View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DCU Online Research Access Service

Rotationally Invariant 3D Shape Contexts Using Asymmetry Patterns

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 and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland*
{*federico.sukno, paul.whelan*}*@dcu.ie, jwadding@rcsi.ie*

Keywords: 3D geometric descriptors, rotational symmetry, craniofacial landmarks.

Abstract: This paper presents an approach to resolve the azimuth ambiguity of 3D Shape Contexts (3DSC) based on asymmetry patterns. We show that it is possible to provide rotational invariance to 3DSC at the expense of a marginal increase in computational load, outperforming previous algorithms dealing with the azimuth ambiguity. We build on a recently presented measure of approximate rotational symmetry in 2D defined as the overlapping area between a shape and rotated versions of itself to extract asymmetry patterns from a 3DSC in a variety of ways, depending on the spatial relationships that need to be highlighted or disabled. Thus, we define Asymmetry Patterns Shape Contexts (APSC) from a subset of the possible spatial relations present in the spherical grid of 3DSC; hence they can be thought of as a family of descriptors that depend on the subset that is selected. This provides great flexibility to derive different descriptors. We show that choosing the appropriate spatial patterns can considerably reduce the errors obtained with 3DSC when targeting specific types of points.

1 INTRODUCTION

Geometric descriptors for three dimensional (3D) data are important for a wide range of applications, as they constitute a core element for the identification of corresponding points in relation to object retrieval (Tombari et al., 2010), recognition (Frome et al., 2004), surface registration (Bariya et al., 2012) and landmark identification (Creusot et al., 2011; Passalis et al., 2011).

The increased availability of 3D data in the last decade has generated much research in this area and several 3D descriptors have been proposed. Depending on the data that is targeted, the descriptors can be purely geometric (Johnson and Hebert, 1999; Rusu et al., 2009; Zhang, 2009; Chen and Bhanu, 2007) or include additional functions that are attached to the geometry, such as radiometric information (Steder et al., 2011; Zaharescu et al., 2012).

Among purely geometric descriptors, which are the most general type, 3D shape contexts (and extensions derived from them) have attracted considerable interest due to their good performance in diverse applications. Indeed, a recent comparison of geometric descriptors in the context of craniofacial landmark localization highlighted 3D shape contexts as one of the most accurate algorithms (Sukno et al., 2012).

Shape contexts in 3D are based on the distribution of distances with respect to the point of interest, estimated by means of a histogram over a spherical grid (elevation, azimuth and radius). The spherical grid is centered at the point of interest and its North Pole is oriented in the direction of the normal to the surface. This is enough to uniquely determine the elevation and radial bins but leaves unresolved the origin of azimuth bins. Different approaches have been taken to resolve this ambiguity:

- In one of the earliest works (Frome et al., 2004), the 3D Shape Contexts descriptor (3DSC) was introduced, without actually resolving the azimuth ambiguity. The authors compute multiple descriptors to account for all possible rotations (as many as the number of azimuth bins). During matching, when comparing descriptors of different points, all possible rotations are tested and the one that produces the highest similarity score is retained.
- As an alternative that achieves invariance to the azimuth angle, Frome et al. explored the use of Spherical Harmonics. Similarly to other descriptors (Kazhdan et al., 2004), they proposed to keep only the magnitude of the Spherical Harmonic coefficients, which are rotationally invariant. We will refer to this approach as Harmonic Shape Contexts (HSC) (Frome et al., 2004).

- A third option (Kortgen et al., 2003; Tombari et al., 2010) consists of performing Singular Value Decomposition (SVD) on the support region (i.e. all points within the considered sphere) to identify the principal axes and disambiguate the sign by considering the heaviest tail of each axis as the positive direction. Thus, a unique axis can be identified to set the azimuth origin, obtaining the Unique Shape Contexts (USC) descriptor.

It would be desirable to avoid the evaluation of multiple descriptors as done by (Frome et al., 2004). Such a strategy increases the computational load during matching, can suffer from false matches (due to an unfortunate rotation of the descriptor of a non-corresponding point) and adds considerable complexity to the application of machine learning techniques that can be useful upon the availability of a training set. Despite the above efforts to obtain shape context descriptors without azimuth ambiguity, the best performance is still obtained by using 3DSC (i.e. computing multiple descriptors).

The performance of HSC was found comparable to 3DSC in some cases (Frome et al., 2004) but at the expense of a huge increase in computational load. On the other hand, USC was reported to perform slightly better than 3DSC in terms of precision-recall curves for a task of feature matching on synthetically transformed shapes (Tombari et al., 2010). However, USC was found considerably less accurate than 3DSC when targeting specific points on a craniofacial landmark localization task (Sukno et al., 2012). As we will show, this can be explained by the instability of the sign disambiguation on objects that present a high variability, such as the human face. That is, the direction of the axes determined by the proposed disambiguation step cannot be assured to be consistent across a population of facial scans. Since USC rely on the unique definition of the azimuth bins, the lack of consistency has an important effect on accuracy.

In this paper we present a different approach to resolve the azimuth ambiguity based on asymmetry patterns and show that it is possible to attain rotationally invariant shape contexts that obtain comparable accuracy to 3DSC for the localization of craniofacial landmarks and remarkably outperform 3DSC for specific points like the outer eye corners and nose corners.

We build on a recently presented measure of approximate rotational symmetry in 2D (Guo et al., 2010), defined as the overlapping area between a shape and rotated versions of itself. We show that such a measure can be extended to 3DSC and derive asymmetry based on the absolute differences between overlapping bins of the descriptor and rotated versions of itself. Both measures depend of the rotation angle but not on the selection of the origin of azimuth bins, which allows us to obtain patterns that capture the rotational asymmetry of the descriptor over the azimuth but are invariant to the rotation of its bins.

The asymmetry patterns can be defined in a variety of ways, depending on the spatial relationships that need to be highlighted or disabled. Thus, we define Asymmetry Patterns Shape Contexts (APSC) from a subset of the possible spatial relations present in the spherical support region; hence they can be thought of as a family of descriptors that depend on the subset that is selected.

Concrete examples of APSC are evaluated by defining some of the simplest possible spatial patterns. We show that the performance of APSC depends heavily on the selection of these spatial patterns, which can be useful to target different types of points. Regarding the complexity, the computation of an APSC descriptor is slightly more expensive than a single 3DSC but produces considerable savings at matching time (due to the rotational invariance) and memory (APSC requires half the memory of 3DSC). This computational efficiency contrasts with prior work exploring the use of symmetry in geometric descriptors using Spherical Harmonics (Kazhdan et al., 2004)

In the next section we provide the definition of APSC as well as a brief review of 3DSC. Experimental evaluation is presented in Section 3, followed by a discussion of results (Section 4) and concluding remarks (Section 5).

2 ASYMMETRY PATTERNS SHAPE CONTEXTS

Computation of the APSC descriptor starts by computing a 3DSC descriptor (Frome et al., 2004), from which the asymmetry patterns are later extracted.

2.1 3D Shape Contexts

This descriptor is based on a 3D-histogram computed on a spherical support region centered at the interest point, $\mathbf{v}$, considering a neighborhood $\mathcal{N}_v = \{\mathbf{w} \,|\, \|\mathbf{w} - \mathbf{v}\| \leq r_N\}$, namely all points within a radius r_N. The North pole of the sphere is oriented with the normal vector at the interest point $\mathbf{n}_v$. The default structure has $N_E = 11$ elevation bins and $N_A = 12$ azimuth bins, both uniformly spaced, and $N_R = 15$ radial bins logarithmically spaced as follows:

$$r_k = \exp\left(\ln(r_{min}) + \frac{k}{N_R}\ln\left(\frac{r_N}{r_{min}}\right)\right) \qquad (1)$$

where r_k is the *k-th* radial division from a total of N_R, r_N is the radius of the spherical neighborhood and r_{min} is the radius of the smallest bin.

The logarithmic sampling is aimed at assigning more importance to shape changes that are closer to the interest point. The contribution to the histogram of a point $\mathbf{w}$ that falls in bin (i,j,k) will be:

$$\left(\rho_w \sqrt[3]{V_{i,j,k}}\right)^{-1} \tag{2}$$

where i, j and k are the indices of elevation, azimuth and radial bins, respectively. The normalization by bin volume, $V_{i,j,k}$, accounts for the large difference between bin sizes (especially along radius and elevation), while the inverse local point density ρ_w aims to correct variations in sampling density and is estimated as the count of points in a small sphere centered at $\mathbf{w}$.

As the spherical support region is defined based on $\mathbf{v}$ and $\mathbf{n}_v$, there is an ambiguity in the origin of the azimuth bins. This is dealt with by calculating N_A descriptors per point, 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 the multiple descriptors by choosing the one that yields the smallest Euclidean distance:

$$d(\mathbf{x},\mathbf{y}) = \min_{0\le a<N_A} \sqrt{\sum_{i=0}^{N_E-1}\sum_{j=0}^{N_A-1}\sum_{k=0}^{N_R-1}(x_{i,j+a,k}-y_{i,j,k})^2} \tag{3}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the descriptors to compare and the addition $j+a$ is modulo N_A (i.e circular) so that $\mathbf{x}_{i,j+a,k}$ is an azimuth rotation of $\mathbf{x}_{i,j,k}$ (i.e. about the North-South axis of the sphere) by a bins.

2.2 Rotational Symmetry

In a recent work, Guo et al. (Guo et al., 2010) presented continuous measures of approximate bilateral and rotational symmetry. Specifically, given a shape $\mathbf{m}$ in 2D and a rotation angle ϕ about the z-axis (i.e. perpendicular to the plane containing the shape), they defined the rotational central symmetry degree $\mathcal{S}_c(\mathbf{m},\phi)$ to be the area of intersection of shape $\mathbf{m}$ with a rotated version of itself by an angle ϕ, $R(\mathbf{m},\phi)$, normalized by the area of shape $\mathbf{m}$:

$$\mathcal{S}_c(\mathbf{m},\phi) = \frac{\text{Area}(\mathbf{m}\cap R(\mathbf{m},\phi))}{\text{Area}(\mathbf{m})} \tag{4}$$

Defined in this way, $\mathcal{S}_c(\mathbf{m},\phi)$ measures the degree of rotational symmetry of shape $\mathbf{m}$ from 0 (no symmetry) to 1 (perfect symmetry at the considered angle).

We can adapt this symmetry measure to the 3DSC descriptor by converting the area overlap into the minimum value of overlapping bins. That is, as the support region of the descriptor is spherical, the overlap of rotated shapes (in terms of area or volume) is always perfect, but not the values assigned to the coinciding bins. For example, assume we extract from a 3DSC descriptor $\mathbf{x}$ the ring composed by all the bins at a given elevation i and radius k; this will generate a shape $\mathbf{m}$ represented as a sequence of N_A non-negative values from the corresponding bins:

$$m_j = x_{i,j,k}, \quad m_j \ge 0 \forall j \in [1;N_A] \tag{5}$$

We can define the symmetry degree of the sequence $\mathbf{m}$ as follows:

$$\mathcal{S}(\mathbf{m},a) = \frac{\sum_j \min(m_j, m_{j+a})}{\sum_j m_j} \tag{6}$$

where, as before, $j+a$ is the addition modulo the cardinality of $\mathbf{m}$ (N_A in this example).

Notice that the angular parameter ϕ used in the area-based definition is replaced by the shift parameter a, which represents a discrete azimuth rotation of $2\pi/N_A$. Thus, $\mathcal{S}$ behaves analogously to $\mathcal{S}_c$.

Figure 1 shows an example sequence $\mathbf{m}$ that corresponds to a ring with perfect rotational symmetry every 120 degrees and the sequence resulting from the concatenation of the symmetry degree for all possible discrete rotations:

$$\mathcal{P}_S(\mathbf{m}) = \mathcal{S}(\mathbf{m},1),\mathcal{S}(\mathbf{m},2),\ldots,\mathcal{S}(\mathbf{m},N_A) \tag{7}$$

The sequence $\mathcal{P}_S(\mathbf{m})$ is the symmetry pattern of the sequence $\mathbf{m}$ which indicates how symmetric is the ring that originated $\mathbf{m}$ for different angles of azimuth rotation. However, from the definition of $\mathcal{S}(\mathbf{m},a)$, it is clear that the generated pattern is invariant to the origin chosen for the azimuth bins, i.e.

$$\mathcal{P}_S(\mathbf{m}) = \mathcal{P}_S(R(\mathbf{m},a)), \forall a \in \mathbb{Z} \tag{8}$$

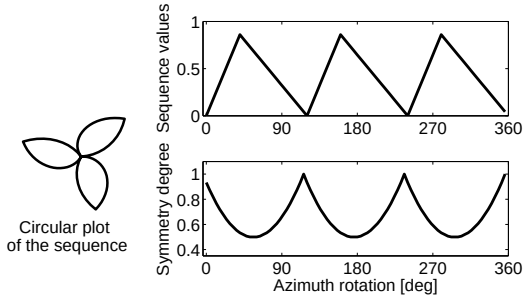

Figure 1: Example of a sequence with perfect rotational symmetry at shifts of 120 degrees. A circular representation is provided on the left by varying the distance to the centre proportionally to the sequence values, which are shown on the top-right plot. The bottom-right plot shows the resulting symmetry pattern.

2.3 Asymmetry Patterns

It is interesting to define asymmetry as the complement of symmetry, also between 0 and 1, as follows:

$$\mathcal{A}(\mathbf{m},a) = 1 - \mathcal{S}(\mathbf{m},a) \tag{9}$$

In the Appendix we show that this definition implies:

$$\mathcal{A}(\mathbf{m},a) = \frac{1}{2}\frac{\sum_j |m_j - m_{j+a}|}{\sum_j m_j} \tag{10}$$

which is the mean of absolute differences between $\mathbf{m}$ and $R(\mathbf{m},a)$ with an appropriate normalization factor. While such normalization is important to facilitate a meaningful interpretation of the asymmetry value in $[0;1]$, it is not desirable in our case as it removes potentially useful information. Thus we define:

$$\mathcal{A}_1(\mathbf{m},a) = \sum_j |m_j - m_{j+a}| \tag{11}$$

As stated above, this measure captures the average difference between bins of the sequence originated from a ring and one azimuth-rotated version of itself. However, other functions of $|m_j - m_{j+a}|$ would be applicable, with the only requisite being that they are summations over all the elements of the sequence. For example, it would be possible to use the central moments of any order (indeed our definition of $\mathcal{A}_1$ is a scaled version of the first-order moment).

We will restrict ourselves to the use of asymmetry as defined in (11). Considering all possible azimuth rotations of the sequence that generate distinct values, we obtain the following asymmetry pattern:

$$\mathcal{P}_A(\mathbf{m}) = \mathcal{A}_1(\mathbf{m},1), \mathcal{A}_1(\mathbf{m},2), \ldots, \mathcal{A}_1(\mathbf{m}, \lfloor \frac{N_A}{2} \rfloor) \tag{12}$$

where $\lfloor x \rfloor$ is the integer part of x. Defined in this way the asymmetry pattern accounts for approximately half the possible rotations. This is because the remaining ones would only generate repeated values (as $\mathcal{A}_1(\mathbf{m},a)$ is an even function with respect to a):

$$\forall a \in [1;N_A] : \mathcal{A}_1(\mathbf{m},a) = \mathcal{A}_1(\mathbf{m},-a) \tag{13}$$

where both addition and substraction are modulo N_A operations. Then it follows that:

$$\forall a \in [1;N_A] : \mathcal{A}_1(\mathbf{m},a) = \mathcal{A}_1(\mathbf{m},N_A - a) \tag{14}$$

Intuitively, this can be understood from the definition as an overlap between $\mathbf{m}$ and a rotated version of itself by an angle ϕ that we can call $\mathbf{m}' = R(\mathbf{m},\phi)$. The overlap between the two would be the same if we rotate both $\mathbf{m}$ and $\mathbf{m}'$ by any angle, for example $-\phi$, which would transform shape $\mathbf{m}$ into $R(\mathbf{m},2\pi-\phi)$ and shape $\mathbf{m}'$ into $\mathbf{m}$. Hence, the overlap between $\mathbf{m}$ and $R(\mathbf{m},\phi)$ is equivalent to the overlap between $\mathbf{m}$ and $R(\mathbf{m},2\pi-\phi)$ and the same applies to both the symmetry and asymmetry measures defined above.

2.4 Spatial Relationships

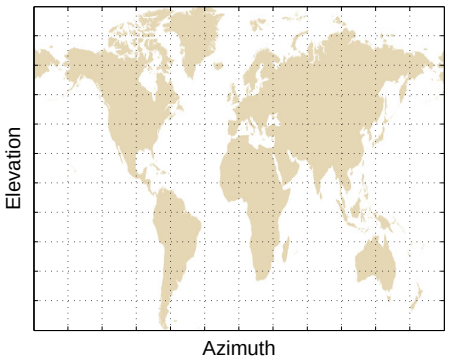

Figure 2: The World map as an example of a shell with (ideally) constant radius. The azimuth (longitude) and elevation (latitude) bins are also indicated.

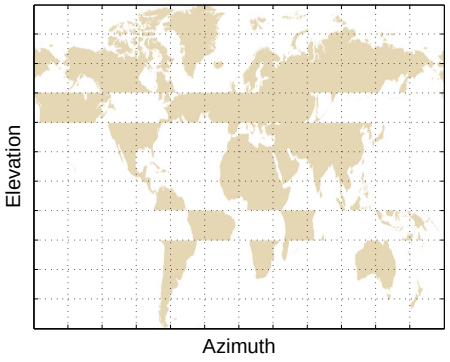

Figure 3: The same spherical shell as in Figure 2 after arbitrary and independent azimuth rotations of two rings with constant elevation (4-th and 8-th bins).

So far we have worked with a sequence $\mathbf{m}$ defined as in (5), namely the bins of a ring at fixed elevation and radius from the spherical support of a 3DSC. This choice seems natural as it allows to transform each $(i\text{-}k)$ ring of a 3DSC descriptor $\mathbf{x}$ into an asymmetry pattern that is invariant to the choice of azimuth.

However, such a choice only takes into account the spatial relationships within each $(i\text{-}k)$ ring. To illustrate this suppose that we consider all bins of $\mathbf{x}$ at a fixed radius. This is a spherical shell and we could represent it on a Cartesian plane similarly to a World map, as the one shown in Figure 2, where the latitude is the elevation and the longitud is the azimuth. If we consider the representation of each ring independently, then any azimuth shift of a ring has no effect in our representation and both the *correct* World map of Figure 2 and the example with shifted rings in Figure

3 will generate the same set of patterns. A similar reasoning can be applied to the relation between shells of different radii. In contrast, when shifted versions of $\mathbf{x}$ are generated to select the best match for 3DSC as in equation (3), the whole sphere rotates at once and all spatial relations are kept.

The above is not necessarily bad and, as we will show, sometimes it might be useful to *disable* certain spatial relationships. However, in the general case this can lead to a loss of discriminant information.

The choice of what spatial relations are considered is related to the definition of $\mathbf{m}$. It is easy to verify that each of the generated sequences must cover all azimuth bins, as otherwise we would lose the invariance of the patterns to azimuth rotations, but there is no restriction regarding the variation of elevation and radius within the sequence. In other words, equation (5) is just a specific choice of $\mathbf{m}$ that leads to one of many possible APSC. A few straightforward alternatives include:

- Considering *diagonals*[1], where the variation of azimuth is accompanied by a variation in elevation and/or radius:

$$m_j = x_{i+j,j,k} \tag{15}$$

$$m_j = x_{i+j,j,k+j} \tag{16}$$

- Jointly considering two (or more) rings that are neighbors:

$$m_{1,j} = x_{i,j,k}, \quad m_{2,j} = x_{i+1,j,k}$$

$$\mathcal{A}_2(\mathbf{m},a) = \sum_j |m_{1,j} - m_{1,j+a}| + |m_{2,j} - m_{2,j+a}| \tag{17}$$

Notice that, when jointly considering two or more rings, the overlap is computed only between rings with the same definition. All additions are circular, modulo the corresponding number of bins (N_E, N_A and N_R respectively for i, j and k).

In principle, the definition of the sequences can be arbitrary and the above are just a few intuitive choices. Thus, APSC can be though of as a family of descriptors with a flexible definition that allows to adapt them to highlight or disable specific spatial relationships. We will discuss the advantages and limitations of this fact in Section 4.

[1]Given that the support region is spherical the resulting sampling pattern is not a diagonal but we use this name in analogy to the shape it would take if the bins were represented in a Cartesian grid.

3 EXPERIMENTAL EVALUATION

In this section we compare the performance of Asymmetry Patterns Shape Contexts to the following three algorithms, that constitute competing alternatives:

- 3D shape contexts (3DSC) (Frome et al., 2004), which generate descriptors that are not invariant to azimuth rotations.
- Harmonic Shape Contexts (HSC) (Frome et al., 2004), which achieve invariance to azimuth rotations by decomposing each spherical shell at fixed radius r_k of a 3DSC descriptor into Spherical Harmonics keeping only the modulus of the resulting coefficients.
- Unique Shape Contexts (USC) (Tombari et al., 2010), which compute a 3DSC with a unique orientation of the spherical support region based on the principal axes in a neighborhood of the interest point and a sign disambiguation step.

In all cases we used the default configuration as indicated in the original papers: $N_E = 11$ elevation bins, $N_A = 12$ azimuth bins and $N_R = 15$ radial bins. The radius of the spherical support region was set to $r_N = 30$ mm and the minimum radius to $r_{min} = 1$ mm (see (1)). Spherical Harmonics were computed up to order $N_{SH} = 16$. Thus, 3DSC and USC had a total of $N_E \times N_A \times N_R = 1980$ bins while HSC had a total of $N_R \times N_{SH} \times (N_{SH}+1)/2 = 2040$ bins.

Regarding APSC, as mentioned before they can be considered a family of descriptors with many possible instances depending on the spatial relations selected to construct the sequences $\mathbf{m}$ from which the asymmetry patterns are derived. We performed tests using the fixed elevation and radius rings (azimuth rings, for short) as defined in (5) and eight other simple patterns resulting from the *diagonals* (i.e. jointly changing the bin indexes of azimuth with radius and/or elevation), adjacent rings (either in elevation or radius) and combinations of diagonals and azimuth rings. From these, we selected 5 representative cases to report, for which we provide the corresponding equations in Table 1. Results for the remaining four are available on-line[2].

All sequences in Table 1 are computed starting from a 3*DSC* descriptor $\mathbf{x}$, whose elements are indexed by $(i,j,k) =$ (elevation, azimuth, radius). We always generate sequences for all possible combinations of i and k (while varying j), which results in a full coverage of the bins of $\mathbf{x}$. In the case of two sequences considered jointly (bottom three rows of the table), they are combined to generate the asymmetry

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

Table 1: Description of some specific spatial patterns for APSC descriptors. In all cases the sequences are generated by varying the azimuth index j.

Abbreviation	Sequence(s) equation	Description
D_{AR}	$m_j = x_{i,j,k+j}$	Azimuth-Radius diagonal
D_{AER}	$m_j = x_{i+j,j,k+j}$	Azimuth-Elevation-Radius diagonal
A+E	$m_{1,j} = x_{i,j,k}, \quad m_{2,j} = x_{i+1,j,k}$	Azimuth ring + Elevation neighbors
A+R	$m_{1,j} = x_{i,j,k}, \quad m_{2,j} = x_{i,j,k+1}$	Azimuth ring + Radial neighbors
A+D_{AER}	$m_{1,j} = x_{i,j,k}, \quad m_{2,j} = x_{i+j,j,k+j}$	Azimuth ring + Azim-Elev-Rad diagonal

pattern as indicated in (17). For each sequence, which has always $N_A = 12$ bins, an asymmetry pattern of length $\lfloor \frac{N_A}{2} \rfloor = 6$ is generated. Thus, each *APSC* descriptor has only $N_E \times N_A \times \lfloor \frac{N_A}{2} \rfloor = 990$ bins.

3.1 Data

We frame our evaluation in the task of craniofacial landmark localization. This landmark-based evaluation has two important advantages with respect to evaluations based on *keypoints* (i.e. points that are considered highly discriminant or *salient* from the point of view of a descriptor): *i*) all descriptors are evaluated in the same set of points which are not necessarily salient and, as in the case of facial landmarks, can include diverse (local) geometries that pose different degrees of challenge to the descriptor; *ii*) the evaluation is done on *real world* examples (e.g. a population of faces where anatomical correspondences have been manually annotated) instead of using synthesized examples obtained by modifying a given example by some set of transformations (Tombari et al., 2010; Bronstein et al., 2010; Steder et al., 2011).

Our test dataset consisted of 144 facial scans acquired by means of a hand-held laser scanner (FastSCAN^{TM}, Colchester, VT, USA). Special care was taken to avoid occlusions due to facial hair. The extracted surfaces were subsampled by a factor of 4 : 1, resulting in an average of approximately 21.3 thousand vertices per mesh. The dataset contains exclusively healthy volunteers who acted as controls in the context of craniofacial dysmorphology research. Each scan was annotated with a set of anatomical landmarks, in accordance with definitions in (Hennessy et al., 2002) (based on (Farkas, 1994)), from which we target the 22 points indicated in Figure 4.

The fact that the test dataset was acquired in the context of clinical research makes it especially suited for tests in localization accuracy. As it can be observed in Figure 4 these are high quality scans, which have been carefully annotated by experts based on anthropometric definitions. Recent studies on manual identification of 3D facial landmarks indicate that the intra- and inter-observer uncertainty of this type of

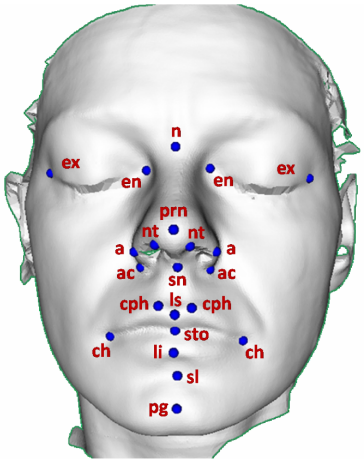

Figure 4: Example of the facial scans from the test dataset with the annotation of the 22 landmarks used in this study: *en* = endocanthion; *ex* = exocanthion; *n* = nasion; *a* = alare; *ac* = alar crest; *nt* = nostril top; *prn* = pronasale; *sn* = subnasale; *ch* = cheilion; *cph* = crista philtrum; *li* = labiale inferius; *ls* = labiale superius; *sto* = stomion; *sl* = sublabiale; *pg* = pogonion; (Hennessy et al., 2002).

annotations are typically between 1 mm and 2 mm (Aynechi et al., 2011; Toma et al., 2009).

3.2 Accuracy discriminated by landmark

In this section we evaluate the performance of each descriptor for the different landmarks on an individual basis. This is done using the expected local accuracy $\bar{e}_L(r_S)$ defined by (Sukno et al., 2012). For each descriptor and landmark that is targeted, $\bar{e}_L(r_S)$ is computed as follows:

1. Start from an annotated set of shapes, in this case facial surfaces represented by meshes $\mathcal{M}_i$.
2. For every vertex $\mathbf{v} \in \mathcal{M}_i$ compute a descriptor *score*, $s(\mathbf{v})$, which measures how similar is the descriptor of vertex $\mathbf{v}$ to that of the landmark being targeted.
3. For every vertex $\mathbf{v} \in \mathcal{M}_i$ compute also the Euclidean distance to the *correct* position of the targeted landmark, say $d(\mathbf{v})$.

Table 2: Expected local accuracy [mm] for the different descriptors and landmarks. If a plateau is found, its value and limits are indicated, otherwise (n.p - no plateau) is indicated. For each landmark (rows), the best descriptor is highlighted in boldface as well as the ones with no statistically significant difference to it. The latter are further highlighted with an asterisk.

Lmk	HSC	USC	3DSC	APSC				
				D_{AR}	D_{AER}	A+E	A+R	A+D_{AER}
en (2)	**1.3*** **(2 - 24)**	1.9 (3 - 25)	1.4 (3 - 25)	1.5 (3 - 25)	1.5 (3 - 25)	**1.4*** **(3 - 24)**	**1.3** **(3 - 25)**	**1.3*** **(3 - 25)**
ex (2)	4.5 (16 - 90)	n.p	4.3 (13 - 88)	**2.9** **(6 - 67)**	3.9 (19 - 48)	5.4 (13 - 88)	4.7 (14 - 89)	**3.1*** **(8 - 88)**
n	1.8 (3 - 200)	4.6 (5 - 12)	**1.5** **(3 - 200)**	**1.6*** **(4 - 200)**	**1.6*** **(4 - 64)**	2.3 (4 - 200)	2.0 (3 - 200)	**1.7*** **(4 - 200)**
a (2)	**1.4*** **(3 - 26)**	n.p	**1.4** **(4 - 27)**	2.9 (6 - 12)	n.p	2.1 (4 - 25)	1.8 (4 - 26)	2.0 (6 - 26)
ac (2)	**2.1*** **(5 - 25)**	5.8 (14 - 25)	4.7 (9 - 25)	9.0 (16 - 24)	n.p	2.3 (7 - 25)	**2.1** **(4 - 11)**	5.1 (14 - 25)
nt (2)	**2.0** **(4 - 8)**	12.2 (14 - 200)	8.0 (14 - 200)	6.9 (12 - 200)	7.5 (11 - 200)	2.3 (5 - 8)	2.2 (5 - 9)	6.6 (11 - 200)
prn	1.4 (3 - 200)	1.4 (2 - 200)	**1.2** **(2 - 200)**	1.3 (3 - 200)	**1.3*** **(2 - 200)**	**1.3*** **(2 - 200)**	**1.3*** **(2 - 200)**	1.3 (3 - 200)
sn	1.8 (4 - 200)	n.p	**1.6** **(4 - 55)**	**1.8*** **(4 - 22)**	2.0 (5 - 16)	1.9 (3 - 200)	1.9 (3 - 200)	1.9 (4 - 200)
ch (2)	3.8 (11 - 22)	2.4 (4 - 42)	**2.1** **(5 - 19)**	2.5 (9 - 29)	2.9 (10 - 39)	2.8 (6 - 18)	2.9 (5 - 20)	**2.3*** **(5 - 28)**
cph (2)	**2.1** **(4 - 9)**	13.3 (20 - 34)	8.4 (18 - 200)	7.1 (17 - 86)	7.0 (16 - 59)	n.p	7.7 (16 - 200)	2.7 (5 - 8)
li	5.0 (16 - 51)	2.7 (7 - 48)	**2.3** **(5 - 10)**	4.4 (16 - 37)	3.4 (11 - 45)	4.9 (10 - 15)	4.8 (9 - 15)	3.8 (15 - 95)
ls	4.1 (6 - 14)	n.p	**2.3*** **(8 - 46)**	2.7 (8 - 13)	**2.2** **(6 - 11)**	5.2 (14 - 200)	5.7 (10 - 54)	3.8 (7 - 200)
sto	**2.7*** **(6 - 14)**	2.9 (8 - 46)	**2.2** **(8 - 78)**	**2.5*** **(7 - 17)**	6.1 (14 - 40)	4.0 (9 - 14)	4.5 (11 - 89)	3.1 (12 - 54)
sl	5.4 (10 - 54)	**3.0** **(10 - 18)**	**3.2*** **(11 - 27)**	5.5 (13 - 79)	7.4 (16 - 29)	4.7 (11 - 77)	6.0 (12 - 84)	6.2 (17 - 62)
pg	7.0 (10 - 200)	11.6 (19 - 120)	**5.4** **(10 - 200)**	7.9 (19 - 200)	7.1 (13 - 200)	7.6 (13 - 26)	**5.6*** **(13 - 23)**	**5.7*** **(10 - 200)**

4. For each $\mathcal{M}_i$ consider a neighborhood of radius r_S around the ground truth position of the targeted landmark and select $\mathbf{v}_i^{max}$ as the vertex with the maximum score in this neighborhood. Its distance to the ground truth is $d(\mathbf{v}_i^{max})$.
5. There is one value of $d(\mathbf{v}_i^{max})$ for each mesh; $\bar{e}_L(r_S)$ is their expected value over the test set:

$$\bar{e}_L(r_S) = E[d(\mathbf{v}_{i,r_S}^{max})] \tag{18}$$

$$\mathbf{v}_{i,r_S}^{max} = \{\mathbf{v} \in \mathcal{M}_i \,|\, d(\mathbf{v}) \leq r_S \wedge \forall \mathbf{w} \neq \mathbf{v}, \; d(\mathbf{w}) \leq r_S, \mathbf{w} \in \mathcal{M}_i : s(\mathbf{v}) \geq s(\mathbf{w})\} \tag{19}$$

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.

We used the negative Euclidean distance to a template as the descriptor score. The template for each landmark was computed as the median of descriptors over a training set. The training and test sets were obtained from the set of 144 facial scans described above by means of 6 fold cross validation.

An indicative example is provided in Figure 5, showing the obtained curves of $\bar{e}_L(r_S)$ for the nose corner using USC, 3DSC and APSC (computing patterns over A+E rings). The three curves show an initial growth of the error with the search radius until they reach a nearly flat region or *plateau*. This is the most important part of the curve because it provides

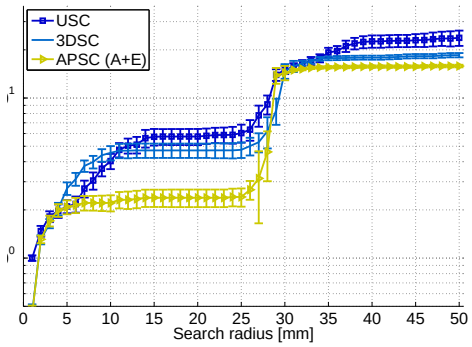

Figure 5: Average accuracy curves of USC, 3DSC and APSC (computing patterns over $A+E$ rings) targeting the nose corners (*ac*). Error bars indicate a 95% confidence interval.

both the accuracy and usable local range of the descriptor for the analyzed landmark. In other words, for search radii at which $\bar{e}_L(r_S)$ is flat the descriptor shows a stable behavior.

Hence, the first plateau is identified as the main feature of the local accuracy curves, allowing to characterize them with just three numbers: the value of $\bar{e}_L(r_S)$ at the plateau and the plateau limits, in terms of r_S (Sukno et al., 2012). Table 2 summarizes the results for all descriptors and landmarks.

Continuing with the example from Figure 5, it is interesting to analyze the behavior of $\bar{e}_L(r_S)$ for radii beyond the plateau: for the three descriptors in the plot there is a sudden increase of the error at radii between 25 and 30 mm. Typically this obeys to the presence of a strong source of false positives (i.e. points with very high score but not too close to the target landmark) at the distance where the error increase is observed. In this case, the source of false positives is the bilaterally symmetric point (i.e. *the other* nose corner), which is indeed typically located at 25 to 30 mm. This explains the strong coincidence in the upper plateau limits shown in Table 2 for nose corners (*ac*) or the inner eye-corners (*en*), as the bilaterally symmetric points are relatively close to each other.

The sources of false positives are not necessarily the symmetric point to the one targeted and depend on the descriptor that is used. There are also two special types of points: *i*) the ones without false positives in the analyzed range (which we set to 200 mm for the human face); *ii*) points that do not show a stable behavior in terms of $\bar{e}_L(r_S)$, which are indicated in Table 2 by *n.p* (no plateau).

From the results in Table 2 we can conclude that:

- For the majority of landmarks, at least one of the specific patterns of APSC that we tested showed comparable performance to the best descriptor.
- For eight landmarks (*ex*(2), *ac*(2), *nt*(2) and *cph*(2)) there were one or more APSC descriptors that significantly outperformed 3DSC. Interestingly, HSC also outperformed 3DSC for *ac*, *nt* and *cph*, but not for *ex*.
- There were four landmarks (*a*(2), *li* and *sl*) for which none of the tested APSC achieved sufficient performance when compared to 3DSC.
- The performance of APSC descriptors depending strongly on the spatial patterns that are considered. Jointly considering two rings produced lower errors than APSC derived from single rings.

3.3 Overall Accuracy

While the description of local accuracy curves based on the first plateau allows to simplify the comparison on a per-landmark basis, inferring the overall performance of a descriptor from Table 2 is not straightforward as the radii of the plateaus vary considerably for each landmark and descriptor.

Hence, in Figure 6 we provide curves of $\bar{e}_L(r_S)$ averaged over all 22 landmarks for each descriptor. Observe that 3DSC, HSC and the three APSC using patterns of two rings (A+E, A+R and A+D_{AER}) show very similar overall accuracy. Although we do not show error bars (to keep the plot as clear as possible), it is evident that the differences between these five descriptors are not statistically significant. On the other hand, USC and the two APSC based on a single ring (D_{AE} and D_{AER}) showed poorer performance.

Therefore, the plots in Figure 6 confirm that, in general, considering individual rings (either at constant radius and elevation or in diagonal form) implies a loss of important information, as all spatial relationships between different rings are not taken into account (Section 2.4). Nevertheless, Table 2 shows that for some particular cases this might not have an impact in local accuracy (e.g. *nasion*, *pronasale*, *labiale superius*) or might even be beneficial (*exocanthion*).

3.4 Implementation and complexity

Our implementations of 3DSC and USC are based on the Point Cloud Library (Rusu and Cousins, 2011) with some modifications to improve the computation speed by removing redundant operations and including multi-threading with OpenMP (Dagum and Menon, 1998). Additionally, a trilinear interpolation was included in the construction of the histograms as it was experimentally found to improve the performance of all tested descriptors.

It is interesting to analyze the sign disambiguation step when deriving the axes for USC: the ori-

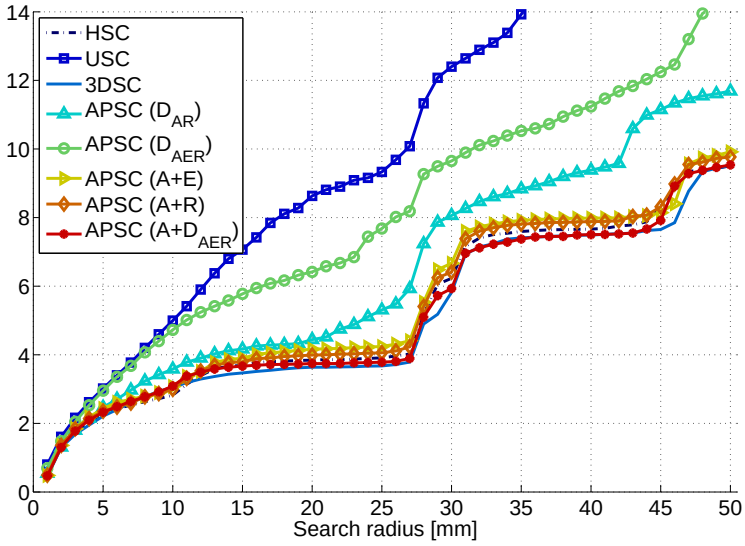

Figure 6: Average accuracy curves of all tested descriptors considering all landmarks together (by averaging).

entations of the generated normals[3] were not consistently pointing inwards or outwards for 30% to 35% of points. This is easy to verify and correct in our case as the input data are facial surfaces. The results reported in this paper include the correction of the reference frame orientation to ensure that all normals were pointing outwards of the object, which reduced the overall error of USC by approximately 10%. The latter suggests that similar inconsistencies might also exist in the sign of the other axes (and hence in the origin of the azimuth bins), which explains the lower accuracy of USC with respect to the other methods that were tested.

Regarding the computational complexity, there are two different aspects to consider: *i*) the computation of the descriptor and *ii*) the point-wise comparisons or matching.

The fastest descriptor to compute is 3DSC, as all the others are built from it plus some additional step. In the case of USC the additional step is dominated by an SVD on a neighborhood of the point of interest. For APSC and HSC the additional step is carried out based on the 3DSC bins and is therefore decoupled from the sampling density of the mesh. However, the computation of the histogram to build the 3DSC descriptor depends on the number of neighbors considered and, therefore, on the density of the mesh.

[3] As indicated in (Tombari et al., 2010) none of the reference axes derived for USC actually coincide with the true normal as the contribution of each point to the covariance matrix is weighted by the distance to the interest point. Nonetheless, this deviation with respect to the true normal is much smaller than the 180 degrees of a sign flip.

Table 3: Computational complexity of the descriptors relative to 3DSC.

Descriptor	Computation	Matching
HSC	11.1	$\frac{N_{SH}(N_{SH}+1)}{2N_E N_A^2}$
USC	1.23	$\frac{1}{N_A}$
APSC (D_{AR})	1.05	$\frac{1}{2N_A}$
APSC (D_{AER})		
APSC (A+E)	1.09	$\frac{1}{2N_A}$
APSC (A+R)		
APSC (A+D_{AER})		

The above hampers for an exact analysis of complexity. Thus, in Table 3 we provide numerical results for the computation time of the descriptors, in relative terms to the computation time of 3DSC, which in our experiments averaged 3.45 seconds on an Intel Xeon E5320 @1.86 GHz. Note that HSC was approximately an order of magnitude slower than all other descriptors as it required the decomposition of each fixed-radius shell into Spherical Harmonics. Assuming that the $N_{SH} \times (N_{SH}+1)/2$ basis functions are pre-computed, we still need to perform the projection into each of them of every shell, which roughly implies $N_A \times N_E$ complex multiplications and additions. Thus, the whole decomposition takes at least:

$$O\left(N_A N_E N_R \frac{N_{SH}(N_{SH}+1)}{2}\right) \qquad (20)$$

The above cost is considerably higher than the cost of computing APSC, which for each ring $\mathbf{m}$ takes

only $O(N_A^2/2)$ additions. Thus, if considering only single rings, the total complexity added by APSC to the computation of 3DSC is:

$$O\left(N_A N_E N_R \frac{N_A}{2}\right) \tag{21}$$

This cost grows linearly with the number of rings jointly considered, so it doubles for the last three rows of Table 3. Note that the complexities in (20) and (21) are not directly comparable as the first one is a lower bound based on complex additions and multiplications while the latter involves only real additions.

The matching time depends exclusively on the bins for all descriptors. Hence, the relative complexity to that of 3DSC can be easily derived and is shown in Table 3. Being the fastest to compute, 3DSC are also the slowest to match as they require to compute the N_A distances that correspond to all possible azimuth rotations, as in equation (3). All other descriptors are azimuth invariant and hence compute a single distance. In the case of HSC, as the number of bins is different to 3DSC, the relative computation time depends on the choice of N_{SH}, but approaches $(1/N_A)$ with the default parameters. Finally, all APSC have just half as many bins as 3DSC and USC which makes them the fastest to match.

4 DISCUSSION

From the results presented in the previous section we can conclude that APSC allows to construct descriptors that perform comparably to 3DSC, in terms of overall accuracy, with little extra load in the computation of the descriptor ($< 10\%$ in our experiments) and run several times faster during matching.

With respect to the previous alternatives to achieve azimuth-invariance in shape contexts, APSC showed similar accuracy to HSC at a much lower computational load (an order of magnitude) and outperformed USC both in terms of accuracy and speed.

However, the greatest potential of APSC is their flexibility to derive different descriptors depending on the spatial patterns that are selected to construct the sequences $\mathbf{m}$, from which asymmetry is extracted. As shown in Table 2, specific choices of spatial patterns might produce considerably lower errors than those obtained with 3DSC for certain landmarks.

The spatial patterns that were tested correspond to some straightforward definitions from a large set of possibilities. While the wrong choice of spatial patterns might negatively affect the performance, it would be expected that more elaborated strategies to choose these patterns, such as feature selection, would bring further improvement. While feature selection strategies would also be possible in 3DSC, the issue of azimuth ambiguity can considerably complicate the search of an optimal solution.

It might be argued that none of the tested APSC was optimal for all landmarks and a potential reduction of the error generalized through the majority of points would require different APSC to targeted different landmarks. Nonetheless such a strategy is possible and is analogous to previous works in landmark localization that adopt different features to localize each facial landmark (Gupta et al., 2010; Segundo et al., 2010). Moreover, combining two or more APSC can be far more efficient than combining other different descriptors, as the extra computation required would be rather marginal (all spatial patterns are extracted from the same 3DSC, which would be computed only once). For example, from Table 3 we see that all five APSC descriptors tested in this paper can be computed together with less than 1.4 times the computational load of a single 3DSC descriptor.

5 CONCLUSIONS

In this paper we present a new family of 3D geometric descriptors, Asymmetry Patterns Shape Contexts (APSC). These descriptors provide invariance to azimuth rotations to the popular 3D Shape Contexts (3DSC) by adapting a simple measure of rotational symmetry that has been recently proposed based on the overlap of a shape with rotated version of itself.

The asymmetry patterns are computed from sequences of bins extracted from a 3DSC descriptor by varying the azimuth index and, optionally, the radial and/or elevation bins. This allows to define different APSC descriptors to highlight or disable some of the spatial patterns present in the spherical grid of a 3DSC, which can be used to specialize the descriptor for different types of points.

We evaluated five examples of APSC in terms of local accuracy by targeting 22 craniofacial landmarks on a set of 144 facial scans. The accuracy was measured in terms of distance to ground truth consisting of expert annotations. Our results showed that APSC can achieve comparable overall accuracy to 3DSC, providing invariance to azimuth rotations at the expense of a small overhead in the computation of the descriptor, which did not exceed 10%. On the other hand the rotation invariance reduces the time required for matching two descriptors by twice the number of azimuth bins. APSC were also showed to perform better than previous approaches that provided azimuth invariance to shape contexts.

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues in the Face3D Consortium (www.face3d.ac.uk), and the financial support provided from the Wellcome Trust (grant 086901/Z/08/Z) and the Marie Curie IEF programme (grant 299605, SP-MORPH).

REFERENCES

Aynechi, N., Larson, B., Leon-Salazar, V., et al. (2011). Accuracy and precision of a 3D anthropometric facial analysis with and without landmark labeling before image acquisition. *Angle Orthod*, 81(2):245–252.

Bariya, P., Novatnack, J., Schwartz, G., et al. (2012). 3D geometric scale variability in range images: features and descriptors. *Int J Comput Vis*, 99(2):232–255.

Bronstein, A., Bronstein, M., Castellani, U., et al. (2010). SHREC 2010: robust correspondence benchmark. In *Eurographics Workshop on 3D Object Retrieval*.

Chen, H. and Bhanu, B. (2007). 3D free-form object recognition in range images using local surface patches. *Pattern Recogn Lett*, 28(10):1252–1262.

Creusot, C., Pears, N., and Austin, J. (2011). Automatic keypoint detection on 3D faces using a dictionary of local shapes. In *Proc. 3DIMPVT*, pages 204–211.

Dagum, L. and Menon, R. (1998). OpenMP: an industry standard API for shared-memory programming. *IEEE Computat Sci Eng*, 5(1):46–55.

Farkas, L. (1994). *Anthropometry of the head and face*. Raven Press (New York), 2nd ed.

Frome, A., Huber, D., Kolluri, R., et al. (2004). Recognizing objects in range data using regional point descriptors. In *Proc. ECCV*, pages 224–237.

Guo, Q., Guo, F., and Shao, J. (2010). Irregular shape symmetry analysis: Theory and application to quantitative galaxy classification. *IEEE T Pattern Anal*, 32(10):1730–1743.

Gupta, S., Markey, M., and Bovik, A. (2010). Antopometric 3D face recognition. *Int J Comput Vis*, 90(3):331–349.

Hennessy, R., Kinsella, A., and Waddington, J. (2002). 3D laser surface scanning and geometric morphometric analysis of craniofacial shape as an index of cerebro-craniofacial morphogenesis. *Biol Psychiat*, 51(6):507–514.

Johnson, A. and Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE T Pattern Anal*, 21(5):433–449.

Kazhdan, M., Funkhouser, T., and Rusinkiewicz, S. (2004). Symmetry descriptors and 3D shape matching. In *Eurograph. Symp. on Geometry process*, pages 115–123.

Kortgen, M., Park, G., Novotni, M., et al. (2003). 3D shape matching with 3D shape contexts. In *Central Europ Seminar on Comput Graph*.

Passalis, G., Perakis, N., Theoharis, T., et al. (2011). Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE T Pattern Anal*, 33(10):1938–1951.

Rusu, R., Blodow, N., and Beetz, M. (2009). Fast point feature histograms (FPFH) for 3D registration. In *Proc. ICRA*, pages 3212–3217.

Rusu, R. and Cousins, S. (2011). 3D is here: Point cloud library (PCL). In *Proc. ICRA*, pages 1–4,

Segundo, M., Silva, L., Bellon, O. P., et al. (2010). Automatic face segmentation and facial landmark detection in range images. *IEEE T Syst Man Cy B*, 40(5):1319–1330.

Steder, B., Rusu, R., Konolige, K., et al. (2011). Point feature extraction on 3D range scans taking into account object boundaries. In *Proc. ICRA*, pages 2601–2608.

Sukno, F., Waddington, J., and Whelan, P. (2012). Comparing 3D descriptors for local search of craniofacial landmarks. In *Proc. ISVC*, pages 92–103.

Toma, A., Zhurov, A., Playle, R., et al. (2009). Reproducibility of facial soft tissue landmarks on 3D lasser-scanned facial images. *Orthod Craniofac Res*, 12(1):33–42.

Tombari, F., Salti, S., and Stefano, L. D. (2010). Unique shape context for 3D data description. In *Proc. ACM Workshop on 3D object retrieval*, pages 57–62.

Zaharescu, A., Boyer, E., and Horaud, R. (2012). Keypoints and local descriptors of scalar functions on 2D manifolds. *Int J Comput Vision*, 99(2):232–255.

Zhang, Y. (2009). Intrinsic shape signatures: a shape descriptor for 3D object recognition. In *Proc. ICCV Workshops*, pages 689–696.

APPENDIX

Starting from the definition of $\mathcal{A}$ and $\mathcal{S}$:

$$\mathcal{A}(\mathbf{m},a) = 1 - \mathcal{S}(\mathbf{m},a) = 1 - \frac{\sum_j \min(m_j, m_{j+a})}{\sum_j m_j} = \frac{\sum_j m_j - \sum_j \min(m_j, m_{j+a})}{\sum_j m_j} \tag{22}$$

Now we use the following equality:

$$2\sum_j m_j = \sum_j \big(\max(m_j, m_{j+a}) + \min(m_j, m_{j+a})\big) \tag{23}$$

which holds because for every pair (m_j, m_{j+a}) one element is the maximum and the other one the minimum, hence adding both guarantees to include each element of $\mathbf{m}$ exactly twice in the summation (recall that $j+a$ is an addition module the cardinality of $\mathbf{m}$). Then, in (22):

$$\begin{aligned} &\sum_j m_j - \sum_j \min(m_j, m_{j+a}) = \\ &= \frac{1}{2}\sum_j \big(\max(m_j, m_{j+a}) - \min(m_j, m_{j+a})\big) \\ &= \frac{1}{2}\sum_j |m_j - m_{j+a}| \end{aligned} \tag{24}$$

which directly leads to our final result:

$$\mathcal{A}(\mathbf{m},a) = \frac{1}{2}\frac{\sum_j |m_j - m_{j+a}|}{\sum_j m_j} \tag{25}$$