

Asymmetry Patterns Shape Contexts to Describe the 3D Geometry of Craniofacial Landmarks

Federico M. Sukno^{1,2(✉)}, John L. Waddington², and Paul F. Whelan¹

¹ Centre for Image Processing and Analysis,
Dublin City University, Dublin 9, Ireland
`federico.sukno@gmail.com`

² Molecular and Cellular Therapeutics,
Royal College of Surgeons in Ireland, Dublin 2, Ireland

Abstract. We present a new family of 3D geometry descriptors based on asymmetry patterns from the popular 3D Shape Contexts (3DSC). Our approach resolves the azimuth ambiguity of 3DSC, thus providing rotational invariance, at the expense of a marginal increase in computational load, outperforming previous algorithms dealing with 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. The possibility to define APSC descriptors by selecting diverse spatial patterns from a 3DSC has two important advantages: (1) choosing the appropriate spatial patterns can considerably reduce the errors obtained with 3DSC when targeting specific types of points; (2) Once one APSC descriptor is built, additional ones can be built with only incremental cost. Therefore, it is possible to use a pool of APSC descriptors to maximize accuracy without a large increase in computational cost.

Keywords: 3D geometric descriptors · Rotational symmetry · Craniofacial landmarks

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 [26], recognition [9], surface registration [2] and landmark identification [5, 16].

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 [4, 13, 17, 28] or include additional functions that are attached to the geometry, such as radiometric information [20, 27].

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. A recent comparison of geometric descriptors in the context of craniofacial landmark localization highlighted 3D shape contexts as one of the most accurate algorithms [21].

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 [9], 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 based on symmetry [14], 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) [9].
- A third option [15, 26] 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 [9]. 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 [9] 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 [26]. However, USC was found considerably less accurate than 3DSC when targeting

specific points on a craniofacial landmark localization task [21]. 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, it cannot be assured that the directions determined by the proposed disambiguation step are consistent across a population of facial scans. Since USC rely on the unique definition of azimuth bins, the lack of consistency has an important effect on accuracy [23].

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 [10], 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 on 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. Thus, we define Asymmetry Patterns Shape Contexts (APSC) [23] 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. The computation of an APSC descriptor is slightly more expensive than a single 3DSC but produces considerable savings in matching time 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 [14].

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

2 Asymmetry Patterns Shape Contexts

Computation of the APSC descriptor starts by computing a 3DSC descriptor [9], 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 $\mathcal{N}_v = \{\mathbf{w} \mid \|\mathbf{w} - \mathbf{v}\| \leq r_N\}$, 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:

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

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 each point is normalization by bin volume and sampling density.

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:

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

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. [10] presented continuous measures of approximate bilateral and rotational symmetry. 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} :

$$\mathcal{S}_c(\mathbf{m}, \phi) = \frac{\text{Area}(\mathbf{m} \cap R(\mathbf{m}, \phi))}{\text{Area}(\mathbf{m})} \quad (3)$$

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:

$$m_j = x_{i,j,k}, \quad m_j \geq 0 \forall j \in [1; N_A] \quad (4)$$

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

$$\mathcal{S}(\mathbf{m}, a) = \frac{\sum_j \min(m_j, m_{j+a})}{\sum_j m_j} \quad (5)$$

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 .

By considering asymmetry as the complement of symmetry, we obtain a particularly simple formulation:

$$\begin{aligned} \mathcal{A}(\mathbf{m}, a) &= 1 - \mathcal{S}(\mathbf{m}, a) = \frac{\sum_j m_j - \sum_j \min(m_j, m_{j+a})}{\sum_j m_j} \\ &= \frac{\frac{1}{2} \sum_j (\max(m_j, m_{j+a}) - \min(m_j, m_{j+a}))}{\sum_j m_j} = \frac{\sum_j |m_j - m_{j+a}|}{2 \sum_j m_j} \end{aligned} \quad (6)$$

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 [23].

The asymmetry degree $\mathcal{A}(\mathbf{m}, a)$ defined in (6) 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:

$$\mathcal{A}_1(\mathbf{m}, a) = \sum_j |m_j - m_{j+a}| \quad (7)$$

2.3 Asymmetry Patterns

We obtain asymmetry patterns by considering all possible azimuth rotations of the sequence that generate distinct values:

$$\mathcal{P}_A(\mathbf{m}) = \mathcal{A}_1(\mathbf{m}, 1), \mathcal{A}_1(\mathbf{m}, 2), \dots, \mathcal{A}_1(\mathbf{m}, \lfloor \frac{N_A}{2} \rfloor) \quad (8)$$

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

$$\forall a \in [1; N_A] : \mathcal{A}_1(\mathbf{m}, a) = \mathcal{A}_1(\mathbf{m}, -a) = \mathcal{A}_1(\mathbf{m}, N_A - a) \quad (9)$$

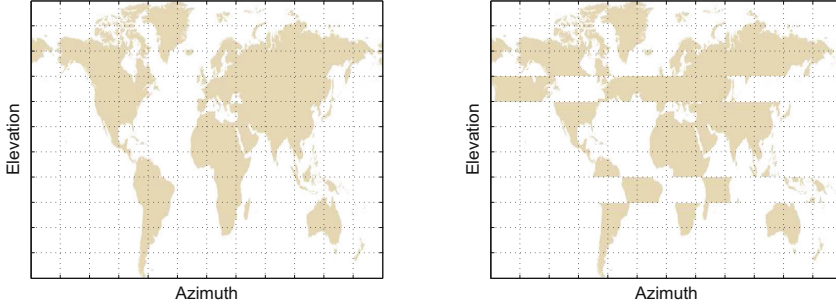


Fig. 1. Left: the World map as an example of a shell with (ideally) constant radius. The azimuth (longitude) and elevation (latitude) bins are also indicated. Right: the same spherical shell after arbitrary and independent azimuth rotations of two rings with constant elevation (4-th and 8-th bins).

where both addition and subtraction are modulo N_A operations. Intuitively, this can be understood from the definition as an overlap between \mathbf{m} and a rotated version of itself by an angle $\phi = 2\pi a/N_A$ 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 $R(\mathbf{m}, 2\pi - \phi)$.

Thus, the sequence $\mathcal{P}_A(\mathbf{m})$ is the asymmetry pattern of the sequence \mathbf{m} , indicating how asymmetric is the ring that originated \mathbf{m} for different angles of azimuth rotation. However, from the definition of $\mathcal{A}(\mathbf{m}, a)$, it is clear that the generated pattern is invariant to the origin chosen for the azimuth bins, i.e.

$$\mathcal{P}_A(\mathbf{m}) = \mathcal{P}_A(R(\mathbf{m}, a)), \forall a \in \mathbb{Z} \quad (10)$$

2.4 Spatial Relationships

So far, we have worked with a sequence \mathbf{m} defined as in (4), 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 $(i - k)$ ring of a 3DSC descriptor \mathbf{x} into an asymmetry pattern that is invariant to the choice of azimuth.

However, such a choice takes into account only the spatial relationships within each $(i - k)$ 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 shown in Fig. 1, where latitude is the elevation and longitude 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 Fig. 1(left) and the example with shifted rings in Fig. 1(right) will generate the same set of patterns. A similar reasoning can be applied to the relation between shells of different radii. While this is not

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
A	$m_j = x_{i,j,k}$	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
$A + E$	$m_{1,j} = x_{i,j,k}, \quad m_{2,j} = x_{i+1,j,k}$	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
$A + D_{AE}$	$m_{1,j} = x_{i,j,k}, \quad m_{2,j} = x_{i+j,j,k}$	Azimuth ring + Az-Elev diagonal
$A + D_{AR}$	$m_{1,j} = x_{i,j,k}, \quad m_{2,j} = x_{i,j,k+j}$	Azimuth ring + Az-Rad diagonal
$A + D_{AER}$	$m_{1,j} = x_{i,j,k}, \quad m_{2,j} = x_{i+j,j,k+j}$	Azimuth ring + Az-Elev-Rad diag

necessarily negative (e.g. it might be useful to *disable* certain spatial relations), in the general case it can lead to a loss of discriminant information.

The choice of what spatial relations are considered is related to the definition of \mathbf{m} . Straight-forward alternatives include considering adjacent rings (either in elevation or radius) or, if relaxing the rotational invariance requirement, *diagonal rings*, i.e. jointly changing the bin indexes of azimuth with radius and/or elevation. In Table 1 we indicate the sequences to generate eight other simple patterns resulting from combinations of diagonals, adjacent rings and azimuth rings. When jointly considering two rings, e.g. $A + E$, $A + D_{AR}$, the overlap is computed only between rings with the same definition:

$$\mathcal{A}_2(\mathbf{m}, a) = \sum_j |m_{1,j} - m_{1,j+a}| + |m_{2,j} - m_{2,j+a}| \quad (11)$$

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 thought of as a family of descriptors with a flexible definition that allows their adaptation so as to highlight or disable specific spatial relationships.

3 Experimental Evaluation

In this section we compare the performance of APSC to the following three algorithms, which constitute competing alternatives:

- 3DSC [9], which generate descriptors that are not invariant to azimuth rotations.
- HSC [9], 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.

- USC [26], 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.

For APSC, we test the 9 descriptors indicated in Table 1, 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 full coverage of the bins of \mathbf{x} . For each sequence of length $N_A = 12$ bins, an asymmetry pattern of length $\lfloor N_A/2 \rfloor = 6$ is generated. Thus, each APSC descriptor has only $N_E \times N_A \times 6 = 990$ bins.

3.1 Craniofacial Landmarks

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 instances obtained by modifying a given example by some set of transformations [3, 20, 26].

Our test dataset consisted of 144 facial scans acquired by means of a hand-held laser scanner¹. 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 [12] (based on [8]), from which we target the 22 points indicated in Fig. 2(a).

The fact that the test dataset was acquired in the context of clinical research makes it especially suited for tests of localization accuracy. As can be observed in Fig. 2(a), these are high quality scans that have been carefully annotated by experts based on anthropometric definitions. Recent studies indicate that the intra- and inter-observer uncertainty of this type of annotation are typically between 1 mm and 2 mm [1, 25].

¹ Polhemus FastSCANTM, Colchester, VT, USA. An example is available at http://www.cipa.dcu.ie/videos/face3d/Scanning_DCU_RCSI.avi [Accessed on 20.05.2013].

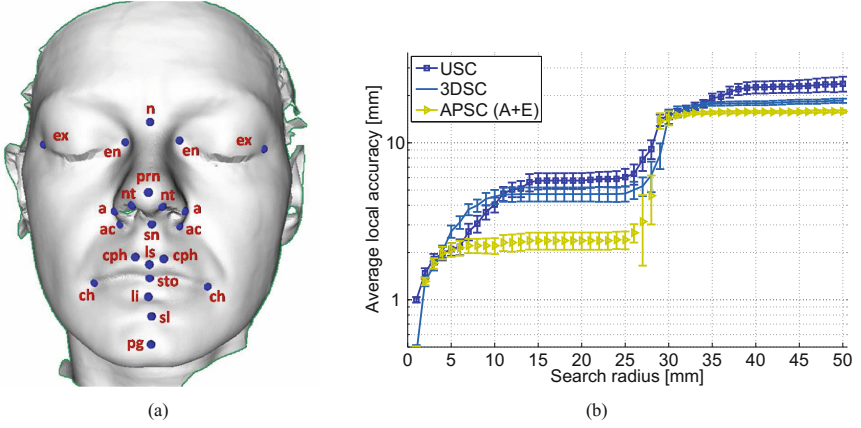


Fig. 2. (a) A scan from the clinical dataset with the 22 landmarks used in this study: *en* = endocanthion; *ex* = exocanthion; *n* = nasion; *a* = alare; *ac* = alar crest; *nt* = nostril top; *prn* = pronasale; *sn* = subnasale; *ch* = cheilion; *cph* = crista philtrum; *li* = labiale inferius; *ls* = labiale superius; *sto* = stomion; *sl* = sublabiale; *pg* = pogonion; [12]; (b) average accuracy curves of USC, 3DSC and APSC (using *A + E* rings) targeting the nose corners (*ac*). Error bars indicate a 95 % confidence interval.

3.2 Accuracy

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 $\bar{e}_L(r_S)$, as defined in [21]. For each descriptor and landmark that is targeted, $\bar{e}_L(r_S)$ is computed as follows:

1. Start from an annotated set of 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 targeted landmark.
3. For every vertex $\mathbf{v} \in \mathcal{M}_i$ compute also the Euclidean distance to the *correct* position of the targeted landmark, say $d(\mathbf{v})$.
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 highest 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:

$$\bar{e}_L(r_S) = E[d(\mathbf{v}_{i,r_S}^{max})] \quad \mathbf{v}_{i,r_S}^{max} = \underset{\mathbf{v} \in \mathcal{N}_{i,r_S}}{\operatorname{argmax}} (s(\mathbf{v})) \quad (12)$$

$$\mathcal{N}_{i,r_S} = \{\mathbf{v} \in \mathcal{M}_i \mid d(\mathbf{v}) \leq r_S\} \quad (13)$$

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 Fig. 2(b), 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 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 stable behavior. Hence, the first plateau is identified as the main feature of the local accuracy curves, allowing their characterization with just three numbers: the value of $\bar{e}_L(r_S)$ at the plateau and the plateau limits in terms of r_S [21]. Table 2 summarizes the results for HSC, USC, 3DSC and 5 selected APSC patterns².

Continuing with the example from Fig. 2(b), 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 in the error at radii between 25 and 30 mm. Typically this is caused by the presence of a strong source of false positives (i.e. points with a very high score but not close to the targeted 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), 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); (ii) points that do not show 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 strongly depends on the spatial patterns that are considered. Jointly considering two rings produced lower errors than APSC derived from single rings.

² The complete results for all patterns listed in Table 1 are available at http://www.cipa.dcu.ie/pubs_full.html [Accessed on 15.07.2013].

Table 2. Expected local accuracy [mm]. 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 those with no statistically significant difference from 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	1.3*	1.9	1.4	1.5	1.5	1.4*	1.3	1.3*
(2)	(2–24)	(3–25)	(3–25)	(3–25)	(3–25)	(3–24)	(3–25)	(3–25)
ex	4.5	n.p	4.3	2.9	3.9	5.4	4.7	3.1*
(2)	(16–90)		(13–88)	(6–67)	(19–48)	(13–88)	(14–89)	(8–88)
n	1.8	4.6	1.5	1.6*	1.6*	2.3	2.0	1.7*
	(3–200)	(5–12)	(3–200)	(4–200)	(4–64)	(4–200)	(3–200)	(4 – 200)
a	1.4*	n.p	1.4	2.9	n.p	2.1	1.8	2.0
(2)	(3–26)		(4–27)	(6–12)		(4–25)	(4–26)	(6–26)
ac	2.1*	5.8	4.7	9.0	n.p	2.3	2.1	5.1
(2)	(5–25)	(14–25)	(9–25)	(16–24)		(7–25)	(4–11)	(14–25)
nt	2.0	12.2	8.0	6.9	7.5	2.3	2.2	6.6
(2)	(4–8)	(14–200)	(14–200)	(12–200)	(11–200)	(5–8)	(5–9)	(11–200)
prn	1.4	1.4	1.2	1.3	1.3*	1.3*	1.3*	1.3
	(3–200)	(2–200)	(2–200)	(3–200)	(2–200)	(2–200)	(2–200)	(3–200)
sn	1.8	n.p	1.6	1.8*	2.0	1.9	1.9	1.9
	(4–200)		(4–55)	(4–22)	(5–16)	(3–200)	(3–200)	(4–200)
ch	3.8	2.4	2.1	2.5	2.9	2.8	2.9	2.3*
(2)	(11–22)	(4–42)	(5–19)	(9–29)	(10–39)	(6–18)	(5–20)	(5–28)
cph	2.1	13.3	8.4	7.1	7.0	n.p	7.7	2.7
(2)	(4–9)	(20–34)	(18–200)	(17–86)	(16–59)		(16–200)	(5–8)
li	5.0	2.7	2.3	4.4	3.4	4.9	4.8	3.8
	(16–51)	(7–48)	(5–10)	(16–37)	(11–45)	(10–15)	(9–15)	(15–95)
ls	4.1	n.p	2.3*	2.7	2.2	5.2	5.7	3.8
	(6–14)		(8–46)	(8–13)	(6–11)	(14–200)	(10–54)	(7–200)
sto	2.7*	2.9	2.2	2.5*	6.1	4.0	4.5	3.1
	(6–14)	(8–46)	(8–78)	(7–17)	(14–40)	(9–14)	(11–89)	(12–54)
sl	5.4	3.0	3.2*	5.5	7.4	4.7	6.0	6.2
	(10–54)	(10–18)	(11–27)	(13–79)	(16–29)	(11–77)	(12–84)	(17–62)
pg	7.0	11.6	5.4	7.9	7.1	7.6	5.6*	5.7*
	(10–200)	(19–120)	(10–200)	(19–200)	(13–200)	(13–26)	(13–23)	(10–200)

In global terms, averaging $\bar{e}_L(r_S)$ over all 22 landmarks for each descriptor, 3DSC, HSC and the APSC descriptors using patterns of two rings showed very similar overall accuracy. On the other hand, USC and the two APSC based on a single ring showed poorer performance [23]. This confirms that, in general, considering individual rings implies a loss of important information, as all spa-

tial relationships between different rings are not taken into account (Sect. 2.4). Nevertheless, Table 2 shows that for some particular cases this might not have an impact in local accuracy (e.g. n , prn , ls) or might even be beneficial (ex).

3.3 Implementation and Complexity

Our implementations of 3DSC and USC are based on the Point Cloud Library [18] with some modifications to improve the computation speed by removing redundant operations and including multi-threading with OpenMP [7]. 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 orientations of the generated normals were not consistently pointing inwards or outwards for 30 %–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 from 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 computational complexity, there are two aspects to consider: (i) computation of the descriptor and (ii) 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.

The above hampers our ability to produce an exact analysis of complexity. Thus, in Table 3 we provide numerical results for the computation time of the descriptors, relative to the computation time of 3DSC, which in our experiments averaged 3.45 s 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 project each shell into each basis function, which roughly implies $N_A \times N_E$ complex multiplications and additions. Thus, the whole decomposition takes at least

$$O\left(N_A N_E N_R \frac{N_{SH} (N_{SH} + 1)}{2}\right) \quad (14)$$

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

Table 3. Computational complexity of the descriptors relative to 3DSC = 1.

Descriptor	Computation	Matching
HSC	11.1	$\frac{N_{SH}(N_{SH}+1)}{2 N_E N_A^2}$
USC	1.23	$\frac{1}{N_A}$
1-ring APSC	1.05	$\frac{1}{2 N_A}$
2-ring APSC	1.09	$\frac{1}{2 N_A}$

single rings, the complexity added by APSC to the computation of 3DSC is $O(N_E N_R N_A^2/2)$. This cost grows linearly with the number of rings jointly considered, so it approximately doubles for the last row of Table 3. Note that the complexity in (14) is not directly comparable to that of APSC, 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 computation of the N_A distances that correspond to all possible azimuth rotations, as in Eq. (2). All other descriptors 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 construction of descriptors that perform comparably to 3DSC in terms of overall accuracy, with little extra load in 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.

It might be argued that none of the tested APSC was optimal for all landmarks; a potential reduction in the error generalized through the majority of points would require different APSC to target different landmarks. Nonetheless such a strategy is possible and prior work in landmark localization has indeed adopted different features to localize each of the targeted landmarks [6, 11, 19].

Table 4. Best-performing APSC descriptors for each landmark (i.e. the one with the lowest error and all those with no statistically significant difference from it) for three different annotated dataset, as indicated by the makers: clinical (★), FRGC with GTA-1 (▼) and FRGC with GTA-2 (▲).

APSC *	en	ex	n	prn	ac	ch	ls	li	pg
A	★ ▲			★ ▲	▼ ▲	▲			★
D_{AE}								★	
D_{AR}	▼	★ ▼	★ ▼ ▲	★ ▼ ▲	▲	★			▼
D_{AER}	▼ ▲		★	★ ▼ ▲			★ ▼ ▲	★ ▼ ▲	▼ ▲
$A+E$	★ ▲			★ ▲	★ ▼ ▲	▲			
$A+R$	★ ▲			★ ▲	★ ▼ ▲	▲			★ ▼
$A+D_{AE}$	▼	▼		▲					
$A+D_{AR}$	★ ▼	★ ▼ ▲		★ ▼ ▲	▼ ▲	★ ▼	▼		
$A+D_{AER}$	★	★	★ ▼	★ ▼ ▲	▲	★ ▼		★	★ ▼ ▲

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 due to all spatial patterns being extracted from the same 3DSC, which would be computed only once. For example, all five APSC descriptors listed in Table 2 can be computed together with less than 1.4 times the computational load of a single 3DSC descriptor.

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 performance, it would be expected that more elaborate 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 for an optimal solution.

Another important aspect is how stable would the selected patterns be for different databases. To investigate this, we selected a subset of 100 scans from the Face Recognition Grand Challenge (FRGC) database, with two independent sets of manual Ground Truth Annotations (GTA):

- GTA-1: Made available by Szeptycki et al. [24], with some additions and corrections introduced by Creusot et al. [6], which are available on line³.
- GTA-2: Our own manual annotations [22], also available on line⁴.

The difference between these two sets is that the annotations in GTA-1 were marked up based on 2D images, while those in GTA-2 were annotated directly in 3D and were shown to be considerably more precise [22]. Hence, these sets allow us to test the influence of *noisy* ground truth in the selection of descriptors.

³ <http://clementcreusot.com/phd/> [Accessed on 08.07.2013].

⁴ http://www.cipa.dcu.ie/pubs_full.html [Accessed on 15.07.2013].

Table 4 summarizes the results obtained by testing all APSC descriptors defined in Table 1. We targeted the 11 landmarks that are common to three annotated datasets: clinical (Sect. 3), FRGC with GTA-1 and FRGC with GTA-2 (the latter two correspond to the first 100 scans of FRGCv1; see [22] for details). For each landmark in Table 4, we indicate with a marker all best-performing descriptors, i.e. the one with the lowest error and all those with no statistically significant difference from it.

It is observed that, for a majority of the landmarks tested, the selection of best performing descriptors correlates reasonably well across the three datasets, highlighting the following set as the most stable choices: $ex \rightarrow A + D_{AR}$; $n \rightarrow D_{AR}$; $prn \rightarrow \{D_{AR} \text{ or } D_{AER} \text{ or } A + D_{AR} \text{ or } A + D_{AER}\}$; $ac \rightarrow \{A + E \text{ or } A + R\}$; $ls \rightarrow D_{AER}$; $li \rightarrow D_{AER}$; $pg \rightarrow A + D_{AER}$. On the other hand, the selection of descriptors for en and ch showed a strong dependency on the dataset that is employed. This was true both for changes in the 3D data (from clinical data to FRGC) and for changes in the annotations (GTA-1 vs GTA-2).

5 Conclusions

In this paper we present a new family of 3D geometric descriptors based on a simple measure of rotational symmetry that has recently been proposed based on the overlap of a shape with a rotated version of itself [10]. We compute asymmetry patterns on sequences of bins extracted from a 3DSC descriptor by varying the azimuth index and, optionally, the radial and/or elevation bins. This allows: (i) definition of APSC descriptors with azimuth invariance, (ii) highlighting or disabling some of the spatial patterns present in the spherical grid of a 3DSC that can be used to specialize the descriptor for different types of points.

We evaluated nine 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 provide invariance to azimuth rotations at the expense of a small overhead in computation of the descriptor, relative to 3DSC, which did not exceed 10%. On the other hand the rotation invariance reduces the time required for matching two descriptors by a factor of twice the number of azimuth bins. APSC were also shown to perform better than previous approaches that provided azimuth invariance to shape contexts.

The greatest potential of APSC is their flexibility to derive different descriptors at an incremental computational cost. By appropriately selecting the patterns to extract, we can build descriptors especially tuned to produce highly accurate detections of one or several landmarks. While in general, the selection of these patterns depends on the dataset, our tests suggest that such variability is limited; for a majority of the landmarks that we tested, it was possible to identify APSC descriptors with stable performance across different datasets.

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

References

1. Aynechi, N., Larson, B.E., Leon-Salazar, V., Beiraghi, S.: Accuracy and precision of a 3D anthropometric facial analysis with and without landmark labeling before image acquisition. *Angle Orthod.* **81**(2), 245–252 (2011)
2. Bariya, P., Novatnack, J., Schwartz, G., Nishino, K.: 3D geometric scale variability in range images: features and descriptors. *Int. J. Comput. Vis.* **99**(2), 232–255 (2012)
3. Bronstein, A.M., Bronstein, M.M., Castellani, U., Dubrovina, A., Guibas, L.J., Horaud, R.P., Kimmel, R., Knossow, D., von Lavante, E., Mateus, D., Ovsjanikov, M., Sharma, A.: SHREC 2010: robust correspondence benchmark. In: *Proceedings of the Eurographics Workshop on 3D Object Retrieval* (2010)
4. Chen, H., Bhanu, B.: 3D free-form object recognition in range images using local surface patches. *Pattern Recogn. Lett.* **28**(10), 1252–1262 (2007)
5. Creusot, C., Pears, N., Austin, J.: Automatic keypoint detection on 3D faces using a dictionary of local shapes. In: *Proceedings of the 1st Joint Conference on 3D Imaging, Modeling, Processing, Visualization, and Transmission*, pp. 204–211 (2011)
6. Creusot, C., Pears, N., Austin, J.: A machine-learning approach to keypoint detection and landmarking on 3D meshes. *Int. J. Comput. Vis.* **102**, 146–179 (2013)
7. Dagum, L., Menon, R.: OpenMP: an industry standard API for shared-memory programming. *IEEE Comput. Sci. Eng.* **5**(1), 46–55 (1998)
8. Farkas, L.G.: *Anthropometry of the Head and Face*. Raven Press, New York (1994)
9. Frome, A., Huber, D., Kolluri, R., Bülow, T., Malik, J.: Recognizing objects in range data using regional point descriptors. In: Pajdla, T., Matas, J.G. (eds.) *ECCV 2004. LNCS*, vol. 3023, pp. 224–237. Springer, Heidelberg (2004)
10. Guo, Q., Guo, F., Shao, J.: Irregular shape symmetry analysis: theory and application to quantitative galaxy classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(10), 1730–1743 (2010)
11. Gupta, S., Markey, M.K., Bovik, A.C.: Anthropometric 3D face recognition. *Int. J. Comput. Vis.* **90**(3), 331–349 (2010)
12. Hennessy, R.H., Kinsella, A., Waddington, J.L.: 3D laser surface scanning and geometric morphometric analysis of craniofacial shape as an index of cerebrocraniofacial morphogenesis. *Biol. Psychiatry* **51**(6), 507–514 (2002)
13. Johnson, A.E., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(5), 433–449 (1999)
14. Kazhdan, M., Funkhouser, T., Rusinkiewicz, S.: Symmetry descriptors and 3D shape matching. In: *Proceedings of the Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, pp. 156–164 (2003)
15. Kortgen, M., Park, G.J., Novotni, M., Klein, R.: 3D shape matching with 3D shape contexts. In: *7th Central European Seminar on Computer Graphics* (2003)
16. Passalis, G., Perakis, N., Theoharis, T., Kakadiaris, I.A.: Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(10), 1938–1951 (2011)
17. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (FPFH) for 3D registration. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3212–3217 (2009)
18. Rusu, R.B., Cousins, S.: 3D is here: point cloud library (PCL). In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1–4 (2011)
19. Segundo, M.P., Silva, L., Bellon, O.R.P., Queirolo, C.C.: Automatic face segmentation and facial landmark detection in range images. *IEEE Trans. Syst. Man Cybern. B Cybern.* **40**(5), 1319–1330 (2010)

20. Steder, B., Rusu, R.B., Konolige, K., Burgard, W.: Point feature extraction on 3D range scans taking into account object boundaries. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2601–2608 (2011)
21. Sukno, F.M., Waddington, J.L., Whelan, P.F.: Comparing 3D descriptors for local search of craniofacial landmarks. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Fowlkes, C., Wang, S., Choi, M.-H., Mantler, S., Schulze, J., Acevedo, D., Mueller, K., Papka, M. (eds.) ISVC 2012, Part II. LNCS, vol. 7432, pp. 92–103. Springer, Heidelberg (2012)
22. Sukno, F.M., Waddington, J.L., Whelan, P.F.: Compensating inaccurate annotations to train 3D facial landmark localization models. In: FG Workshop on 3D Face Biometrics (2013)
23. Sukno, F.M., Waddington, J.L., Whelan, P.F.: Rotationally invariant 3D shape contexts using asymmetry patterns. In: Proceedings of the International Conference on Computer Graphics Theory and Applications, pp. 7–17 (2013)
24. Szeptycki, P., Ardabilian, M., Chen, L.: A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking. In: Proceedings of the 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 1–6 (2009)
25. Toma, A.M., Zhurov, A., Playle, R., Ong, E., Richmond, S.: Reproducibility of facial soft tissue landmarks on 3D laser-scanned facial images. *Orthod. Craniofac. Res.* **12**(1), 33–42 (2009)
26. Tombari, F., Salti, S., Di Stefano, L.: Unique shape context for 3D data description. In: Proceedings of ACM Workshop on 3D Object Retrieval, pp. 57–62 (2010)
27. Zaharescu, A., Boyer, E., Horaud, R.: Keypoints and local descriptors of scalar functions on 2D manifolds. *Int. J. Comput. Vis.* **99**(2), 232–255 (2012)
28. Zhang, Y.: Intrinsic shape signatures: a shape descriptor for 3D object recognition. In: Proceedings of the 12th IEEE International Conference on Computer Vision Workshops, pp. 689–696 (2009)

Computer Vision, Imaging and Computer Graphics:
Theory and Applications

International Joint Conference, VISIGRAPP 2013,
Barcelona, Spain, February 21-24, 2013, Revised
Selected Papers

Battiatto, S.; Coquillart, S.; Laramée, R.S.; Kerren, A.;
Braz, J. (Eds.)

2014, XVI, 245 p. 135 illus., Softcover

ISBN: 978-3-662-44910-3