NON-RIGID 3D SHAPE RETRIEVAL USING MULTIDIMENSIONAL SCALING AND BAG-OF-FEATURES

Zhouhui Lian[1,2], *Afzal Godil*[2], *Xianfang Sun*[3,1], *Hai Zhang*[1]

[1]Beihang University, Beijing, China
[2]National Institute of Standards and Technology, Gaithersburg, USA
[3]Cardiff University, Wales, UK

ABSTRACT

Matching non-rigid shapes is a challenging research field in content-based 3D object retrieval. In this paper, we present an image-based method to effectively address this problem. Multidimensional Scaling (MDS) and Principal Component Analysis (PCA) are first applied to each object to calculate its canonical form, which is afterward represented by 66 depth-buffer images captured on the vertices of an unit geodesic sphere. Then, each image is described as a word histogram obtained by the vector quantization of the image's salient local features. Finally, a multi-view shape matching scheme is carried out to measure the dissimilarity between two models. Experimental results on the McGill Articulated Shape Benchmark database demonstrate that, our method obtains better retrieval performance compared to the state-of-the-art.

***Index Terms*—** 3D shape retrieval, Non-rigid 3D shape, Multidimensional Scaling (MDS), Bag-of-Features (BOF)

1. INTRODUCTION

The explosion in the number of 3D models has led to the rapid development of 3D shape retrieval systems that, given a query object, retrieve similar 3D models based on their shapes. Up to now, a large number of algorithms, including statistic-based [1], graph-based [2], transform-based [3], view-based [4], and composite methods [5], have been proposed. For more details about different methods, we refer the reader to some good surveys [6, 7].

Probably due to the complexity of non-rigid shape processing, previous efforts have been mainly devoted to the retrieval of rigid 3D models. Thereby, comparing non-rigid 3D shapes is still a challenging problem in content-based 3D object retrieval. Yet, as we know, non-rigid models are commonly seen in our surroundings. Take Fig. 1(a) for an example, a human being might appear in several distinct postures that could inevitably be identified as different shapes using most existing methods. In order to properly and efficiently measure the dissimilarity between two non-rigid objects, it is preferable that the shapes can be described by some feature vectors which are invariant or approximately invariant under isometric transformations (e.g. bending and articulation).

This work has been supported by the SIMA and the IDUS program.

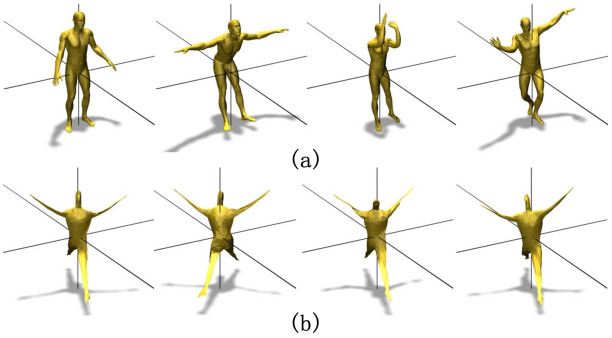

Fig. 1. Non-rigid models (a) and their canonical forms (b).

Recently, Ruggeri and Saupe [8] used the distributions of geodesic distances to create a isometry-invariant shape descriptor, while Mahmoudi and Sapiro [9] discussed six such signatures via the distributions of several geometric distances including diffusion distance, geodesic distance, and a curvature weighted distance, etc. In [10], Ohbuchi etc. reported a articulated-invariant shape descriptor using salient local visual features. They represented a 3D object by a word histogram derived from the vector quantization of salient local descriptors extracted on the depth-buffer views captured uniformly around the object. Elad and Kimmel [11] suggested extracting bending-invariant signatures from embedded 3D surfaces generated by applying Multidimensional Scaling (MDS) techniques. In their paper, a moment-based signature was tested in a simple classification experiment. Gal and Cohen-Or [12] introduced a pose-oblivious shape descriptor which is actually a 2D histogram combining the distributions of Euclidean distances in local regions and the distributions of geodesic distances for the whole object.

Inspired by the papers mentioned above, we develop a novel method for non-rigid 3D shape retrieval, which is largely based on the utilization of Multidimensional Scaling (MDS) [13] and Bag-of-Features (BOF). The key idea of our method is to apply MDS and PCA [14] together to obtain the 3D object's canonical form (see Fig. 1(b)), from which salient local visual features are extracted to generate a discriminative

shape descriptor using BOF. Due to the fact that both the canonical form and the local feature are (or approximately) bending-invariant, the new shape descriptor is expected to be well suited for non-rigid 3D shape matching. As we can see from the experimental results evaluated on a commonly-used articulated shape database, our method significantly outperforms the state-of-the-arts in terms of retrieval accuracy.

2. METHOD DESCRIPTION

In this section, we first present an overview of the method and then elaborate on the implementation details. As depicted in Fig. 2, our method performs step by step as follows:

1. ***Canonical Form Computation:*** Calculate the canonical form for a 3D model based on MDS and PCA.
2. ***Local Feature Extraction:*** Capture 66 depth-buffer views for the canonical form on the vertices of a given geodesic sphere, and then extract salient SIFT descriptors [15] from these views.
3. ***Word Histogram Construction:*** Generate a word histogram by vector quantizing each view's local features against a pre-specified codebook, such that the shape can be represented by a set of histograms.
4. ***Dissimilarity Calculation:*** Carry out an efficient multi-view shape matching (Clock Matching) scheme to measure the dissimilarity between two models by calculating the minimum distance of their 24 matching pairs.

Since our method is mainly based on *Multidimensional Scaling*, *Clock Matching*, and *Bag-of-Features*, for the sake of convenience, we denote the algorithm as "MDS-CM-BOF".

2.1. Canonical Form Computation

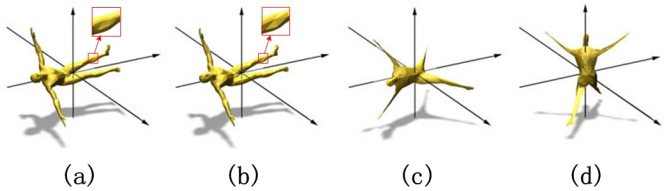

Fig. 3. The procedure of our canonical form computation. The original mesh (a) is first simplified into a coarse version (b) and then MDS is applied to map the simplified mesh into a bending-invariant surface (c), on which we use a PCA-based alignment to obtain the final canonical form (d).

Based on the fact that the geodesic distance between every two points on a surface remains unchanged under isometric transformations, a bending invariant representation can be obtained by applying MDS to map the geometric structure of the surface to a new 3D Euclidean space, in which geodesic distances are approximated by Euclidean ones. This idea is originally proposed in [11], where three different MDS techniques are also compared. To get better results, we here experimentally choose the least squares technique with the SAMCOF algorithm [13], whose source code written in Matlab is publicly available on the web site of the book [16], to implement the MDS embedding.

As the calculation of geodesic distances and the SAMCOF algorithm are both computational expensive, the 3D surface is simplified before the MDS embedding procedure. A reliable source code of mesh simplification can be found in [17] and the number of vertices on the mesh is reduced to about 1000.

Given the embedded surface, we first translate the center of its mass to the origin and then scale the maximum polar distance of the points on the surface to one. Rotation invariance is achieved by applying the PCA technique to find the principal axes and align them to the canonical coordinate frame. Note that, we only employ the information of eigenvectors to fix the positions of three principal axes, namely, the direction of each axis is still undecided and the x-axis, y-axis, z-axis of the canonical coordinate system can be located in all three axes. That means 24 different orientations are still plausible for the canonical form of a 3D object, or rather, 24 matching operations should be carried out when comparing two models. It should also be pointed out that, the exact values of the surface moments used in our PCA-based pose normalization are calculated via the explicit formulae introduced by [18].

To sum up, as illustrated in Fig. 3, our canonical form computation consists of the following three steps: Mesh simplification, MDS embedding, and PCA-based alignment,

2.2. Local Feature Extraction

After the first step, we obtain the canonical forms of 3D models, which have been well aligned to the canonical coordinate frame. Then their 66 depth-buffer views with size 256×256 are captured on the vertices of a given unit geodesic sphere whose mass center is also located in the origin, such that a 3D object can be represented by a set of images from which we extract salient SIFT descriptors, as presented in [15]. The SIFT descriptor is calculated, using the *VLFeat* matlab source code developed by Vedaldi and Fulkerson [19].

2.3. Word Histogram Construction

Directly comparing 3D models by their local visual features is time consuming, especially for the 3D shape retrieval methods that use large numbers of views. To address this problem, we quantize the SIFT descriptors extracted from a depth-buffer image into one word histogram so that the view can be represented in a highly compact and distinctive way.

Before vector quantization, a codebook with N_w visual words is generated via off-line clustering. More specifically,

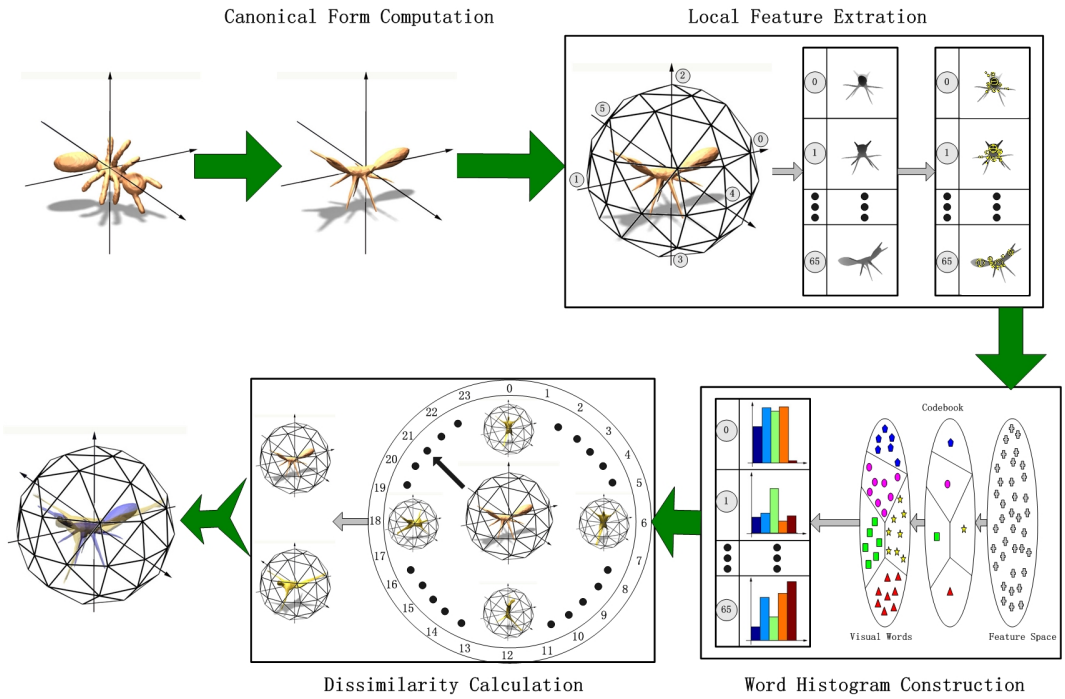

Fig. 2. Overview of our method.

huge numbers of feature vectors are first randomly sampled from the feature set of the target database to form a training set. Then, the training set is clustered into N_w clusters using the K-means method. At last, centers of the clusters are selected as the feature vectors of visual words in the codebook. Here, we choose the *Integer K-means* algorithm [19] to do the clustering and the number of clusters is selected as $N_w = 1500$ according to our experiments.

By searching for the nearest neighbor in the codebook, a local descriptor is assigned to a visual word. Then each view can be represented using a word histogram whose i^{th} bin records the number of ith visual words in the depth-buffer image. We also design a compact data structure for our 3D shape descriptor, where only the information (i.e. bin No. and bin value) of some bins, whose values are not equal to zero, appears in the feature vector.

2.4. Dissimilarity Calculation

The last step of the MDS-CM-BOF algorithm is the dissimilarity calculation for two shape descriptors. The basic idea of our multi-view shape matching (*Clock Matching*) scheme is that, after we get the principal axes of an object, instead of completely solving the problem of fixing the exact positions and directions of these three axes to the canonical coordinate frame, all possible poses are taken into account during the shape matching stage.

The dissimilarity between the query model q and the source model s is defined as,

$$Dis_{q,s} = \min_{0 \le i \le 23} \sum_{k=0}^{65} D\left(FV_q(p_0^{'}(k)), FV_s(p_i^{'}(k))\right),$$

where $FV_m = \{FV_m(k) | 0 \le k \le 65\}$ denotes the shape descriptor of 3D object m, $FV_m(k)$ stands for the feature vector of view k, the permutations $p_i^{'} = \{p_i^{'}(k) | 0 \le k \le 65\}$, $0 \le i \le 23$ indicate the arrangements of views for all (24) possible poses of a canonical form, and $D(\cdot, \cdot)$ measures the distance between two histograms H_1, H_2 with N_w bins by the formula,

$$D(H_1, H_2) = 1 - \frac{\sum_{j=0}^{N_w-1} \min(H_1(j), H_2(j))}{\max(\sum_{j=0}^{N_w-1} H_1(j), \sum_{j=0}^{N_w-1} H_2(j))}.$$

More details of this shape matching scheme can be found in our previous paper [5].

3. EXPERIMENTAL RESULTS

The goal of this section is to evaluate the retrieval performance of our MDS-CM-BOF algorithm and compare it with other state-of-the-art methods. We carry out experiments on the McGill Articulated Shape Benchmark database (255 articulated models with 10 categories) and evaluate the retrieval accuracy by Precision-recall plots as well as the following four quantitative measures (see [6] for their explicit definitions): Nearest neighbor (NN), First-tier (1-Tier), Second-tier (2-Tier), and Discounted Cumulative Gain (DCG).

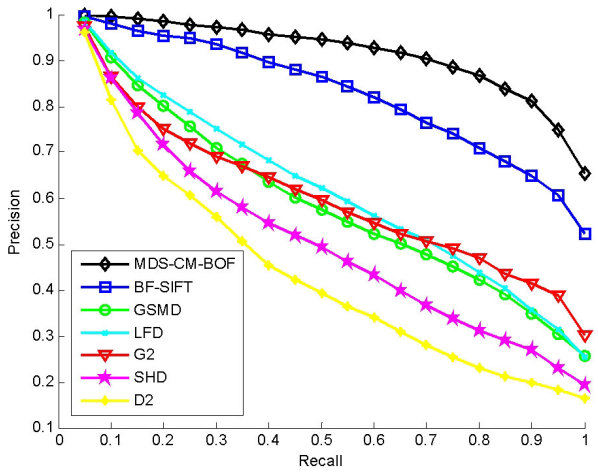

Fig. 4. Precision-recall plots of our method (MDS-CM-BOF) and other six approaches evaluated on the McGill database.

Table 1. Comparing retrieval results of our method (first row) with the state-of-the-art on the McGill database.

	NN	1-Tier	2-Tier	DCG
MDS-CM-BOF	99.6%	84.7%	95.5%	97.2%
BF-SIFT	97.3%	74.6%	87.0%	93.7%
LFD	91.0%	52.8%	69.7%	83.7%

In Fig. 4, the Precision-recall curves of our method and other well-known approaches, including BF-SIFT [10], GSMD [5], LFD [4], G2 [9], SHD [3], and D2 [1], are displayed. Among them, BF-SIFT produces the best result we can find so far for non-rigid 3D shape searching and LFD is the top ranked rigid shape descriptor in [6]. As we can see more clearly from Table 1, our method (MDS-CM-BOF) obtains excellent results and markedly outperforms all existing methods for the application of non-rigid 3D shape retrieval. Moreover, the new shape descriptor is compact (about 4800 byte), the comparison of two feature vectors takes less than 1.0 millisecond, and the feature extraction for an object can be calculated within 15 seconds on average using a common PC (2.66GHz CPU and 4.0GB memory) under Windows XP.

4. CONCLUSION

In this paper, using Multidimensional Scaling and Bag-of-Features, a practical and powerful method was developed for the retrieval of non-rigid 3D objects. By taking advantage of the canonical form and salient local features, the proposed method obtains excellent results when searching for non-rigid shapes. Experiments on a commonly-used articulated shape benchmark database verified that, our method results in performance which is superior to the state-of-the-art.

5. REFERENCES

[1] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM TOG*, vol. 21, no. 4, pp. 807–832, 2002.

[2] H. Sundar, D. Silver, N. Gavani, and S. Dickinson, "Skeleton based shape matching and retrieval," in *Proc. SMI*, 2003, pp. 130–139.

[3] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Proc. SGP*, 2003, pp. 156–164.

[4] D-Y. Chen, X-P. Tian, Y-T. Shen, and M. Ouhyoung, "On visual similarity based 3D model retrieval," in *Proc. Eurographics 2003*, 2003, vol. 22, pp. 223–232.

[5] Z. Lian, P. L. Rosin, and X. Sun, "Rectilinearity of 3D meshes," *IJCV* , vol. 89, no. 2-3, pp. 130–151, 2010.

[6] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The princeton shape benchmark," in *Proc. SMI*, 2004, pp. 167–178.

[7] J. W. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," *Multimedia Tools and Applications*, vol. 39, no. 3, pp. 441–471, 2008.

[8] M. R. Ruggeri, G. Patane, M. Spagnuolo, and D. Saupe, "Spectral-driven isometry-invariant matching of 3D shapes," *IJCV*, vol. 89, no. 2-3, pp. 248–265, 2010.

[9] M. Mahmoudi and G. Sapiro, "Three-dimensional point cloud recognition via distributions of geometric distances," *Graphical Models*, vol. 71, pp. 22–31, 2009.

[10] R. Ohbuchi, K. Osada, T. Furuya, and T. Banno, "Salient local visual features for shape-based 3D model retrieval," in *Proc. SMI*, 2008, pp. 93–102.

[11] A. Elad and R. Kimmel, "On bending invariant signatures for surface," *PAMI*, vol. 25, no. 10, pp. 1285–1295, 2003.

[12] R. Gal, A. Shamir, and D. Cohen-Or, "Pose-oblivious shape signature," *TVCG*, vol. 13, pp. 261–271, 2007.

[13] I. Borg and P. Groenen, *Modern Multidimensional Scaling-Theory and Applications*, Springer, 1997.

[14] D. V. Vranić, D. Saupe, and J. Richter, "Tools for 3D-object retrieval: Karhunen-loeve transform and spherical harmonics," in *Proc. IEEE Fourth Workshop on Multimedia Signal Processing*, 2001, pp. 293–298.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[16] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, *Numerical geometry of non-rigid shapes*, Springer, 2008.

[17] MeshLab1.1.0, "http://meshlab.sourceforge.net/," 2008.

[18] S. A. Sheynin and A. V. Tuzikov, "Explicit formulae for polyhedra moments," *Pattern Recognition Letters*, vol. 22, pp. 1103–1109, 2001.

[19] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," *http://www.vlfeat.org/*.