# Spherical Harmonics and Distance Transform for Image Representation and Retrieval

Atul Sajjanhar[1], Guojun Lu[2], Dengsheng Zhang[2], Jingyu Hou[1], and Yi-Ping Phoebe Chen[1]

[1] School of Information Technology
Deakin University
221 Burwood Highway
Burwood, VIC 3125
Australia
{atuls,jingyu,phoebe}@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 have proposed a method for 2D image retrieval based on object shapes. The method relies on transforming the 2D images into 3D space based on distance transform. Spherical harmonics are obtained for the 3D data and used as descriptors for the underlying 2D images. The proposed method is compared against two existing methods which use spherical harmonics for shape based retrieval of images. MPEG-7 Still Images Content Set is used for performing experiments; this dataset consists of 3621 still images. Experimental results show that the performance of the proposed descriptors is significantly better than other methods in the same category.

**Keywords:** Spherical harmonics, content based image retrieval.

## 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 images in 3D space [1][2]. In this paper, we propose a new method which relies on 3D modeling. The process is twofold. First, it is shown how to represent a 2D shape 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 spherical harmonics

E. Corchado and H. Yin (Eds.): IDEAL 2009, LNCS 5788, pp. 309–316, 2009.

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

The 2.5D method and the connectivity method are described in Section 2. The proposed method is described in Section 3. Experimental Setup and Results are presented in Section 4. Finally, Discussion is presented in Sections 5.

## 2 Related Work

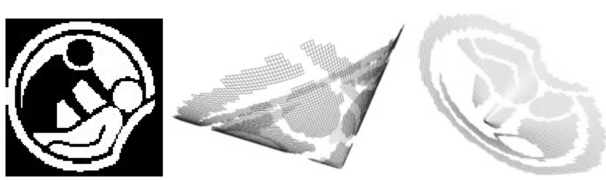
The 2.5D method proposed by Pu and Ramani [1] is closely related to the proposed method. The 2D image is placed in the equator plane of a sphere which encloses the image. Rays are generated which start from the centre of the sphere and located in the xy plane where the image lies. Each intersection point $p_i$ is represented in polar coordinates $p_i=f(\theta_i, d_i)$ where $\theta_i$ is the angle with respect to the x-axis and $d_i$ is the distance of the point from the centre of the sphere. However, for a single $\theta_i$ there might be multiple intersection points. Each intersection point $p_i$ is transformed into unique 3D spherical representation using the function $p_i= (\theta_i, \varphi_i, d_i)$. $\varphi_i$ is obtained as shown below.

$$\Phi_i = \arctan\left(\frac{d_i}{r}\right) \tag{1}$$

where, $r$ is the radius of the sphere.

Visualisation of $d_i$ is shown in Fig. 1. The values $d_i$, are computed, for the image shown in Fig. 1(a) and the resulting point cloud is triangulated into a mesh as shown in Fig. 1(b) and Fig. 1(c). Spherical harmonics transformation is performed on the 3D spherical representation to obtain signatures for 2D images.

The method proposed by Sajjanhar et al. [2] uses the concept of connectivity for transformation of a 2D image into 3D space. Connectivity captures the pixel neighbourhood information. The state of the nearest 8-neighbours is computed for each *OFF* pixel. An *OFF* pixel is a dark pixel i.e. has intensity below a predefined threshold. Connectivity of an *OFF* pixel is obtained as the number of *OFF* pixels amongst the nearest 8-neighbours. Connectivity information is represented in the



(a) (b) (c)

**Fig. 1.** (a) Original Image (b) and (c) 3D Space

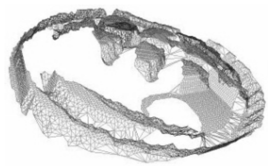


**Fig. 2.** Connectivity in 3D Space

z-axis of 3D Cartesian coordinates. For each *OFF* pixel within the image, the connectivity can take values 0 through 8. A connectivity of 0 indicates that none of the nearest 8-neighbours are *OFF*. A connectivity of 8 indicates that all of the nearest 8-neighbours are *OFF*. Connectivity is computed for the image shown in Fig. 1(a) which is triangulated into a polygon mesh as shown in Fig. 2.

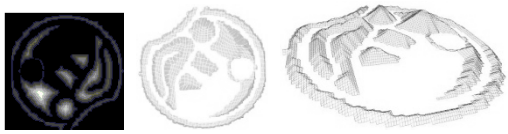
Connectivity method is found to be sensitive. Preliminary experiments showed that some images performed poorly with this method; however, inverting the images resulted in significant improvement. Gaussian smoothing of images also improved the results.

## 3 Proposed Method

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 greyscale image in Fig. 3(a). The pixel intensity of each pixel in Fig. 3(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. 3(b) and Fig. 3(c).

The mesh shown in Fig. 3 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.

$$f(\theta,\phi)=\sum_{n=0}^{\infty}\sum_{m=-n}^{m=n}a_{n,m}Y_n^m(\theta,\phi) \tag{2}$$



(a) (b) (c)

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

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

$$Y_n^m(\theta,\phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos(\theta)) e^{im\phi} \tag{3}$$

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.

$$Y_n = \{Y_n^n, Y_n^{n-1} \dots Y_n^{-n}\} \tag{4}$$

The spherical harmonics descriptor (SHD) is shown in Eqn. 5, this descriptor is invariant to rotation. $\|SH_n(\theta,\phi)\|$ is the $L_2$- norm of $SH_n(\theta,\phi)$

$$SHD = \left\{ \frac{\|SH_1(\theta,\phi)\|}{\|SH_0(\theta,\phi)\|}, \frac{\|SH_2(\theta,\phi)\|}{\|SH_0(\theta,\phi)\|} \dots \right\} \tag{5}$$

$$SH_n(\theta,\phi) = \sum_{m=-n}^{m=n} a_{n,m} Y_n^m(\theta,\phi) \tag{6}$$

$$SHD(i) = \frac{\|SH_i(\theta,\phi)\|}{\|SH_0(\theta,\phi)\|} \tag{7}$$

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

$$f(\theta,\phi) \approx \sum_{n=0}^{N} \sum_{m=-n}^{m=n} a_{n,m} Y_n^m(\theta,\phi) \tag{8}$$

$$SHD \approx \left\{ \frac{\|SH_1(\theta,\phi)\|}{\|SH_0(\theta,\phi)\|}, \cdots \frac{\|SH_N(\theta,\phi)\|}{\|SH_0(\theta,\phi)\|} \right\} \tag{9}$$

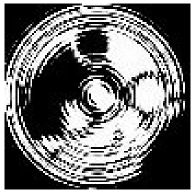


**Fig. 4.** Concentric Spheres on Voxel Grid

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 $L2$-difference. The $L2$-difference between the harmonic representations of two spherical functions is a lower bound for the minimum of the $L2$-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. 3) 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. 4.

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. The main steps for calculating the spherical harmonics of the voxel grid are:

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

Experimental results are shown in the recall-precision plots Fig. 5 below.

Experiments are performed on Item number S8 within the MPEG-7 Still Images Content Set; this is a collection of trademark images which was originally provided by the Korean Industrial Property Office. S8 consists of 3621 still images. It is divided into sets A1, A2, A3, A4 to test the robustness of methods to geometric and perspective transformations. For the proposed method, point clouds are generated for 2D images. Within the legends, *conn* represents the connectivity method, *2.5d* represents the 2.5D method and *dt* represents the proposed method which is based on distance transform.

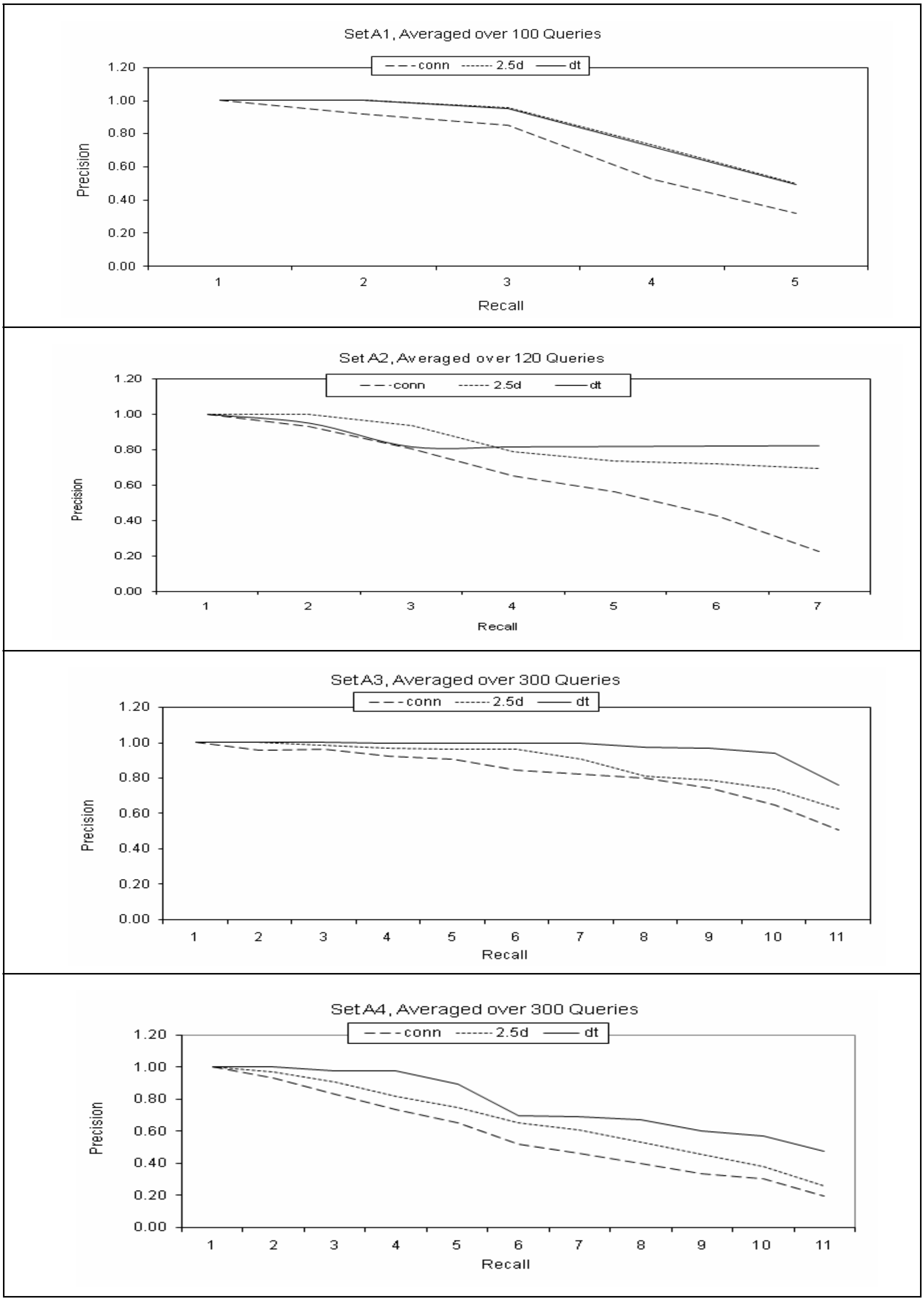



**Fig. 5.** Recall-Precision Plots

## 5 Discussion

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 Salience Distance Transform (SDT) which has greater stability. In this approach the distances from the edges are weighted by the salience of the edges. Salience 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.

Use of spherical harmonics contributes to the effectiveness of the proposed method. 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.

Computational expense of the proposed method also needs to be addressed. The proposed method requires substantial processing compared with other techniques for 2D image retrieval. Processing overheads of the proposed method include: distance transformation, 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. For applications where accuracy of retrieval is important, the improvement in effectiveness may outweigh the processing complexity. Experiments in Section 4, show the feasibility of the method for trademark images. In the future, we will apply the approach for protein matching.

**Acknowledgement.** We are grateful to Michael Kazhdan for the code to obtain Spherical Harmonics Descriptors.

## 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, June 2003, pp. 167–175 (2003)
4. Kazhdan, M.: Shape Representations and Algorithms for 3d Model Retrieval, PhD Thesis, Princeton University (June 2004)
5. Borgefors, G.: Distance transformations in digital images, Computer Vision. Graphics and Image Processing 34(3), 344–371 (1986)
6. Rosin, P.L., West, G.A.W.: Salience 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-poiny 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. Breu, H., 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. Novotni, M., Klein, R.: Shape retrieval using 3D Zernike descriptors. Computer-Aided Design 36, 1047–1062 (2004)