Eurographics Workshop on 3D Object Retrieval (2013)
S. Biasotti, I. Pratikakis, U. Castellani, T. Schreck, A. Godil, and R. Veltkamp (Editors)

Geometric Histograms of 3D Keypoints for Face Identification with Missing Parts

Stefano Berretti[1], Naoufel Werghi[2], Alberto del Bimbo[1] and Pietro Pala[1]

[1]Department of Information Engineering, University of Firenze, Italy
[2]Department of Electrical and Computer Engineering, Abu Dhabi, United Arab Emirates

Abstract

In this work, an original solution to 3D face identification is proposed, which supports recognition also in the case of probes with missing parts. Distinguishing traits of the face are captured by first extracting 3D keypoints of a face scan, then measuring how the face surface changes in the keypoints neighborhood using a local descriptor. To this end, an adaptation of the meshDOG algorithm to the case of 3D faces is proposed, together with a multi-ring geometric histogram descriptor. Face similarity is then evaluated by comparing local keypoint descriptors across inlier pairs of matching keypoints between probe and gallery scans. Experiments have been performed to assess the keypoints distribution and repeatability. Recognition accuracy of the proposed approach has been evaluated on the Bosphorus database, showing competitive results with respect to existing 3D face biometrics solutions.

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

Accuracy of automatic identity recognition based on faces still suffers from many factors, such as pose changes, illumination variations, facial expressions and occlusions. To solve these problems, face recognition using 3D scans of the face has been recently proposed as an alternative or complementary solution to conventional 2D face recognition approaches using still images or videos, so as to allow accurate face recognition also in real-world applications with unconstrained acquisition. Confirming this recent research trend, several 3D face recognition approaches have been proposed and experimented in the last few years (see the survey in [BCF06], and the literature review in [BDP10, WLT10] for a thorough discussion). However, many of the works appeared in this field proposed conventional face recognition experiments, where both probe and gallery scans are assumed to be acquired cooperatively in a controlled environment in which the whole face is precisely captured and represented. These methods mainly focussed on face recognition in the presence of expression variations, reporting very high accuracy on benchmark databases, like the FRGC v2.0 [PFS*05]. Solutions enabling face recognition in uncooperative scenarios are now attracting an increasing interest [PPTK11]. In such a case, probe scans are acquired in unconstrained conditions that may lead to *missing parts* (non-frontal pose of the face), or to *occlusions* due to hair, glasses, scarves, hand gestures, etc. These difficulties are further sharpened by the recent advent of 4D scanners (3D plus time), capable of acquiring temporal sequences of 3D scans. In fact, the dynamics of facial movements captured by these devices can be useful for many applications, but also increases the acquisition noise and the variability in subjects' pose. In summary, techniques supporting face recognition using *3D partial face scans* are gaining importance in making existing techniques deployable in more general contexts and, in perspective, in scenarios where dynamic 3D acquisition is performed. However, the research in this context is still preliminary also due to the limited number of face databases that also comprise partial acquisitions of 3D faces.

In this work, we propose an original 3D face recognition approach which is capable to perform subject identification also in the case parts of the face scans are missing. We rely on the observation that describing the face with local geometric information extracted at the neighbors of keypoints allows partial face comparison by sparse keypoints matching. According to this, the contribution of our approach and

DOI: 10.2312/3DOR/3DOR13/057-064

its novelty over existing solutions using a similar framework [MBO08], [LC11], and [MFK*10, SKVS13] is as follows: Original adaptation of the meshDOG detector to the case of face meshes (Sect. 2); Definition of the *multi-ring* Geometric Histogram as local descriptor at the keypoints (Sect. 3); 3D keypoints matching that also encompasses outliers removal using RANSAC (Sect. 4); An original investigation of keypoints detection addressing distribution and repeatability of keypoints (Sect. 5.1), and cohesiveness of local descriptors of neighboring keypoints (Sect. 5.2). A comprehensive evaluation and comparison of our work with state of the art solutions is also carried out on the *Bosphorus* dataset (Sect. 5.3). Finally, results are discussed and future research directions are outlined in Sect. 6.

2. 3D Keypoints of the Face

Among 3D keypoint detectors [BBB*11, STS12], the recently proposed meshDOG algorithm [ZBVH09] has proven its effectiveness in locating repeatable extrema on 3D meshes [BBB*11]. However, so far meshDOG keypoints were used for matching generic objects; In the following, we present the adaptation of the method for extracting keypoints of 3D face meshes.

Keypoints detection starts by defining and computing a scalar function f on a 3D mesh S. In principle, the function f can be any scalar function $f(v) : S \rightarrow R$, that for any vertex $v \in S$ returns a scalar value. In our case, we used the *mean* curvature at vertex v, computed according to [Pey09], as value of the function $f(v)$. Though such function is not completely intrinsic, and therefore not completely invariant to local isometric deformations, in practice the keypoints detected using mean curvature in our framework turned out to be more stable on 3D face data than keypoints obtained using Gaussian curvature. Similar results were reported in [STS12] for meshDOG keypoints, and in [MFK*10, SKVS13] for meshSIFT. Once the function f is computed for each vertex of the mesh, the keypoints detection proceeds in three subsequent steps. In the first step, a *scale-space* representation of the scalar function f is constructed. At each scale, the function f is convolved with a Gaussian kernel:

$$g_\sigma(x) = \frac{\exp(-x^2/2\sigma^2)}{\sigma\sqrt{2\pi}}, \qquad (1)$$

where σ is the standard deviation of the Gaussian (set equal to $\sigma = 2^{1/3} e_{avg}$ in our experiments, being e_{avg} the average edge length); and, at any vertex v_i, x is the distance between neighboring vertices and v_i, that is $\|\mathbf{v}_j - \mathbf{v}_i\|$.

The scale-space of f is built incrementally on N+1 levels, so that: $f_0 = f$, $f_1 = f_0 * g_\sigma$, $f_2 = f_1 * g_\sigma$, $\ldots, f_N = f_{N-1} * g_\sigma$. The N *Difference of Gaussian* (DOG) are then obtained by subtracting adjacent scales, e.g., $DOG_1 = f_1 - f_0$, $DOG_2 = f_2 - f_1$, $\ldots, DOG_N = f_N - f_{N-1}$. In so doing, it is relevant to note that the geometry of the face does not change, but the different scalar functions f_k and DOG_k defined on the mesh. Once the scale-space is computed, the extrema are selected as the maxima of the *DOG* across scales. In particular, a vertex is an extremum at a given scale k if its DOG_k value is the maximum with respect to the DOG_k values in the 1-ring neighbourhood at the same scale. The extrema of the scale space obtained at the previous step are then sorted according to their magnitude. Only the top 1% of the sorted vertices are retained as extrema in our setting. The last step aims to remove unstable extrema, by retaining only those with corner characteristics, according to the Hessian computed at each vertex v of the mesh [Low04]. The ratio between the largest λ_{max} and the lowest λ_{min} eigenvalues of the Hessian matrix is a good indication of a corner response, which is independent of the local coordinate frame (we used $\lambda_{max}/\lambda_{min} = 4$ as a minimum value of threshold responses).

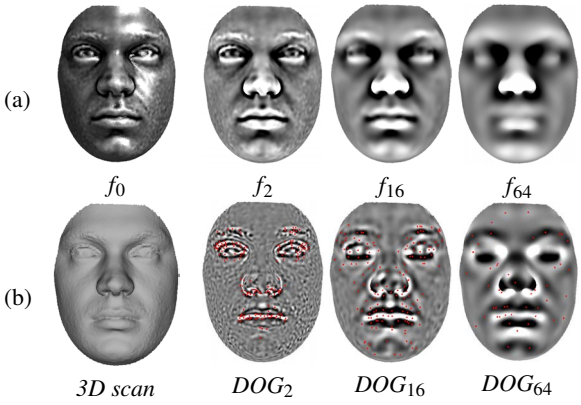


Figure 1: *(a) Values of f_k at different scales (f_0 is the mean curvature); (b) 3D face scan and DOG_k values at different scales with the 3D keypoints detected at that scale (in red).*

An example of the scale-space construction is reported in Fig. 1. In (a), f_k values at different scales (f_0 being the mean curvature) are shown for a sample face. In (b), gray levels are used to represent the DOG values at different scales (scales 2, 16, and 64 are reported). The derivation of multiple DOG scales allows the identification of more stable keypoints, which are typically located at highest scales, whereas keypoints detected in the first DOG scales are likely to be unstable and more affected by noise, as shown in Fig. 1(b). At the first level of the scale-space (see DOG2), the keypoints are mainly located in the mouth and eyes regions (quite unstable with expressions) and around the nose and the eyebrows (more stable regions under expression changes). As the scale increases, keypoints tend to be more distributed on the face (see for example DOG64). At these latter scales, some keypoints are located in the *forehead*, *cheekbone* and *chin*, with some keypoints close to the *pronasal* and *nasion* (regions of these keypoints are much less affected by expression variations). In the case keypoints are detected at multiple different scales only the keypoint occurring at the highest

scale is retained. Based on these observations, only the keypoints detected in the last 64 DOG scales (out of 96 total scales) are considered for local descriptor computation.

3. Multi-Ring Geometric Histogram

The geometric histogram (GH) is a local geometric descriptor proposed in [AFRW98] and employed in surface alignment and matching. Basically, it is a 2D accumulator, or frequency table, that counts the frequencies of two geometrical measurements, namely, the angle and the distance between pairs of facets in a given neighborhood of a keypoint. In the following, we propose a variation of the GH, which develops on the idea of constructing the GH descriptor at a given keypoint in an incremental way, by accounting for an ordered sequence of ring facets defined around the keypoint.

We used the Ordered Ring Facets (ORF) method to identify the facets of the mesh which are comprised in the neighborhood of a keypoint [WRK12]. In this approach, the neighbourhood around a central (root) facet t_c is constructed through a sequence of concentric rings of facets emanating from t_c. The facets are arranged circular-wise within each ring. The size of the neighbourhood is simply controlled by the number of rings. This mechanism allows an easy analysis of the GH variability, and thus of the local geometry evolution, as the size of the neighbourhood increases. When the triangular mesh is regular and the facets are nearly equilateral, the ORF rings form an approximation of iso-geodesic rings around the central facet t_c. The ORF construction has a linear complexity. Fig. 2 depicts examples of ORF's with increasing number of rings and their related GH's. In the experiments reported in Sect. 5, we obtained good results by using 8 ORF as neighborhood of the keypoints.

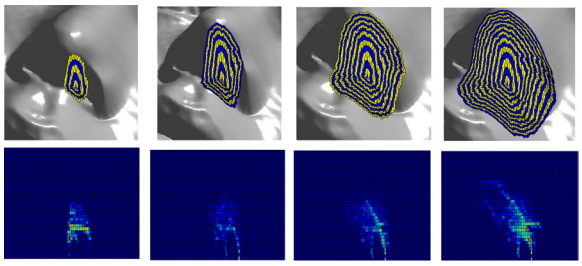


Figure 2: *ORF neighbourhoods with different sizes constructed at a facial keypoints, and their corresponding GHs.*

The Geometric Histogram definition starts by considering a triangular mesh approximation $\hat{S} = \{t_1, ..., t_M\}$ of an object surface. The discrete geometric distribution is constructed for each triangular facet t_i in a given mesh which describes its pairwise relationship with each of the other surrounding facets within a predefined neighbourhood. The range of the neighbourhood controls the degree to which the representation is a local description of shape. Here, we choose a neighbourhood range that encompasses the facets that share one or two vertices with the central triangular facet (Fig. 3(b)). The distribution is defined such that it encodes the surrounding shape geometry in a manner which is invariant to rigid transformations of the surface and which is stable in the presence of surface clutter and missing data.

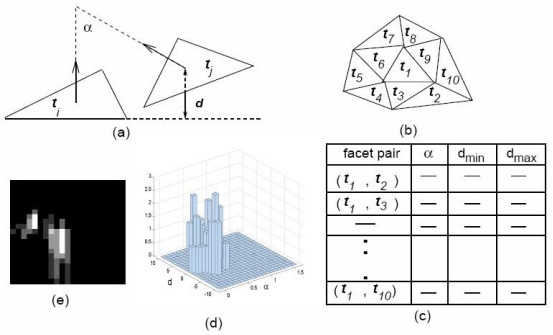


Figure 3: *GH computation: (a) Geometric measurements used to characterise the relationship between two facets t_i and t_j. (b) A facet t_1 and its neighbour facets. (c) For each pair (t_1, t_s) in (b), the angle α between the two facets' normals, the minimal and the maximal of the perpendicular distance from the plane of t_1 to the facet t_s are computed. (d) The pairs (α, d) derived from these measurements are entered in a 2D accumulator, thus obtaining a geometric distribution; (e) The geometric distribution can be visualized with a gray level mapping.*

Fig. 3(a) shows the measurements used to characterise the relationship between facet t_i and one of its neighbouring facets t_j. These measurements are the relative angle, α, between the facet normals, and the range of perpendicular algebraic distances, d, from the plane in which facet t_i lies to all points on the facet t_j. The range of perpendicular algebraic distances is defined by $[d_{min}, d_{max}]$, where d_{min} and d_{max} are the minimal and the maximal of the distance from the plane in which t_i lies to the facet t_j. These values are simply obtained by calculating the distances to three vertices of the facet t_j and then selecting the minimal and the maximal distance. Since the distance measurement is a range, a single value $d_{min} \leq d \leq d_{max}$ can be derived, based on the amplitude of the range $[d_{min}, d_{max}]$ and the resolution adopted for the distance quantization d. The group of pairs (α, d), extracted from the measurements related to a given facet and its neighbours (Fig. 3(b)-(c)), are entered to a 2D discrete frequency accumulator that encodes the perpendicular distance d and the angle α (Fig. 3(d)). This accumulator has size $N \times M$, where N and M are the number of bins in the axis α and d, respectively. The values of the accumulated matrix are also normalized so as to sum up to 1. The accumulator can be visualized in a 2D plotting using a gray level colormap (Fig. 3(e)), and stored in a matrix for subsequent processing. This representation only depends upon the surface shape

and not on the placement of facets over the surface. This independence on the placement of the facets is important as it guarantees the invariance of the correspondence with respect to geometric transformations. In our experiments, we considered the computation referred to the central facet t_c, using $N = 8$ and $M = 20$. With respect to the computation of the central GH, we introduced a variant which is related to the definition of Ordered Ring Facets (ORF) [WRK12]. In particular, in our approach, a GH is constructed on each of the rings that constitute the ORF of a keypoint: This means that the GH descriptor is actually given by a set of GH, constructed on the sequence of rings which surround the keypoint (an 8-rings ORF has been used). This improves the descriptiveness of GH by capturing information on how the local characteristic of the surface changes when the distance from the keypoint increases. This multi-ring structure is also exploited during the match. In particular, the normalized GH can be viewed as a probability density function, and thus can be adapted to probabilistic matching paradigms. To this end, the Bhattacharyya distance (d_B) is used as metric for evaluating the similarity between GHs at each ring. According to this, given two GHs in the form of 1D arrays of $K = N \times M$ elements, $A(l) = \{a_1, \ldots, a_K\}$ and $B(l) = \{b_1, \ldots, b_K\}$, their distance at ring-l is computed as:

$$d_B(A(l), B(l)) = \sqrt{1 - \sum_{k=1}^{K} \sqrt{(a_k \cdot b_k)}}. \qquad (2)$$

The overall distance between two multi-ring GH, computed on L rings is then obtained by accumulating the distances between the GHs at different rings.

4. Face Comparison

Given two face scans, their comparison is performed by matching the multi-ring GH descriptor of keypoints under the constraint that a consistent spatial transformation exists between inliers pairs of matching keypoints.

To this end, local shape descriptors at the keypoints detected in probe and gallery scans are compared using Eq. (2), so that for each keypoint in the probe, a candidate corresponding keypoint in the gallery is identified. In particular, a keypoint k_p in the probe is assigned to a keypoint k_g in the gallery, if they match each other among all keypoints, that is, if and only if k_p is closer to k_g than to any other keypoint in the gallery, and k_g is closer to k_p than to any other keypoint in the probe. In so doing, it is also required the second best match is significantly worse than the best one (i.e., a match is accepted if the ratio between the distance of the best and the second best matches is lower than 0.7).

As a result of this match a candidate set of keypoint correspondences is identified. The actual set of keypoint correspondences is then obtained as outcome of a constraint targeting the consistent spatial transformation between corresponding keypoints in the probe and gallery scans. The RANSAC algorithm [ZKM05, FB81] is used to identify outliers in the candidate set of keypoint correspondences. This involves generating transformation hypotheses using a minimal number of correspondences and then evaluating each hypothesis based on the number of inliers among all features under that hypothesis. In our case, we modeled the problem of establishing correspondences between sets of keypoints detected on two matching scans as that of identifying points in $\Re^3$ that are related via a *rotation*, *scaling* and *translation* transformation (RST transformation). According to this, at each iteration, the RANSAC algorithm validates sampled pairs of matching keypoints under the current RST hypothesis, updating at the same time the RST transformation according to the sampled points. In this way, corresponding keypoints whose RST transformation is different from the final RST hypothesis are regarded as outliers and are removed from the match. The number of keypoints matches is then used as similarity measure between scans. Matching examples are reported in Fig. 4(a)-(b), for scans of same and different subjects, respectively. In the figure, detected keypoints are highlighted (in blue) with a "+"; Corresponding keypoints based on descriptors matching are connected by green lines; Finally, the inlier matching passing the RANSAC algorithm, are shown with a red line connection. It can be observed as applying RANSAC, just the matches that show a coherent RST transformation among each other are retained, thus avoiding matches of keypoints that are located in different parts of the face in two scans.

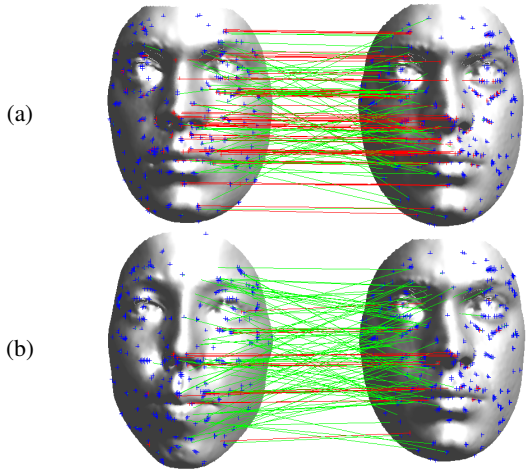


Figure 4: *Comparison of scans of same (a), and different subjects (b). All the detected keypoints are shown with "+". Lines indicate matching keypoints (in green), and inliers matching after RANSAC (in red).*

5. Experimental Results

The performance of the proposed approach have been evaluated in a comprehensive set of experiments, which include three parts:

1. The goal of the first session of experiments was to investigate the keypoints number, distribution and repeatability (see Sect. 5.1);
2. In the second session, we provide more insights on the relationship between the position of the keypoints and the corresponding descriptors (see Sect. 5.2);
3. Finally, the third session of experiments (see Sect. 5.3), reports the recognition performance of the proposed approach on the *Bosphorus* database. This dataset has been used in other works on 3D face recognition, thus permitting a direct comparison of our approach with state of the art solutions.

5.1. Keypoints Distribution and Repeatability

The idea of representing the face by extracting local descriptors from the neighbor of a set of 3D keypoints, relies on the assumption of *intra-subject* keypoints repeatability: Keypoints extracted from different facial scans of the same individual are expected to be located approximately in the same positions of the face. Since keypoints detection only depends on the geometry of the face surface through its mean curvature (see Sect. 2), these keypoints are not guaranteed to correspond to specific meaningful landmarks of the face. For the same reason, the detection of keypoints on two face scans of the same individual should yield to the identification of the same points of the face, unless the shape of the face is altered by major occlusions or non-neutral facial expressions.

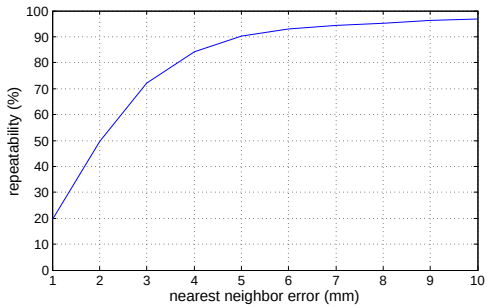


Figure 5: *Repeatability of keypoints.*

To test the repeatability of keypoints detection, we used a subset of the 3D scans of the *Binghantom University* 3D face database (BU-3DFE) [YWS*06]. We considered 80 subjects, each with 25 scans showing six different facial expressions at four gradation (from moderate to exaggerated), plus the neutral one (2000 scans in total). We followed the approach proposed in [MBO08], and measured the correspondence of the location of keypoints detected in two face scans by performing ICP registration. Accordingly, the 3D faces belonging to the same individual are automatically registered and the errors between the nearest neighbors of their keypoints (one from each face) are recorded. Figure 5 shows the results of our keypoint repeatability experiment, by reporting the cumulative rate of repeatability as a function of increasing values of the distance. The repeatability reaches a value of 90% for frontal faces with neutral and non-neutral expressions at a distance error of 5*mm* (with an average number of 360 keypoints detected per scan). We remark that these results, and those reported in the following about the number of detected keypoints, have been obtained by computing 96 DOG scales, and retaining the unique keypoints that are detected in the last 64 DOG scales (see also Sect. 2).

Table 1 also reports the number of keypoints detected on the face scans of the BU-3DFE. In particular, separate values are given for the *average*, *minimum* and *maximum* number of keypoints. No remarkable differences are observed for the number of keypoints detected on left or right probes. Non-neutral expressions have a small impact on the number of detected keypoints, which remains comparable to that obtained for frontal neutral scans (in some cases, an increase in the number of keypoints is observed).

dataset		*number of keypoints*		
name	scans	avg	min	max
BU-3DFE *neutral*	80	327	265	402
BU-3DFE *expressive*	1920	361	292	464
BU-3DFE *total*	2000	360	265	464

Table 1: *Number of detected keypoints per scan (average, min and max).*

From Tab. 1, it results that the number of detected keypoints is quite large. In fact, an important trait of a keypoint detector is the amount of repeatable keypoints it can provide to the subsequent modules of an application. Detecting a small number of keypoints can not be enough to apply geometrical verification or outliers removal steps, whereas too many may waste computational resources [STS12]. In the case of meshDOG, the number of detected keypoints is the result of the thresholds involved in the detection algorithm (see Sect. 2). Of course, making these thresholds more selective, the number of keypoints can be reduced. In our experiments, the number of keypoints reported in Tab. 1, represented a good compromise between computational cost and accuracy of recognition. A number of detected keypoints on 3D face scans of the order of hundreds is also reported for the keypoints detector defined by Mian et al. [MBO08], and for the meshSIFT detector [SKVS13]. For example, in the meshSIFT, an average number of 560 keypoints is reported, with a number of matching at rank-1 of about 97. The recent survey on 3D keypoint detectors [STS12], also reported that meshDOG tend to extract a high number of keypoints, that accumulate around areas characterized by high local curvature. These results seem to support our findings.

5.2. Keypoints Clustering

Observing the keypoints distribution across different facial scans it results that they are not fully dispersed but, on the

opposite, a considerable portion of them share quite close locations. This spatial clustering aspect of the keyppoint distribution is attractive as it has the potential of reducing the combinatorial number of matches between face keypoints, thus speeding-up the matching process. In our context, a cluster of keypoints S_{k_c} is defined as the group of keypoints that are within the spherical neighborhood of radius r of a keypoint k_c of the set, that is:

$$S_{k_c} = \{k_i : |k_c - k_i| < r,\ i = 1,\ldots,n_{k_c},\ n_{k_c} > 1\}. \quad (3)$$

So, according to Eq. (3), a set with a single keypoint cannot be considered as cluster, but a cardinality greater than one is required. The radius r is set to $r = \rho \cdot d$, where d is the mean of the edge length in the mesh and ρ is an integer. The computation of the clusters is performed with a neighborhood grouping procedure. The hypothesis we consider here is that a cluster of keypoints should produce similar or at least quite close descriptors, so that just the central keypoint k_c can be considered in the matching rather than all the keypoints in the cluster. The degree of compliance of a given cluster of keypoints with this assumption can then be adopted as a validity criterion on whether or not to accept that cluster in the matching. For instance, if a cluster obtained with a low r has quite disparate descriptors at its keypoints, it might result in conflicting matching. Following this intuition, we studied how such hypothesis sounds for the scans data of the BU-3DFE. To this end, we examined the behavior of the clusters, as the sphere radius increases, in terms of: *(i)* Variation of their number; *(ii)* Homogeneity with respect to the GH descriptor, that is the extent to which the local descriptors computed at keypoints in a cluster keep close as r increases (ideally, the descriptors should be identical at all the keypoints within a given cluster).

To verify the hypothesis above, an experiment has been conducted on the BU-3DFE facial scans. For each face scan, the following quantities have been computed (with ρ ranging from 1 to 5, and thus for increasing radius r):

- The ratio $\alpha = (N - (\eta + \chi))/N$, where N, η and χ are the numbers of keypoints, clusters, and single keypoints, respectively. The numerator represents the number of keypoints that would be used in the matching if a cluster is substituted by its central keypoint. So, the ratio α can be regarded as the amount of keypoints reduction due to keypoints clustering;
- The mean μ of the pair-wise distances between the GH descriptors in each cluster. Depending on the μ value, clusters are then divided into four groups of increasing μ, as reported in Tab. 2. The threshold 0.2 has been chosen upon a statistics in which we estimate the maximum distance between GH descriptors computed at the same locations for different scans of a same person. This subdivision reflects the homogeneity of the GH descriptors within a cluster: Clusters in the *group-1* are the most homogenous; At the other extreme, clusters of *group-4* exhibit the highest disparity and thus should be discarded from the match.

Group	μ
1	$\mu \leq 0.2$
2	$0.2 < \mu \leq 0.5$
3	$0.5 < \mu \leq 0.8$
4	$0.8 < \mu$

Table 2: *Groups of clusters based on the mean μ of the pair-wise distances between GH descriptors in each cluster.*

Results of this analysis are summarized in Fig. 6 and Fig. 7. In Fig. 6, the mean and standard deviation of the ratio α are plotted against the radii of the spherical neighborhood (values of ρ from 1 to 5 have been used). We notice that for $\rho = 1$, that is for clusters practically confined within a facet and its adjacent neighbors, the average percentage rate of keypoints reduction is around 30%. This value increases up to more than 50% for $\rho = 2$. This is very encouraging if we assume that the local shape is not expected to change too much in a such reduced neighborhood.

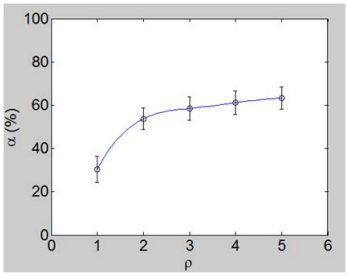
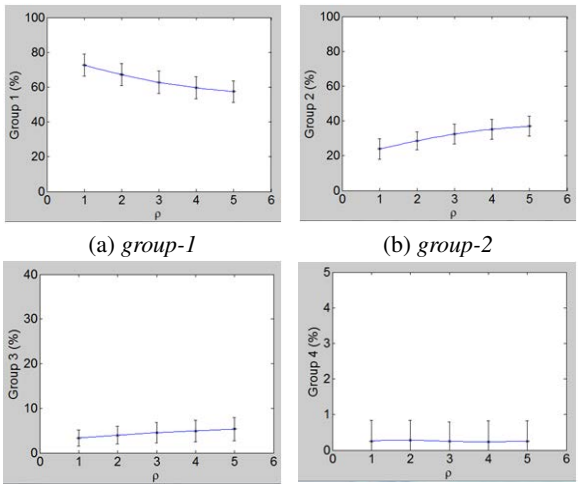


Figure 6: *Mean ratio α, in percentage, with respect to the spherical neighborhood size (parameterized by ρ). The vertical bars also report the standard deviation at each ρ.*

Figures 7(a)-(d) report the percentage of clusters belonging to the four groups listed in Tab. 2, with respect to spherical neighborhood size ρ. We notice that the average number of clusters having a mean of pair-wise distances μ between GH descriptors remains above 60% up to the third spherical neighborhood (*group-3*). This is encouraging as it means that the homogeneity, and thus the trustworthiness of the clusters, for a considerable number of clusters, is not compromised when the extent of the neighborhood increases. On the other hand, we notice that the number of non-homogenous groups remains less than 30% up to the second neighborhood size, especially for *group-2*. Extremely non-homogenous clusters (*group-3* and *group-4*) have very low proportions, yet they are present across all the neighborhood sizes. These can be viewed as outliers or instable clusters (*group-4* in particular),

for which the descriptors show large disparity. Therefore, including in the match the clusters belonging to *group-3* and *group-4* can jeopardize the accuracy of the results.



(a) *group-1* (b) *group-2*

(c) *group-3* (d) *group-4*

Figure 7: *Percentage of clusters belonging to groups 1, 2, 3, and 4 (reported from (a) to (d), respectively), for increasing spherical neighborhood size (ρ).*

This statistical analysis provides insights on the clustering aspects of the keypoints, evidencing the potential of exploiting this characteristic for reducing the combinatorial space in keypoints matching. The final goal of this is improving the plausibility of the match by eliminating unreliable keypoints. Based on the results above, the keypoints can be ranked in the following way: *(i)* Keypoints belonging to clusters of *group-1*; *(ii)* Individual keypoints; *(iii)* Keypoints belonging to clusters of *group-2*.

5.3. Recognition Results

Recognition experiments have been performed on the *Bosphorus* database. This dataset has been collected at the Boğaziçi University and made available during 2008 [SAD*08]. It consists of the 3D facial scans and images of 105 subjects acquired under different expressions and various poses and occlusion conditions. Occlusions are given by hair, eyeglasses or predefined hand gestures covering one eye or the mouth. Many of the male subjects have also beard and moustache. The majority of the subjects are Caucasian aged between 25 and 35, with a total of 60 males and 45 females. The database includes a total of 4,666 face scans, with the subjects categorized as follows:

- 34 subjects with up to 31 scans per subject (including 10 expressions, 13 poses, 4 occlusions and 4 neutral);
- 71 subjects with up to 54 different face scans. Each scan is intended to cover one pose and/or one expression type, and most of the subjects have only one neutral face, though some of them have two. Totally, there are 34 expressions, 13 poses, 4 occlusions and one or two neutral faces. In this set, 29 subjects are professional actors/actresses, which provide more realistic and pronounced expressions.

This variability of the scans in terms of subjects' pose, expressions and occlusions, motivated us in selecting this dataset for our experiments. In addition, this dataset has been used by many state of the art solutions for 3D face recognition, thus permitting a direct comparison with our approach.

In our experiments, we used the same protocol proposed in [LC11] and [SKVS13], thus allowing a direct comparison of the results. For each subject, the first neutral scan was included in the gallery, whereas the probe scans have been organized in different classes as reported in Tab. 3 (the number of probes per class is also indicated). The first class groups probes according to their facial expression, distinguishing between neutral probes and expressive probes, plus some not-classified probes. Probes where subjects exhibit Face Action Units are accounted in the second class, by considering scans with Lower Face Action Unit (LFAU), Upper Face Action Unit (UFAU), and Combined Action Unit (CAU). Finally, the last class reports probes with missing parts due to Yaw Rotation (YR), Pitch Rotation (PR) and Cross Rotation (CR), plus probes with Occlusions (O). For methods in [LC11] and [SKVS13], the rank-1 RR is reported as appear in the respective publications.

	rank-1 RR		
Probes (#)	Li et al.	Smeets et al.	**this work**
Neutral (194)	100.0%	-	97.9%
Anger (71)	88.7%	-	85.9%
Disgust (69)	76.8%	-	81.2%
Fear (70)	92.9%	-	90.0%
Happy (106)	95.3%	-	92.5%
Sad (66)	95.5%	-	93.9%
Surprise (71)	98.6%	-	91.5%
other (18)	-	-	100.0%
LFAU (1549)	97.2%	-	96.5%
UFAU (432)	99.1%	-	98.4%
CAU (169)	98.8%	-	95.6%
YR (735)	78.0%	-	81.6%
PR (419)	98.8%	-	98.3%
CR (211)	94.3%	-	93.4%
O (381)	99.2%	-	93.2%
All (4561)	94.1%	93.7%	93.4%

Table 3: *Rank-1 RR for different probe classes. Our approach is compared with [LC11] and [SKVS13].*

Results show that our approach has overall performance which are very close to state of the art solutions, and for some categories are even better. In particular, our solution performs particularly well in recognizing scans with missing parts (see for example the YR category). More in detail, our

approach achieves an accuracy of 45.7% on scans with $\pm 90^{\circ}$ left/right yaw rotations. Result for these scans is not directly reported in [SKVS13].

6. Discussion and Future Work

In this work, we have proposed an original approach to 3D face recognition based on the idea of capturing local information of the face surface around a set of 3D keypoints detected at multiple scales according to differential surface measurements. The approach, first detects 3D keypoints of the face mesh, then local descriptors are extracted at each keypoint and used to find keypoint correspondences during the match. To improve the accuracy of keypoints correspondences, a spatial constraint is introduced using the RANSAC algorithm. In summary, the proposed approach presents some new solutions in the perspective to make 3D face recognition deployable in real non-cooperative contexts of use: The approach is fully-3D, and does not require any costly normalization or alignment; The meshDOG keypoints combined with the multi-ring GH, provide a good compromise between robustness to expression changes and missing parts of the face; The inclusion of a statistical technique for outliers removal largely improves the recognition results.

The comparative evaluation carried out on the Bosphorus database showed that our solution can compete with state of the art works evidencing a clear advantage in the case of probes with large missing parts.

In perspective, the proposed framework could be easily adapted to include texture information of the face, so as to define a multi-modal solution that combines together, in the function used for meshDOG detection, 2D and 3D data.

References

[AFRW98] ASHBROOK A., FISHER R., ROBERTSON C., WERGHI N.: Finding surface correspondance for object recognition and registration using pairwise geometric histograms. In *Proc. European Conference on Computer Vision* (Friburg, Germany, June 1998), pp. 674–686. 3

[BBB*11] BOYER E., BRONSTEIN A. M., BRONSTEIN M. M., BUSTOS B., DAROM T., HORAUD R., HOTZ I., KELLER Y., KEUSTERMANS J., KOVNATSKY A., LITMAN R., REININGHAUS J., SIPIRAN I., SMEETS D., SUETENS P., VANDERMEULEN D., ZAHARESCU A., ZOBEL V.: Shrec 2011: robust feature detection and description benchmark. In *Proc. of Eurographics Workshop on 3D Object Retrieval (3DOR 2011)* (Llandudno, UK, April 2011). 2

[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

[BDP10] 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

[FB81] FISCHLER M. A., BOLLES R. C.: Random sample consensus. *Communications of the ACM 24*, 6 (June 1981), 381–395. 4

[LC11] LI H., CHEN L.: SHREC'11 track: Salient points. In *Eurographics Workshop on 3D Object Retrieval* (Llandudno, UK, April 2011), pp. 89–95. 2, 7

[Low04] LOWE D.: Distinctive image features from scale-invariant key points. *International Journal of Computer Vision 60*, 2 (Nov. 2004), 91–110. 2

[MBO08] MIAN A. S., BENNAMOUN M., OWENS R.: Keypoint detection and local feature matching for textured 3D face recognition. *International Journal of Computer Vision 79*, 1 (Aug. 2008), 1–12. 2, 5

[MFK*10] MAES C., FABRY T., KEUSTERMANS J., SMEETS D., SUETENS P., VANDERMEULEN D.: Feature detection on 3D face surfaces for pose normalisation and recognition. In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (Washington D.C., USA, Sept. 2010), pp. 1–6. 2

[Pey09] PEYRE G.: Toolbox graph. In *MATLAB Central File Exchange Select* (2009). 2

[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

[PPTK11] PASSALIS G., PERAKIS P., THEOHARIS T., KAKADIARIS I. A.: Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 10 (October 2011), 1938–1951. 1

[SAD*08] SAVRAN A., ALYÜZ N., DIBEKLIOĞLU H., ÇELIKTUTAN O., GÖ B., SANKUR B., AKARUN L.: Bosphorus database for 3D face analysis. In *Proc. First COST 2101 Workshop on Biometrics and Identity Management* (May 2008). 7

[SKVS13] SMEETS D., KEUSTERMANS J., VANDERMEULEN D., SUETENS P.: meshSIFT: Local surface features for 3D face recognition under expression variations and partial data. *Computer Vision and Image Understanding 117*, 2 (February 2013), 158–169. 2, 5, 7, 8

[STS12] SALTI S., TOMBARI F., STEFANO L. D.: Performance evaluation of 3D keypoint detectors. *International Journal of Computer Vision to appear* (2012). 2, 5

[WLT10] WANG Y., LIU J., TANG X.: Robust 3D face recognition by local shape difference boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 12 (Oct. 2010), 1858–1870. 1

[WRK12] WERGHI N., RAHAYEM M., KJELLANDER J.: An ordered topological representation of 3D triangular mesh facial surface: Concept and applications,. *EURASIP Journal on Advances in Signal Processing to appear* (July 2012). 3, 4

[YWS*06] YIN L., WEI X., SUN Y., WANG J., ROSATO M.: A 3D facial expression database for facial behavior research. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition* (Southampton, UK, Apr. 2006), pp. 211–216. 5

[ZBVH09] ZAHARESCU A., BOYER E., VARANASI K., HORAUD R.: Surface feature detection and description with applications to mesh matching. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition* (Miami Beach, FL, June 2009), pp. 373–380. 2

[ZKM05] ZULIANI M., KENNEY C. S., MANJUNATH B. S.: The multiransac algorithm and its application to detect planar homographies. In *IEEE International Conference on Image Processing* (Sep 2005). 4