

Automatic Keypoint Detection on 3D Faces Using a Dictionary of Local Shapes

Clement Creusot
Dept. of Computer Science
University of York
York, U.K.
Email: creusot@cs.york.ac.uk

Nick Pears
Dept. of Computer Science
University of York
York, U.K.
Email: nep@cs.york.ac.uk

Jim Austin
Dept. of Computer Science
University of York
York, U.K.
Email: austin@cs.york.ac.uk

Abstract—Keypoints on 3D surfaces are points that can be extracted repeatably over a wide range of 3D imaging conditions. They are used in many 3D shape processing applications; for example, to establish a set of initial correspondences across a pair of surfaces to be matched. Typically, keypoints are extracted using extremal values of a function over the 3D surface, such as the descriptor map for Gaussian curvature. That approach works well for salient points, such as the nose tip, but can not be used with other less pronounced local shapes. In this paper, we present an automatic method to detect keypoints on 3D faces, where these keypoints are locally similar to a set of previously learnt shapes, constituting a ‘local shape dictionary’. The local shapes are learnt at a set of 14 manually-placed landmark positions on the human face. Local shapes are characterised by a set of 10 shape descriptors computed over a range of scales. For each landmark, the proportion of face meshes that have an associated keypoint detection is used as a performance indicator. Repeatability of the extracted keypoints is measured across the FRGC v2 database.

Keywords-3D face; keypoint detection; local 3D descriptors;

I. INTRODUCTION

Finding correspondences between 3D surfaces is a crucial step in many 3D shape processing applications, such as surface landmarking [2], surface registration [3], 3D object retrieval [4] and 3D face recognition [5]. The difficulty in solving this correspondence problem is dependent on the type of shapes that need to be matched. If we define a ‘keypoint’ as an *unlabelled* point of interest on the 3D surface, then we would like a set of keypoints that have local shapes sufficiently different from their neighbour’s local shapes, so that they can be extracted repeatedly under a wide range of 3D imaging conditions. In addition, the keypoint’s local shapes should ideally be different from each other, to facilitate the keypoint matching process across the two surfaces. Here we are focusing on finding a method of keypoint detection that works across a highly varied set of soft, organic shapes, such as 3D faces. This is a difficult challenge, particularly if the goal is to solve the problem with real world data for which changes in pose and possible occlusions should be taken into account.

In this paper, we present a method to detect unlabelled 3D keypoints (interest points) whose local descriptors are statistically similar to a dictionary of learnt landmark shapes.

The learnt shape descriptors are those computed over a region of interest, at a given scale, around a manually-placed facial landmark. The ideal output of our algorithm is a set of keypoints that are localised close to all of the visible landmark locations, with a limited number of additional detections. In practice, we expect to miss some landmarks due to spurious local data. To determine keypoint positions, a likelihood map is computed over every vertex of a facial mesh. The selected keypoints are the points reaching local maxima over this map (see figure 1). This paper discusses the best way to compute this likelihood map from a given set of local descriptors. In this discussion, we describe how to select features (descriptors) and their appropriate scales for the detection of a specific landmark. Note that, although a keypoint is generated when a 3D surface point has a local shape similar to some labelled landmark, the index label of that landmark (eg. 5, ‘nose tip’) is not retained and is not an output of the system presented here.

II. RELATED WORK

Keypoint Detection on Faces: Keypoints on 3D faces are typically used in recognition applications. In such cases, the desired keypoints are repeatable for a given identity but differ from one individual to another. In [14], Mian et al. use a coarse curvature-related descriptor to detect those keypoints, while in [15] and [16], they are computed using Scale-Invariant Feature Transform (SIFT[17]) on 2D depth-maps. In our case, the term ‘keypoint’ is justified as we try to detect unlabelled, repeatable points of interest. However, our approach differs, as the scope for the targeted repeatability is larger. Our technique should be able to detect repeatable point of interest across the population and not only for several captures of the same individual.

Keypoint Detection on Other Objects: Computing keypoints in order to determine correspondences is useful for all kind of objects. In [4], keypoints are computed using a coarse curvature descriptor to localise objects in scenes with occlusions. A similar technique using Harris Operator is proposed in [18]. In [19], two kinds of curvature-related descriptor (Shape Index and Willmore Energy) are combined to detect the keypoints on archaeological objects in order to detect regions matching a given pattern. This last paper

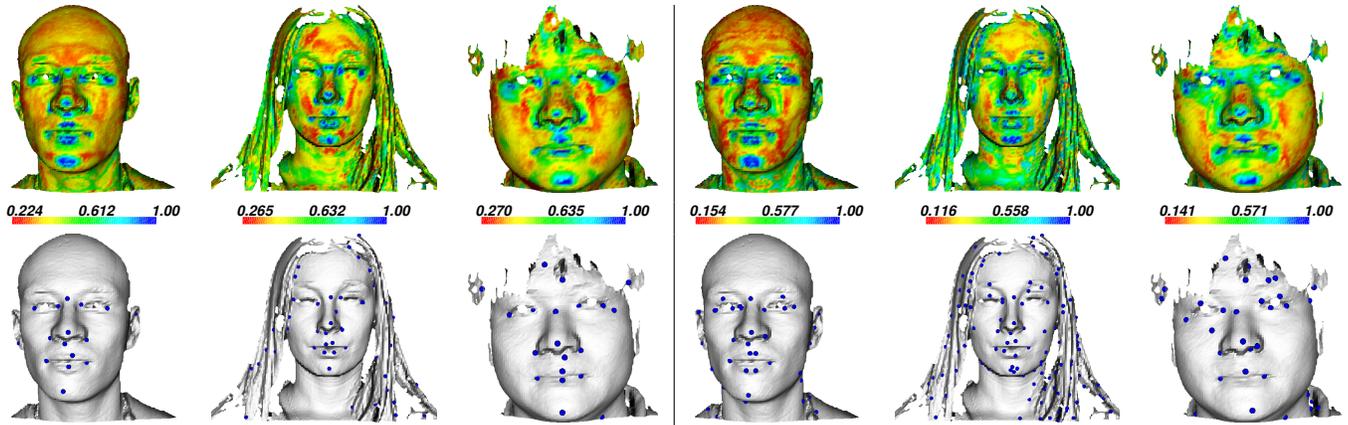


Figure 1. Examples of extracted keypoints on faces from the FRGC v2 database[1] using our system configuration 1 (left) and 2 (right). The first row shows the final likelihood map where blue vertices represent the highest scores. In the second row, the associated detected keypoints are shown.

is one of the rare cases where more than one descriptor is used for keypoint candidate selection. Besides, when several descriptors are used, combining them is usually done using fixed coefficients. In this paper, a framework is presented to determine what descriptors should be used and how they can be combined for the particular problem of 3D facial keypoint detection. However, the general learning framework can be extended to other classes of object.

Landmark Localisation on Faces: Much of the research on 3D facial landmark localisation is linked to recognition applications for which correspondences need to be made in preprocessing. All of those techniques need at least one point (usually the tip of the nose, e.g. [6]) and many of them require at least three points, which enables the computation of a rigid registration. The most common triplet is composed of the tip of the nose and the two inner corner of the eyes because of their extremal curvature.

The selection of landmark candidates is done using different kinds of descriptor map:

- Curvature and/or volume-related maps: assuming the tip of the nose or the inner-corners of the eyes are the most salient points ([7] [8] [9] [6][10] [11] [12]).
- Directional projection extrema: assuming the tip of the nose is the closest point to the 3D camera, or more generally the most extreme point along a particular direction ([7][9]).
- Slice and/or profile point detection: detection on curves generated by intersecting the mesh with planes ([13][5][11])

In order to be able to deal with large pose variations and occlusion, no assumptions should be made on what landmarks will be present in the facial mesh. Therefore, a robust system should be able to detect a greater number of landmarks than the inner corner of the eyes and the tip of the nose. Unfortunately, other points on the face that may be used as landmarks are far less salient than these three.

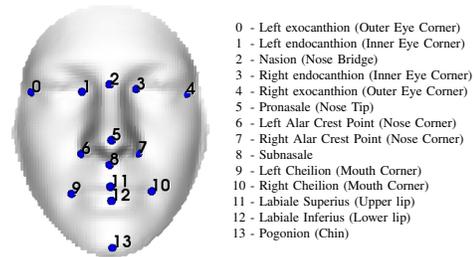


Figure 3. Position of the 14 landmarks used as “shape of interest” for the training part of the system.

Detecting less salient points using a single local descriptor would be extremely difficult. In this paper, we show that a linear combination of simple descriptors can be used to reliably detect these points.

III. MAIN PROCESS

Figure 2 explains the main workflow used in our experiments. The framework is composed of an offline process and an online process. The offline part is used to teach the system what is considered to be a shape of interest. For that purpose, 14 landmarks over the training set are used to define a dictionary of local shapes from which statistical distributions of descriptors are learnt. The 14 landmarks used are shown in figure 3.

The online part is composed of the following steps:

- D descriptors are computed for all vertices
- For each learnt local shape (14):
 - D matching score maps are computed by projecting the descriptor values of each vertex against the associated learnt distributions of the target landmark. The scores generated are between 0 and 1.
 - A linear combination of those D maps is produced using landmark-specific weights learnt in the offline part. The result is one mixed descriptor map

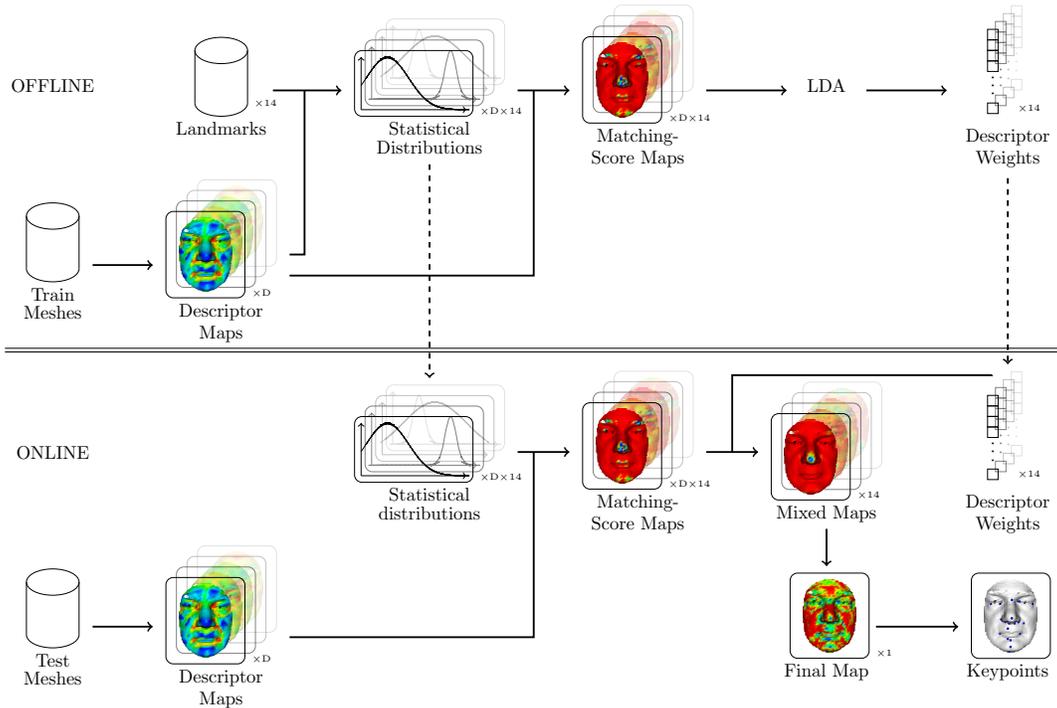


Figure 2. Offline process and online process. The offline process provides the online system with parameters of the descriptor distributions for a set of landmarks and the weights to linearly combine individual matching score maps. In the online process, D descriptor maps are computed from the input mesh, each value is matched against the 14 learnt descriptor distributions to get score maps with values between 0 and 1. For each landmark, the D descriptor score maps are combined using the learnt weights. The 14 normalised mixed descriptor maps are combined into a single final map, using the maximal value (of 14 values) at each vertex. The output keypoints are the local maxima detected on this final map that are above some given threshold (0.85).

per landmark. The mixed map is then normalised over the mesh to ensure the matching scores of different local shapes have the same impact on the final decision.

- All the mixed descriptor maps are combined into one final map, by using the maximum value (over all 14 landmark dictionary shapes) for each vertex.
- The keypoints are defined as the strong local maxima on this final map. We use an empirical threshold of 0.85 to discard weak candidates.

The variation of the descriptors at known landmark locations is learnt in the offline part by fitting an idealised distribution (most of the time a Gaussian) to the training data. The weights to combine matching score maps are defined using Linear Discriminant Analysis (LDA) over a population of neighbouring and non-neighbouring vertices, relative to the relevant landmark. The population of neighbouring vertices is defined as those at a distance less than 5 mm from the specified landmark on all facial meshes in the training set. The population of non-neighbouring vertices is constituted of those between 15 and 45 mm from the same landmark (see figure 4 for the upper-lip landmark). Those empirical radii have been selected to get two populations of manageable size on faces.

LDA applied to these two classes (neighbouring and non-neighbouring) returns, as its first component, the direction in D -dimensional descriptor space that best separates the two sets. Combining weak features for consistent detection

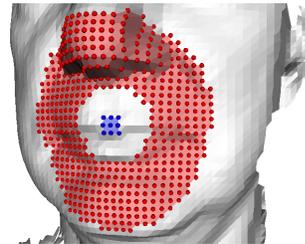


Figure 4. Example of class generation for LDA. Showing neighbouring (blue) and non-neighbouring (red) vertices for the upper-lip landmark on one face of the training set.

can of course be done with more complex and non-linear tools such as kernel methods and boosting techniques. We decided to first investigate the use of a simple LDA as a proof of concept.

IV. EXPERIMENTS

Databases: For the experiments, we use the FRGC v2 database [1] which contains 4950 faces of 568 subjects with some variations in facial expression.

The database is split into two disjoint sets ‘train’ and ‘test’. The training set is composed of 200 neutral faces of 200 different subjects. The test set is composed of the remaining 4750 faces.

Hand-placed landmarks are available on all faces to train the automatic keypoint detector, and on all faces of the test

set to check the results. The landmarks used are a mixture of contributions from [12] and [10].

Mesh Resolution: Resolution is reduced in each facial mesh by replacing each block of 4×4 raw 3D data points with its average. Meshes are constructed from those points to facilitate the computation of neighbourhoods and normals. Two triangle faces are defined for every group of four adjacent vertices \mathcal{V} .

Preprocessing: No cropping, spike removal, hole-filling or smoothing is performed on the data. Firstly, because smoothing and the size of the neighbourhood used for the local descriptors are intimately linked. Secondly, because it might be hard to compare results with other researchers who use different preprocessing techniques. Spike and hole removal would probably improve the results of our system. Cropping the facial region (automatically or otherwise) would certainly reduce the time of computation.

Local Descriptors: In this paper, 10 common local shape descriptors are evaluated with varying neighbourhood size and bin size (if relevant).

- Spin image histograms (SIH): computed using the method described in [20], the radius of the spin image is divided in 9 bins, and the height in 18 bins (9 above, 9 beneath the 3D point).
- Spherical histograms (SH): similar to spin images but with spherical bins defined by two consecutive radii. 9 bins are used.
- Local volume (Vol): computed as the sum of the tetrahedron signed volume in a given neighbourhood. The top of all the tetrahedra is the barycentre of the peripheral points of the neighbourhood, while the bases are the triangles inside the local area.
- Distance to local plane (DLP): computed as the distance between the current point and the plane that best fits its local neighbourhood [10].

Several products and by-products of curvature:

- Principal curvatures (k_1) and (k_2): computed using Goldfeather’s adjacent-normal cubic approximation method [21].
- Gaussian curvature (K): $k_1 k_2$
- Mean curvature (H): $\frac{k_1 + k_2}{2}$
- Shape Index (SI): $\frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_1 + k_2}{k_1 - k_2}$
- Log Curvedness (LC): $\frac{2}{\pi} \log \sqrt{\frac{k_1^2 + k_2^2}{2}}$

Of course, there are many local surface descriptors and this list is not comprehensive. We tried to choose pose invariant descriptors that are mature and very common in the literature, and we accept that many of them may be highly correlated (this aspect can be dealt with within the LDA machinery). Note that two descriptors have been deliberately omitted from this list: the local Discrete Willmore Energy [22] because it seems too sensitive to the triangulation of the mesh, and the x/y Wavelet Transform (as used in [9]) because it is not pose invariant.

Descriptor Sizes: For a given mesh resolution, the value of $k_1, k_2, H, K, SI, LC, Vol$ and DLP will only depend on the scale of the local neighbourhood over which they are computed. The histograms (SIH, SH) will only depend on the size of the bins, as the number of bins is fixed.

To determine the best scales and bin sizes to employ, five different sizes of neighbourhood radius (5, 15, 30, 45 and 60 mm) and four different bin sizes (2.5, 5.0, 7.5 and 10 mm) are evaluated .

Descriptor Distributions: In the offline part of the algorithm, distributions of the descriptors collected from training data are well approximated with Gaussians (unimodal, non-mixture), except for the Shape Index (SI) which is bounded between 0 and 1. Here, we notice that an inverse Gaussian distribution fits the training data better.

A. Evaluation

A compromise between different objectives has to be found for this particular problem. Firstly, a small number of points should be detected, as returning the whole set of vertices wouldn’t be useful. Secondly, as landmarks have been used as shapes of interest in the training, the system should be able to detect keypoints at those very locations. Thirdly, keypoints, by definition, should have a high intra-class (same subject’s face) repeatability. We decided to fix the first objective to a constant, by selecting at most 1% of the total number of vertices. This represents a few tens of points per face mesh, which have on average a few thousand vertices.

Descriptor and Size Selection: Using all the descriptors at all the scales and resolutions (Configuration 1) would be very time consuming and unpractical for real-world applications. The most costly parts of the computation are the principal curvatures k_1 and k_2 over different scales, the computation of the histograms ($\mathcal{O}(n^2)$) and the computation of the matching scores, as the number of maps is proportional to both the number of descriptors and the number of landmarks. In table I, the weights returned by LDA for each landmark are given. The higher the value is, the more discriminative the descriptor is. From this table, the best descriptor and the best size of neighbourhood can be observed. They are represented by dark blue cells in the column corresponding to a given landmark. There is no common set of descriptors that works for all the landmarks. For example, not very salient points like the upper and lower lips (11 and 12) will prefer histogram descriptors to curvature descriptors, the corners of the mouth (9 and 10) will prefer to use the Shape Index with a small neighbourhood (15 mm), while the tip of the nose which is salient will best be detected with bigger neighbourhood (60 mm). Figure II shows the mean results per neighbourhood size. It can be seen that the best neighbourhood size for our data resolution is around 15 mm. For the spherical histogram (SH) and the spin image histogram (SIH), the best bin dimension is between 5 and

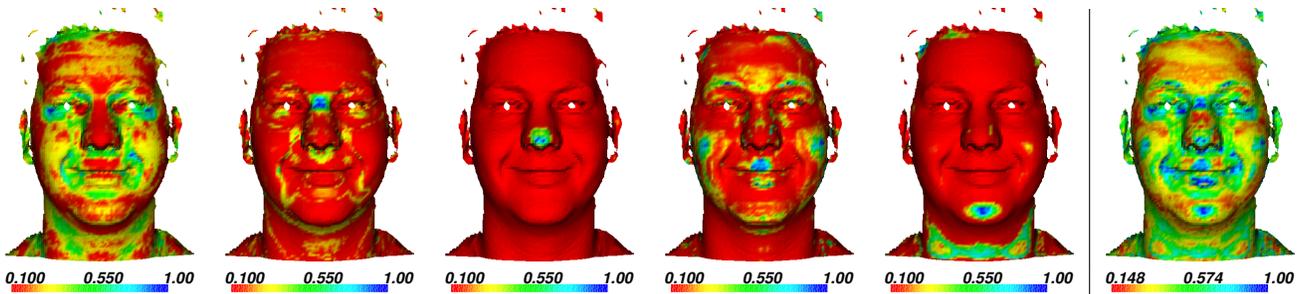


Figure 5. Example of normalised mixed maps for landmarks 0 (left outer eye corner), 2 (nose bridge), 5 (nose tip), 11 (upper lip) and 13 (chin). The final map, computed for the same subject using the 14 mixed descriptor maps, is shown on the right.

7.5 mm. Regarding the type of descriptor, it can be seen in table III that the distance to local plane (DLP) often gives bad support to the distinction between neighbouring and non-neighbouring vertices when looking at relatively flat landmark local shape. The spin-image histogram (SIH), the Shape Index (SI) and the local volume are usually the more supportive descriptors for this set of landmarks.

From these observations, a new experiment (Configuration 2) is conducted using only one neighbourhood size (15 mm) for all scalar descriptors, and one bin size for the histogram descriptors (5 mm). In total, this comprises 10 descriptor maps and 140 matching maps (far less than the original 672). The computation of the online process takes on average 7.75 seconds on a laptop processor Intel Core I3 M350 (the mean number of vertices in our meshes being 6392). The more computational part being the neighbourhood and curvature computation (2.04 seconds) and the computation and projection of the histograms for each vertex (3.85 seconds). Using fewer descriptors on smaller meshes (using automatic face cropping) may be a good way to reduce computation time.

V. RESULTS

Landmark Retrieval: To evaluate the rate at which keypoints are localised near defined landmarks, the percentage of face meshes in which a keypoint is present in a sphere of radius R from the landmark is computed. As there is no clear definition about what distance error should be considered for a match, this percentage is computed for an increasing acceptance radius ranging from $R = 2.5$ mm to $R = 25$ mm. Results for configuration 1 and 2 are given in figure 6. With configuration 2, at 10 mm, the nose tip is present in the detected keypoints 99.47% of the time, the left and right inner eye corners in 90.50% and 92.56% of the cases. It can be seen that this method will not succeed in detecting all potential landmarks in all facial meshes. However, we aim to provide an initialisation for further face processing that doesn't rely on a small, specific set (eg. triplet) of target landmarks. In figure 7, it can be seen that the mean number of correctly selected landmark candidates is around 12 already for a radius of 10 mm. This summarises

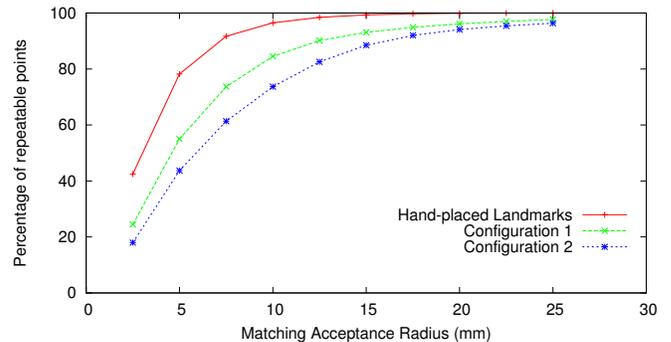


Figure 8. Percentage of points repeatable after registration at an increasing matching acceptance radius. The measure of the human hand-placed landmarks is used as a baseline.

our approach: by detecting more points in a looser way, we decrease the single-point-of-failure risk that candidate landmark selection systems often have.

Repeatability: The intra-class repeatability is measured on the FRGC v2 database for which registration of the faces to a common pose has been computed using the Iterative Closest Point method (ICP [23]) on the cropped meshes. The transformation matrices describing those registrations are available on the first author's webpage. For each pair of faces of the same subject, the two sets of keypoints are cropped and registered to a common frame of reference. The proportion of points in the smallest set that have a counterpart in the second set at a distance R is computed. The repeatability using configuration 1 and 2 is given in figure 8 and compared with the repeatability of the hand-placed landmarks. It can be seen that at 10 mm the proportion of repeatable points is around 75% (configuration 2) on average. An aspect of our system is that it cannot discriminate between shapes of interest that are linked to identity from the ones linked to change in expression or other variations. We think that the best way to use those detected points will be to label them (for example with [2]) to get a more sparse and consistent set of points.

Table I

WEIGHTS CORRESPONDING TO THE FIRST PRINCIPAL DIRECTION OF THE LDA FOR EACH OF THE 14 LANDMARKS DESCRIBED IN FIGURE 3 (0 TO 13). DARK BLUE CELL REPRESENT THE HIGHEST CONTRIBUTION PER COLUMN. RED CELLS REPRESENT NEGATIVE CONTRIBUTIONS.

Size	Desc.	00	01	02	03	04	05	06	07	08	09	10	11	12	13	All
2.5 mm	SIH	.03	.08	.00	.10	.02	.04	.00	.04	.00	.03	.03	.01	.03	.11	.04
	SH	.00	.02	.00	.02	.02	.19	.00	.01	.02	.00	.01	.00	.00	.05	.01
5 mm	SIH	.04	.02	.04	.05	.04	.02	.04	.00	.09	.00	.00	.28	.44	.29	.09
	SH	.01	.00	.00	.01	.01	.00	.02	.00	.00	.04	.02	.01	.01	.37	.03
7.5 mm	SIH	.02	.03	.01	.02	.01	.00	.04	.01	.11	.02	.00	.40	.00	.09	.04
	SH	.05	.00	.01	.01	.04	.15	.01	.01	.00	.01	.00	.00	.07	.23	.04
10 mm	SIH	.02	.00	.08	.00	.00	.15	.02	.04	.05	.05	.05	.05	.20	.08	.04
	SH	.08	.01	.01	.00	.05	.12	.03	.03	.00	.01	.00	.00	.00	.00	.00
5 mm	k ₁	.00	.02	.02	.04	.00	.00	.05	.05	.01	.06	.10	.03	.09	.00	.02
	k ₂	.02	.04	.01	.01	.01	.00	.03	.03	.01	.02	.02	.01	.03	.00	.00
	H	.07	.02	.02	.00	.07	.00	.05	.00	.06	.01	.00	.05	.04	.01	.01
	K	.03	.06	.02	.05	.03	.00	.02	.02	.03	.03	.01	.01	.00	.00	.01
	SI	.01	.03	.01	.04	.01	.01	.02	.02	.01	.00	.04	.00	.00	.01	.00
	LC	.04	.00	.02	.00	.04	.00	.01	.06	.01	.09	.04	.01	.01	.00	.02
	Vol	.04	.00	.00	.01	.02	.00	.25	.04	.00	.02	.02	.00	.04	.00	.03
	DLP	.10	.10	.01	.02	.12	.00	.00	.03	.03	.02	.06	.07	.11	.14	.00
15 mm	k ₁	.01	.07	.05	.06	.00	.14	.00	.02	.04	.06	.06	.00	.17	.04	.02
	k ₂	.03	.39	.03	.14	.05	.00	.17	.22	.07	.07	.07	.02	.07	.01	.06
	H	.04	.00	.04	.02	.06	.01	.05	.22	.04	.02	.00	.00	.01	.04	.02
	K	.05	.07	.17	.04	.04	.00	.03	.02	.23	.07	.06	.01	.18	.00	.01
	SI	.08	.11	.08	.28	.08	.00	.06	.06	.00	.51	.57	.03	.03	.00	.13
	LC	.02	.03	.03	.02	.04	.00	.00	.03	.03	.04	.04	.00	.08	.00	.01
	Vol	.13	.01	.02	.12	.12	.16	.07	.11	.01	.11	.09	.03	.09	.02	.08
	DLP	.08	.00	.00	.04	.08	.00	.06	.13	.00	.08	.07	.00	.14	.00	.01
30 mm	k ₁	.00	.01	.02	.03	.01	.01	.00	.02	.04	.00	.01	.03	.05	.18	.01
	k ₂	.00	.02	.04	.08	.01	.01	.02	.02	.00	.03	.05	.05	.12	.01	.00
	H	.15	.09	.01	.10	.12	.00	.09	.11	.04	.05	.04	.05	.12	.08	.02
	K	.12	.00	.00	.01	.17	.03	.00	.01	.02	.02	.00	.02	.01	.20	.01
	SI	.17	.05	.08	.00	.15	.00	.05	.08	.02	.03	.02	.01	.04	.06	.04
	LC	.04	.01	.05	.00	.03	.01	.03	.02	.04	.06	.09	.08	.07	.00	.01
	Vol	.10	.03	.01	.08	.11	.00	.05	.05	.06	.06	.04	.01	.06	.02	.04
	DLP	.07	.41	.03	.33	.11	.05	.03	.05	.02	.04	.00	.01	.08	.01	.05
45 mm	k ₁	.05	.04	.01	.06	.05	.01	.00	.02	.00	.05	.05	.04	.01	.01	.00
	k ₂	.02	.00	.00	.02	.02	.00	.07	.05	.05	.01	.03	.08	.01	.01	.01
	H	.02	.01	.00	.03	.02	.01	.01	.02	.04	.12	.10	.01	.19	.00	.03
	K	.00	.00	.00	.00	.01	.00	.05	.02	.03	.01	.01	.06	.08	.01	.01
	SI	.04	.03	.04	.06	.02	.00	.01	.05	.00	.00	.00	.00	.02	.00	.02
	LC	.01	.03	.02	.03	.00	.00	.16	.09	.00	.15	.09	.10	.18	.02	.06
	Vol	.00	.01	.00	.01	.00	.00	.04	.06	.03	.02	.02	.03	.05	.01	.02
	DLP	.03	.14	.03	.02	.06	.24	.01	.01	.03	.15	.13	.03	.04	.00	.00
60 mm	k ₁	.08	.02	.04	.00	.08	.00	.05	.04	.00	.00	.00	.00	.04	.02	.01
	k ₂	.00	.15	.01	.09	.00	.00	.04	.07	.02	.03	.03	.01	.03	.04	.03
	H	.01	.03	.00	.00	.02	.00	.00	.02	.00	.02	.04	.00	.04	.00	.00
	K	.01	.05	.01	.04	.01	.00	.03	.00	.01	.00	.03	.00	.04	.02	.00
	SI	.09	.02	.04	.06	.08	.00	.01	.00	.02	.00	.00	.02	.05	.04	.01
	LC	.02	.02	.03	.01	.03	.00	.04	.07	.02	.03	.05	.02	.03	.04	.01
	Vol	.06	.00	.00	.00	.05	.02	.01	.03	.03	.05	.07	.01	.07	.00	.03
	DLP	.02	.03	.01	.02	.04	.33	.03	.00	.00	.05	.07	.01	.11	.01	.00

Table II

SUMMED VALUE PER NEIGHBOURHOOD SIZE OF THE WEIGHTS RETURNED BY LDA (SEE TABLE I). THE DESCRIPTIVENESS OF THE DESCRIPTORS PEAKS AROUND 15 MM FOR THE SELECTED LANDMARKS.

Size	00	01	02	03	04	05	06	07	08	09	10	11	12	13	All
5 mm	.01	.10	.10	.07	.01	.01	.21	.01	.04	.10	.12	.07	.07	.02	.01
15 mm	.28	.41	.45	.54	.27	.33	.34	.51	.45	.53	.60	.09	.08	.03	.34
30 mm	.22	.38	.18	.20	.24	.12	.18	.14	.21	.01	.06	.08	.21	.12	.16
45 mm	.21	.01	.05	.10	.22	.26	.07	.12	.00	.15	.15	.10	.20	.01	.08
60 mm	.03	.14	.00	.01	.04	.35	.01	.07	.08	.04	.06	.04	.01	.05	.06

Table III
SUMMED VALUE PER DESCRIPTOR OF THE WEIGHTS RETURNED BY LDA (SEE TABLE I).

	00	01	02	03	04	05	06	07	08	09	10	11	12	13	All
DLP	.07	.19	.00	.35	.13	.14	.18	.24	.04	.00	.08	.08	.27	.02	.02
H	.11	.05	.08	.11	.12	.00	.10	.38	.10	.18	.09	.01	.39	.14	.01
K	.03	.17	.20	.12	.12	.04	.02	.05	.21	.08	.10	.07	.04	.19	.01
SH	.15	.01	.04	.03	.12	.20	.03	.03	.02	.04	.01	.01	.06	.55	.09
SI	.23	.14	.18	.23	.19	.02	.16	.17	.03	.48	.60	.06	.09	.10	.19
SIH	.07	.15	.14	.18	.09	.23	.12	.08	.26	.10	.09	.73	.67	.22	.22
k ₁	.15	.03	.16	.01	.16	.14	.11	.07	.05	.06	.10	.05	.18	.12	.07
k ₂	.03	.56	.00	.14	.05	.00	.08	.25	.03	.00	.02	.03	.03	.01	.09
LC	.15	.11	.15	.06	.16	.00	.13	.13	.08	.10	.05	.14	.32	.02	.10
Vol	.35	.07	.02	.20	.32	.19	.44	.31	.13	.29	.27	.08	.32	.01	.21

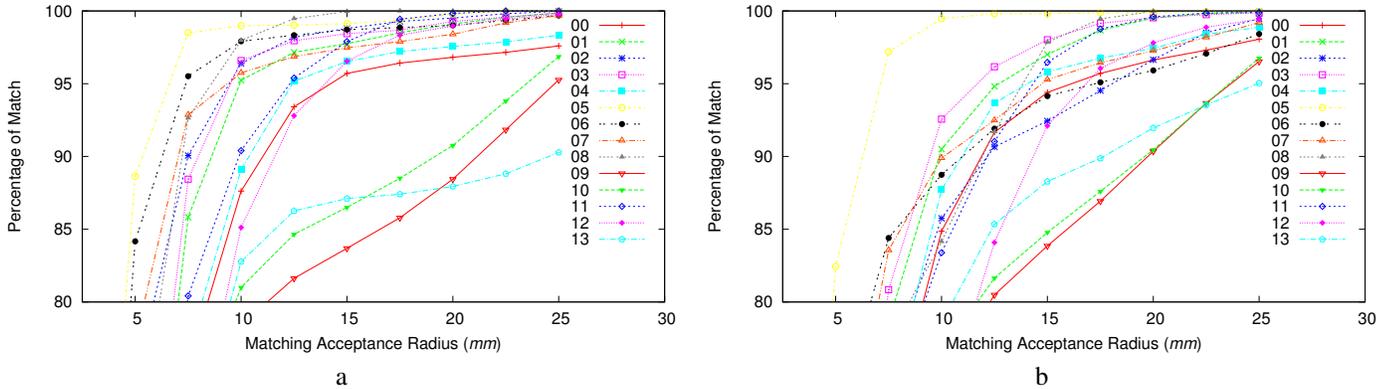


Figure 6. Matching percentage per landmark (0-13) with an increasing matching acceptance radius on the FRGC v2 test set. (a) using all descriptors (Configuration 1), (b) using a subset of descriptors (Configuration 2).

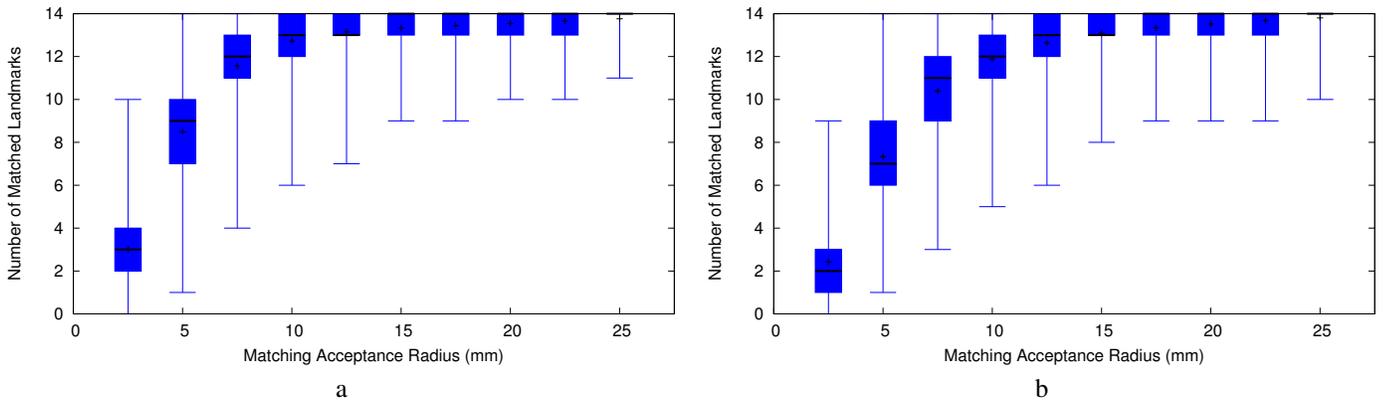


Figure 7. Number of matching landmark per file on the test part of the FRGC v2 database. (a) using all descriptors (Configuration 1), (b) using a subset of descriptors (Configuration 2).

VI. CONCLUSION

A simple method has been proposed to deal with the problem of combining descriptors to detect unlabelled points of interest (keypoints) on 3D objects. This method gives interesting results on 3D faces, but presents some limitations in its current state, like the fact that only linear combinations are used for the descriptors or that it is computationally expensive if too many descriptors are considered. However, a good point of our method is that it doesn't assume that the detected points should have an extremal value over a descriptor map. Instead, it assumes that the matching score

of this descriptor against a learnt distribution should be maximal. Furthermore, this technique is very generic, as the set of descriptors, the sizes of the neighbourhoods and the dictionary of local shapes can be changed easily. Our method could therefore be used on other types of 3D objects without any modification.

In this paper, we have also studied the behaviour of different local descriptors competing with each other at different scales. It provides us better indications of what descriptor should be used with which parameters to detect each of the 14 common landmarks used as shapes of interest

in our experiments (see table I). Further work will look at computing correspondences between the detected keypoints both with and without using label retrieval.

REFERENCES

- [1] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 947–954, June 2005.
- [2] C. Creusot, N. Pears, and J. Austin, "3D face landmark labelling," in *Proceedings of the ACM workshop on 3D object retrieval*, ser. 3DOR '10. Firenze, Italy: ACM, 2010, pp. 27–32.
- [3] T. J. Hutton, B. F. Buxton, and P. Hammond, "Dense surface point distribution models of the human face," in *MMBIA '01: Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2001, pp. 153–160.
- [4] A. Mian, M. Bennamoun, and R. Owens, "On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes," *International Journal of Computer Vision*, vol. 89, no. 2, pp. 348–361, Sep. 2010.
- [5] A. S. Mian, M. Bennamoun, and R. A. Owens, "Automatic 3d face detection, normalization and recognition," in *3DPVT*, 2006, pp. 735–742.
- [6] N. Pears, T. Heseltine, and M. Romero, "From 3d point clouds to pose-normalised depth maps," *International Journal of Computer Vision*, vol. 89, no. 2, pp. 152–176, Sep. 2010.
- [7] K. I. Chang, K. Bowyer, and P. Flynn, "Multiple nose region matching for 3d face recognition under varying facial expression," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 10, pp. 1695–1700, Oct. 2006.
- [8] D. Colbry, G. Stockman, and A. Jain, "Detection of anchor points for 3d face verification," in *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 25-25 2005, pp. 118 –118.
- [9] J. D'Hose, J. Colineau, C. Bichon, and B. Dorizzi, "Precise localization of landmarks on 3d faces using gabor wavelets," *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pp. 1–6, Sept. 2007.
- [10] M. Romero and N. Pears, "Landmark localisation in 3d face data," in *Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on*, Sept. 2009, pp. 73–78.
- [11] M. Segundo, C. Queirolo, O. Bellon, and L. Silva, "Automatic 3d facial segmentation and landmark detection," *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pp. 431–436, Sept. 2007.
- [12] P. Szeptycki, M. Ardabilian, and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," in *International Conference on Biometrics: Theory, Applications and Systems*, Sep. 2009, pp. 32–37.
- [13] T. Faltemier, K. Bowyer, and P. Flynn, "Rotated profile signatures for robust 3d feature detection," in *Automatic Face Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on*, 17-19 2008, pp. 1–7.
- [14] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3d face recognition," *Int. J. Comput. Vision*, vol. 79, no. 1, pp. 1–12, 2008.
- [15] A. D. B. S. Berretti and P. Pala., "Recognition of 3d faces with missing parts based on profile networks," in *1st ACM Workshop on 3D Object Retrieval (ACM 3DOR'10)*, Firenze, Italy, October 2010, pp. 81–86.
- [16] M. Mayo and E. Zhang, "3d face recognition using multiview keypoint matching," in *Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on*, Sept. 2009, pp. 290 –295.
- [17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [18] I. Sipiran and B. Bustos, "A robust 3d interest points detector based on harris operator," in *Eurographics 2010 Workshop on 3D Object Retrieval (3DOR'10)*. Eurographics Association, 2010, pp. 7–14.
- [19] A. Itskovich and A. Tal, "Surface partial matching & application to archaeology," *Computers & Graphics*, vol. In Press, Accepted Manuscript, pp. –, 2010.
- [20] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 1, pp. 433 – 449, May 1999.
- [21] J. Goldfeather and V. Interrante, "A novel cubic-order algorithm for approximating principal direction vectors," *ACM Trans. Graph.*, vol. 23, no. 1, pp. 45–63, 2004.
- [22] E. I. Bobenko, P. Schrder, and T. B. Caltech, "Discrete willmore flow," in *In Eurographics Symposium on Geometry Processing*, 2005, pp. 101–110.
- [23] P. Besl and N. McKay, "A method for registration of 3d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.