Eurographics Workshop on 3D Object Retrieval (2012)
M. Spagnuolo, M. Bronstein, A. Bronstein, and A. Ferreira (Editors)

Local Goemetrical Feature with Spatial Context for Shape-based 3D Model Retrieval

S. Kawamura[1], K. Usui[1], T. Furuya[2], and R. Ohbuchi[1]

[1]University of Yamanashi, Japan
[2]Nisca Corporation, Japan

Abstract

With recent popularity of 3D models, retrieval and recognition of 3D models based on their shape has become an important subject of study. This paper proposes a 3D model retrieval algorithm that is invariant to global deformation as well as to similarity transformation of 3D models. The algorithm is based on a set of local 3D geometrical features combined with bag-of-features approach. The algorithm employs a novel local feature, which is a combination of local geometrical feature enhanced with its spatial context computed as histogram of diffusion distance computed over mesh surface. Experimental evaluation of retrieval accuracy by using benchmark databases showed that adding positional context significantly improves retrieval accuracy.

Categories and Subject Descriptors (according to ACM CCS): H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing, I.3.m [Computer Graphics]: Miscellaneous

1. Introduction

As three-dimensional (3D) shape model become ubiquitous, effective and efficient management of them, especially via content based retrieval by their shape, has become quite important.

In this paper, we propose a 3D model retrieval algorithm that accepts 3D models as query and has invariance to global deformation and similarity transformation of 3D models. Invariance to global deformation is important for almost all animal and many mechanical parts. Proposed algorithm employs *Bag-of-Features* (BoF) integration of a set of simple, 3D, local, geometrical features, so a 3D model is described by an integrated feature vector per 3D model. If each local geometrical feature is invariant to 3D similarity transformation, BoF integration of a set of such features that discards (3D Euclidian) coordinate of the features made the algorithm invariant to global deformation. The BoF integration also make comparison among 3D models more efficient, as each model is described by a feature vector, instead of a set of local features.

While quite popular and powerful, BoF has a shortcoming; it discards all the positional information of local feature so that no contextual information is used. The proposed algorithm attempts to better BoF by augmenting each local geometrical feature with its spatial context. As a spatial context of a point on a 3D model, the algorithm employs statistics of diffusion distance (e.g., [SOG09]) measured over the surface of a 3D model. Such statistics is invariant to articulation and/or deformation of the mesh surface as well as to similarity transformation. Experimental evaluation of the proposed 3D model retrieval algorithm has shown that adding spatial context to local geometrical feature do improve retrieval accuracy.

We will briefly review related work in the next section. Section 3 and Section 4 will present the proposed algorithm and its evaluation results. Summary and future work will be presented in Section 5.

2. Related work

A predominant approach in the field of image recognition and retrieval is Bag-of-Features (BoF) integration of local visual features (e.g., [CDF*04]). In such an approach, local features are combined into a feature vector per image without regard to absolute or relative positions of each local feature. However, importance of context in object recognition is well known [OT07]. Thus, many recent works have tried to incorporate contextual information for object recognition. e.g., matching of (sets of) local features organized in grid [LSP06] or graph [DJP11].

Bag-of-features approach has also become popular in the field of 3D object retrieval (e.g., [OOF*08][FO09] [TDV*11]). A set of 2D image local features [OOF*08] [FO09] or 3D local features [FSB09] are extracted from a 3D model to be compared, and then integrated into a feature vector per 3D model by using bag-of-features approach. These algorithms, however, do not employ contextual information. One of possible reasons is difficulty in finding canonical orientation and parameterization for a 3D model. A photograph typically has a global orientation, an important cue for spatial context, and parameterization of the image is relatively easy (e.g., as pixel grid.). In 3D model recognition and retrieval, however, invariance against 3D similarity transformation is expected. Furthermore, invariance to global deformation or articulation is often required. In such a case, finding canonical orientation of a 3D model and finding global parameterization of local features are quite difficult.

DOI: 10.2312/3DOR/3DOR12/055-058

One of the ways to establish intrinsic coordinate system on a deformable 3D mesh surfaces is local statistics of geodesic-like distances computed on the surface, e.g., [SOG09]. Such local statistics of distances among points on the mesh may be used by itself as a feature for 3D object retrieval. Our proposed method uses it as spatial or positional context that enhances local geometrical feature.

3. Algorithm

Proposed algorithm follows the steps below to retrieve 3D models.

1. **Remeshing:** A 3D model is remeshed into a singly-connected graph *G* having N_s vertices.
2. **FoG extraction:** At each of N_k vertices ($N_k \le N_s$) on graph *G*, which are called *keypoints*, Local Geometrical Feature (LGF) $\mathbf{K}_i$ and its positional context Local Distance Feature (LDF) $\mathbf{L}_i$ are extracted. A *Feature on Geodesics* (*FoG*) feature vector $\mathbf{h}_i$ at interest point *j* (*j*=1...N_s) is a concatenation $\mathbf{h}_i$=($\mathbf{K}_i$, $\mathbf{L}_i$) of $\mathbf{K}_i$ and $\mathbf{L}_i$ (Figure 1). Both LGF and LDF have finite support defined in 3D Euclidian space.
3. **Bag-of-words integration:** A set of N_f FoG features are integrated into a feature vector per 3D model by using Bag-of-Features (BoF) approach.
4. **Distance computation:** Given a feature vector of a query 3D model, features vectors of 3D models in a database are ranked by their similarity to the query.

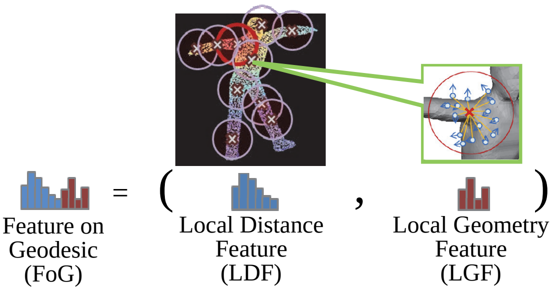

Figure 1. A *Feature-on-Geodesics*, or *FoG*, combines a local geometrical feature with its positional context.

3.1 Remeshing

The remeshing step is added so that the algorithm could compare models in polygon soup representation or models having highly non-uniform sampling densities, e.g., models found in Princeton Shape Benchmark [SMK*04]. If the algorithm deals only with 3D models represented as densely and uniformly sampled manifold mesh, the remeshing should not be used, as remeshing introduces errors. For example, in Figure 2b, tip of a rotar blade is connected to the tail by remeshing. Such unintended edges would degrade retrieval accuracy.

To remesh, the surface based 3D model is first sampled by N_s oriented points placed at locations on the faces determined by using *Sobol's* low-discrepancy sequence (LDS) [PTV*92]. An LDS, or quasi-random sequence, is a deterministic sequence that produces sample points more uniformly distributed than a pseudo-random sequence. Surface normal of a face becomes the orientation of points on the face. Points are generated so that the number of point per unit area of faces is uniform over the 3D model. These points are connected by their proximity into a singly connected graph *R* per 3D model.

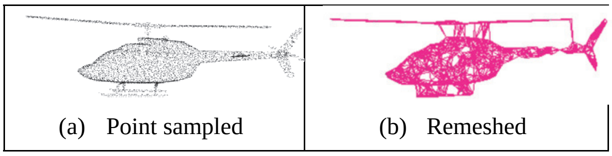

Figure 2. A surface-based 3D model is sampled by oriented points (a) and remeshed (b).

3.2 Feature Extraction

FoG features are computed on mesh *G* at N_k keypoints, each of which is a vertex on *G*. Keypoints are generated by sub-sampling all the vertices of *G* using, again, low-discrepancy sequence.

3.2.1 *Local Geometrical Features* (*LGF*) *Extraction*

Any local geometrical feature invariant to 3D similarity transformation can be used as LGF. We used what we call *Localized Absolute Angle Distance* (*LAAD*) feature as LGF. An LAD centered at vertex x_i (*i*=1...*m*) is computed from a set of *p* oriented points within a sphere (in 3D Euclidian distance) of influence of radius *R* centered at vertex *i*. Assume that a FoG (and LAAD) keypoint is $\mathbf{x}_\mathrm{i}$ and its normal vector is $\mathbf{n}_\mathrm{i}$. Assume also that point $\mathbf{x}_\mathrm{j}$ with normal vector $\mathbf{n}_\mathrm{j}$ is a point within radius *R* of $\mathbf{x}_\mathrm{i}$. Using $\mathbf{x}_\mathrm{i}$ and $\mathbf{x}_\mathrm{j}$, a 2-tuple $(\alpha, \beta) = \left(\mathbf{n}_\mathrm{i} \cdot \mathbf{n}_\mathrm{j}, \left|\mathbf{x}_\mathrm{j} - \mathbf{x}_\mathrm{i}\right|\right)$ is computed. If there are *p* vertices in the sphere, (*p*-1) tuples are computed, and their α and β values are accumulated into a 2D joint histogram. For most of the experiments described below, number of bins for the LAAD are 7 for both α and β, resulting in a 49-dimensional LAAD vector. The histogram is normalized so that the sum of all the bins is 1.0.

The radius *R* of the sphere of influence defines locality of LAAD. The smaller the *R*, the local the LAAD feature is. The value *R* is defined as $R = Dr$, where *r* is a parameter and *D* is the raduis of the smallest sphere centered at the barycenter of the 3D model that encloses the model. Thus a parameter *r*=1.0 means that the support of LAAD has the size of enclosing sphere of the 3D model.

3.2.2 *Local Distance Feature* (*LDF*) *Extraction*

LDF $\mathbf{L}_i$ is a 1D histogram of distances from the FoG interest point $\mathbf{x}_\mathrm{i}$ to all the sample points on mesh *G*. Each distance is a geodesic distance between a pair of points over the graph *G*, compass a diffusion distance by using *Manifold Ranking* (*MR*) algorithm [ZBL*03][WHY*07]. The interest point $\mathbf{x}_\mathrm{i}$ is used as the source of diffusion of ranking score for the MR. Ranking scores at every sample points on G computed by MR are accumulated into 1D histogram having D_G bins. In the experiments below, we used D_G=100, which is determined by preliminary experiments. LDF histogram is normalized so that values in the bins add up to 1.0. $\mathbf{L}_i$ is computed at the same keypoints $\mathbf{x}_\mathrm{i}$ as $\mathbf{K}_i$. MR has two parameters σ, and μ. The parameter σ is a scale parameter for diffusion. A large σ makes the rank value to diffuse faster and farther, so that LDF captures larger scale feature. A small σ, on the other hand, makes diffusion slower so that small scale feature is captured. Cost of MR increases with number of sample points N_s. MR requires $O(N_s^2)$ of storate and $O(N_s^3)$ to compute ranking rank values by means of matrix inversion.

3.3 Bag-of-Features Integration

A set of FoG features $\mathbf{h}_i$ is integrated into a feature vector $\mathbf{F}_s$ for 3D model *s* by using BoF approach. In the following, LDF, LGF, and FoG integrated by using BoW approach are called BoLDF, BoLGF, and BoFoG, respectively. For the integration, FoG (or LDF or LGF) features extracted from the 3D model are vector quantized into visual words by using a pre-formed codebook of vocabulary size N_v. Resulting visual words are accumulated into a (1D) histogram having N_v bins. This histogram becomes the feature vector of the 3D models. The codebook is pre-computed by clustering FoG features obtained from the database to be retrieved. Each cluster center corresponds to a code vector, that is, a visual word. We employed *k*-means clustering algorithm for the clustering, by giving vocabulary size N_v as number of clusters.

3.4 Distance Computation

Distance between a pair of feature vectors $\mathbf{F}_p$ and $\mathbf{F}_q$ is computed by using a symmetric version of *Kullback-Leibler Divergence* (*KLD*);.

$$D_{KLD}\left(\mathbf{F}_p, \mathbf{F}_q\right) = \sum_{k=1}^{m}\left(\mathbf{F}_{p,k} - \mathbf{F}_{q,k}\right)\log\left(\mathbf{F}_{p,k} / \mathbf{F}_{q,k}\right) \quad (1)$$

BoFoG features may be computed at multiple scales, that is, multiple values of σ. Combining BoFoG features at multiple scales may improve retrieval accuracy. We combined multi-scale BoFoG features by their distance. That is, an overall distance is a sum of inter-3D model distances computed by using different values of σ.

4. Experiments and Results

We performed experiment to quantify (1) the effect of adding positional context (i.e., LDF) to LGF to form FoG, (2) the effect of multi-scale LDF (Section 3.4), and (3) to compare retrieval accuracy with the other algorithms. Retrieval experiments are performed by using two benchmark databases: *McGill Shape Benchmark* (MSB) [SZM*08] for highly articulated (non-rigid) but less geometrically detailed shapes, and *Princeton Shape Benchmark* (PSB) [SMK*04] for a set of quite diverse, rigid, and relatively detailed shapes. MSB contains models represented as densely sampled closed manifold mesh, while PSB contains 3D models represented by using polygons soup, meshes having high variance in sampling density, etc. PSB test set is used for evaluation. We use performance index R-Precision, which is a ratio, in percentile, of the models retrieved from the desired class C_k (i.e., the same class as the query) in the top R retrievals, in which R is the size of the class $|C_k|$. Throughout the experiments presented below, LGF is 7×7=49 dimensional, LDF is D_G=100 dimensional, and FoG, which is a concatenation of LGF and LDF vectors, is 149 dimensional.

4.1 FoG v.s. LDF and LGF

First experiment compares BoLGF, BoLDF, and BoFoG using PSB and MSB. Parameters used are N_s=3,000, N_k=500 R=0.5, σ = 2, μ = 0.01, and N_w=500. Suffix "_S" for BoFoG means that it is a single-resolution version computed using single value of σ. Figure 3 shows that BoFoG_S outperformed both BoLDF and BoLGF. Adding positional context did improve retrieval accuracy. Note that, for mostly rigid models of PSB, BoLGF worked better than BoLDF. On the other hand, for deformable models of MSB, BoLDF worked better than BoLGF.

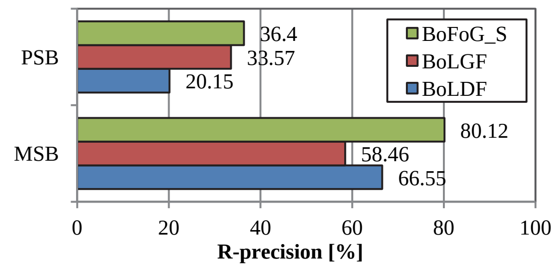

Figure 3. Retrieval accuracies of BoFoG is better than BoLDF and BoLGF.

4.2 Comparison with Other Methods

This experiment compares retrieval accuracy of BoFoG with several previous algorithms; *Light Field Descriptor* (LFD) [CTS*03], *Spherical Harmonic Descriptor* (SHD) [KFR03] and *Bag-of-Features Dense SIFT* (BF-DSIFT) [FO09]. Executables for LFD is downloaded from the author's web site. We implemented BF-DSIFT. We used both single-scale and multi-scale versions of BoFoG. For single-scale version, we usedσ = 2. Multi-scale version used seven values σ= {0.25, 0.5, 1, 2, 4, 8, 16} to obtain seven distances, which are then added with equal weight to produce an overall distance. All the other parameters are the same as in Section 4.1.

Table 1 shows retrieval accuracy (R-Precision) for the methods compared. Figure 4a and Figure 4b show their recall-precision plots of 5 algorithms. In the table and plots, BoFoG_S and BoFoG_M indicate, respectively, single-scale and multi-scale versions of BoFoG algorithm. Multi-scale BoFoG_M has small but consistent advantage over single scale BoFoG_S.

For MSB database, BoFoG algorithm performed the best among those compared. As MSB contains highly deformable models represented as densely sampled manifold mesh, our proposed algorithm is expected to do well. BoFoG algorithm did not fare well for the PSB database, however. Possible reasons include lack of descriptive power of local feature LAAD, and degradation of geometrical feature due to remeshing. Note that BF-DSIFT is a very powerful contender for PSB-like benchmark, judging from SHREC 2012 Generic 3D Track results.

We entered SHREC 2011 Non-Rigid 3D Watertight Meshes track [LGB*11] with BoFoG_S and BoFoG_S combined with distance metric learning. For the distance metric learning, we used MR [ZBL*03] in its original form. In the track, BoFoG_S and its MR version placed at about 4th among 9 entrants.

Table 1. Retrieval accuracy of various algorithms.

Methods	R-precision [%]	
	MSB	PSB
SHD [KFR03]	55.6	39.6
LFD [CTS*03]	55.5	44.7
BF-DSIFT [FO09]	75.4	**54.1**
BoFoG_S	80.1	36.4
BoFoG_M	**82.1**	40.2

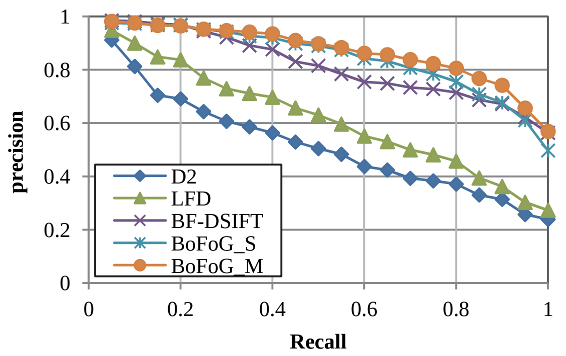

Figure 4a. Recall-precision plot for MSB.

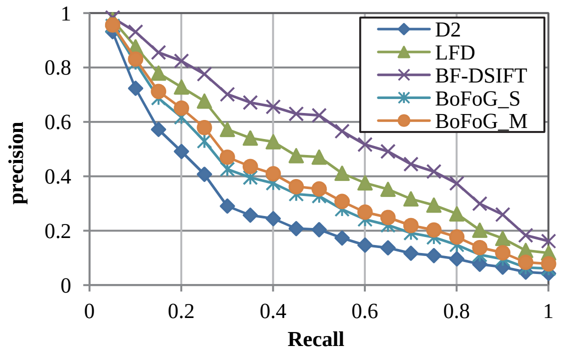

Figure 4b. Recall-precision plot for PSB.

5. Conclusion

This paper proposed a 3D model retrieval algorithm that is invariant to global deformation and similarity transformation of 3D models. Basis of the algorithm is a popular bag-of-features approach combined with 3D local geometrical feature. Novelty of the proposed algorithm is an addition of spatial context to the local geometrical feature. The positional context is a histogram of distribution of diffusion distances between points on the 3D model. Experiments showed that adding spatial context significantly improves retrieval accuracy, especially for deformable models of MSB [SZM*08]. For rigid and diverse models of PSB [SMK*04], which contains polygon soup models, the algorithm did not perform well.

In the future, we'd like to evaluate proposed approach without the remeshing step, assuming dense manifold mesh models as input. We'd also like to try more powerful local geometrical feature coupled with proposed positional context.

Acknowledgements

This research has been funded in part by the Ministry of Education, Culture, Sports, Sciences, and Technology of Japan (No. 18300068).

References

[CDF*04] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual Categorization with Bags of Keypoints, *Proc. ECCV '04 workshop on Statistical Learning in Computer Vision*, 59-74, (2004)

[CTS*03] D-Y. Chen, X-P. Tian, Y-T. Shen, M. Ouh-young.: On Visual Similarity Based 3D Model Retrieval. *Computer Graphics Forum*, **22**(3), 223-232, 2003.

[DJP11] O. Duchenne, A. Joulin, J. Ponce: A graph-matching kernel for object categorization, IEEE ICCV 2011, 6-13, 1792 - 1799, 2011.

[FSB09] J. Fehr, A. Streicher, H. Burkhardtm, A Bag of Features Approach for 3D Shape Retrieval, ISVC 2009, LNCS 5875, 34-43, 2009.

[FO09] T. Furuya, R. Ohbuchi: Dense Sampling and Fast Encoding for 3D Model Retrieval Using Bag-of-Visual Features. *ACM CIVR* 2009, 2009.

[KFR03] M. Kazhdan, T. Funkhouser, S. Rusinkiewicz.: Rotation Invariant Spherical Harmonics Representation of 3D Shape Descriptors. *SGP* 2003, 167-175, 2003.

[LSP06] S. Lazebnik, C. Schmid, and J. Ponce: eyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, IEEE CVPR 2006, vol. II, 2169-2178, (2006).

[LGB*11] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoué, H.V. Nguyen, R. Ohbuchi, Y. Ohkita, Y. Ohishi, F. Porikli, M. Reuter, I. Sipiran, D. Smeets, P. Suetens, H. Tabia, D. Vandermeulen, Shape Retrieval on Non-rigid 3D Watertight Meshes, *Eurographics 3DOR* 2011, (2011).

[OT07] A. Oliva, A. Torralba: The role of context in object recognition, *Trends in Cognitive Science*, **11**(12), 520-527, 2007.

[OOF*08] R. Ohbuchi, K. Osada, T. Furuya, T. Banno.: Salient local visual featuers for shape-based 3D model retrieval. *Proc. IEEE SMI* 2008, 93-102, 2008.

[PTV*92] W.H. Press et al. *Numerical Recipes in C –The art of Scientific Computing, Cambridge Universe Press, Cambridge*, UK, 1992, pp.309-315.

[SMK*04] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser.: The Princeton Shape Benchmark. *SMI '04*, 167-178, 2004. http://shape.cs.princeton.edu/benchmark/.

[SZM*08] K. Siddiqi, J. Zhang, D. Macrini, A. Shokoufandeh, S. Bouix, S. Dickinson.: Retrieving Articulated 3D Models Using Medial Surfaces. *Machine Vision and Applications*, **19**(4), 261-275, 2008. http://www.cim.mcgill.ca/~shape/benchMark/.

[SOG09] J. Sun, M. Ovsjanikov, L. J. Guibas.: A concise and provably informative multi-scale signature based on heat diffusion. *SGP*, 2009.

[WHY*07] M. Wang, X-S. Hua, X. Yuan, Y. Song, L-R. Dai.: Optimizing Multi-Graph Learning: Towards A Unified Video Annotation Scheme. *ACM Multimedia* 2007.

[ZBL*03] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf.: Learning with Local and Global Consistency. *NIPS* 2003.