

3D Object Recognition by Fast Spherical Correlation between Combined View EGIs and PFT*

Donghui Wang, Hui Qian

Institute of Artificial Intelligence, Zhejiang University, China
dhwang@zju.edu.cn, qianhui@zju.edu.cn

Abstract

This paper proposes a method to recognize 3D object from probed range image under arbitrary pose by fast spherical correlation. First, all view EGIs under different viewpoints are extracted and combined onto a Gaussian sphere to form a feature description for each object. Second, the probed range image at arbitrary pose is represented as a PFT feature by phase-encoded Fourier transform and the PFT feature is mapped onto the Gaussian hemisphere by coordinates transforms and intensity scaling. Third, the spherical correlation algorithm based on spherical harmonic functions is used to do matching and similarity measurement between mapped PFT and combined view EGIs. The spherical correlation peak can output both of the recognition result and pose estimation. The experimental results proved that the proposed method can not only recognize totally different objects but also has enough discriminating capability for scalable dataset of similar objects.

1. Introduction

As an important spherical representation of global geometric feature of 3D object, the extended Gaussian image (EGI) has been proved it's effective for 3D object recognition [1], pose estimation [2] and 3D registration [3] and has been extended to some EGI-like methods, e.g. CEGI, DEGI and SAI [4]. Comparing with some typical recognition methods, such as ICP-like method [5], local feature based method (SPIN image) [6, 7] or others [8], the EGI-like

methods have some intrinsic advantages: 1) it is not affected by the translation of the object and rotates with the object. By normalization, it's also invariable to the scaling of object; 2) don't need to do initial registration; 3) recognition process is easy to implement by spherical correlation; 4) do recognition and pose estimation simultaneously. Otherwise, the EGI-like methods also need to face some problems: non-convex objects and partial obscured objects. In this paper, we propose a combined view EGIs that is an improved representation for our recognition scheme.

All EGI-like representations are based on the computation of surface normal or surface curvature [1]. For the recognition between two intact 3D objects, we can extract the EGI-like representation offline before doing recognition. But in the most of recognition applications, the input is a range image of 3D object from arbitrary view point and the recognition should be performed by matching an intact 3D object with a range image under particular view point. In such case, we need a fast mapping algorithm from range image to corresponding EGI representation. Unfortunately, there is lack of such algorithm introduced in literatures based on our knowledge. In this paper, we propose a fast algorithm that can map range image to view EGI representation by phase encoded Fourier transform (PFT) and recognize 3D object by fast spherical correlation between combined view EGIs and PFT. The formulas transformed among view EGI, combined view EGIs and PFT of range image are given in following section and experimental results of recognition show that the proposed method is effective and available.

The paper is structured as follows: First, the combined view EGIs feature of 3D object is introduced. Next, the phase encoded Fourier transform of range image and transformation between view EGI and PFT are covered. Then, a fast spherical correlation

*This work is partly supported by NSFC (60502029), Department of Education of Zhejiang Province (20060701), Science and Technology Department of Zhejiang Province (2007C23103, 2006C13096).

algorithm based on spherical harmonic functions is addressed. Finally, experimental results are presented to demonstrate the capabilities of proposed method.

2. Combine view EGIs onto Gaussian Sphere

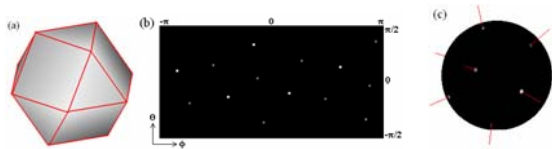


Figure 1 (a) A convex polyhedron. (b) EGI on θ - ϕ Coordinate. (c) Spike model of EGI on Gaussian sphere.

For convex polyhedron object, EGI is exhibited as a spike model on Gaussian sphere, see figure 1. For convex curved objects, EGI associates a point on the Gaussian sphere with a given point on a surface by finding the point on the sphere which has the same surface normal. In practice, most 3D objects (e.g. 3D face) are not convex and the Gaussian curvature for some points is negative. In this case, EGI can be extended to two different representations: 1) Adopt the sum of the absolute values of the inverses of the Gaussian curvature at all points having the same surface orientation. This will lead to that more than one point on 3D object surface, even when some of these points are obscured by other parts, will contribute to same point on the Gaussian sphere. 2) Use view EGIs instead of whole EGI and combine all view EGIs onto Gaussian sphere. Here, the view EGI is defined as a partial EGI that is dependent on the viewpoint and is produced from those parts of the object surface visible from a particular viewpoint. For 3D object recognition, we usually capture a range image by 3D sensing (e.g. laser scan, stereo vision, 3D range camera) as input from an arbitrary viewpoint and need to match this input data with 3D object indexed in database. The input range image just includes the visible parts of 3D object surface from single view line, so the view EGI is more suitable for such 2.5D-to-3D matching process.

An example of how to combine view EGIs onto Gaussian sphere is given in Figure 2. Here a re-sampling process to object surface from particular viewpoint is performed for generating each view EGI. If the combined view EGIs feature has been extracted, we can use it to recognize 3D object by computing the spherical correlation with single view EGI generated from any input range image at arbitrary viewpoint. In this recognition scheme, there are two key steps: 1) Fast production of view EGI from range image. 2) Fast

spherical correlation. The increasing precision on modeling and representation of 3D object model will conduce to extract refined EGI feature that approximates to truth and causes to more computation concomitantly.

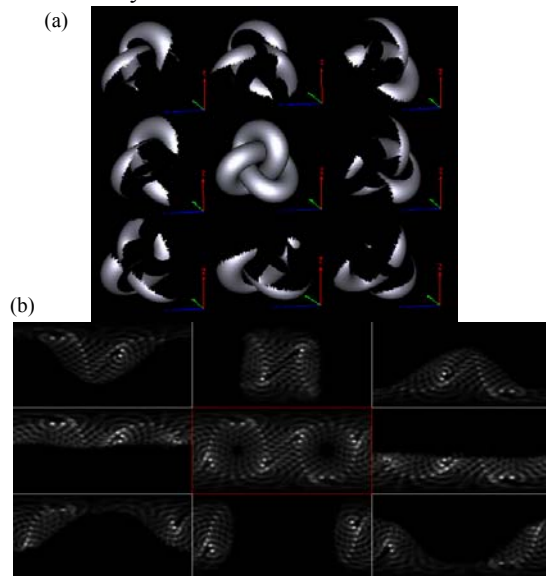


Figure 2: (a) Torus object (center) and its visible surfaces at 8 particular viewpoints rotated around x-axis. (b) Corresponding view EGIs and their combination on θ - ϕ Coordinate (center)

3. Phase-Encoded Fourier Transform of Range image

A range image stores 3D geometrical information of object in a 2D image from given view line, which defines the depth map $z(x, y)$ at sampling points (x, y) . The depth value comes between sampling points can be interpolated linearly or non-linearly by the z values of vertexes of covered facet. In the case of linear interpolation, the 3D object shape can be approximated by a polygonal mesh. The phase encoded Fourier transform (PFT) of a range image can map a facet of such polygonal mesh in spatial domain into a peak of spectral domain. The position and distribution of the peak represent the orientation and the boundary of the facet, respectively.

Let consider a range image which includes N individual facets with boundary function $s_i(x, y)$ defined as

$$z(x, y) = \sum_{i=1}^N (a_i \cdot x + b_i \cdot y + c_i) \cdot s_i(x, y) \quad (1)$$

where $s_i(x, y) = 1$ if the point (x, y) belongs to the i^{th} facet and zero otherwise. The normal vector of the i^{th} facet is $(a_i, b_i, 1)$. The PFT of the range image is given by

$$PFFT(u, v) = \mathcal{F}\{e^{i-wz(x, y)}\} = \mathcal{F}\left\{e^{i-w\sum_{i=1}^N (a_i x + b_i y + c_i) s_i(x, y)}\right\} \quad (2)$$

$$= \sum_{i=1}^N \left[e^{iwc_i} \delta(u - \gamma a_i, v - \gamma b_i) \otimes S_i(u, v) \right]$$

where w is a scaling factor, $\gamma = w/2\pi$, $\mathcal{F}\{\}$ is the Fourier transform, $\delta()$ is the delta function, \otimes is the convolution operator and $S_i(u, v) = \mathcal{F}\{s_i(x, y)\}$. $PFFT(u, v)$ is the cumulative sum of the PFT of all facets. For i^{th} facet, its PFT is centered in a peak $(\gamma a_i, \gamma b_i)$ given by the orientation of the normal to the facet and shape of the peak is given by $S_i(u, v)$.

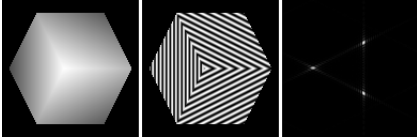


Figure 3: Range image of a cube (left), Phase-encoded map of range image (middle) and PFT of range image (right).

The range image, phase-encoded map and PFT of a cube object are shown in Figure 3. The salient three peaks in PFT are corresponding to three visible sides of the cube and the shape of each peak is corresponding to the FT of boundary of each side.

The intensity of the PFT exhibits a crucial property: 1) it is invariant to arbitrary translations of the 3D object; 2) it is rotated with 3D object.

4. Transform between view EGI and PFT

Let consider a facet its normal vector is $(a, b, 1)$ and related spherical coordinates on \mathbb{S}^2 are (θ, ϕ) . The relationship between normal vector $(a, b, 1)$ and (θ, ϕ) is given by:

$$\begin{pmatrix} a & b \end{pmatrix} = \begin{pmatrix} \tan(\phi) & \tan(\theta)/\cos(\phi) \end{pmatrix} \quad (3)$$

θ has a range of $(0, \pi)$ and ϕ has a range of $(0, 2\pi)$. We adopt a uniform sampling grid (x, y) to represent spherical coordinates (θ, ϕ) on the sphere \mathbb{S}^2 . The basic relationship between spherical coordinates (θ, ϕ) and (u, v) coordinates in Phase-encoded Fourier transform image can be described as:

$$(u - N_u/2, v - N_v/2) = (N_u \times a/2\pi, N_v \times b/2\pi) \quad (4)$$

$$= \left(\frac{\tan(\phi)}{2\pi} \times N_u, \frac{\tan(\theta)}{2\pi \cos(\phi)} \times N_v \right)$$

$$= \left(\frac{\tan(2\pi \times x/L_x - \pi)}{2\pi} \times N_u, \frac{\tan(\pi \times y/L_y - \pi/2)}{2\pi \cos(2\pi \times x/L_x - \pi)} \times N_v \right)$$

Where N_u, N_v are sampling resolutions of PFT image at u and v directions, L_x, L_y are sampling resolutions

of EGI image at x and y directions. The inverse transform can be described as:

$$(x - L_x/2, y - L_y/2) = (L_x \times \phi/2\pi, L_y \times \theta/\pi) \quad (5)$$

$$= (L_x \times \tan^{-1}(a)/2\pi, L_y \times \tan^{-1}(b \cos(\tan^{-1}(a)))/\pi)$$

$$\tan^{-1}\left(\frac{2\pi u}{N_u}\right) \quad \tan^{-1}\left(\frac{2\pi v}{N_v} \cos\left(\tan^{-1}\left(\frac{2\pi u}{N_u}\right)\right)\right)$$

$$= \left(\frac{1}{2\pi} \times L_x, \frac{1}{\pi} \times L_y \right)$$

The intensity change is given as:

$$PFFT(u, v) = EGI(\theta, \phi) \cdot \cos(\theta) \cdot \cos(\phi) \quad (6)$$

For a face range image, its range image, phase-encoded map, PFT are shown in Figure 4(a). The view EGI transformed from PFT and combined view EGIs are shown in Figure 4(b). From Figure 4, we can find that the view EGI can be obtained by simple Fourier transform of phase encoded range image instead of complex surface curvature computation in section 2. Because the Fourier transform can be implemented not only by DSP hardware but also optical system [9], that means above transform can be very fast and it's very useful for practical application.

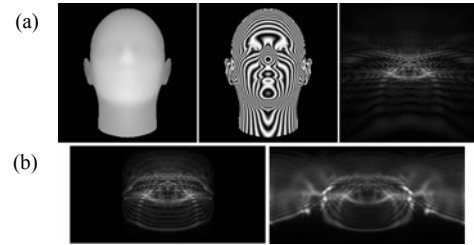


Figure 4: (a) Range image, Phase-encoded map and PFT. (b) Transform PFT to view EGI (left) and combined view EGIs of 3D face (right)

5. Fast Spherical Correlation based on spherical harmonic functions

We can regard both of combined view EGIs and view EGI from PFT as two feature functions on unit sphere \mathbb{S}^2 . Then the recognition task of 3D object from an input of range image can be converted to a spherical correlation between two spherical functions. By detecting the global maximal peak in correlation result, we can decide if the input range image matches with a 3D object in database.

Let consider two feature functions on unit sphere are $f(\omega)$ and $h(\omega)$ respectively. We can define the correlation on unit sphere as:

$$K(\alpha, \beta, \gamma) = \int_{\omega \in \mathbb{S}^2} f(\omega) \Lambda_{g(\alpha, \beta, \gamma)} h(\omega) d(\omega) \quad (7)$$

Where, $\omega = (\theta, \phi)$ is the spherical coordinate, $\Lambda_{g(\alpha, \beta, \gamma)}$ is a rotation arithmetic operator on spherical function, $g(\alpha, \beta, \gamma)$ is an element in rotation basis $SO(3)$, and $0 \leq \alpha, \gamma \leq 2\pi$, $0 \leq \beta < \pi$, we have:

$$g(\alpha, \beta, \gamma) = R_z(\alpha) \cdot R_y(\beta) \cdot R_z(\gamma) \quad (8)$$

where $R_z(A)$, $R_y(A)$ represent the rotation matrix along the z-axis and y-axis respectively.

As in the planar case, the correlation on unit sphere can be performed on the frequency domain by using spherical Fourier transforms [10]. Here the spherical harmonic functions are used to do the Fourier expansion instead of linear exponential functions. The total complexity of spherical correlation can be reduced from $O(M^3N^2)$ to $O(L^3 \log^2 L)$ by using a rapid discrete implementation SOFT in [11], where M is the number of samples in each dimension of $SO(3)$, N is related to the size of the spherical histogram and L is the bandwidth of the spherical signal.

6. Experiments

We designed two experiments to test the performance of proposed method. One experiment is to test if the feature is robust to sensing noise and another is for the test of discriminating capability. Figure 5 demonstrates the test result of noise effect that usually happens in range sensing step. From Figure 5(c), we found that the PFT feature is robust to surface noise and it can reach enough stable property by adopting simple surface smoothing technology.

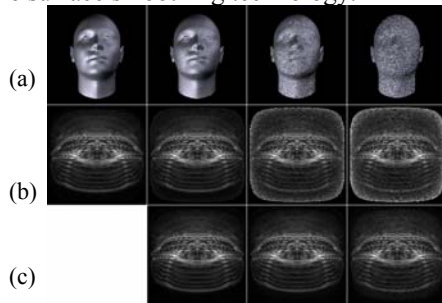


Figure 5 (a) From left to right: input range image without noise, with 0.5% noise, 2% noise and 5% noise. (b) PFT-transformed view EGI of (a) without surface smoothing. (c) After Gaussian surface smoothing.

Figure 6 gives another test result of discriminating capability of our proposed method. Here, we implemented an experimental system of 3D object recognition for both of general 3D objects and 3D faces. The combined view EGIs for each object is extracted offline and indexed in database. For input range image under arbitrary pose, we use above proposed method to test the capability for recognition and find that the intensity of the correlation peak represents the similarity of two objects as well as the location of the correlation peak gives the pose estimation. The results proved that the proposed

method can recognize not only totally different objects but also similar objects such as 3D faces that show in Figure 6(a). The distance matrix of recognizing 150 3D faces is shown in Figure 6(b) and we can find that the proposed method has enough discriminating capability for scalable dataset of similar objects.

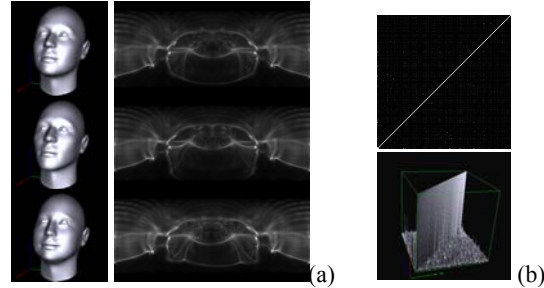


Figure 6 (a) Three similar 3D faces and combined view EGIs. (b) Distance matrix of 150 3D faces.

References

- [1] B.K.P. Horn. Extended Gaussian Images. *IEEE* (72), No.12, pp.1656-1678, December 1984.
- [2] Y. Someya, K. Kawamura, K. Hasegawa, K. Ikeuchi: Localization of Objects in Electric Distribution Systems by Using Segmentation and 3D Template Matching with M-Estimators. *ICPR 2000*: 1725-1729
- [3] Ameesh Makadia, Alexander IV Patterson, Kostas Daniilidis, Fully Automatic Registration of 3D Point Clouds, *CVPR 2006*, p.1297-1304, June 17-22, 2006
- [4] A. Adan, C. Cerrada, V. Feliu. Global shape invariants: a solution for 3D free-form object discrimination / identification problem. *Pattern Recognition* (34), No. 7, July 2001, pp. 1331-1348.
- [5] Yang Chen, Gerard Medioni, Object modelling by registration of multiple range images, *Image and Vision Computing*, v.10 n.3, p.145-155, April 1992
- [6] Chen, H., Chen, H., and Bhanu, B. 2007. 3D free-form object recognition in range images using local surface patches. *Pattern Recognition Letter*. 28, 10 (Jul. 2007), 1252-1262.
- [7] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 5, May, 1999, pp. 433 - 449.
- [8] Richard J. Campbell, Patrick J. Flynn, A survey of free-form object representation and recognition techniques, *Computer Vision and Image Understanding*, v.81 n.2, p.166-210, February 1, 2001
- [9] S. Chang, M. Rioux and C. P. Grover. Range face recognition based on the phase Fourier transform. *Optics Communications*, 222:143-153, 2003.
- [10] S. Kunis, D. Potts. Fast spherical Fourier algorithms. *J. Comput. Appl. Math.* 161, 75-98 (2003).
- [11] P. J. Kostelec and D. N. Rockmore. FFTs on the Rotation Group. In *Working Papers Series, Santa Fe Institute*, 2003