Boosting 3D Object Retrieval by Object Flexibility

Boqing Gong[1]   Chunjing Xu[1,2]   Jianzhuang Liu[1,2]   Xiaoou Tang[1,2]
[1]Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China
[2]Multimedia Lab, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China
{gbq008, cjxu6, jzliu, xtang}@ie.cuhk.edu.hk

ABSTRACT

In this paper, we propose a novel feature, called *object flexibility*, at a point of a 3D object to describe how the neighborhood of this point is massively connected to the object. We show that this feature is stable to the deformation of objects' articulations, in addition to commonly concerned linear transforms, i.e., translation, scale, and rotation. A shape descriptor is obtained based on this feature using the bag-of-words model. As an application, the descriptor is used to perform 3D object retrieval. Extensive experiments demonstrate its superiority over a variety of existing 3D shape descriptors in the retrieval of articulated objects, as well as its enhancement of other shape descriptors to retrieve generic 3D objects.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*

General Terms

Algorithms, Design

Keywords

3D object retrieval, features, object flexibility

1. INTRODUCTION

3D data are now widely recognized as the upcoming wave of digital media. 3D object retrieval rapidly becomes a key issue in this new multimedia content processing, and attracts more and more research interests. Some 3D object benchmarks and experimental retrieval systems have been made available, such as the Princeton shape benchmark and its associated search engine [9], and the NTU 3D model benchmark and its corresponding retrieval system [1]. The reader is referred to [10] for a comprehensive survey of this research.

The shape of a 3D object can be with arbitrary scale, location, and orientation. For generic 3D object retrieval, a retrieval method has to either perform a pose normalization process or use shape descriptors that are inherently invariant to linear transformations (translation, rotation, and scale). Much work has been presented in solving these problems so far. Shilane et al. compared twelve shape descriptors in [9]. Bimbo and Pala included five descriptors in their experimental analysis, which fall into statistics-based, geometry-based, and view-based categories [2].

On the other hand, less attention has been paid to the problem of non-linear shape deformations caused by objects' articulations. Zhang et al. carried out the first work of 3D articulated object retrieval using medial surfaces and their graph spectra, and provided a 3D articulated object database, the McGill 3D shape benchmark [11]. The main problem in the graph-based method is that it is sensitive to topological changes which are common in generic 3D models. Jain and Zhang tried to achieve articulation invariance in [5] by using the spectral embedding of an affinity matrix. Ion et al. used the continuous eccentricity transform to make their method insensitive to shape articulations [4]. However, these two methods [5][4] are sensitive to objects with disconnected parts or outliers.

In this paper, we propose a novel feature, called *object flexibility*, at a point of a 3D object to describe how the neighborhood of this point is massively connected to the object. This feature is stable to both linear transformations and non-linear deformations caused by objects' articulations. Based on this object flexibility, we propose a new shape descriptor for 3D object retrieval. Extensive experiments show that it outperforms a variety of existing shape descriptors in the retrieval of articulated 3D objects, which are often natural objects like animals, plants, and humans. Besides, combined with existing shape descriptors, it also helps to obtain better performance of retrieving generic 3D objects.

2. OBJECT FLEXIBILITY

In this section, we first define the object flexibility mathematically, and then we discuss how to compute it and to form the final shape descriptor based on it.

2.1 Definition

Definition 1. Given a radius r, let $C_{p,r} \subset \mathcal{O}$ be the set of points within a sphere S_r^3 centered at a point p of a 3D object $\mathcal{O}$. The object flexibility at p is defined as:

$$\rho_r(p) = \frac{Eig_2(X^T X)}{r}, \tag{1}$$
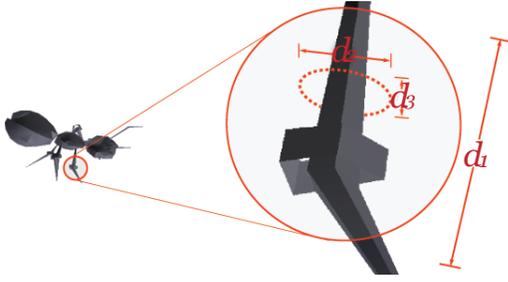
Figure 1: Illustration of the flexibility.
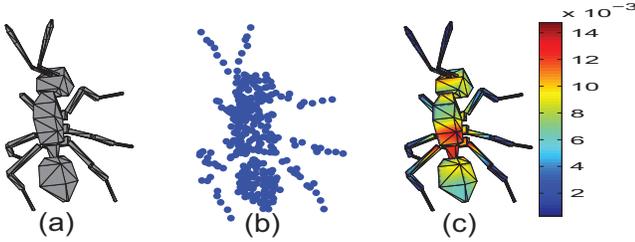
Figure 2: (a) A 3D model of an ant. (b) Sampled points of the ant. (c) The flexibility distribution on the ant.

where X is the data matrix consisting of the 3D coordinates of all the points in $C_{p,r}$, one point per row, and $Eig_2(\cdot)$ is a function that returns the second largest eigenvalue of a square matrix.

Consider a small part enclosed by S_r^3, centered at a point p on the leg of an ant in Fig. 1. Suppose that the three dimensions denoted by d_1, d_2, and d_3 are arranged as $d_1 \geq d_2 \geq d_3$ according to the variance on each dimension. In the definition, we estimate the normalized variance in the direction d_2, which can be a potential measure about how the local part (a segment of the ant's leg) is massively connected to the object (the ant), or how easily that part of the object can be bent. The smaller $\rho_r(p)$ is, the more tenuous that part is with more "bending ability".

Note that we do not use the variance in d_1 because it tends to be constant for different parts of an object when r is fixed. We also discard the third eigenvalue since it may degenerate to zero when the points in $C_{p,r}$ lie in a plane.

2.2 Computation of the Flexibility

Before computing the flexibility, we need to determine the radius r and the points of an object where their flexibilities are computed.

Since the flexibility describes local shape characteristics, we have to select enough points of a 3D model to obtain a complete shape description. For an object represented by voxels, we select all its surface points, each of which has less than 13 non-zero voxels among its 26 neighbor voxels. In our experiments, 590 surface points of a 3D model are left on average after filtering out inner points in the McGill database [11], where each model is represented by 128^3 voxels. For an object represented by meshes in the Princeton shape benchmark [9], 2000 surface points are sampled using a scheme presented in [8] in the first round, and then 500 points are randomly selected from them. We use all the 2000 points to compute the flexibilities of the 500 points. One example
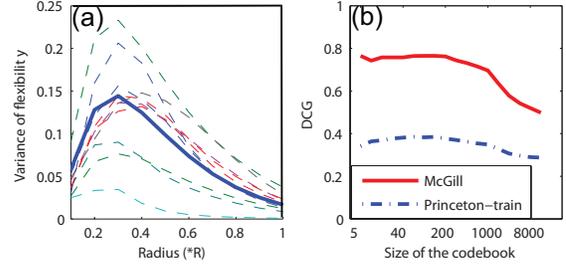
Figure 3: (a) Flexibility variance versus radius. (b) DCG versus size of the codebook.

of the sampled points from a mesh model is shown in Fig. 2(b).

Next, a proper radius r is determined for each selected point. It is easy to see that

$$\lim_{r \to 0} \rho_r(p) = 0, \quad \lim_{r \to \infty} \rho_r(p) = 0, \quad (2)$$

for $p \in \mathcal{O}$, which further result in

$$\lim_{r \to 0} \eta(r) = 0, \quad \lim_{r \to \infty} \eta(r) = 0, \quad (3)$$

where

$$\eta(r) = Var_{p \in \mathcal{O}}(\rho_r(p)). \quad (4)$$

Therefore, there exists a r^* such that the variance $\eta(r^*)$ is maximized. The goal of maximizing $\eta(r)$ is to obtain the richest descriptor on flexibility for a given 3D model.

Fig. 3(a) shows the flexibility variances $\eta(r)$ versus different radii, $0.1R_i, 0.2R_i, \cdots, 1.0R_i$, $1 \leq i \leq 10$, for 10 randomly selected objects (dashed lines), where R_i is the radius of the i-th 3D model defined as the average distance from all the surface points to the center of mass. The solid curve in Fig. 3(a) denotes the average variances of 100 randomly selected objects. This figure shows that the maxima of the curves mainly fall into a relatively small range of r, indicating that the discriminative ability of $\rho_r(p)$ is insensitive to the choice of r. In practice, one can compute the flexibility of a point p using some $r \in [0.2R, 0.4R]$. Fig. 2(c) shows the flexibility distribution of an ant computed using $r = 0.3R$. Alternatively, we can define a more elaborate measure to explore the flexibility characteristic thoroughly by using a series of radii to form a *flexibility vector* at a point. We discuss it in the next section.

2.3 A Flexibility Descriptor

Let $\mathbf{r} = [r_1, r_2, \cdots, r_K]^T$, $r_1 < r_2 < \cdots < r_K$, denote the radii of a series of concentric spheres centered at some point p. A flexibility vector $\boldsymbol{\rho_r}(p)$ at p is then obtained by

$$\boldsymbol{\rho_r}(p) = [\rho_{r_1}(p), \rho_{r_2}(p), \cdots, \rho_{r_K}(p)]^T. \quad (5)$$

Note that a 3D model generates a number of flexibility vectors and two models usually have different numbers of such vectors. To organize these vectors into a global shape descriptor, the bag-of-words model [3] is adopted here. We use the k-means algorithm to cluster the flexibility vectors from half of the objects in a database to N clusters, the centers of which compose a codebook $\{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_N\}$. Each flexibility vector is then represented by a codeword $\mathbf{c}_i$ if it is closest to $\mathbf{c}_i$, and a 3D model is described by a histogram over

$\{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_N\}$ obtained by counting the codewords representing all the flexibility vectors of the 3D model. Finally, the object flexibility descriptor is formed by normalizing the histogram.

To measure the dissimilarity between two flexibility descriptors P and Q, we use the symmetric Kullback–Leibler divergence defined as

$$dis_{Fl}(P||Q) = D_{KL}(P||Q) + D_{KL}(Q||P), \qquad (6)$$

where

$$D_{KL}(P||Q) = \sum_{i=1}^{N} P(i) \log \frac{P(i)}{Q(i)}. \qquad (7)$$

With this descriptor, we can conduct 3D object retrieval by computing the dissimilarities between a query and every object in a database. Using every object as the query in the McGill database [11] and the Princeton training set [9], Fig. 3(b) shows the average retrieval performance (evaluated by the discounted cumulative gain (DCG) [9]) of the flexibility descriptor. We can see that DCG remains consistent in quite a wide range of the size of the codebook. We choose $N = 10$ for the McGill benchmark and $N = 60$ for the Princeton benchmark in our experiments.

2.4 Enhancing Existing Methods

The flexibility descriptor performs very well in retrieving articulated 3D objects (see Section 3). In addition, it can enhance existing shape descriptors when combined with them to retrieve generic 3D objects. A majority of existing shape descriptors represent a 3D shape without explicit geometric meanings, while the flexibility descriptor measures a particular geometric characteristic, the flexibility. So a more complete shape description of a 3D model can be obtained by combining it with other descriptors.

A natural way to combine two shape dissimilarity measures is

$$dis = \alpha \cdot dis'_{Fl} + (1 - \alpha) \cdot dis'_{Other}, \qquad (8)$$

where dis'_{Fl} is a normalized version of dis_{Fl} defined in (6) such that it is in $[0, 1]$, dis'_{Other} is some other dissimilarity measure, also normalized into $[0, 1]$, and α is a weighting factor to balance the two measures with $0 \leq \alpha \leq 1$.

3. EXPERIMENTS

Three groups of experiments are carried out in this paper, each with emphasis upon different requirements for a shape descriptor. In these experiments, we compare our flexibility descriptor (FD) with four other shape descriptors, the source codes or executable programs of which have been provided by the authors. They are the statistics-based shape distribution D2 [8] and one of its extended descriptors, the mutual absolute-angle distance histogram (AAD) [7], the geometry-based spherical harmonic descriptor (SHD) [6], and the view-based light field descriptor (LFD) [1]. The retrieval results are quantified using the Princeton shape benchmark (PSB) evaluation tools of first tier (FT), second tier (ST), e-measure (EM), discounted cumulative gain (DCG), and the precision-recall plot [9].

The first group of experiments is designed to measure the robustness of the shape descriptors to nonlinear shape transformations caused by objects' articulations, with the McGill articulated shape database. This database consists of 255

Table 1: Retrieval performance of different shape descriptors for retrieving 3D articulated objects.

	FD	LFD	SHD	AAD	D2
FT	**0.560**	0.508	0.478	0.439	0.419
ST	**0.720**	0.697	0.641	0.624	0.605
EM	**0.530**	0.497	0.464	0.433	0.414
DCG	**0.844**	0.831	0.804	0.768	0.764
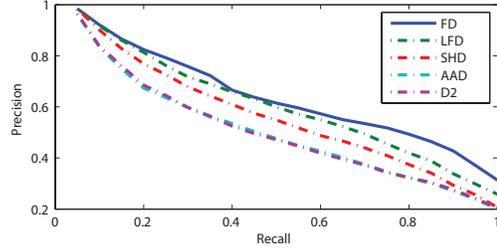
Figure 4: Precision-recall plots of different shape descriptors for retrieving 3D articulated objects.

models in 10 classes. Every model is used as the query. Table 1 shows the average retrieval results of the five descriptors evaluated by FT, ST, EM, and DCG. Fig. 4 is the precision-recall plots of the five descriptors. We see that our FD outperforms the other descriptors.

In the second group of experiments, we use the whole McGill shape benchmark (MSB) which includes the 255 articulated objects used in the previous experiments and the other 200 objects with few or no articulations. Our FD itself does not perform very well in retrieving such generic objects. However, significant improvements can be achieved when it is combined with other descriptors using (8) ($\alpha = 0.5$ is chosen here). Fig. 5 shows the retrieval results of LFD, SHD, AAD, and D2 with and without our FD combined, evaluated by FT, ST, EM, and DCG. With the improved performance, all the four shape descriptors are enhanced by our FD under different evaluations. The improvements are due to the fact that the four shape descriptors describe 3D models from a general viewpoint only, without considering particular geometry properties, but the fusion of them with our FD provides a more complete description.

The reader may wonder if the combination of two of the previous descriptors can also give similar or even better results than the combination of our FD with one of the previous. Since LFD works best among the previous four descriptors in our experiments as well as in [9] and [2], we use it as the baseline and combine each of the other descriptors with it. The retrieval results are given in Fig. 6(a), which are the precision-improvement-recall plots obtained from the precision-recall plots by subtracting the precision of LFD from the precisions of the four combinations. Fig. 6(a) indicates that our FD with LFD not only outperforms the other combinations, but also has improvement over the original LFD in a wide range of the recall.

The third group of experiments is conducted on the Princeton shape benchmark (PSB) training set [9], which contains 907 models in 90 classes. The majority of the models are rigid, man-made objects without much requirements for a shape descriptor to be articulation invariant. Even though it seems that the articulation invariant property of our FD is not necessary in such a case, it is still able to enhance
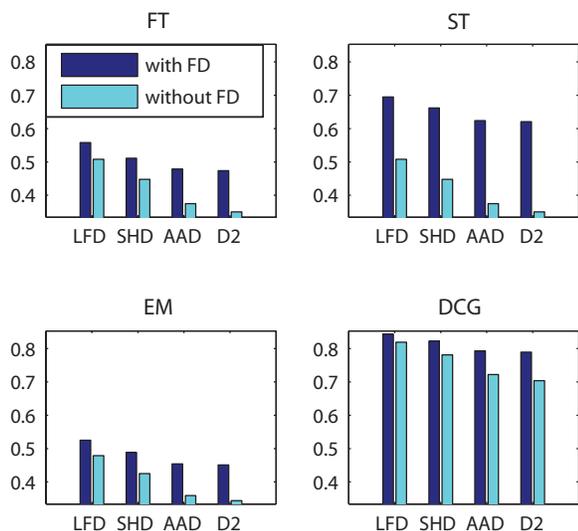
Figure 5: Retrieval performance of different shape descriptors with and without FD on MSB.

Figure 6: Precision improvement comparison when FD, SHD, AAD, or D2 is combined with LFD on MSB (a) and on the PSB training set (b).

the other shape descriptors to some extent. Fig. 6(b) shows the precision-improvement-recall plots by fusing FD, SHD, AAD, or D2 with LFD tested on the PSB training set. Obviously, LFD combined with FD improves the performance of LFD itself.

More specifically, there are improvements in 72 classes in the PSB training set where LFD with FD performs better than LFD itself. Among them, 17 classes are with more than 5% improvements. These improvements mainly happen to natural objects such as tree, apatosaurus, and face, and objects with large intra-class variances such as city, shoe, rectangular, roman building, and antique car.

The extensive experiments demonstrate that our flexibility descriptor is suited for retrieving articulated objects, especially natural objects. For generic object retrieval, it provides a favorable complementary to other shape descriptors.

The feature extraction times of FD, LFD, SHD, AAD, and D2 are about 1.12s, 2.69s, 2.25s, 0.46s, and 0.23s, respectively, on average for each object in the PSB training set. The programs are run on an Intel Pentium(R) 3.20GHz CPU with 2 GB RAM.

4. CONCLUSIONS

We have proposed a new feature, called object flexibility, to measure local shape characteristics of an object about how a local part is massively connected to the object. Based on this feature, a new shape descriptor is obtained, which is stable to shape deformations caused by articulations. Extensive experiments show that this shape descriptor outperforms four previous popular shape descriptors in retrieving articulated objects. For generic 3D object retrieval, it can be combined with them to obtain better performance.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] D. Chen, X. Tian, Y. Shen, and M. Ouhyoung. On visual similarity based 3D model retrieval. *Computer Graphics Forum*, 22(3):223–232, 2003.

[2] A. D. Bimbo and P. Pala. Content-based retrieval of 3D models. *ACM Trans. on Multimedia Computing, Communications, and Applications*, 2(1):20–43, 2006.

[3] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *CVPR*, pages 524–531, 2005.

[4] A. Ion, N. Peyre, G. Marmol, S. Kropatsch, W. Cohen, and P. Laurent. 3D shape matching by geodesic eccentricity. In *CVPR Workshop*, pages 1–8, 2008.

[5] V. Jain and H. Zhang. A spectral approach to shape-based retrieval of articulated 3D models. *Computer-Aided Design*, 39(5):398–407, 2007.

[6] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Symposium on Geometry Processing*, pages 156–164, 2003.

[7] R. Ohbuchi, T. Minamitani, and T. Takei. Shape-similarity search of 3D models by using enhanced shape functions. *Int'l J. of Computer Applications in Technology*, 23(2):70–85, 2005.

[8] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Trans. on Graphics*, 21(4):807–832, 2002.

[9] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The Princeton shape benchmark. In *Shape Modeling Int'l*, pages 167–178, 2004.

[10] J. Tangelder and R. Veltkamp. A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications*, 39(3):441–471, 2008.

[11] J. Zhang, K. Siddiqi, D. Macrini, A. Shokoufandeh, and S. Dickinson. Retrieving articulated 3-D models using medial surfaces and their graph spectra. In *Int'l Workshop on Energy Minimization Methods in CVPR*, 2005.