Comparing 3D descriptors for local search of craniofacial landmarks

Federico M. Sukno[1,2], John L. Waddington[2], and Paul F. Whelan[1]

[1] Centre for Image Processing & Analysis, Dublin City University, Dublin 9, Ireland
[2] Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland

Abstract. This paper presents a comparison of local descriptors for a set of 26 craniofacial landmarks annotated on 144 scans acquired in the context of clinical research. We focus on the accuracy of the different descriptors on a per-landmark basis when constrained to a local search. For most descriptors, we find that the curves of expected error against the search radius have a plateau that can be used to characterize their performance, both in terms of accuracy and maximum usable range for the local search. Six histograms-based descriptors were evaluated: three describing distances and three describing orientations. No descriptor dominated over the rest and the best accuracy per landmark was strongly distributed among 3 of the 6 algorithms evaluated. Ordering the descriptors by average error (over all landmarks) did not coincide with the ordering by most frequently selected, indicating that a comparison of descriptors based on their global behavior might be misleading when targeting facial landmarks.

1 Introduction

We address the comparison of local geometry descriptors for highly accurate localization of 3D facial landmarks, in the context of craniofacial research [1, 2]. In contrast to applications on facial biometrics, which emphasize robustness to challenging conditions [3], medical applications tend to have a greater focus on the highly accurate localization of landmarks, as they constitute the basis for the analysis, often aimed at detecting quite small shape differences. Depending on the author, localization and repeatability errors are considered clinically relevant when they exceed 1 mm [4] or 2 mm [5]. Acquisition conditions are therefore carefully controlled to minimize holes and other artifacts.

In this context, we aim to evaluate the performance of a variety of local descriptors for the purpose of facial landmark localization. There are two key elements that motivate an interest in such a study: 1) the popularity of local descriptors for the detection of 3D facial landmarks, as opposed to global cues, and 2) the fact that previous comparisons of 3D descriptors have been focused on the detection and reproducibility of generic keypoints, rather than on the accuracy of specific ones (e.g. landmarks).

The first element derives from the difficulty in detecting individual points globally on the human face. In general, facial landmarks can be distinguished from their neighboring points based on some geometric properties. For example, the nose tip can be detected as a curvature *peak* or *cap*, while eye corners are *pits* or *cups*. However, the chin tip is also a *peak* and the mouth corners are also *pits* [6, 7]. Unfortunately, as we will show, these coincidences are not exclusive of simple descriptors, such as curvature, but they persist for more elaborate ones. Thus, landmark descriptors do not work satisfactorily on a global basis and one usually combines local search with higher level constraints based on the spatial relationships between sets of landmarks [8–10]. Hence, it is of interest to assess how different descriptors perform when constrained to a local search.

On the other hand, prior work on the evaluation of 3D descriptors has not focused on the accuracy of specific points. It is common practice to report the detection and reproducibility of *keypoints* defined generically as those that are most distinct from the rest for the descriptor of choice. The number of detected keypoints and their reproducibility are used as measures of quality [11]. Accuracy is secondary, indirectly addressed by means of the acceptance radius (the maximum distance at which two different points are considered to match).

Bronstein et al. [12] put more emphasis on accuracy by evaluating the identification of dense correspondences with different keypoint detection algorithms. They report the average geodesic distances to the true correspondences; this is possible because they constrain the matching pairs to different instances of the same object, after some synthetic transformations.

Closer to the present study are the evaluations reported by Romero & Pears [14] and Creusot et al. [13]. However, in both cases the evaluation is performed based on the joint search of all points under the global constraints provided by a graph-matching scheme. Furthermore, descriptors are not compared individually but, rather, combined together. While Creusot et al. provide the resulting weights for the descriptor combinations using Linear Discriminant Analysis (targeting 14 landmarks on a set of 200 scans), these do not constitute an optimal criteria to assess the performance of each descriptor individually.

In this work we present a comparison of local descriptors for a set of 26 facial landmarks relevant in the context of craniofacial dysmorphology [1]. The evaluation is performed on a per-landmark basis with a focus on the accuracy that can be achieved when the descriptors are constrained to search in a local neighborhood of the targeted landmark. We empirically show that, for a good descriptor, the curves of overall accuracy against the search radius have a plateau that is indicative of the descriptor's accuracy and usable range (Section 2). The set of descriptors that are evaluated are detailed in Section 3, results are presented in Section 4 and conclusive remarks are provided in Section 5.

2 Analysis of local accuracy

We start from a set of annotated facial surfaces in 3D, organized in meshes $\mathcal{M}$ described by sets of vertices and triangles. We will indicate that a vertex $\mathbf{v}$

belongs to the vertices of mesh $\mathcal{M}$ simply by $\mathbf{v} \in \mathcal{M}$. Evaluation will be based on the (Euclidean) distance from a given vertex $\mathbf{v}$ to the ground truth (manual location of the considered landmark) and will be denoted by $d(\mathbf{v})$.

The model-instance of a descriptor for a given landmark will be referred to as the *template*; for example, the average of the spin images [16] of the nose tip (for a set of *training* scans) is a descriptor template for the nose tip using the spin image descriptor. The value resulting from the evaluation of a descriptor *template* at a given vertex $\mathbf{v}$ will be denoted as descriptor *score* $s(\mathbf{v})$.

We define the expected local accuracy $\overline{e}_L(r_S)$ over the distances from $\mathbf{v}_i^{max}$ (the vertices that obtain the maximum score in each mesh) to the ground truth position of the targeted landmark, evaluated only on a neighborhood composed of vertices whose distances do not exceed the search radius parameter r_S,

$$\overline{e}_L(r_S) = E[d(\mathbf{v}_{i,r_S}^{max})] \tag{1}$$

$$\mathbf{v}_{i,r_S}^{max} = \{\mathbf{v} \in \mathcal{M}_i \,|\, d(\mathbf{v}) \leq r_S \wedge \forall \mathbf{w} \neq \mathbf{v}, d(\mathbf{w}) \leq r_S, \mathbf{w} \in \mathcal{M}_i \,:\, s(\mathbf{v}) \geq s(\mathbf{w})\} \tag{2}$$

where $E[x]$ is the expected value of x. That is, given a target landmark, for each mesh $\mathcal{M}_i$ we consider a neighborhood of radius r_S around the ground truth position of the landmark and select $\mathbf{v}_i^{max}$ as the vertex with the maximum score in this neighborhood. We are interested in the expected distance of these maximum-score vertices to the targeted landmark.

It is evident that $\overline{e}_L(r_S) \leq r_S$. However, a useful descriptor should also beat chance (i.e. random selection), which would be equivalent to a uniform distribution of the scores over the neighborhood (i.e. a fully uninformative descriptor):

$$\overline{e}_L^{rand}(r_S) = \int_{A(r_S)} r\, P(r)\, dA = \int_{A(r_S)} \frac{r}{A(r_S)}\, dA$$

$$A(r) = \pi r^2, \quad dA = 2\pi r\, dr \quad \Rightarrow \quad \overline{e}_L^{rand}(r_S) = \int_0^{r_S} \frac{2\pi r^2}{\pi r_S^2} dr = \frac{2}{3} r_S \tag{3}$$

Hence, we require that $\overline{e}_L(r_S) < \frac{2}{3} r_S$. Fig. 1 shows three examples of $\overline{e}_L(r_S)$, selected to illustrate the different behaviors observed in the landmarks used for this study:

- In the first example, $\overline{e}_L(r_S)$ initially increases with r_S until reaching a flat region or *plateau*. This means that, except for relatively small r_S, the descriptor produces, on average, a maximum score consistently at the same distance from the target.
- In the second example, $\overline{e}_L(r_S)$ behaves similarly to the first example up to a certain r_S, after which there is a sudden increase, which typically reaches a second plateau.
- In the third example, $\overline{e}_L(r_S)$ does not show any plateau (at least for the range of interest). Although it is below the theoretical limit of $\frac{2}{3} r_S$, its value constantly increases.

The descriptor from the first example is the most useful one, because it could be used for global search; however the second one is the most frequent. The reason for this is the presence of highly similar points from the viewpoint of the descriptor that is used. For example, both inner eye corners are evidently similar to each other (*twin* points), hence when targeting one of them we will find an increase in $\overline{e}_L(r_S)$ at the average distance between inner eye corners. Any descriptor will show such an increase of the error due to twin points but there can also be strong similarities between different landmarks (e.g. this is typical between the mouth and nose corners).

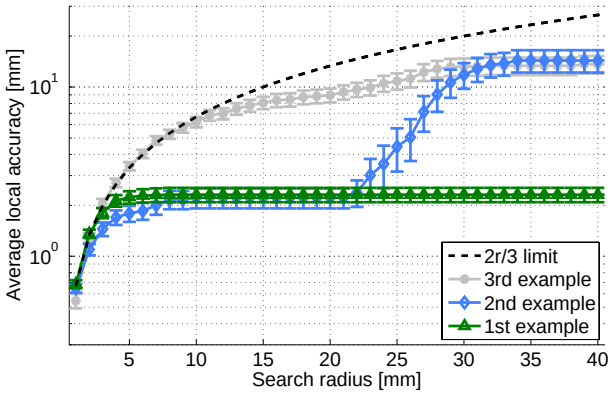

Fig. 1. Expected local accuracy, $\overline{e}_L(r_S)$, for some facial landmarks. The theoretical value for random choice ($\frac{2r_S}{3}$) is also provided for reference.

The third example is the least useful one, because it indicates somewhat erratic behavior: the descriptor finds confusing points roughly at any distance, indicating that the targeted landmark is not distinctive enough.

Thus, the curves of expected local accuracy can be very informative about the performance of a descriptor for a given landmark, allowing us to compare between different alternatives. Nonetheless, proceeding this way would require one plot per landmark, each containing several curves (one per descriptor). A more practical solution is to focus on the analysis of the first plateau, which is the most important part of the curve. This can be done with just three numbers: the value of $\overline{e}_L(r_S)$ and the plateau limits, in terms of r_S.

Two important considerations when identifying the plateau are the stability of $\overline{e}_L(r_S)$ and the range of interest. Due to the presence of outliers, we used the median to estimate $\overline{e}_L(r_S)$. Regarding the range of interest, it is evident from the second example that we should focus our analysis before the transition from the first to the second plateau (around 25 mm in this case), as it indicates the presence of a strong source of false positives, possibly due to a twin point. However, sometimes there is no plateau before this sudden increase, and it is

useful to analyze the curve generated by the difference to the theoretical limit (random choice), which we might regard as the accuracy gain, $\overline{G}_L(r_S)$:

$$\overline{G}_L(r_S) = \frac{2}{3} r_S - \overline{e}_L(r_S) \tag{4}$$

In general, the accuracy gain is a monotonically increasing curve. However, in the presence of a sudden increase of $\overline{e}_L(r_S)$, the accuracy gain will drop and we will find a local maximum. Therefore, the search for the plateau should be constrained up to the first local maximum of $\overline{G}_L(r_S)$. Note that this criterion applies only to curves like the one in example 2, as the curves in examples 1 and 3 will not produce local maxima on $\overline{G}_L(r_S)$.

3 Evaluated descriptors

In this section we briefly review the descriptors that were evaluated, including the default parameters that were used in each case. We choose three descriptors based on histograms of distances and three based on signatures of orientations (histograms of relative orientations of the normal vectors).

The descriptors are computed for each vertex $\mathbf{v}$, whose normal vector is $\mathbf{n}_v$, considering a neighborhood $\mathcal{N}_v = \{\mathbf{w} \,|\, \|\mathbf{w} - \mathbf{v}\| \leq r_N\}$, namely all points within a radius r_N. Except for the spin image approach, the implementations used in this paper are based on the Point Cloud Library [15].

Spin Images (SI) [16]: This descriptor is computed as a bi-dimensional histogram of distances. One axis encodes the unsigned distance to the normal vector and the other encodes the signed distance projected in the direction of the normal vector. That is, the histogram is generated from the following pairs:

$$\big(\alpha, \beta\big) = \Big(\sqrt{\|\mathbf{w} - \mathbf{v}\|^2 - (\mathbf{n}_v \cdot (\mathbf{w} - \mathbf{v}))^2}\,,\, \mathbf{n}_v \cdot (\mathbf{w} - \mathbf{v})\Big) \tag{5}$$

By default, the histograms contain 15×15 bins and the contribution of each point is calculated with bilinear interpolation. While this descriptor has been proposed more than a decade ago, it is still very widespread and is often used as an indicator of baseline performance.

3D Shape Contexts (3DSC) [17]: This descriptor is based on a 3D-histogram computed on a spherical support region centered at the interest point, $\mathbf{v}$, and with its North pole oriented with the normal, $\mathbf{n}_v$. The default structure has 11 elevation bins and 12 azimuth bins, both uniformly spaced, and 15 radial bins logarithmically spaced so that more importance is assigned to shape changes that are closer to the interest point. Similar to spin images, these descriptors are based on distances, as the value for each bin is based on the number of points that fall within its boundaries. The contribution of each point is weighted by 1) the inverse of its local density (to account for uneven sampling) and 2) the

inverse of the cube root of the bin volume, due to the large difference between bin sizes, especially along radius and elevation.

As the spherical support region is defined based only on $\mathbf{v}$ and $\mathbf{n}_v$, there is an ambiguity on the azimuth origin. This is dealt with by calculating as many descriptors per point as the number of azimuth bins, covering all possible shifts. The computation of multiple descriptors is done for the model (i.e. during training), so that during matching only one descriptor is computed and matched to multiple descriptors by choosing the one that yields the highest score.

Unique Shape Context (USC) [18]: This descriptor is analogous to the 3DSC but without the ambiguity in the azimuth direction, thanks to the definition of a local reference frame consisting of 3 unit vectors that replace the orientation of the North pole with the normal and are computed by a distance-weighted eigen-decomposition, followed by a sign disambiguation step [19].

The use of a local reference frame reduces the computational complexity during point matching and might also improve accuracy by reducing false positives due to spurious similarities that arise from inconvenient azimuth rotations (i.e. those that make the descriptors of different points become too similar). Nonetheless, errors or instabilities in the computation of the reference frame (e.g. due to noise) might impair the performance of this descriptor.

Signature of Histograms of OrienTations (SHOT) [19]: This descriptor is based on a histogram of orientations, rather than distances as for the three previously presented. A spherical domain is defined based on the local reference frame as described above for USC, with a coarse and isotropic grid of 8 azimuth, 2 elevation and 2 radial divisions. For each of these grid divisions, an 11-bin histogram is defined encoding the cosine of the angle between the normals of points within the grid division and the reference point, $\mathbf{v}$.

The use of the cosine, divided in equally spaced bins, results in a non-uniform division of the angular space, favoring directions nearly perpendicular to $\mathbf{n}_v$. Tombari et al. argue that such directions are more informative than those close to $\mathbf{n}_v$, which are more likely to appear in nearly-planar regions (e.g. due to noise). Quadrilinear interpolation is used to construct the local histograms to avoid boundary effects and the whole descriptor is normalized to sum the unit for robustness with respect to sampling density.

Point Feature Histograms (PFH) [20]: This descriptor is based on histograms of 3 angles and, optionally, one distance (not used in this paper). Given the points in the considered neighborhood $\mathbf{w} \in \mathcal{N}_v$, all possible pairs of points are analyzed. For each pair $(\mathbf{w}_i, \mathbf{w}_j), i \neq j$ with normals $(\mathbf{n}_i, \mathbf{n}_j), i \neq j$, the following angles are computed:

$$\alpha = \big((\mathbf{w}_j - \mathbf{w}_i) \times \mathbf{n}_i\big) \cdot \mathbf{n}_j, \qquad \phi = \frac{\mathbf{n}_i \cdot (\mathbf{w}_j - \mathbf{w}_i)}{\|\mathbf{w}_j - \mathbf{w}_i\|}$$

$$\theta = \arctan \frac{\big(\mathbf{n}_i \times \big((\mathbf{w}_j - \mathbf{w}_i) \times \mathbf{n}_i\big)\big) \cdot \mathbf{n}_j}{\mathbf{n}_i \cdot \mathbf{n}_j} \tag{6}$$

The resulting angles are used to construct a three-dimensional histogram (5 bins for each angle were used, yielding a descriptor of length 125). Note that, due to the evaluation of all possible pairs within $\mathcal{N}_v$, the computational complexity for this descriptor is much higher than all other ones evaluated here.

Fast Point Feature Histograms (FPFH) [21]: This is the fast variant of PFH, and is computed in two steps. Firstly, a Simplified Point Feature Histogram (SPFH) is constructed for each point, as described for PFH but considering only the relations between the reference point and each of the neighbors. Later on, this estimation is refined to obtain the final descriptor by weighting the simplified histograms by the inverse of their Euclidean distance to the reference point. The authors also point out the sparseness of the resulting three-dimensional histogram and propose, instead, to compute separate histograms for each of the three angles and simply concatenate them (11 bins per angle are used, resulting in a descriptor of length 33).

4 Performance comparison

4.1 Data

Our test dataset consisted of 144 facial scans acquired by means of a hand-held laser scanner[3]. This type of scanner allows acquisition of a three dimensional surface by smoothly sweeping a scanning wand over an object, in a manner similar to spray painting. The whole facial surface was acquired, up to (and including) the ears. Special care was taken to avoid occlusions due to facial hair. There is some heterogeneity regarding the extent to which neck and shoulders were included.

A unique surface is reconstructed by combining the different sweeps, which allows coverage of multiple viewpoints. Thus, the complete facial surface can be obtained irrespective of the head pose and possible self-occlusions. This is an important advantage compared to single-view scanners used in other databases.

The dataset contains exclusively healthy volunteers acting as controls in the context of craniofacial dysmorphology research. The mesh resolution varies between 1.2 and 2.4 mm and, on average, there are 24.2 thousand vertices per mesh. Each scan was annotated with 26 anatomical landmarks, in accordance with definitions in [22] (based on [23]), as indicated in Fig. 2.

4.2 Results

We computed the expected local accuracy curves, as defined in Section 2, for each descriptor-landmark pair, varying the search radius r_S from 1 to 200 mm.

[3] Cobra Wand 192 (FastSCAN^{TM}, Colchester, VT, USA).

Table 1. Expected local accuracy for neighborhood radius $r_N = 30$ mm. If a plateau is found, its value and limits are indicated, otherwise (n.p - no plateau) only the limit based on the first peak of $\overline{G}_L$ is indicated. For each landmark (rows), the best descriptor is highlighted in boldface and those that do not differ significant from it are indicated with an asterisk. The neighborhood radius for which we obtained the best performance is also indicated with a symbol: 20 mm (↓), 30 mm (−) or 40 mm (↑).

Landmark	SI	3DSC	USC	SHOT	PFH	FPFH
en	**1.7** ↑	2.2 ↑	2.8 ↓	6.2 ↓	4.6 −	2.4 ↑
(2)	**(5 - 23)**	(5 - 24)	(6 - 23)	(8 - 21)	(8 - 21)	(5 - 23)
ex	4.0 ↑ *	3.8 − *	n.p −	**3.8** −	6.1 −	6.6 ↓
(2)	(11 - 37)	(11 - 86)	(< 23)	**(6 - 23)**	(11 - 68)	(14 - 52)
n	3.4 ↑	**1.9** ↑	3.4 ↑	3.1 −	5.1 ↓	2.4 −
	(6 - 200)	**(5 - 200)**	(6 - 18)	(5 - 17)	(7 - 200)	(5 - 200)
a	**1.5** −	1.6 ↓ *	2.5 ↓	4.8 ↓	6.2 ↓	5.7 ↓
(2)	**(4 - 25)**	(3 - 27)	(4 - 16)	(6 - 26)	(12 - 17)	(9 - 14)
ac	**2.5** ↑	3.9 ↓	n.p −	4.3 ↓	5.6 −	6.4 ↓
(2)	**(14 - 23)**	(10 - 25)	(< 105)	(6 - 21)	(12 - 22)	(13 - 22)
nt	n.p ↑	n.p ↑	13.2 ↓	8.5 ↑	7.6 ↓	**7.0** −
(2)	(< 8)	(< 9)	(14 - 200)	(15 - 200)	(11 - 200)	**(12 - 200)**
prn	2.8 ↑	**1.3** ↑	1.9 ↑	3.0 −	4.3 ↓	1.8 −
	(4 - 200)	**(2 - 200)**	(3 - 200)	(4 - 200)	(5 - 200)	(3 - 200)
sn	2.0 −	**1.7** −	n.p ↓	2.6 −	6.7 ↓	2.6 −
	(5 - 60)	**(3 - 200)**	(< 111)	(4 - 200)	(10 - 200)	(5 - 200)
ch	2.7 ↓	3.1 ↑	3.4 −	**2.2** ↓	5.7 ↑	3.7 ↑
(2)	(8 - 23)	(7 - 14)	(12 - 30)	**(4 - 43)**	(10 - 24)	(9 - 40)
cph	n.p ↓	n.p ↑	13.9 ↓	8.5 ↑ *	**7.9** −	8.3 − *
(2)	(< 9)	(< 10)	(21 - 30)	(16 - 39)	**(12 - 200)**	(15 - 200)
li	n.p ↓	2.7 ↑	2.5 −	**1.9** ↑	8.1 −	5.1 ↑
	(< 11)	(12 - 42)	(8 - 30)	**(4 - 49)**	(9 - 200)	(7 - 200)
ls	10.8 ↑	**2.6** ↑	10.5 ↑	3.6 ↑	5.0 −	4.1 −
	(21 - 38)	**(13 - 200)**	(20 - 34)	(13 - 123)	(6 - 200)	(8 - 200)
sto	6.3 ↑	2.7 ↑ *	**2.4** −	2.8 ↑	6.1 −	5.3 ↑
	(14 - 84)	(7 - 58)	**(6 - 27)**	(7 - 15)	(8 - 200)	(9 - 200)
sl	9.4 ↑	2.9 ↓ *	n.p ↑	**2.6** ↑	5.9 − *	6.3 − *
	(13 - 21)	(9 - 200)	(< 16)	**(4 - 200)**	(10 - 97)	(11 - 18)
pg	18.7 ↑	**4.7** ↑	13.4 ↑	5.3 ↑ *	7.1 −	4.9 ↑ *
	(23 - 200)	**(9 - 200)**	(19 - 200)	(9 - 200)	(9 - 200)	(12 - 200)
t	n.p −	n.p −	n.p −	**7.4** −	13.4 ↓	8.1 ↓ *
(2)	(< 58)	(< 125)	(< 142)	**(20 - 96)**	(23 - 89)	(25 - 100)
oi	**7.9** ↑	n.p −	n.p −	12.3 ↑	15.1 ↓	9.1 ↑
(2)	**(17 - 22)**	(< 17)	(< 129)	(23 - 30)	(25 - 37)	(17 - 27)

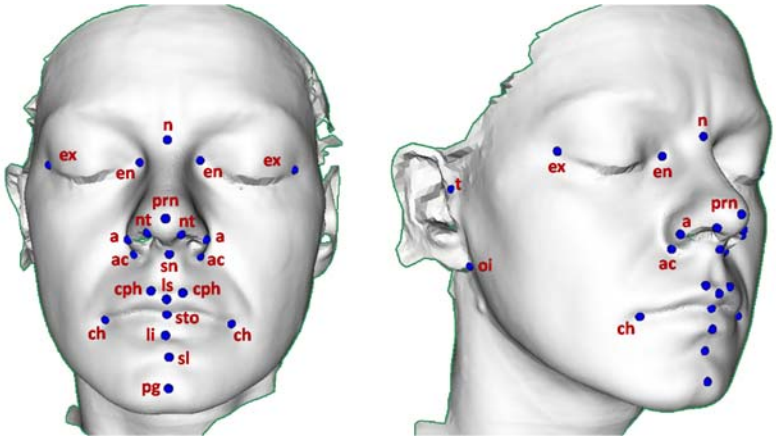

Fig. 2. The 26 landmarks used in this study: *en* = endocanthion; *ex* = exocanthion; *n* = nasion; *a* = alare; *ac* = alar crest; *nt* = nostril top; *prn* = pronasale; *sn* = subnasale; *ch* = cheilion; *cph* = crista philtrum; *li* = labiale inferius; *ls* = labiale superius; *sto* = stomion; *sl* = sublabiale; *pg* = pogonion; *t* = tragion; *oi* = otobasion inferius [22].

Landmarks with bilateral symmetry (left and right) were merged together by considering each as a separate instance of the same test.

The descriptor template for each landmark was computed as the median descriptor from a training set, created by means of 6-fold cross-validation. To compute the scores $s(\mathbf{v})$, the descriptor template was compared with the one of each vertex using (minus) the Euclidean distance (i.e. considering each descriptor as a point in N-dimensional space, N being the descriptor length). The only exception was the case of spin images, where we used the (2D) cross-correlation, as suggested in the original paper [16]. Nonetheless, we should mention that results using Euclidean distance (with the descriptor normalized to sum the unit) were similar to those using cross-correlation.

Table 1 summarizes the results. Each cell describes the first plateau of the expected local accuracy curve: the number on the top indicates its median value and the ones below (in parentheses) indicate its limits. Recall that the plateau is only searched for r_S values below the first peak of $\overline{G}_L$. The plateau range was determined as the region for which $\overline{e}_L$ did not vary by more than 10%.

The best descriptor for each landmark is highlighted in boldface and those not significantly different from it are indicated with an asterisk[4]. For example, the best descriptor for the inner-eye corners (*en*) is SI; if we constrain the search to a radius below 22 mm we can expect to locate each of the inner-eye corners at 1.7 mm from their correct (ground truth) position. Clearly, the great majority of landmarks must be constrained to a local search range for all six descriptors and only a few of them could be used globally (e.g. *n*, *prn*, *pg*, *sn*).

[4] $p > 0.05$ on a paired Wilcoxon signed rank test.

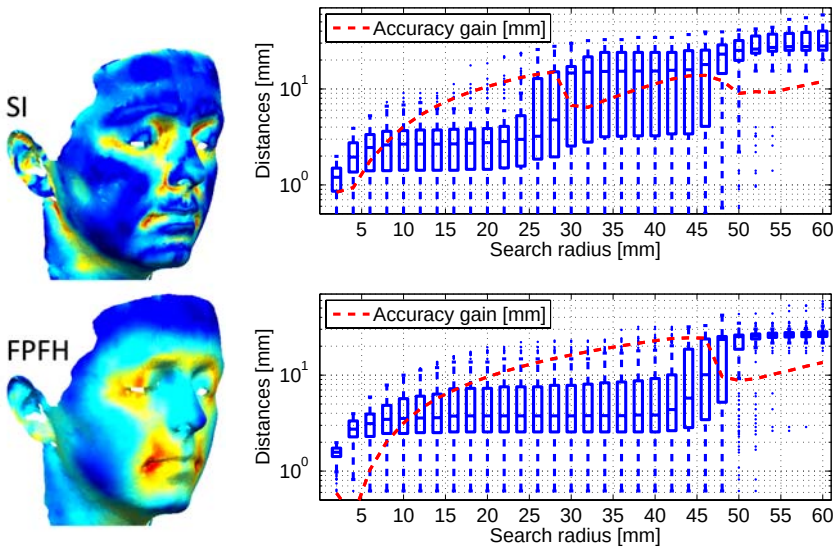

Fig. 3. Left: a facial scan with descriptor scores color-coded (red = high, blue = low) for SI and FPFH, targeting the mouth corners (ch). Right: boxplots of the distances of the highest-score vertices (for the whole set of 144 meshes) for different search radii, r_S. The expected local accuracy, $\overline{e}_L$ is estimated as the median; the discontinuous lines indicate the accuracy gain, $\overline{G}_L$.

It is interesting to analyze the consistency of the plateau ranges. We observed strong agreement for the different descriptors on symmetric landmarks for which the twin point is relatively nearby (e.g. en, a, ac), especially for the upper limit, which is the most important one. On the other hand, symmetric landmarks that are further apart showed significant variations in their plateau limits across descriptors. This is due to the presence of strong sources of false positives different from the twin points. To illustrate this, Fig. 3 shows the descriptor scores obtained by SI and FPFH for the mouth corners (ch) color-coded on a facial scan. While SI tends to produce high scores on the nose corners (ac), which are approximately 20 mm apart from the mouth corners, FPFH does not show high scores on the nose and the upper plateau limit is therefore extended up to the twin point (the other mouth corner), at about 40 mm, as indicated on the right of the figure. Note also that both descriptors show a peak of $\overline{G}_L$ at about 45 mm, but for SI there is an earlier one at 27 mm, in coincidence with the narrower usable local range of the descriptor for this particular landmark.

We also explored the influence of the neighborhood size used for the computation of the descriptors, testing for $r_N = 20$, 30 and 40 mm. Table 1 shows the results for $r_N = 30$ mm and, for each cell, it is also indicated whether that neigh-

borhood size or one of the two other options was optimal (see table caption). The full tables are available on-line[5].

5 Conclusions

We present a comparison of local geometry descriptors for 3D facial landamrks on a dataset of 144 scans at moderately high resolution, annotated with 26 anatomical points relevant for craniofacial research. To facilitate the analysis we explored the patterns generated when computing the local accuracy at different search radii. It was found that the most useful descriptors present a flat region or *plateau* that can be used to characterize the descriptor's behavior, both in terms of accuracy and maximum range for the local search.

Six histograms-based descriptors were evaluated: three describing distances and three describing orientations. No descriptor dominated over the rest. From the point of view of the overall error (i.e. the average over all landmarks), 3DSC was the best, followed by SHOT, FPFH, SI, USC and PFH. However, 3DSC was best only for 5 out of 26 landmarks (+6 that did not differ significantly from the best), while SHOT did so for 8(+3) landmarks and SI for 8(+2) landmarks. This illustrates how a comparison of descriptors based on their global behavior might be misleading if targeting facial landmarks.

Finally, while for some landmarks the expected accuracy was below 2 mm, some others did not obtain satisfactory results for any of the six descriptors. This is the case for the ear points (*t* and *oi*), typically difficult due to their very complex geometry, but also for symmetric points very close to each otehr (*nt*, *ls*), where the accuracy of the best descriptor was similar to the separation between the twin points and therefore not good enough to distinguish between them.

Acknowledgment

The authors would like to thank their colleagues in the Face3D Consortium (www.face3d.ac.uk), and the financial support provided for it from the Wellcome Trust (grant 086901/Z/08/Z).

References

1. Hennessy, R., Baldwin, P., Browne, D., Kinsella, A., Waddington, J.: Frontonasal dysmorphology in bipolar disorder by 3D laser surface imnaging and geometric morphometrics: Comparison with schizophrenia. Schizophr Res **122** (2010) 63–71
2. Sharifi, A., Jones, R., Ayoub, A., et al.: How accurate is model planning for orthognathic surgery. Int J Oral Max Surg **37** (2008) 1089–1093
3. Bowyer, K., Chang, K., Flynn, P.: A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recogn. Comput Vis Image Und **101** (2006) 1–15

[5] http://fsukno.atspace.eu/Research.htm

4. Plooij, J., Swennen, G., Rangel, F., et al.: Evaluation of reproducibility and reliability of 3D soft tissue analysis using 3D stereophotogrammetry. Int J Oral Max Surg **38** (2009) 267–273
5. Aynechi, N., Larson, B., Leon-Salazar, V., et al.: Accuracy and precision of a 3D anthropometric facial analysis with and without landmark labeling before image acquisition. Angle Orthod **81** (2011) 245–252
6. Dibeklioglu, H., Salah, A., Akarun, L.: 3D faceial landmarking under expression, pose, and occlusion variations. In: Proc. BTAS (2008) 1–6
7. Segundo, M., Silva, L., Bellon, et al.: Automatic face segmentation and facial landmark detection in range images. IEEE T Syst Man Cy B: Cybernetics **40** (2010) 1319–1330
8. Gupta, S., Markey, M., Bovik, A.: Antopometric 3D face recognition. Int J Comput Vision **90** (2010) 331–349
9. Szeptycki, P., Ardabilian, M., Chen, L.: A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking. In: Proc. BTAS (2009) 1–6
10. Passalis, G., Perakis, N., Theoharis, T., et al.: Using facial symm to handle pose variations in real-world 3D face recogn. IEEE T Pattern Anal **33** (2011) 1938–1951
11. Salti, S., Tombari, F., Stefano, L.: A performance evaluation of 3D keypoint detectors. In: Proc. 3DimPVT (2011) 236–243
12. Bronstein, A., Bronstein, M., Castellani, U., et al.: SHREC 2010: robust correspondence benchmark. In: Proc, 3DOR (2010)
13. Creusot, C., Pears, N., Austin, J.: Automatic keypoint detection on 3D faces using a dictionary of local shapes. In: 3DimPVT (2011) 204–211
14. Romero-Huertas, M., Pears, N.: Landmark localisation in 3D face data. In: Proc. AVSS (2009) 73–78
15. Rusu, R., Cousins, S.: 3D is here: Point cloud library (PCL). In: Proc. ICRA (2011) 1–4
16. Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. IEEE T Pattern Anal **21** (1999) 433–449
17. Frome, A., Huber, D., Kolluri, R., et al.: Recognizing objects in range data using regional point descriptors. In: Proc. ECCV (2004) 224–237
18. Tombari, F., Salti, S., Stefano, L.D.: Unique shape context for 3D data description. In: Proc. 3DOR (2010) 57–62
19. Tombari, F., Salti, S., Stefano, L.D.: Unique signature of histograms for local surface description. In: Proc. ECCV (2010) 356–369
20. Rusu, R., Blodow, N., Marton, Z., et al.: Aligning point cloud views using persistent feature histograms. In: Proc. IROS (2008) 3384–3391
21. Rusu, R., Blodow, N., Beetz, M.: Fast point feature histograms (FPFH) for 3D registration. In: Proc. ICRA (2009) 3212–3217
22. Hennessy, R., Kinsella, A., Waddington, J.: 3D laser surface scanning and geometric morphometric analysis of craniofacial shape as an index of cerebro-craniofacial morphogenesis: initial applic to sexual dimorph. Biol Psychiat **51** (2002) 507–514
23. Farkas, L.: Anthropometry of the head and face. Raven Press (New York), 2nd edition (1994)