus enabling the recognition also in the case of missing data. Results of partial matching are given for the 61 left and right side scans of the Gavab database.

Many local approaches are limited by the need to identify facial landmarks used to define the interesting parts in matching faces. Methods that use keypoints of the face promise to solve some of these limitations. In particular, a

DOI: 10.2312/3DOR/3DOR11/117-120

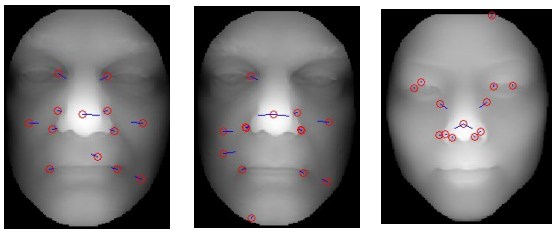
few recent works have shown that local descriptors computed around salient keypoints can be usefully applied to describe 3D objects and faces. Following this idea, in this work we develop on the approach in [BDP10b] to support 3D face recognition in presence of missing parts. The approach is motivated by the observation that SIFT keypoints can detect salient points of the face with high repeatability and these can be complemented with data that model the morphological changes of the face across pair of keypoints. For this purpose, we introduce the *facial curve* to model the depth of the face scan along the surface line connecting two SIFT keypoints. Characterizing traits of the face are captured by considering the SIFT descriptors of detected keypoints as well as the set of facial curves identified by each pair of keypoints. Facial curves are also associated with a measure of salience so as to distinguish those that model characterizing traits of some subjects from those that are frequently observed in the face of many different subjects. In the comparison of two faces, SIFT descriptors are matched to measure the similarity between pairs of keypoints identified on the two range images. Then, the distance between the two faces is derived by composing the individual distances between facial curves (weighted by their salience) that originate from pairs of matching keypoints. The proposed solution is experimented using the FRGC dataset.

The paper is organized as follows. In Sect. 2, the main characteristics of the SIFT are summarized, and its adaptation to our case is described. The extraction of facial curves and the definition of a distance to measure the dissimilarity between curves are discussed in Sect. 3, together with a method to evaluate the salience of facial curves. Experiments carried out with the proposed approach, with results obtained on the FRGC database are reported in Sect. 4.

2. SIFT keypoints of range facial images

The proposed 3D face description approach relies on the detection of a number of keypoints on the 3D face surface and the description of the 3D face surface in correspondence to these keypoints as well as along linear paths connecting pairs of keypoints. We expect the position of keypoints to be influenced by the specific morphological traits of each subject. In particular, assuming that the process of keypoint detection incorporates a measure of the scale associated with each keypoint, we relax the assumption that detected keypoints correspond to meaningful landmarks and exploit the more general assumption of *within subject repeatability*: the position of the most stable keypoints—detected at the coarsest scales—do not change substantially within facial scans of the same subject.

Following this approach, we used the SIFT algorithm for the purpose of keypoints identification and description. SIFT has been defined for 2D gray-scale images and cannot be directly applied to 3D face scans. However, the 3D information of scanned faces can be captured through *range images* that use the gray-scale of image pixels to represent the depth of a face scan. According to this, the SIFT *keypoint detector* is applied to the range images in order to extract image keypoints. Although many possible keypoints at different locations in an image could be detected, only the most distinctive and invariant ones should be retained for matching. These often lay on edges and corners of the image, and can be of many different sizes and orientations as well.



(a) (b) (c)

Figure 1: *Keypoints detected on the range images of two subjects (the orientation of the lines shown at the keypoints also represent the orientation of the main component of the SIFT descriptor). The range images in (a) and (b) represent the same subject, whereas the image in (c) represents a different subject.*

For the detected keypoints, the SIFT *descriptors* are computed. Briefly, a SIFT descriptor for a small image patch, for example of size 4×4, is computed from the gradient vector histograms of the pixels in the patch. There are 8 possible gradient directions, and therefore the total size of the SIFT descriptor is $4 \times 4 \times 8 = 128$ elements. This descriptor is normalized to enhance invariance to changes in illumination (not relevant in the case of range images), and transformed in other ways to ensure invariance to scale and rotation as well. These properties make the SIFT descriptor capable to provide compact and powerful local representation of the range image and, as a consequence, of the face surface. In particular, in our solution SIFT keypoints are extracted, and their scale and orientation angles are retained. The orientation histograms of 4×4 sample regions of each keypoint are used to calculate the SIFT descriptor.

In Fig. 1, the detected SIFT keypoints are shown for the range images of two different subjects. In particular, Figs. 1(a)-(b) show the keypoints for two range images of the same subject, whereas the range image in (c) represents a different subject. In general, the position of keypoints detected on a face scan depends on the morphological traits of the face. This is confirmed by the fact that the spatial arrangement of keypoints detected on different scans of the same subject—with neutral expression—is very similar. However, the information that is captured by combining local SIFT descriptors of the keypoints detected in a face scan is not discriminant enough to support accurate recognition of the identity of the subject. This leads to the definition of *facial curves* that is discussed in detail in the next Section.

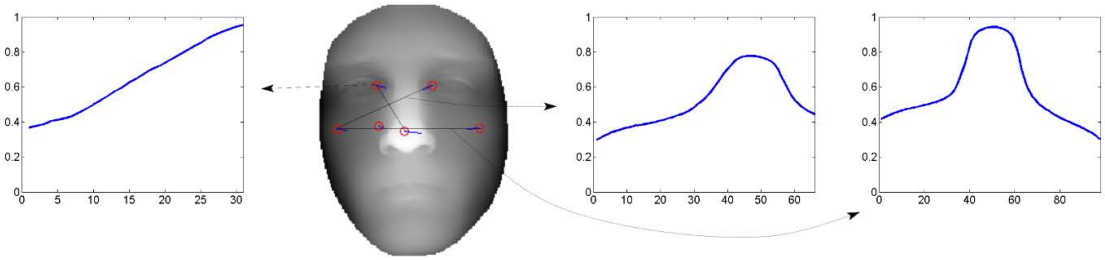


Figure 2: *Facial curves originated by some pairs of keypoints. In each plot, the horizontal axis reports the number of pixels between the two keypoints, whereas the vertical axis is the normalized gray-level of the range image along the curve.*

3. Facial curves between keypoints

Each pair of SIFT keypoints detected on a range image is used to identify a *facial curve*, that is the 1D function of the depth values of the pixels that lay on the segment connecting the two keypoints. More formally, let $I(\mathbf{x})$ with $\mathbf{x} \in \Re^2$ be the range image representing a face scan, $\mathbf{x}_1$ and $\mathbf{x}_2$ two keypoints, then the facial curve identified by the ordered pair $(\mathbf{x}_1, \mathbf{x}_2)$ is defined as:

$$\mathcal{C}^I_{\mathbf{x}_1,\mathbf{x}_2}(t) = I((1-t)\mathbf{x}_1 + t\mathbf{x}_2), \quad t \in [0,1]. \tag{1}$$

As an example, Fig. 2 shows three facial curves derived from the range image of a sample subject. It can be observed as the curves capture the shape of the face across different paths and in particular the nose protrusion.

In the proposed face description model, distinguishing facial traits are captured by retaining SIFT descriptors of keypoints detected on the range image as well as the curves identified by pairs of these keypoints. These data are organized in a graph $\mathcal{G}(N,E)$, where nodes correspond to keypoints and edges to facial curves. Given the graphs of two faces, their dissimilarity is evaluated by first assigning to each node of the first graph its closest node in the second graph, proximity being measured as the Euclidean distance d_N between the 128-dimensional SIFT descriptors associated with the keypoints. Then, for each pair of matching nodes in the two graphs all curves originating from the keypoints are compared so as to identify the two curves with minimum distance. Considering the generic curve $\mathcal{C}_1(t)$, $t \in [0,t_1]$ extracted from the pair of keypoints $(\mathbf{x}_1, \mathbf{x}_2)$ of the face scan I_1, and the curve $\mathcal{C}_2(t)$, $t \in [0,t_2]$ extracted from the pair of keypoints $(\mathbf{x}_3, \mathbf{x}_4)$ of the face scan I_2, the distance between the two curves is measured as:

$$\mathcal{D}(\mathcal{C}_1(t), \mathcal{C}_2(t)) = \int_0^{min\{t_1,t_2\}} |\mathcal{C}_1(t) - \mathcal{C}_2(t)|\, dt. \tag{2}$$

Eventually, the distance between the two face scans is measured by averaging the curve minimum distance values over all pairs of matching nodes. It should be noticed that the matching scheme does not exploit any specific assumption about the parts of the face that are modeled by nodes and edges of the graph. As a consequence, the matching scheme can be used without any change to support matching between a partial scan and a full scan of the face, thus enabling the recognition of faces with missing parts and/or occlusions.

3.1. Salience of facial curves

In the following, a model is defined to associate with a generic curve a value of salience. The higher the salience of the curve y, the lower the uncertainty about the identity of the subject that is represented in the 3D scan M if y is observed in M. Formally, let's assume that the 3D face scans of n different subjects are available and let X be a discrete random variable taking values in the set $\{x_1, \ldots, x_n\}$, representing the identity of the subject. Furthermore, let Y be a continuous random variable representing a sample curve. Given an observation $Y = y$, the uncertainty of X once y is observed can be measured through the *Shannon entropy of the posterior distribution*, that is defined as:

$$H(X|Y=y) = -\sum_{x_i} P(X=x_i|Y=y) \log P(X=x_i|Y=y). \tag{3}$$

Values of $H(X|Y=y)$ are high for those curves y that are observed in the faces of many subjects and low for those curves y that are observed in the faces of just a few subjects.

Operatively, since in a real application context only the gallery scan is available for each subject, estimation of $P(X=x_i|Y=y)$ is prevented. Therefore, in the proposed solution the salience of the curve y is estimated through the formula $\mathcal{S}(y) = e^{-\alpha N}$, being N the occurrences of y in the scans of the gallery. The salience of curves is included in Eq. (2) by weighting the values of the distance with $\mathcal{S}$ so as to penalize the distance to non salient curves. In particular, the distance between a curve $\mathcal{C}_1(t)$, $t \in [0,t_1]$ extracted from a probe face scan I_1, and a curve $\mathcal{C}_2(t)$, $t \in [0,t_2]$ extracted from a gallery face scan I_2 is measured as:

$$\mathcal{D}(\mathcal{C}_1(t), \mathcal{C}_2(t)) = \frac{\int_0^{min\{t_1,t_2\}} |\mathcal{C}_1(t) - \mathcal{C}_2(t)|\, dt}{\mathcal{S}(\mathcal{C}_2(t))}. \tag{4}$$

4. Experimental Results and Discussion

In the following, we present the 3D face recognition results in presence of missing parts obtained on the FRGC v1.0 dataset [PFS*05], which includes 943 3D face scans of 275 individuals showing a neutral facial expression. First, face descriptions are extracted from depth images of the 3D face scans [BDP10b]. According to the proposed approach, SIFT keypoints are automatically extracted from the range images of all the faces. This results in a variable number of keypoints per image, dependent on the specific characteristics of the face surface. On every depth image, only the top 15 keypoints—selected after ordering all keypoints from the coarsest to the finest scale (σ) value—are retained, and facial curves are computed between every pair of these keypoints. On these keypoints and the corresponding incident facial curves the computation of salience is applied as reported in Sect. 3.1.

In the experiments, the first scan of each subject is selected to be included in the gallery. This resulted in a gallery with 275 reference scans, whereas all the scans were used as probes and compared against the gallery. To evaluate the accuracy of the proposed solution with respect to missing parts of the face, each probe face scan is divided into the *left* and *right* parts with respect to the ideal vertical plane of symmetry passing from the nose tip. The curves that originate from only one part of the face are compared to gallery scans, in two separated matching experiments using, respectively, the left and the right part of each probe face scan. The effectiveness of recognition has been measured through the rank-k recognition rate, and presented with *Cumulative Matching Characteristics* (CMC) curves. Fig. 3, reports the CMC curves for recognition using just a half of the face as probe. The recognition accuracy obtained using the full probe face scans without facial curve salience is reported as baseline for comparison (dotted green curve). It can be observed that if the salience of the curves is disregarded, the accuracy of recognition is about 70%. In contrast, if the salience of facial curves is used in the comparison of two scans, recognition accuracy is improved by more than 10%. In particular, the rank-1 recognition accuracy is up to 83% if the full face is used as probe and the salience of facial curves is considered. It is 80% and 79% if just the left and right part of the face is used as probe.

Results on 3D partial face matching are also reported in [FBF08], [PPT*09] and [DBDS10], but using different datasets so that the results are not directly comparable among them and with our work. In [FBF08], a rank-1 recognition rate of 88% and 89.2% on the FRGC v2.0 dataset is reported, respectively, matching regions in the left and right part of the face with the entire scans in the gallery (i.e., all facial regions are used to represent gallery scans). However, their experimental setting implicitly assumes that the matching regions are not affected by missing parts (i.e., regions can miss, but not their parts). In [PPT*09], rank-1 of 69% and

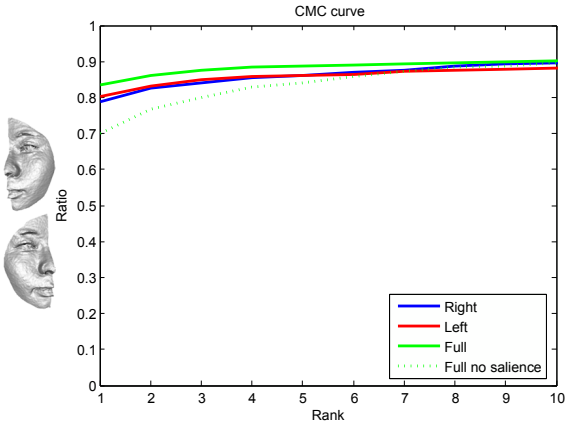


Figure 3: *CMC curves on the FRGC v1.0 dataset. Four curves are shown: the first three correspond to the use of the Full/Left/Right probes with facial curve salience; the last curve is used as baseline and corresponds to the use of the Full probes without facial curve salience.*

44% are reported for left and right probes of the face rotated of 45° and 60° on the FRGC v2.0 gallery. Instead, recognition rates of 70.5% and 86.9% are reported in [DBDS10], respectively, for the right and left side probes of the Gavab database on a gallery of 61 subjects.

References

[BCF06] BOWYER K. W., CHANG K. I., FLYNN P. J.: A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer Vision and Image Understanding 101*, 1 (Jan. 2006), 1–15. 1

[BDP10a] BERRETTI S., DEL BIMBO A., PALA P.: 3D face recognition using iso-geodesic stripes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 12 (Dec. 2010), 2162–2177. 1

[BDP10b] BERRETTI S., DEL BIMBO A., PALA P.: Recognition of 3d faces with missing parts based on profile networks. In *ACM Workshop on 3D Object Retrieval* (Firenze, Italy, Oct. 2010), pp. 81–86. 2, 4

[DBDS10] DRIRA H., BEN AMOR B., DAOUDI M., SRIVASTAVA A.: Pose and expression-invariant 3d face recognition using elastic radial curves. In *British Machine Vision Conference* (Aberystwyth, UK, August 2010), pp. 1–11. 1, 4

[FBF08] FALTEMIER T. C., BOWYER K. W., FLYNN P. J.: A region ensemble for 3D face recognition. *IEEE Transactions on Information Forensics and Security 3*, 1 (Mar. 2008), 62–73. 4

[PFS*05] PHILLIPS P. J., FLYNN P. J., SCRUGGS T., BOWYER K. W., CHANG J., HOFFMAN K., MARQUES J., MIN J., WOREK W.: Overview of the face recognition grand challenge. In *IEEE Workshop on Face Recognition Grand Challenge Experiments* (San Diego, CA, June 2005), pp. 947–954. 1, 4

[PPT*09] PERAKIS P., PASSALIS G., THEOHARIS T., TODERICI G., KAKADIARIS I. A.: Partial matching of interpose 3d facial data for face recognition. In *International Conference on Biometrics: Theory, Applications, and Systems* (Washington, DC, September 2009), pp. 1–8. 1, 4