

3D MODEL RETRIEVAL USING 2D CEPSTRAL FEATURES

Chang-Hsing Lee, Jau-Ling Shih, Chih-Hsun Chou, Kung-Ming Yu, Chuan-Yen Hung

Department of Computer Science and Information Engineering
Chung Hua University, Hsinchu, Taiwan

ABSTRACT

In this paper, we will propose a 3D model retrieval approach using 2D cepstral features. First, six projection planes representing the elevation (depth) value are generated. Then, 2D cepstral features are extracted from each projection plane for searching similar 3D models. Experiments conducted on the Princeton Shape Benchmark (PSB) database have shown that the proposed 2D cepstral features outperforms other state-of-the-art descriptors in terms of the DCG score.¹

Index Terms— 3D model retrieval, 2D cepstral feature

1. INTRODUCTION

The development of computer graphics and computer animations has made 3D models as plentiful as images and video. Therefore, it is necessary to design an automatic 3D model retrieval system which enables the users to search interested 3D models efficiently and effectively. The main challenge to a content-based 3D model retrieval system is how to extract representative features to effectively discriminate the shapes of various 3D models [1].

Vranic et al. applied Fourier transform to the sphere using spherical harmonics to generate embedded multi-resolution 3D shape features [2]. However, pose normalization must be conducted prior to feature extraction to make the extracted feature rotation invariant. Therefore, Funkhouser et al. proposed a modified rotation invariant shape descriptor based on the spherical harmonics in which no pose normalization is needed [3].

The simplest features used to represent 3D models are based on the statistics of geometric characteristics [4]-[7]. Ankerst et al. proposed a method to search similar 3D models using shape histograms which describe the area of intersections of a 3D model with a collection of concentric shells and sectors [4]. The MPEG-7 shape spectrum descriptor (SSD) [5] calculates the histogram of the curvatures of all points on 3D surface. Osada et al. [6] used five geometric features, notated by A3, D1, D2, D3, and D4,

to represent 3D models by the probability distributions of some geometric properties computed from a set of randomly selected points on the surface of the model. These features are sensitive to tessellation of 3D models. Thus, Shih et al. [7] proposed grid D2 (GD2) to improve D2. In GD2, a 3D model is first decomposed into a voxel grid. The distribution of distances between any two randomly selected valid grids is computed to represent a 3D model.

Generally, 3D models can also be characterized by its 2D silhouettes viewed from different directions [8]-[10] such that users can search similar 3D models by 2D shape features. Super and Lu [8] exploited 2D silhouette contours for 3D object recognition. Curvature and contour scale space features are extracted to represent each silhouette. Chen et al. [9] proposed the LightField descriptor (LFD) computed from 10 silhouettes to represent 3D models. Each silhouette is represented by a binary image. In fact, 2D silhouettes cannot describe the altitude (depth) information of 3D models from different views. Thus, Shih et al. [10] proposed the elevation descriptor (ED) to represent the altitude information of a 3D model from six different views.

Kuo and Cheng [11] used principal plane analysis for 3D shape retrieval. The principal plane is defined as the symmetric plane on which the sum of distance of all points projected is minimal. First, each 3D model is projected onto its principal plane. Thus, each 3D model can be represented by a binary image. The feature vectors extracted from the binary shape image are then used for similar model retrieval. Typically, one projected binary image cannot represent a complex 3D model effectively. Therefore, Shih et al. [12] proposed the principal plane descriptor (PPD), in which three binary images are derived to describe each 3D model by projecting it on the principal, second, and third planes. Feature vectors are then extracted from these binary images. Vranic and Saupe proposed a modified principal component analysis (PCA) for pose alignment of 3D models using the triangle areas as weighting factors for covariance matrix computation [13]. The extracted features include the directions of 20 vertices on dodecahedron and the distances computed from the center point to the farthest intersections.

In this paper, we will propose a 3D model retrieval method using 2D cepstral features. In Section 2, the proposed 3D model retrieval system will be described. Section 3 gives some experimental results to show the

¹ This research was supported in part by the National Science Council of R.O.C. under contract NSC-100-2221-E-216-035.

effectiveness of the proposed features. Finally, conclusions are given in Section 4.

2. PROPOSED 3D MODEL RETRIEVAL SYSTEM USING 2D CEPSTRAL FEATURES

First, each 3D model is decomposed into a number of voxels. Second, the principal planes method [12] is used for pose alignment of each 3D model. Third, the elevation (depth) value of each voxel will be projected onto six viewing planes. Finally, 2D cepstral features extracted from these projection planes will be used for 3D model retrieval.

2.1. 3D Model Alignment

First, the pose of the input 3D model is aligned by the principal planes method [12]. The smallest bounding cube that circumscribes the 3D model is then decomposed into a voxel grid of size $128 \times 128 \times 128$. If there is a mesh located within the voxel located at coordinates (x, y, z) , this voxel is defined as an opaque voxel, denoted by $V(x, y, z) = 1$; otherwise, this voxel is defined as a transparent voxel, denoted by $V(x, y, z) = 0$. To make the extracted features robust to translation and scaling, we transform the 3D model such that its mass center becomes $(0, 0, 0)$ and the average distance from all opaque voxels to the mass center is 32.

Once the pose of a 3D model is aligned, the elevation (depth) value of each opaque voxel will be projected onto six planes indicating the six different views of the 3D model (see Fig. 1). The elevation value describes the distance from the opaque voxel to the viewing plane. Each projection plane is represented by a gray level image from which 2D cepstral features will be extracted to represent a 3D model.

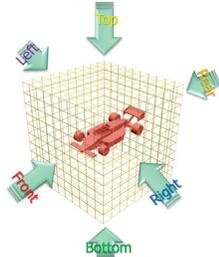


Fig. 1. The six views of a 3D model.

2.2. Elevation Projection

The elevation value can capture the depth information of the model's surface to each viewing plane. For an opaque voxel located at (x, y, z) , the elevation value is measured as its distance from the projection plane. Let the six projection planes be denoted by $I_k, k = 1, 2, \dots, 6$. Then, the gray value, indicating the elevation value, of each pixel on these projected images is computed as follows:

$$I_1(x, y) = \max_{0 \leq z \leq 64} ((65 - z)V(x, y, z)), -64 \leq x, y \leq 64 \quad (1)$$

$$I_2(x, z) = \max_{0 \leq y \leq 64} ((65 - y)V(x, y, z)), -64 \leq x, z \leq 64 \quad (2)$$

$$I_3(y, z) = \max_{0 \leq x \leq 64} ((65 - x)V(x, y, z)), -64 \leq y, z \leq 64 \quad (3)$$

$$I_4(x, y) = \max_{-64 \leq z \leq 0} ((65 + z)V(x, y, z)), -64 \leq x, y \leq 64 \quad (4)$$

$$I_5(x, z) = \max_{-64 \leq y \leq 0} ((65 + y)V(x, y, z)), -64 \leq x, z \leq 64 \quad (5)$$

$$I_6(y, z) = \max_{-64 \leq x \leq 0} ((65 + x)V(x, y, z)), -64 \leq y, z \leq 64 \quad (6)$$

Once the six projection images are generated, 2D cepstral features will then be extracted from each projection image.

2.3. 2D Cepstral Feature Extraction

Fig. 2 shows the flow diagram for 2D cepstral feature extraction. First, 2D discrete Fourier transform (2D-DFT) of each 2D projection image $I(x, y)$ is computed as follows

$$F(u, v) = \text{DFT}(I(x, y)) \quad (7)$$

where (u, v) is 2D spectral frequency index. The 2D spectrum is then decomposed into $M \times N$ subbands $B_{\rho, \theta}$ ($0 \leq \rho \leq M-1, 0 \leq \theta \leq N-1$) along the radial and angular directions. The energy of each subband is then computed as follows:

$$E(\rho, \theta) = \sum_{(u, v) \in B_{\rho, \theta}} |F(u, v)|^2, 0 \leq \rho \leq M-1, 0 \leq \theta \leq N-1 \quad (8)$$

In this paper, two subband decomposition methods, called generic subband decomposition (GSD) and complement subband decomposition (CSD), will be employed to divide the 2D spectrum into several subbands (please see Fig. 3). In GSD, $(u, v) \in B_{\rho, \theta}$ if

$$\frac{64\rho}{M} \leq \sqrt{u^2 + v^2} < \frac{64(\rho + 1)}{M} \quad (9)$$

and

$$\frac{2\pi\theta}{N} \leq \tan^{-1} \frac{v}{u} < \frac{2\pi(\theta + 1)}{N} \quad (10)$$

In CSD, $(u, v) \in B_{\rho, \theta}$ if

$$\frac{64\rho}{M} \leq \sqrt{u^2 + v^2} < \frac{64(\rho + 1)}{M} \quad (11)$$

and

$$\frac{2\pi\theta + \pi}{N} \leq \tan^{-1} \frac{v}{u} < \frac{2\pi(\theta + 1) + \pi}{N} \quad (12)$$

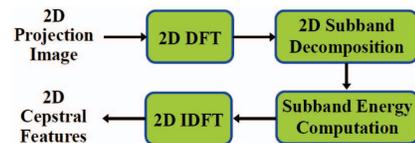


Fig. 2. Flow diagram for 2D cepstral feature extraction.

The 2D cepstrum $C(p, q)$ can be derived by taking the inverse DFT of the subband energy spectrum $E(\rho, \theta)$:

$$C(p, q) = \text{DFT}^{-1}(E(\rho, \theta)), 0 \leq p \leq M-1, 0 \leq q \leq N-1 \quad (13)$$

where (p, q) is 2D cepstral quefrequency index, DFT^{-1} is 2D inverse DFT. The magnitudes of these $M \times N$ cepstral coefficients will constitute the 2D cepstral feature vector:

$$\begin{aligned} \mathbf{f} &= [f(1), f(2), \dots, f(MN)]^T \\ &= [|C(0, 0)|, \dots, |C(0, N-1)|, \dots, \\ &\quad |C(M-1, 0)|, \dots, |C(M-1, N-1)|]^T \end{aligned} \quad (14)$$

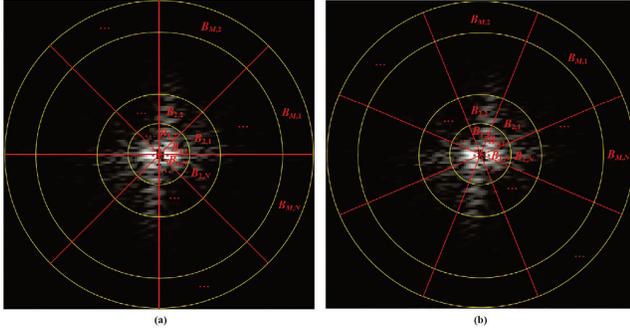


Fig. 3. Subband decomposition of 2D DFT spectrum (a) generic subband decomposition (b) complement subband decomposition

In this paper, the 2D cepstral feature vectors extracted by using GSD and CSD techniques will be combined to search similar 3D models. Let $\mathbf{f}^s = [(\mathbf{f}_1^s)^T, \dots, (\mathbf{f}_6^s)^T]^T$ and $\mathbf{f}^c = [(\mathbf{f}_1^c)^T, \dots, (\mathbf{f}_6^c)^T]^T$ denote respectively the feature vectors extracted from the six projection planes of the query model corresponding to GSD and CSD techniques. Similarly, let $\mathbf{y}^s = [(\mathbf{y}_1^s)^T, \dots, (\mathbf{y}_6^s)^T]^T$ and $\mathbf{y}^c = [(\mathbf{y}_1^c)^T, \dots, (\mathbf{y}_6^c)^T]^T$ denote the feature vectors extracted from the matching model in the database. The distance between the query model and the matching model corresponding to GSD and CSD are defined as follows:

$$d^s(\mathbf{f}^s, \mathbf{y}^s) = \sum_{k=1}^6 \|\mathbf{f}_k^s - \mathbf{y}_k^s\| = \sum_{k=1}^6 \sum_{i=1}^{MN} |f_k^s(i) - y_k^s(i)| \quad (15)$$

$$d^c(\mathbf{f}^c, \mathbf{y}^c) = \sum_{k=1}^6 \|\mathbf{f}_k^c - \mathbf{y}_k^c\| = \sum_{k=1}^6 \sum_{i=1}^{MN} |f_k^c(i) - y_k^c(i)| \quad (16)$$

The combined distance between the query model and the matching model is defined as the sum of these two distances:

$$d(\mathbf{f}, \mathbf{y}) = d^s(\mathbf{f}^s, \mathbf{y}^s) + d^c(\mathbf{f}^c, \mathbf{y}^c) \quad (17)$$

The matching models that have the minimum combined distances will be regarded as the retrieved similar models.

3. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the proposed feature, some experiments have been conducted on the Princeton Shape Benchmark (PSB) database [15]. The PSB database contains 1814 models (161 classes) which are divided into 907 training models (90 classes) and 907 test models (92 classes). The discounted cumulative gain (DCG) [16], will be employed to compare the performance of different descriptors. DCG at the r -th rank is defined as follows:

$$\text{DCG}_r = \begin{cases} \text{DCG}_{r-1} + \frac{L_r}{\log_2(r)} & r \geq 2 \\ L_1 & r = 1 \end{cases} \quad (18)$$

where $L_r = 1$ if the r -th model in the ranked retrieval list and the query one belong to the same class; otherwise, $L_r = 0$. The overall DCG score for a query model is defined as DCG_{kmax} , where $kmax$ is the total number of models in the database. It is clear that if the models appearing in the head of the retrieval list have the same class label as the query one, a larger DCG score can be obtained. On the other hand, if the models with identical class label to the query one appear in the tail of the retrieval list, a small DCG score will be obtained.

In our experiments, each model in the database will be presented as a query one to evaluate the average DCG score. Table 1 compares the retrieval results of the proposed descriptors with other state-of-the-art descriptors in terms of DCG score. In this table, 2DCEP-G($M \times N$), 2DCEP-C($M \times N$), and 2DCEP-GC($M \times N$) denote the proposed 2D cepstral descriptors extracted by using GSD, CSD, and their combination, respectively. In this table, M and N are the number of decomposed segments in radial and angular directions. The experimental results show that 2DCEP-GC($M \times N$) outperforms 2DCEP-G($M \times N$) and 2DCEP-C($M \times N$). In addition, (16 \times 16) 2D subband decomposition outperforms (8 \times 8) subband decomposition. Further, the combination of (16 \times 16) and (8 \times 8) subband decomposition, denoted by 2DCEP-GC(8 \times 8)+2DCEP-GC(16 \times 16), yields the best performance. It can also be seen that each of the proposed 2D cepstral descriptors outperform the other descriptors in terms of DCG score.

4. CONCLUSIONS

In this paper, 2D cepstral features are proposed for 3D model retrieval. First, six projection planes, represented as gray-level images, will be generated to describe the altitude (depth) value of each 3D voxel from six different views. 2D cepstral features are then used to extract the feature vector from each projection plane. Experiments on PSB database have shown that the proposed 2D cepstral descriptors outperform other state-of-the-art descriptors in terms of DCG score.

Table 1. Comparison of the proposed descriptors with other descriptors in terms of the DCG score (%). Note that the approaches marked with * are implemented by Akgul et al. [16].

Descriptor	DCG
SH [3]	58.35
SSD [5]	48.07
GD2 [7]	60.91
LF [9]	64.30
ED[10]	67.04
RISH [11]*	58.40
PPD [12]	65.86
CRSF [14]	66.80
DBF [16]	65.90
DSR+DBF [16]	70.20
AED [17]	70.29
DED[18]	66.92
CED[18]	68.04
EGI [19]	43.80
SH-GEDT [20]	58.40
DBI [21]*	66.30
DSR [21]*	66.50
SIL [21]*	59.70
SWD [22]*	65.40
3DHT [23]*	57.70
CAH [24]*	43.30
REXT [25]*	60.10
2DCEP-G(8×8)	71.26
2DCEP-C(8×8)	70.82
2DCEP-GC(8×8)	71.70
2DCEP-G(16×16)	72.48
2DCEP-C(16×16)	71.84
2DCEP-GC(16×16)	72.63
2DCEP-GC(8×8)+2DCEP-GC(16×16)	72.88

5. REFERENCES

- [1] J. W. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," *Multimedia Tools Applicat.*, vol. 39, pp. 441-471, 2008.
- [2] D. V. Vranic, D. Saupe, and J. Richter, "Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics," in *Proc. IEEE Workshop Multimedia Signal Process.*, 2001, pp. 293-298.
- [3] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs, "A search engine for 3D models," *ACM Trans. Graphics*, vol. 22, no. 1, pp. 83-105, 2003.
- [4] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Trans. Graphics*, vol. 21, no. 4, pp. 807-832, Oct. 2002.
- [5] S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7 Multimedia Content Descriptor Interface*. John Wiley & Sons, 2002.
- [6] M. Ankerst, G. Kastentmüller, H. P. Kriegel, and T. Seidl, "3D shape histograms for similarity search and classification in spatial databases," in *Proc. 6th Int. Symp. Advances in Spatial Databases*, 1999, pp. 207-226.
- [7] J. L. Shih, C. H. Lee, and J. T. Wang, "3D object retrieval system based on grid D2," *Electron. Lett.*, vol. 41, no. 4, pp. 23-24, Feb. 2005.
- [8] B. J. Super and H. Lu, "Evaluation of a hypothesizer for silhouette-based 3-D object recognition," *Pattern Recognition*, vol. 36, no. 1, pp. 69-78, Jan. 2003.
- [9] D. Y. Chen, X. P. Tian, Y. T. Shen, and M. Ouhyoung, "On visual similarity based 3D model retrieval," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 223-232, Sep. 2003.
- [10] J. L. Shih, C. H. Lee, and J. T. Wang, "A new 3D model retrieval approach based on the elevation descriptor," *Pattern Recognition*, vol. 40, no. 1, pp. 283-295, Jan. 2007.
- [11] C. T. Kuo and S. C. Cheng, "3D model retrieval using principal plane analysis and dynamic programming," *Pattern Recognition*, vol. 40, no. 2, pp. 742-755, Feb. 2007.
- [12] J. L. Shih and W. C. Wang, "A 3D model retrieval approach based on the principal plane descriptor," in *Proc. 2nd Int. Conf. Innovative Computing, Inform. Control*, 2007.
- [13] D. V. Vranic and D. Saupe, "3D model retrieval," in *Proc. Spring Conf. Comput. Graph. Applicat.*, 2000, pp. 89-93.
- [14] P. Papadakis, I. Pratikakis, S. Perantonis, and T. Theoharis, "Efficient 3D shape matching and retrieval using a concrete radicalized spherical projection representation," *Pattern Recognition*, vol. 40, no. 9, pp. 2437-2452, Sep. 2007.
- [15] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape benchmark," in *Proc. Shape Modeling*, 2004, pp. 167-178.
- [16] C. B. Akgul, B. Sankur, Y. Yemez, and F. Schmitt, "3D model retrieval using probability density-based shape descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1117-1133, June 2009.
- [17] J. L. Shih and H. Y. Chen, "A 3D model retrieval approach using the interior and exterior 3D shape information," *Multimedia Tools Applicat.*, vol. 43, no. 1, pp. 45-62, 2009.
- [18] J. L. Shih, T. Y. Huang, and Y. C. Wang, "A 3D model retrieval system using the derivative elevation and 3D-ART," in *Proc. IEEE Asia-Pacific Services Computing Conf. (APSSC'2008)*, 2008, pp. 739-744.
- [19] B. K. P. Horn, "Extended Gaussian images," *Proc. IEEE*, vol. 72, no. 12, pp. 1671-1686, Dec. 1984.
- [20] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Proc. Eurographics/ACM SIGGRAPH Symp. Geometry Process.*, 2003, pp. 156-164.
- [21] D. V. Vranic, "3D model retrieval," Ph.D. Dissertation, University of Leipzig, Department of Computer Science, 2004.
- [22] H. Laga, H. Takahashi, and M. Nakajima, "Spherical wavelet descriptors for content-based 3D model retrieval," in *Proc. IEEE Int. Conf. Shape Modeling Applicat.*, 2006.
- [23] T. Zaharia and F. J. Preteux, "Shape-based retrieval of 3D mesh models," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2002, vol. 1, pp. 437-440.
- [24] E. Paquet and M. Rioux, "Nefertiti: A query by content software for three-dimensional models databases Management," in *Proc. Int. Conf. Recent Advances in 3D Digital Imaging and Modeling*, 1997, pp. 345-352.
- [25] D. V. Vranic, "An improvement of rotation invariant 3D shape descriptor based on functions on concentric spheres," in *Proc. IEEE Int. Conf. Image Process.*, 2003, pp. 757-760.