

Novel Spectral Descriptor for Object Shape

Atul Sajjanhar^{1,*}, Guojun Lu², and Dengsheng Zhang²

¹ School of Information Technology
Deakin University
221 Burwood Highway
Burwood, VIC 3125
Australia

atuls@deakin.edu.au

² Gippsland School of Information Technology
Monash University
Northways Road
Churchill, VIC 3842
Australia

{guojun.lu, dengsheng.zhang}@infotech.monash.edu.au

Abstract. In this paper, we propose a novel descriptor for shapes. The proposed descriptor is obtained from 3D spherical harmonics. The inadequacy of 2D spherical harmonics is addressed and the method to obtain 3D spherical harmonics is described. 3D spherical harmonics requires construction of a 3D model which implicitly represents rich features of objects. Spherical harmonics are used to obtain descriptors from the 3D models. The performance of the proposed method is compared against the CSS approach which is the MPEG-7 descriptor for shape contour. MPEG-7 dataset of shape contours, namely, CE-1 is used to perform the experiments. It is shown that the proposed method is effective.

Keywords: shape descriptor, content based image retrieval, feature extraction.

1 Introduction

Approaches for shape representation and retrieval can be broadly classified into contour based and region based [14]. Some of the region based methods are geometric moments, moments constructed from orthogonal functions and generic Fourier descriptors. Some of the contour based methods are polygonal approximation, autoregressive model, Fourier Descriptors and distance histograms.

Recently, region based methods have been proposed which rely on representation of 2D objects in 3D space [1][2]. These methods, however, do not exploit the full extent of shape representation in 3D space. The proposed descriptor manifests the topology of the image in 3D space. The process is twofold. First, it is shown how to represent a 2D object in 3D space. Second, a 3D modeling technique is adopted for representation of the 3D model obtained in step 1. Rotation invariant spherical harmonics are effective for representation and retrieval of 3D models [3]. Hence, we use

* PhD candidate, Monash University, Australia.

spherical harmonics for representation of 3D models which are obtained a priori from 2D objects. 2.5D method [1] and connectivity method [2] also use spherical harmonics for the retrieval of 2D objects. The key difference between the 2.5D method, connectivity method and the proposed methods are the features represented in 3D space. The proposed method is found to be significantly better than other methods in the same category.

Motivation for the proposed method is described in Section 2. The proposed method is described in Section 3. Experimental Results are presented in Section 4. Discussion is presented in Section 5. Conclusion is given in Section 6.

2 Related Work

Our motivation is to use spherical harmonics for 2D region-based shape representation. The motivation stems from the successful use of the spectral domain for contour-based shape representation. Spherical harmonics are the Fourier coefficients for the group $SO(3)$ acting on the 2-sphere. The two coefficients are addressed by Healy et al. [11]. Funkhouser et al. [5] have described a 2D analog method of spherical harmonics [4] to extract a series of rotation invariant signatures by dividing a 2D silhouette shape into multiple circular regions.

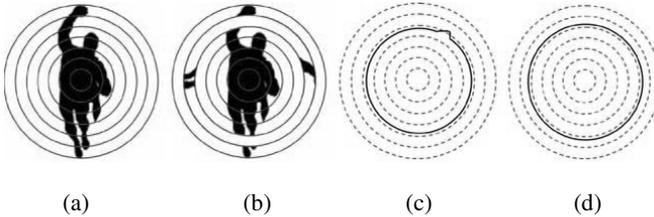


Fig. 1. (a) and (b) have similar descriptors and (c) and (d) have larger dissimilarity between their descriptors. Source [10].

The drawbacks of the method have been pointed out by Pu and Ramani [1]. First, consider the rotation of the circular regions, as shown in Fig. 1 (a) and (b). The spherical harmonics of the circular functions remain unchanged because they are rotation invariants. Hence, the same set of descriptors will correspond to different shapes i.e. descriptors are unable to discriminate between shapes. Another drawback is that a small perturbation can result in large dissimilarity between images. Consider the shapes in Fig. 1 (c) and (d); the similarity between the two shapes is measured by computing the summation of the squared differences between the frequencies of the corresponding circular regions. Due to the small perturbation in Fig. 1(c), the squared differences are large. Therefore, the two shapes in Fig. 1 (c) and (d) will be considered different although they are perceptually similar. In order to overcome the limitations of spherical harmonics descriptors for 2D shapes, we propose to obtain spherical harmonics for representation of shape in 3D space.

3 Proposed Method

In this section, we propose a method for obtaining 3D models of 2D objects. Distance transform method is described below. Given an image containing a set of features (e.g. edge pixels, lines, points etc.), the distance transform calculates, for each pixel, the distance to the nearest feature. Distance transform have been widely used for encoding of metric information associated with images. It has been used in computer vision for a number of applications such as shape decomposition [8] and skeletonisation [9], thickening and thinning of binary objects [10]. Since, distance transform is a global operation, the computational requirement using the naïve approach is proportional to the size of the image ($O(nm)$) [15]. However, efficient algorithms have been developed that require only two passes of the image [13]. A pixel p_i is represented as $p_i = (x_i, y_i)$ in 2D where x_i, y_i are the 2D cartesian coordinates. Each pixel is transformed into 3D space of the form $p_i = (x_i, y_i, z_i)$ where z_i is the distance transform of the pixel. The distance transform is applied to the binary image in Fig 1(a) to generate the grayscale image in Fig. 2(a). The pixel intensity of each pixel in Fig. 2(a) reflects the distance of the pixel to the nearest edge. The distance to the nearest edge is computed for each pixel and represented in the z-axis of 3D Cartesian coordinates. The point cloud thus obtained is triangulated into a mesh as shown in Fig. 2(b) and Fig. 2(c). Features which are implicitly represented in the 3D model are:

- i. The ridges in 3D space which represent the skeleton.
- ii. The valleys which represent the contours in the object
- iii. The height of the ridge indicates the thickness of the shape. Higher the ridge, greater is the distance from the nearest edge.
- iv. The height of a point between the valley and the ridge represents the distance of the point from the contour.

The mesh shown in Fig. 2 is represented as a 3D model and spherical harmonics descriptors obtained thereafter are used for shape representation. According to the theory of spherical harmonics, a spherical function $f(\theta, \phi)$ can be decomposed as the sum of its harmonics $Y_n^m(\theta, \phi)$ as shown in Eqn. 2.

$$f(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{m=n} a_{n,m} Y_n^m(\theta, \phi) \quad (1)$$

$a_{n,m}$ are the coefficients in the frequency domain, $Y_n^m(\theta, \phi)$ is defined as

$$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos(\theta)) e^{im\phi} \quad (2)$$

where $P_n^m(x)$ is Legendre polynomial.

The key property of this decomposition is that if we restrict to some frequency n , and define the subspace functions as shown below then the subspace Y_n is invariant under the operations of the full rotation group and it is irreducible.

$$Y_n = \{Y_n^n, Y_n^{n-1} \dots Y_n^{-n}\} \quad (3)$$

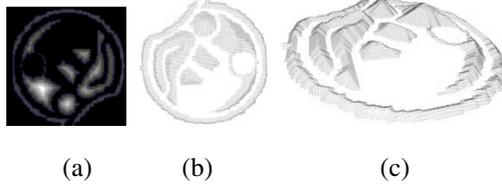


Fig. 2. (a) Distance Transform (b) and (c) Distance Transform in 3D Space

The spherical harmonics descriptor (SHD) is shown in Eqn. 4, this descriptor is invariant to rotation. $\|SH_n(\theta, \phi)\|$ is the L_2 - norm of $SH_n(\theta, \phi)$

$$SHD = \left\{ \frac{\|SH_1(\theta, \phi)\|}{\|SH_0(\theta, \phi)\|}, \frac{\|SH_2(\theta, \phi)\|}{\|SH_0(\theta, \phi)\|}, \dots \right\} \quad (4)$$

$$SH_n(\theta, \phi) = \sum_{m=-n}^{m=n} a_{n,m} Y_n^m(\theta, \phi) \quad (5)$$

$$SHD(i) = \frac{\|SH_i(\theta, \phi)\|}{\|SH_0(\theta, \phi)\|} \quad (6)$$

Approximate reconstruction of spherical function $f(\theta, \phi)$ is:

$$f(\theta, \phi) \approx \sum_{n=0}^N \sum_{m=-n}^{m=n} a_{n,m} Y_n^m(\theta, \phi) \quad (7)$$

$$SHD \approx \left\{ \frac{\|SH_1(\theta, \phi)\|}{\|SH_0(\theta, \phi)\|}, \dots, \frac{\|SH_N(\theta, \phi)\|}{\|SH_0(\theta, \phi)\|} \right\} \quad (8)$$

The steps to obtain the spherical harmonics descriptors are summarized as: first, decomposition of spherical function into its harmonics; second, summing the harmonics within each frequency; third, obtaining the norm of each frequency component. The spherical harmonics are compared using the L_2 -difference. The L_2 -difference between the harmonic representations of two spherical functions is a lower bound for the minimum of the L_2 -difference between the two functions, taken over all possible orientations. Spherical harmonics method can be extended to voxel descriptors [3][4]. We briefly describe how Funkhouser et al [5] obtained voxel descriptors from spherical harmonics. Polygons within the 3d model (refer Fig. 2) are rasterized into a voxel grid. A voxel is assigned a value of 1 if it is within one voxel width of the polygonal surface, and assigned a value of 0 otherwise. The model is normalized for translation and scale. Voxel grid is treated as a binary function defined in spherical coordinates; it is restricted to a collection of concentric spheres as shown in Fig. 3.



Fig. 3. Concentric Spheres on Voxel Grid

Each spherical restriction is represented in terms of a function, which gives a collection of spherical functions. Each spherical function is represented as the sum of its different frequencies i.e. spherical harmonics representation. Rotation invariant signature is obtained for each radius as a collection of scalars from the spherical harmonics representation. Rotation invariant signatures for different radii are combined to obtain the spherical harmonics descriptor (SHD) for the 3D model. We use spherical harmonics of the voxel grid as shape descriptors. Distance between two shapes is computed as the Euclidean distance between their SHD.

4 Experiments and Results

For the proposed method, point clouds are generated for 2D images. Point clouds are represented by xyz files and include distance transform for each pixel. Polygonal meshes are generated from the xyz file using Delaunay triangulations. The polygonal meshes are converted into PLY format (introduced by Stanford University) [17]. PLY format represents the 3D model and is used to generate SHD as described in Section 3. Distance between two shapes is computed as the Euclidean distance between their SHD.

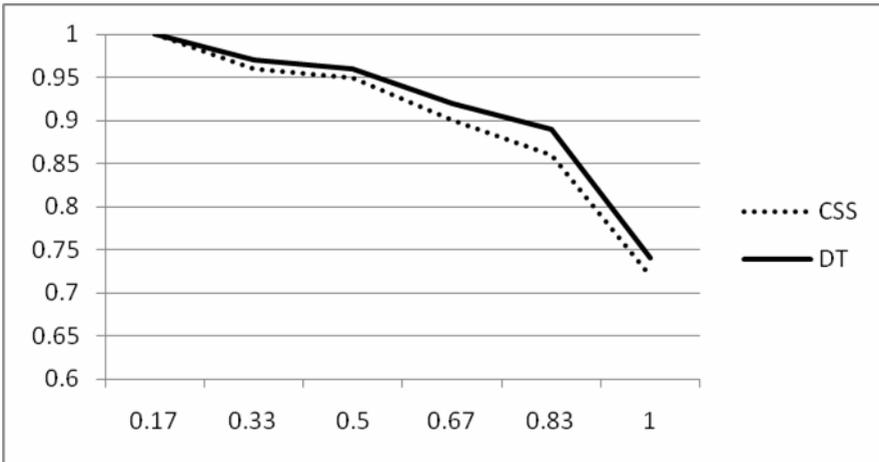


Fig. 4. Recall-Precision plots for Set A1

In this experiment, we use MPEG-7 CE-1 dataset of shapes [16]. The dataset is divided into three sets, namely, A, B and C. Set A is divided into sets A1 and A2. Set A1 is used for testing the robustness to scaling. Set A2 is used for testing the robustness to rotation. The results for Set A1 and Set A2 are shown below.

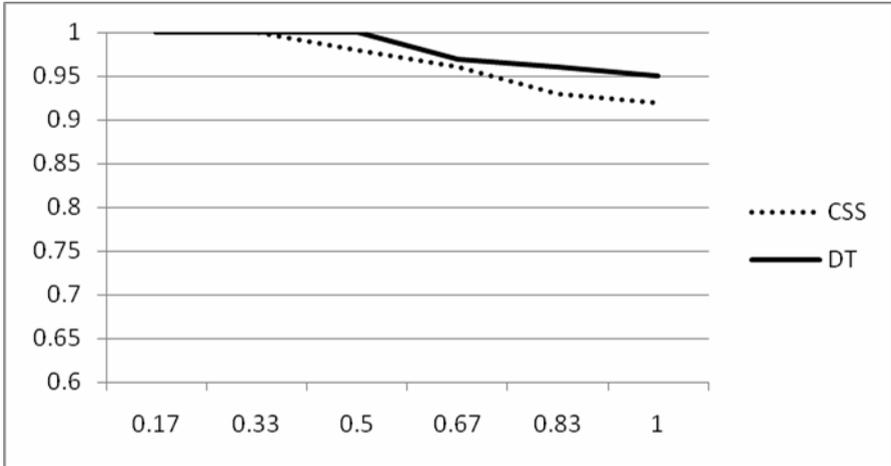
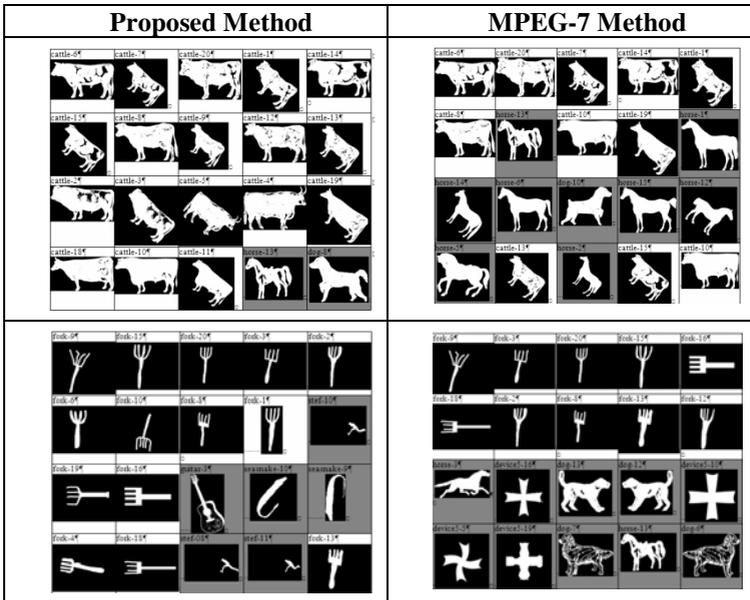


Fig. 5. Recall-Precision plots for Set A2



Set B is the most important set so we show the outcome of few queries in the figures below. In Fig. 6, the top left corner is the query shape and the grey background indicates that the shape is not relevant to the query. Left column in Fig. 6 shows results obtained for the distance transform method and the right column show results obtained for CSS.

The recall-precision plots for sets B and C are shown in the figures below.

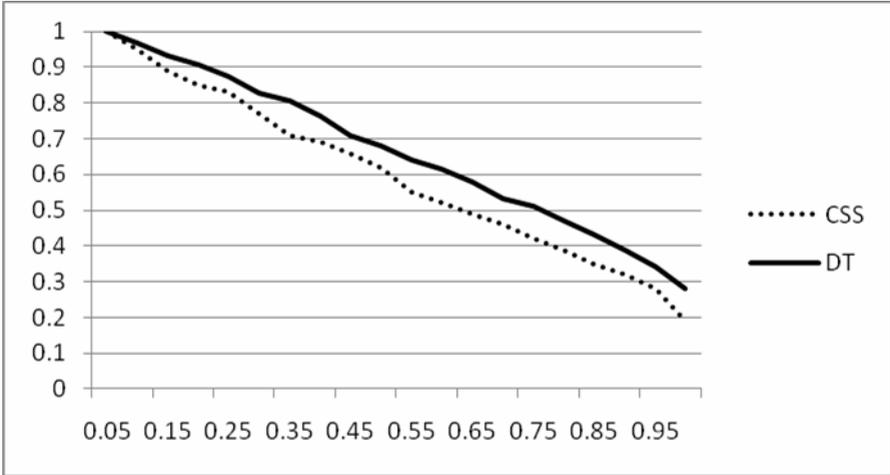


Fig. 7. Average Recall-Precision for MPEG-7 database, CE-1, Set B using the Proposed Method and CSS method

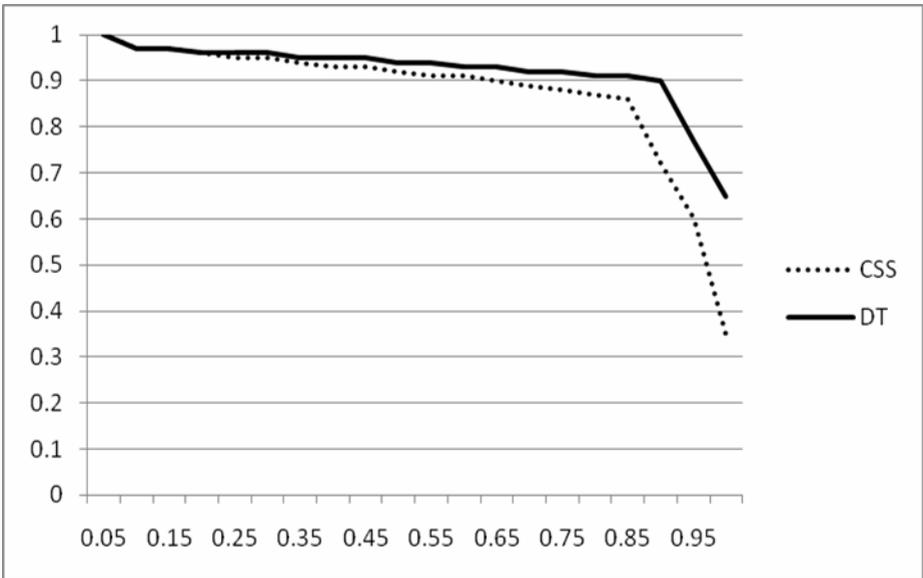


Fig. 8. Average Recall-Precision for MPEG-7 database, CE-1, Set C using the Proposed Method and CSS method

The proposed method and CSS are able to achieve good results for Sets A1, A2. Hence, both the methods are sufficiently robust to scaling and rotation. In Set A1, it is not possible to obtain a precision of 100% [16]. The reason is that 17 shapes in the set are too small; these are the shapes obtained by scaling with a factor 0.1. Some examples of these shapes are shown below.

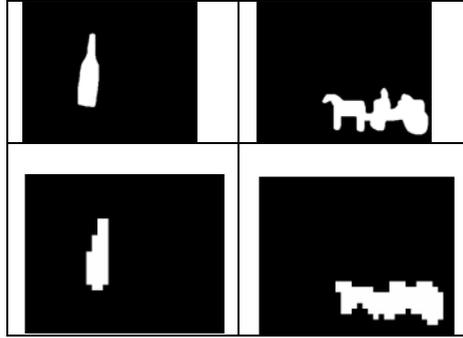


Fig. 9. Shapes from Set A1. Top Row shows Basic Shapes. Bottom Row shows Shapes Obtained by Scaling with factor 0.1.

The best possible precision for Set A1 is reduced to 93% because of the 17 shapes which are scaled beyond recognition [16]. Considering that the best possible precision is 93%, the three methods perform near optimal for Set A1. The precision of retrieval for Set A2 is near 100% for the proposed method. Hence, the proposed method is robust to rotation. CSS is slightly less robust to rotation.

The experiments on Set B of CE-1 show that the proposed method performs significantly better than CSS. Set B is the most challenging. The precision of retrieval is lower than Set A and Set C. For each query, there are 20 relevant shapes however some shapes in each class are similar to shapes in other classes.

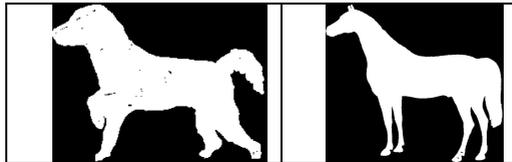


Fig. 10. Shapes from dog class and horse class which are perceptually similar

Only a single query is used in Set C, namely, bream-000. 14 bream fish do not have shape similar to bream-000 which is the query shape. The best possible precision for this dataset is 93% (186/200) [16]. Both methods have high precision for low recall; however, the precision of CSS drops significantly for high values of recall.

We have used 512 dimensions in the feature vectors. The shape is decomposed into 32 concentric spheres. For each sphere, 16 coefficients are computed. Hence, the feature vector has $32 \times 16 = 512$ dimensions.

5 Discussion

A reason contributing to the effectiveness of the proposed method is the use of spherical harmonics. Spherical harmonics can be regarded as Spherical Fourier Transform. Fourier transform has been shown to be effective for pattern recognition. Spherical harmonics represent spherical functions in the spectral domain; spherical functions are obtained a priori for a series of concentric spheres on the voxel grid. The inherent nature of obtaining spherical harmonics which uses concentric spheres to sample image features provides a balanced approach for feature extraction. In contrast, a regular grid distribution in Cartesian coordinates tends to undersample the image in the centre and oversample the image away from the centre. Another advantage of spherical harmonics is representation of image features in the spectral domain. Spectral analysis of images has been widely used for image retrieval. There are two advantages of spectral features. First, they are robust compared with spatial features. Second, spectral features are inherently multiresolutional and this property can be leveraged to determine the degree of detail encoded during indexing.

In the case of fine shapes, SHD is not robust. Sparse point clouds result in inaccurate mesh generation when Delaunay triangulation is applied. SHD can be used adaptively i.e. the pixel cloud may be built for the shape complement to improve the performance of fine shapes.

Computational expense of the proposed method also needs to be addressed. The proposed method requires more processing compared with other techniques for 2D image retrieval. Processing overheads of the proposed method include: triangulations of the point cloud to generate a mesh and 3D modeling of the mesh. Efficient methods for computing distance transforms require only two passes of the image [13]. Efficient methods for Delaunay triangulations have computational complexity $O(n^2)$ [7]. Efficient methods for computing spherical harmonics of a spherical function have been developed [11][12] which have complexity $O(b^2 \log b^2)$ sampled on a regular $O(b^2)$ grid.

The method described above can be extended to grayscale images. A point cloud is constructed as shown in Fig. 1 for grayscale images. The z-axis, however, specifies the pixel intensity in the grayscale image. Voxel is built from the point cloud and SHD is used to represent energy information of concentric spheres on the voxel.

Distance Transforms are not robust for some images. Spurious edges detected by the edge detector and the undetected edges missed by the edge detector causes performance deterioration [6]. Varying the threshold slightly can generate large changes in the number of false positives and false negatives. Rosin and West [6] proposed the Saliency Distance Transform (SDT) which has greater stability. In this approach the distances from the edges are weighted by the saliency of the edges. Saliency of edges is determined by various criteria such as edge magnitude, curve length and local curvature. Applying the edge detection over a range of scales also improves the stability of the method.

6 Conclusion

In this paper, we proposed a shape descriptor. Distance transform approach is used to obtain the 3D model. Spherical harmonics are obtained from the 3D model. The shape descriptor is tested against standard dataset and it is shown that the proposed descriptor is more effective than MPEG-7 descriptors for shape contour, namely, Curvature

Scale Space. The proposed descriptor is a generic descriptor i.e. it can be applied to shape regions and shape contours.

Acknowledgement

We thank Michael Kazhdan of John Hopkins University for the spherical harmonics code provided.

References

1. Pu, J., Ramani, K.: On visual similarity based 2D drawing retrieval. *Computer-Aided Design* 38, 249–259 (2006)
2. Sajjanhar, A., Lu, G., Zhang, D.: Spherical Harmonics Descriptor for 2d-Image Retrieval. In: *IEEE International Conference on Multimedia and Expo.*, pp. 105–108 (2005)
3. Kazhdan, M., Funkhouser, T., Rusinkiewicz, S.: Rotation Invariant Spherical Harmonic Representation of 3d Shape Descriptors. In: *Symposium on Geometry Processing*, pp. 167–175 (2003)
4. Kazhdan, M.: *Shape Representations and Algorithms for 3d Model Retrieval*, PhD Thesis, Princeton University (2004)
5. Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., Jacobs, D.: A search engine for 3d models. In: *ACM transactions on Graphics*, pp. 83–105 (2003)
6. Rosin, P.L., West, G.A.W.: Saliency distance transforms. *Graphical Models Image Processing* 57, 483–521 (1995)
7. Attali, D., Boissonnat, J.: A Linear Bound on the Complexity of the Delaunay Triangulation of Points on Polyhedral Surfaces. *Discrete and Computational Geometry* 31(3), 369–384 (2004)
8. Toriwaki, J.I., Saitoh, T., Okada, M.: Distance transformation and skeleton for shape feature analysis. In: *International Workshop on Visual Form*, pp. 547–563 (1992)
9. Tsang, P.W.M., Yuen, P.C., Lam, F.K.: Classification of partially occluded objects using 3-pointy matching and distance transformation. *Pattern Recognition* 27, 27–40 (1994)
10. Paglieroni, D.W.: Distance Transforms: Properties and machine vision applications. *CVGIP: Graphical Models Image Processing* 54, 56–74 (1992)
11. Healy, D., Kostelec, P., Moore, S.: FFTs for the 2-sphere-improvements and variations. *Journal of Fourier Analysis and Applications* 9(4), 341–385 (2003)
12. Driscoll, J., Healy, D.: Computing Fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics* 15, 202–250 (1994)
13. Borgefors, G.: Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing* 34(3), 344–371 (1986)
14. Zhang, D.S., Lu, G.: Review of Shape Representation and Description Techniques. *Pattern Recognition* 37(1), 1–19 (2004)
15. Heinz, B., Gil, J., Kirkpatrick, D., Werman, M.: Linear Time Euclidean Distance Transform Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(5), 529–533 (1995)
16. Shape data for the MPEG-7 core experiment CE-Shape-1, <http://www.cis.temple.edu/~latecki/TestData/mpeg7shapeB.tar.gz> (last accessed December 9, 2009)
17. http://www.cc.gatech.edu/projects/large_models/