

# Rotationally Invariant 3D Shape Contexts Using Asymmetry Patterns

Federico M. Sukno<sup>1,2</sup>, John L. Waddington<sup>2</sup> and Paul F. Whelan<sup>1</sup>

<sup>1</sup>Centre for Image Processing & Analysis, Dublin City University, Dublin 9, Ireland

<sup>2</sup>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: Supplementary materials from the paper in GRAPP 2013.

## COMPLETE RESULTS FOR APSC BASED ON ADJACENT AND DIAGONAL PATTERNS

Below we provide additional results from the ones reported in the paper, corresponding to all Asymmetry Patterns Shape Contexts (APSC) 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. A brief explanation is provided below; more details can be found in the main paper (Sukno et al., 2013).

### Asymmetry Patterns

The asymmetry patterns are obtained by considering all possible azimuth rotations of the sequence that generate distinct values:

$$P_A(\mathbf{m}) = A_1(\mathbf{m}, 1), A_1(\mathbf{m}, 2), \dots, A_1(\mathbf{m}, \lfloor \frac{N_A}{2} \rfloor) \quad (1)$$

where  $N_A$  is the number of azimuth bins and  $\lfloor x \rfloor$  is the integer part of  $x$ . Defined in this way the asymmetry pattern accounts for approximately half the possible rotations, because the remaining ones would only generate repeated values. Also,  $A_1(\mathbf{m}, a)$  is the asymmetry of sequence  $\mathbf{m}$  after a cyclic shift of  $a$ , equivalent to an azimuth rotation of  $2\pi/N_A$ , which can be computed as follows:

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

The sequence  $\mathbf{m}$  to generate the asymmetry patterns is always obtained starting from a 3D Shape Context (3DSC) descriptor  $\mathbf{x}$ . The simplest example

is to use 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:

$$m_j = x_{i,j,k}, \quad m_j \geq 0 \forall j \in [1; N_A] \quad (3)$$

The definition of  $\mathbf{m}$  controls what spatial relations in the 3DSC are considered and which are discarded. 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 (3) is just a specific choice of  $\mathbf{m}$  that leads to one of many possible APSC. A few straightforward alternatives include:

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

$$m_j = x_{i+j,j,k} \quad (4)$$

$$m_j = x_{i+j,j,k+j} \quad (5)$$

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

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

$$A_2(\mathbf{m}, a) = \sum_j |m_{1,j} - m_{1,j+a}| + |m_{2,j} - m_{2,j+a}| \quad (6)$$

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  $k$ ).

<sup>1</sup>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.

Table 1: Expected local accuracy [mm] for the different APSC 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	1-ring APSC				2-ring APSC				
	A <sub>0</sub>	D <sub>AE</sub>	D <sub>AR</sub>	D <sub>AER</sub>	A+E	A+R	A+D <sub>AE</sub>	A+D <sub>AR</sub>	A+D <sub>AER</sub>
en (2)	<b>1.3</b> (3 - 25)	1.7 (4 - 23)	1.5 (3 - 25)	1.5 (3 - 25)	1.4 (3 - 24)	<b>1.3*</b> (3 - 25)	1.5 (5 - 24)	<b>1.3*</b> (3 - 25)	<b>1.3*</b> (3 - 25)
ex (2)	4.8 (14 - 90)	n.p (< 9)	<b>2.9</b> (6 - 67)	3.9 (19 - 48)	5.4 (13 - 88)	4.7 (14 - 89)	3.7 (8 - 63)	<b>3.1*</b> (6 - 87)	<b>3.1*</b> (8 - 88)
n	2.1 (4 - 200)	3.1 (13 - 34)	<b>1.6</b> (4 - 200)	<b>1.6*</b> (4 - 64)	2.3 (4 - 200)	2.0 (3 - 200)	2.0 (5 - 200)	1.9 (4 - 200)	<b>1.7*</b> (4 - 200)
a (2)	<b>1.9*</b> (4 - 27)	n.p (< 77)	2.9 (6 - 12)	n.p (< 200)	2.1 (4 - 25)	<b>1.8*</b> (4 - 26)	n.p (< 26)	<b>1.8</b> (4 - 24)	2.0 (6 - 26)
ac (2)	2.4 (7 - 26)	n.p (< 23)	9.0 (16 - 24)	n.p (< 22)	2.3 (7 - 25)	<b>2.1</b> (4 - 11)	4.8 (15 - 24)	4.9 (13 - 26)	5.1 (14 - 25)
nt (2)	<b>2.2</b> (4 - 8)	9.0 (12 - 45)	6.9 (12 - 200)	7.5 (11 - 200)	<b>2.3*</b> (5 - 8)	<b>2.2*</b> (5 - 9)	7.4 (12 - 200)	n.p (< 8)	6.6 (11 - 200)
prn	<b>1.2</b> (2 - 200)	1.5 (3 - 200)	<b>1.3*</b> (3 - 200)	<b>1.3*</b> (2 - 200)	<b>1.3*</b> (2 - 200)	<b>1.3*</b> (2 - 200)	1.5 (3 - 200)	<b>1.3*</b> (3 - 200)	<b>1.3*</b> (3 - 200)
sn	2.0 (4 - 200)	2.4 (5 - 14)	<b>1.8</b> (4 - 22)	2.0 (5 - 16)	<b>1.9*</b> (3 - 200)	<b>1.9*</b> (3 - 200)	2.3 (6 - 200)	<b>1.8*</b> (3 - 200)	<b>1.9*</b> (4 - 200)
ch (2)	2.9 (6 - 17)	5.3 (14 - 19)	<b>2.5*</b> (9 - 29)	2.9 (10 - 39)	2.8 (6 - 18)	2.9 (5 - 20)	2.9 (13 - 38)	<b>2.4*</b> (5 - 25)	<b>2.3</b> (5 - 28)
cph (2)	n.p (< 10)	7.3 (16 - 33)	7.1 (17 - 86)	7.0 (16 - 59)	n.p (< 11)	7.7 (16 - 200)	8.0 (18 - 200)	7.2 (14 - 200)	<b>2.7</b> (5 - 8)
li	5.0 (13 - 34)	<b>3.8*</b> (10 - 24)	4.4 (16 - 37)	<b>3.4</b> (11 - 45)	4.9 (10 - 15)	4.8 (9 - 15)	4.6 (17 - 47)	5.1 (19 - 76)	<b>3.8*</b> (15 - 95)
ls	5.5 (14 - 200)	3.8 (12 - 37)	2.7 (8 - 13)	<b>2.2</b> (6 - 11)	5.2 (14 - 200)	5.7 (10 - 54)	5.1 (15 - 200)	3.8 (7 - 12)	3.8 (7 - 200)
sto	4.4 (12 - 200)	9.9 (14 - 26)	<b>2.5*</b> (7 - 17)	6.1 (14 - 40)	4.0 (9 - 14)	4.5 (11 - 89)	5.4 (20 - 80)	<b>2.5</b> (8 - 17)	3.1 (12 - 54)
sl	5.5 (11 - 85)	n.p (< 18)	5.5 (13 - 79)	7.4 (16 - 29)	<b>4.7</b> (11 - 77)	6.0 (12 - 84)	<b>5.3*</b> (11 - 87)	<b>5.0*</b> (11 - 36)	6.2 (17 - 62)
pg	<b>6.1*</b> (12 - 22)	8.0 (15 - 68)	7.9 (19 - 200)	7.1 (13 - 200)	7.6 (13 - 26)	<b>5.6</b> (13 - 23)	7.0 (12 - 200)	7.1 (18 - 200)	<b>5.7*</b> (10 - 200)

## Data

Our test dataset consisted of 144 facial scans acquired by means of a hand-held laser scanner (FastSCAN<sup>TM</sup>, 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) from which we target the 22 points indicated in Figure 1.

## Accuracy evaluation

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  $M_i$ .
2. For every vertex  $\mathbf{v} \in M_i$  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  $\mathbf{v} \in M_i$  compute also the Euclidean distance to the *correct* position of the tar-

Table 2: 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
$A_0$	$m_j = x_{i,j,k}$	Azimuth ring
$D_{AE}$	$m_j = x_{i+j,j,k}$	Azimuth-Elevation diagonal
$D_{AR}$	$m_j = x_{i,j,k+j}$	Azimuth-Radius diagonal
$D_{AER}$	$m_j = x_{i+j,j,k+j}$	Azimuth-Elevation-Radius diagonal
$A+E$	$m_{1,j} = x_{i,j,k}, m_{2,j} = x_{i+1,j,k}$	Azimuth ring + Elevation neighbors
$A+R$	$m_{1,j} = x_{i,j,k}, m_{2,j} = x_{i,j,k+1}$	Azimuth ring + Radial neighbors
$A+D_{AE}$	$m_{1,j} = x_{i,j,k}, m_{2,j} = x_{i+j,j,k}$	Azimuth ring + Azim-Elev diagonal
$A+D_{AR}$	$m_{1,j} = x_{i,j,k}, m_{2,j} = x_{i,j,k+j}$	Azimuth ring + Azim-Rad diagonal
$A+D_{AER}$	$m_{1,j} = x_{i,j,k}, m_{2,j} = x_{i+j,j,k+j}$	Azimuth ring + Azim-Elev-Rad diagonal

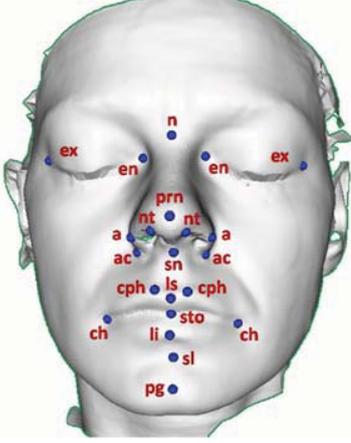


Figure 1: Example of the facial scans from the test dataset with the annotation of the 22 landmarks used in this study:  $en$  = endocanthion;  $ex$  = exocanthion;  $n$  = nasion;  $a$  = alare;  $ac$  = alar crest;  $nt$  = nostril top;  $prn$  = pronasale;  $sn$  = subnasale;  $ch$  = cheilion;  $cph$  = crista philtrum;  $li$  = labiale inferius;  $ls$  = labiale superius;  $sto$  = stomion;  $sl$  = sublabiale;  $pg$  = pogonion; (Hennessy et al., 2002).

geted landmark, say  $d(\mathbf{v})$ .

- For each  $M_i$  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  $d(\mathbf{v}_i^{max})$ .
- 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:

$$\bar{e}_L(r_S) = E[d(\mathbf{v}_{i,r_S}^{max})] \quad (7)$$

$$\mathbf{v}_{i,r_S}^{max} = \{\mathbf{v} \in M_i \mid d(\mathbf{v}) \leq r_S \wedge \forall \mathbf{w} \neq \mathbf{v}, d(\mathbf{w}) \leq r_S, \mathbf{w} \in M_i : s(\mathbf{v}) \geq s(\mathbf{w})\} \quad (8)$$

where  $E[x]$  is the expected value of  $x$ . That is, given a target landmark, for each mesh  $M_i$  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.

Typically, for a good descriptors the curves of overall accuracy against the search radius have one or more *plateaus*, the first of which 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 1 summarizes the results for all descriptors and landmarks.

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.

## Experimental settings

In all cases we used the default configuration for 3DSC (Frome et al., 2004):  $N_E = 11$  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.

We performed tests using the fixed elevation and radius rings (azimuth rings, for short) as defined in (3) 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. The corresponding equations are provided in Table 2. Results are summarized in Table 1.

All sequences in Table 2 are computed starting from a 3DSC descriptor  $\mathbf{x}$ , whose elements are indexed by  $(i, j, k) = (\text{elevation, azimuth, radius})$ . 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 se-

quences considered jointly (bottom five rows of the table), they are combined to generate the asymmetry pattern as indicated in (6). 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.

## REFERENCES

- Frome, A., Huber, D., Kolluri, R., et al. (2004). Recognizing objects in range data using regional point descriptors. In *Proc. ECCV*, pages 224–237.
- 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.
- Sukno, F., Waddington, J., and Whelan, P. (2012). Comparing 3D descriptors for local search of craniofacial landmarks. In *Proc. ISVC*, pages 92–103.
- Sukno, F., Waddington, J., and Whelan, P. (2012). Rotationally Invariant 3D Shape Contexts Using Asymmetry Patterns. In *Proc. GRAPP 2013*, in press.