

SPHERICAL HARMONICS DESCRIPTOR FOR 2D-IMAGE RETRIEVAL

Atul Sajjanhar
School of Information Technology
Deakin University
221 Burwood Highway
Burwood, VIC 3125
Australia
atuls@deakin.edu.au

Guojun Lu, Dengsheng Zhang
Gippsland School of Comp. & Info. Tech.
Monash University
Northways Road
Churchill, VIC 3842
Australia
{guojun.lu, dengsheng.zhang}@infotech.monash.edu.au

ABSTRACT

In this paper, spherical harmonics are proposed as shape descriptors for 2d images. We introduce the concept of connectivity; 2d images are decomposed using connectivity which is followed by 3d model construction. Spherical harmonics are obtained for 3d models and used as descriptors for the underlying 2d shapes. Difference between two images is computed as the Euclidean distance between their spherical harmonics descriptors. Experiments are performed to test the effectiveness of spherical harmonics for retrieval of 2d images. Item S8 within the MPEG-7 Still Images Content Set is used for performing experiments; this dataset consists of 3621 still images. Experimental results show that the proposed descriptors for 2d images are effective.

1. INTRODUCTION

Approaches for shape representation and retrieval can be broadly classified into contour based and region based. Some of the region based methods are geometric moments, moments constructed from orthogonal functions and grid based method [1]. Recently, Generic Fourier Descriptors (GFD) was proposed by Zhang and Lu [2] for region based matching of shapes. Some of the contour based methods are polygonal approximation, autoregressive model, Fourier Descriptors and distance histograms [1].

In this paper, 3d modeling technique is proposed as a region based method for 2d shape retrieval. The process is twofold. First, the concept of *connectivity* is introduced and it is shown how *connectivity* can be used to construct a 3d model of a 2d image. Second, a 3d modeling technique is adopted for representation of the 3d model

obtained in step 1. Spherical harmonics are effective for representation and retrieval of 3d models [4]. Hence, we use spherical harmonics for representation of 3d models which are obtained a priori from 2d images. Salient features of spherical harmonics are addressed in Section 5.

It has been shown that the performance of GFD is comparable with other contemporary techniques [2]. Hence, we compare the proposed method with GFD. GFD is described in Section 2. The proposed method is described in Section 3. Experimental Setup and Results are presented in Section 4. Finally, Discussion and Conclusion are presented in Sections 5 and 6 respectively.

2. GENERIC FOURIER DESCRIPTORS

Generic Fourier Descriptors (GFD) is a region-based method for image retrieval [1][2]. In GFD, feature vectors are created by extracting spectral information in the frequency domain. Fourier transform is applied to the polar raster sampled shape image. Consider the image shown in Figure 1. To obtain the GFD for the image, the image is first plotted in polar space. The polar image of Figure 1(a), is shown in Figure 1(b).

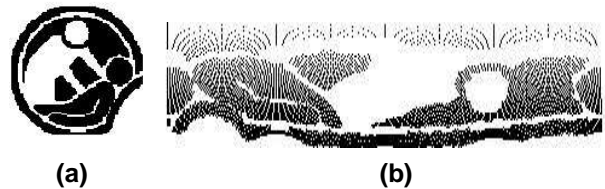


Figure 1. (a) An Image in Cartesian Coordinates
(b) Polar Image

Before obtaining the polar image, the image is normalized for scale. 2d DFT is applied to the rectangular region in polar coordinates to obtain Fourier coefficients

which are used to construct feature vectors for shape representation and similarity measure [1][2].

Polar coordinates (r, θ) are obtained from the 3d Cartesian co-ordinates (x, y) as shown below.

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2} \quad (1)$$

$$\theta = \arctan\left(\frac{y - y_c}{x - x_c}\right) \quad (2)$$

where, (x_c, y_c) is the centroid of the 2d Cartesian image.

Feature vectors are constructed from the polar coordinates by computing the 2d DFT. 2d DFT of the polar coordinates is defined as below.

$$PF(\rho, \tau) = \sum_r \sum_\theta f(r, \theta) e^{-j2\pi\left(\frac{r}{R}\rho + \frac{\theta}{T}\tau\right)} \quad (3)$$

where, R and T is the radial and angular resolution. r, θ is obtained from Eqn. 1 and Eqn. 2. Feature vectors are represented as shown below.

$$F : \left(\frac{F(0,1)}{F(0,0)}, \dots, \frac{F(0,T-1)}{F(0,0)}, \dots, \frac{F(1,0)}{F(0,0)}, \dots, \frac{F(R-1,T-1)}{F(0,0)} \right)$$

where, R and T is the radial and angular resolution as used in Eqn. 3.

The difference between two images is computed as the Euclidean distance between their feature vectors as shown in Eqn. 4.

$$Dist(F_1, F_2) = \sqrt{\sum_{i=0}^{RT-1} (f_{1,i} - f_{2,i})^2} \quad (4)$$

where, $f_{x,i}$ is a descriptor within the feature vector of image x . $0 < i < RT$, where R, T is the radial and angular resolution.

3. 3D-MODELING

First, we introduce connectivity which was proposed in [7]. An analogy is drawn from Color Coherence Vectors (CCV) proposed by Pass and Zabih [3]. CCV is used for image retrieval based on color. Pass et al [3] defined color coherence of pixels as the degree to which pixels of that color are members of a large similarly colored region. Pixels are classified as coherent or incoherent. Coherent pixels are part of a sizable contiguous region of similar color while incoherent pixels are not.

In the case of shape representation, *connectivity* of pixels is defined. 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. Figure 2 provides additional information for the image in Figure 1(a). Connectivity information is added in Cartesian

coordinates. Hence, z-axis is obtained which provides information regarding connectivity of pixels. 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*.

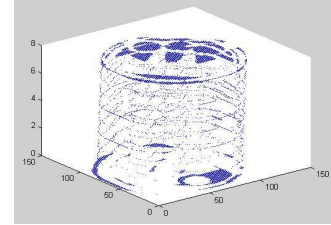


Figure 2. Connectivity Information for Image in Figure 1

Connectivity for the image in Figure 1(a) is shown by the point cloud in Figure 2. The point cloud in 3d Cartesian coordinates contains connectivity information along the z-axis. The point cloud is triangulated into a polygon mesh as shown in Figure 3.

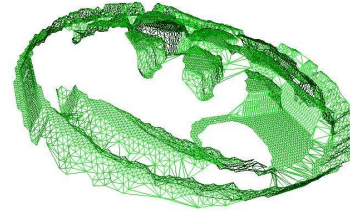


Figure 3. Polygon Mesh for 2d Shape

A 3d model descriptor is used for representation of the mesh (such as Figure 3). Several descriptors have been used to represent 3d models. The process is two fold:

- i. normalizing the model
- ii. representing the model with a transformation invariant descriptor.

Commonly used shape descriptors use a spherical function or voxel grid to represent 3d shapes. We use voxel description which describes a model by computing the negative exponential of its Euclidean Distance Transform [4][5]. Voxel description is obtained from rotation invariant spherical harmonics proposed by Funkhouser et al [5] for matching 3d models. This method describes a spherical function in terms of the amount of energy it contains at different frequencies. Information at larger frequencies corresponds to higher resolution information. By construction, the harmonic representation is rotation invariant, as it does not store information that depends on the alignment of the model.

The steps to obtain the spherical harmonics are summarized as follows. First, decompose the spherical function into its harmonics. Second, sum the harmonics

within each frequency. Third, obtain the norm of each frequency component.

Spherical harmonics method can be extended to voxel descriptors. The main steps for calculating the spherical harmonics of the voxel grid are:

- i. Polygons within the 3d model (refer Figure 3) are rasterized into a $2R \times 2R \times 2R$ 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. Translation normalization is achieved by moving the model so that the center of mass lies at the point (R, R, R) . Scale normalization is achieved by scaling the model such that the average distance between the non-zero voxels and the centre of mass is $R/2$.
- ii. Voxel grid is treated as a binary function defined in spherical coordinates as:

$$f(r, \theta, \phi) = \text{Voxel}(r \sin \theta \cos \phi + R, r \cos \theta + R, r \sin \theta \sin \phi + R)$$

where $r \in [0, R]$, $\theta \in [0, \pi]$, $\phi \in [0, 2\pi]$. Voxel grid is restricted to a collection of concentric spheres, as shown in Figure 4.



Figure 4. Concentric Spheres on Voxel Grid

Each spherical restriction is represented in terms of a function. This gives a collection of spherical functions

$$\{f_0, f_1, \dots, f_R\} \text{ with } f_r(\theta, \phi) = f(r, \theta, \phi)$$

- iii. Each function is represented as the sum of its frequencies as shown in Eqn. 5.

$$f_r(\theta, \phi) = \sum_m f_r^m(\theta, \phi) \quad (5)$$

where

$$f_r^m(\theta, \phi) = \sum_{n=-m}^{n=m} a_{mn} \sqrt{\frac{(2m+1)(m-n)!}{4\pi(m+n)!}} P_{mn}(\cos \theta) e^{im\phi}$$

P_{mn} are associated Legendre polynomials (refer [11]).

- iv. Finally, the L_2 -norm of each frequency component is computed, at each radius. The resultant feature vector is 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. EXPERIMENTAL RESULTS

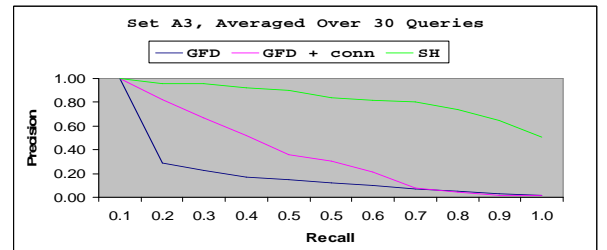
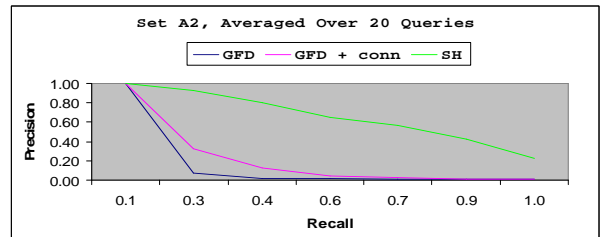
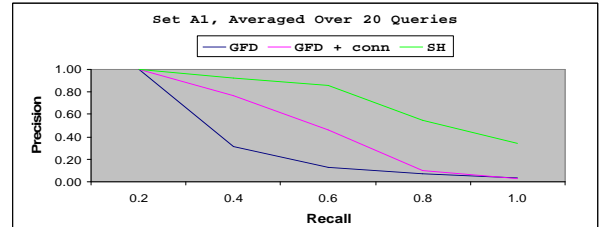
Experiments are performed on Item number S8 within the MPEG-7 Still Images Content Set; this is a collection of

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. Point clouds are represented by xyz files and include connectivity information. Polygonal meshes are generated from the xyz file using Delaunay triangulations. The polygonal meshes are converted into PLY format (introduced by Stanford University) [6]. 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.

Queries are performed using the proposed method. Another set of queries are performed using GFD. In Figure 5, average recall-precision has been plotted for each method. GFD is represented by 'GFD' within the legends. Spherical harmonics method is represented by 'SH'.

Another method is represented by 'GFD + conn' within the legends. In this method, an image is indexed by deriving GFD for each value of connectivity (0 through 8). Hence, a set of nine descriptors based on GFD are used to index an image. The distance between two images is computed as the sum of differences between GFD for each value of connectivity.



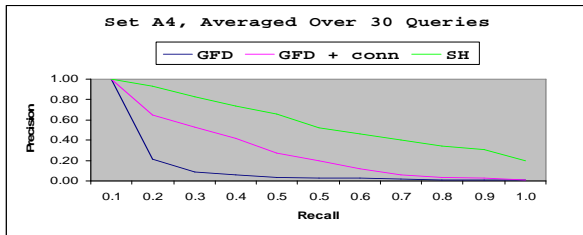


Figure 5. Average Recall-Precision Plots

5. DISCUSSION

There are two factors which contribute to the relative improvement of the proposed method when compared with GFD. First, additional information is captured by connectivity; descriptors which encode connectivity are able to discriminate better between shapes [7]. We note that the dataset does not contain fine contours. In Figure 2, we see that the pixel density is high for connectivity=0 and connectivity=8. We believe that the relative improvement in the effectiveness of the proposed method will be more with an increase in pixel densities for intermediate values of connectivity. In the future, we will perform experiments on different datasets to test the veracity of the statement above.

The second reason contributing to the effectiveness of the proposed method is the use of spherical harmonics. 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: decomposing image by connectivity, triangulations of the point cloud to generate a mesh and 3d modeling of the mesh. Computation of connectivity has complexity $O(n)$ where n is the number of foreground pixels in the image. Efficient methods for Delaunay triangulations have computational complexity $O(n^2)$ [8]. Efficient methods for computing spherical harmonics of a spherical function have been developed [9][10][11] which

have complexity $O(b^2 \log b^2)$ sampled on a regular $O(b^2)$ grid. The intense processing requirements may be prohibitive for some applications. However, for applications where accuracy of retrieval is important, the improvement in effectiveness may outweigh the processing complexity.

6. CONCLUSION

In this paper, we have used spherical harmonics descriptors to represent 2d images which are decomposed by connectivity. Spherical harmonics are used because of the proven accuracy of this method. Experiments have been performed on the MPEG-7 Still Images Content Set. Experimental results prove that the proposed method improves accuracy of retrieval significantly when compared with Generic Fourier Descriptors. The proposed method may be modified to incorporate other 3d modeling techniques, however, this will need further investigation.

REFERENCES

- [1] D. S. Zhang, "Image Retrieval Based on Shape", PhD Thesis, Monash University, Australia, 2002.
- [2] D. S. Zhang and G. Lu, "Generic Fourier Descriptors for Shape-based Image Retrieval", IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland, August 26-29, 2002.
- [3] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval", IEEE Workshop on Applications of Computer Vision, pp. 96-102, December 1996.
- [4] M. Kazhdan, T. Funkhouser and S. Rusinkiewicz, "Rotation Invariant Spherical Harmonic Representation of 3d Shape Descriptors", Symposium on Geometry Processing, pp. 167-175, June 2003.
- [5] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin and D. Jacobs, "A search engine for 3d models", ACM transactions on Graphics, pp. 83-105, 2003.
- [6] http://www.cc.gatech.edu/projects/large_models/
- [7] A. Sajjanhar, G. Lu and D. S. Zhang, "Discriminating Shape Descriptors Based on Connectivity", IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, 2004.
- [8] D. Attali and J. Boissonnat, "A Linear Bound on the Complexity of the Delaunay Triangulation of Points on Polyhedral Surfaces", Discrete and Computational Geometry. Vol. 31, No. 3, pp. 369-384, 2004.
- [9] J. Driscoll and D. Healy, "Computing Fourier transforms and convolutions on the 2-sphere", Advances in Applied Mathematics, Vol. 15, pp. 202-250, 1994.
- [10] D. Healy, D. Rockmore, P. Kostelec and S. Moore, "FFTs for the 2-sphere – improvements and variations", Journal of Fourier Analysis and Applications, Vol. 9, pp. 341-285, 2003.
- [11] Michael Kazhdan, "Shape Representations and Algorithms for 3d Model Retrieval", PhD Thesis, Princeton University, June 2004.