SHREC'08 Entry: Local Volumetric Features for 3D Model Retrieval

Kunio Osada, Takahiko Furuya, Ryutarou Ohbuchi

University of Yamanashi[†]

ABSTRACT

In this paper, we describe a method of shape-based 3D model retrieval that employs a set of 3D, local, multi-scale features extracted from a voxel representation of a 3D model to be compared. The method first convert a surface based 3D model into a voxel model. Then, a novel 3D extension of the popular 2D image feature, the Scale Invariant Feature Transform by David Lowe, is applied to extract a set of 3D local features. The 3D feature is invariant to rotation, uniform scaling, and translation in 3D space. A 3D model typically yields thousands of such local 3D features. Our method extracts thousands of 3D local features from a model to compare its shape. Our evaluation showed that the method is quite effective, achieving Means First Tier of 58% for the Query Set 1 of the SHREC'08 Generic Models Track.

KEYWORDS: Content-based retrieval, multi-scale feature, local shape descriptor, voxel representation, Scale-Invariant Feature Transform.

INDEX TERMS: H.3.3 [Information Search and Retrieval]: Information filtering. I.3.5 [Computational Geometry and Object Modeling]: Surface based 3D shape models. I.4.8 [Scene Analysis]: Object recognition.

1 INTRODUCTION

Appearance based comparison of 3D shapes has an advantage of being able to compare anything that can be rendered regardless of the shape representation the model employs.

The *Individual Match SIFT* (IM-SIFT) we described in [5] employs 2D local visual features. The IM-SIFT generates multiple-view set of depth images by viewing the model placed at the origin from multiple (e.g., 42) directions. A local feature set is extracted from each depth image by using the *Scale Invariant Feature Transform* (*SIFT*) [3] algorithm, which yields features invariant to translation, scaling, and rotation in the 2D image plane. As a depth image yields more than a dozen features, a 3D model typically has a set of thousands of local visual features. A distance among a pair of models is computed as a sum of distance between pairs of SIFT features. The distance computation can be costly; if a model has n features on average, the distance computation takes $O(n^2)$ time.

Despite its relative success, the IM-SIFT method has several issues to be solved. One is the cost of distance computation. As the IM-SIFT depends on multiple view set of images to achieve invariance about 2 out of 3 rotational axes, the n could become thousands if a high degree of rotational invariance is required. (The remaining 1 rotational degree of freedom is taken care of by the SIFT feature.) As the number of views increases, the cost of distance computation increased rapidly.

Our *BoF-SIFT* algorithm described in [6] offered an answer to the cost of distance computation; it fused thousands of features into single feature vector by using the *bag-of-features* (*BoF*) approach. Using the BoF approach, even if the number of views is increased beyond current 42, the cost of distance computation won't change.

The method used in this paper offers an alternative answer to achieve rotational invariance, that of *making a local feature rotation invariant in 3D*. As the new algorithm works on a gray-level volumetric data, we call it *Volumetric-SIFT*, or *VSIFT* for short. There has been so-called "3D" or "*n*D" SIFT algorithms [1, 7]. However, they do not achieve rotation invariance in 3D as they treat one of the axes specially, e.g., to establish temporal causality in a video retrieval.

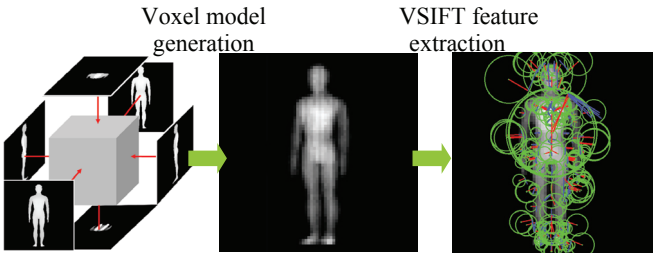

Figure 1. A surface-based model is converted to a voxel model before extracting a VSIFT feature.

2 METHOD

The VSIFT algorithm for 3D shape comparison proceeds as follows (See Figure 1).

1. **Voxel Model Generation:** An input surface based model is converted to a voxel model by re-projection of multiple range image rendered from the input model. We used the method by Karabassi, et al. [1], with an extension to increase the number of views. We wanted the voxel model to be "filled" even if the model has hollow interior. We thus used the re-projection of multi-view depth images, instead of 3D scan-conversion of surfaces. The algorithm renders depth images of the model placed at the coordinate origin from 180 viewpoints approximately-uniformly spaced in the solid angle. For the experiment, we used the voxel buffer of size 72^3, and the input model was scaled so that there is some margin around the model in the voxel buffer.
2. **Volumetric-SIFT Feature Extraction:** Compute VSIFT features from the voxelized 3D model in the manner similar to the 2D SIFT algorithm. As in the case of the (2D) SIFT algorithm [3], a multi-scale set of voxel models are computed by repeatedly blurring the original voxel model. Difference of Gaussian and other operators detects scale and orientation of local gray-level changes of the voxel model. We divided an octave in scale space into 35 sub-bands. The scale, orientation, etc. of the gray level change is encoded

[†]4-3-11 Takeda, Kofu-shi, 400-8511, Japan. osada.researchAT gmail.com, t03kf030AT yamanashi.ac.jp, ohbuchiAT yamanashi.ac.jp.

into a VSIFT feature. A VSIFT feature is a 1,280 dimensional vector, although the dimension varies according to several parameters, e.g., the number of sub-bands in the scale-space.

The rightmost image in Figure 1 shows examples of VSIFT features. The radius of the circle indicates the scale of a feature, and the line emanating from the center of the circle indicates the orientation of the feature. (Note that both position and orientation exist in 3D space; they are projected onto 2D image plane for visualization.)

3. **Feature Dimension Reduction:** We apply the PCA-SIFT algorithm [3] to reduce the dimension VSIFT feature down to 50. We trained the PCA algorithm with 10,000 VSIFT features, and chose the dimension at which the retrieval performance is maximized.
4. **Distance Computation:** As in the IM-SIFT algorithm, the distance among model A and B is computed as the sum of the minimum of distances from a feature of the model A to all the features in the model B. Temporal computational cost of the distance computation is $O(n^2)$ assuming that the average number of local feature per model is n. Unlike the IM-SIFT [6] algorithm in 2D, however, the IM-VSIFT does not consider "context" or relative position of the feature in the pose-normalized coordinate space.

We did perform experiments using the BoF approach [6] to distance computation, which we call *BoF-VSIFT*. The BoF-VSIFT is much faster in terms of distance computation than the IM-VSIFT. However, the IM-VSIFT did better than the BoF-VSIFT in terms of retrieval performance. Due to the space limitation, we include the algorithms and results from IM-VSIFT only in this paper.

3 Experiments and Results

We used the *SHREC'08 Generic Models Track* (*GMT*) for the benchmark. The SHREC'08 GMT contains two query sets; the *Query set 1* (*Q1*) is identical to SHREC 2006 query set, while the *Query set 2* (*Q2*) is new to the SHREC 2008 GMT.

Table 1 shows the performance of the proposed IM-VSIFT algorithm along with that of our *Semi-Supervised Dimension Reduction* (SSDR) based method [8]. The IM-VSIFT performed very well; in fact, the IM-VSIFT won the GMT in terms of retrieval performance. The IM-VSIFT produced the FT=51% for the Query set 1, and FT=46% for the query set 2.

For the Q1, the IM-VSIFT with FT=50% trailed the SSDR with FT=58%. For the Q2, however, the IM-VSIFT won with FT=46%; it is followed by the method by Thibault Napoléon [5] with FT=45%, and then by the SSDR-based method with the FT=36%. If the scores of Q1 and Q2 are averaged, the IM-VSIFT came in 1st with FT=48%, followed by the SSDR with FT=47%, and then by the method by Thibault Napoléon with FT=45%. The significant drop in performance for the SSDR-based method can be explained by the fact that the method is trained by the classes of Q1 but not those of Q2.

The high retrieval performance of the IM-VSIFT comes with a price; it has a very high cost of feature extraction and feature comparison. The IM-VSIFT requires about 30s per model for feature extraction. More significantly, comparing a pair of features takes a bit less than 1s. The time for feature comparison especially is prohibitively high if a database containing a nontrivial number of models is considered. At this time, among the SHREC'08 GMT participants, either the SSDR-based method [8] or the method by Thibault Napoléon et al. [5] may be a better choice for a practical application due to their speed.

4 Conclusion

The IM-VSIFT method achieved very high retrieval performance, taking the 1st place overall in the Generic Models Track of the SHREC'08 with the Mean First Tier of 48% if the Query set 1 and Query set 2 combined.

The IM-VSIFT, however, has a very high cost of feature extraction and comparison. We are intending to investigate methods of acceleration for the IM-VSIFT, both in terms of feature extraction and feature comparison. A possible avenue of exploration is the use of Graphics Processing Unit algorithms.

References

[1] W. Cheung G. Hamarneh, n-SIFT: n-dimensional Scale Invariant Feature Transform for Matching Medical Images. Proc. IEEE ISBI, pp. 720-723, (2007).

[2] E.A. Karabassi, G. Papaioannou, T. Theoharis, A Fast Depth Buffer Based Voxelization Algorithm, Journal of Graphics Tools, ACM, 4(4), pp.5-10, (1999).

[3] Y. Ke, R. Sukthankar, PCA-SIFT: A more distinctive representation for local image descriptors, *Proc. CVPR2004*, (2004).

[4] David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *Int'l Journal of Computer Vision*, **60**(2), (2004).

[5] T. Napoléon, T. Adamek, F. Schmitt, N. E. O'Connor, SHREC'08 Entry: Multi-view 3D retrieval using multi-scale contour representation, *in this proceedings.*

[6] R. Ohbuchi, K. Osada, T. Furuya, T. Banno, Salient local visual features for shape-based 3D model retrieval, *Proc. SMI '08*, (2008).

[7] P. Scovanner, S. Ali, M. Shah, A 3-Dimensional SIFT Descriptor and its Application to Action Recognition, *Poster, Proc. ACM Multimedia*, pp. 357-260, (2007).

[8] A. Yamamoto, M. Tezuka, T. Shimizu, R. Ohbuchi, SHREC'08 Entry: Semi-Supervised Learning for Generic 3D Model Retrieval, *in this proceedings*.

Table 1. Retrieval performances of the IM-VSIFT and the Semi-Supervised Dimension Reduction (SSDR)-based method [8].

Supervised	Features	Query	AP-HR	FT-HR [%]	DAR	NCG @25	NDCG @25
No	IM-VSIFT	Q1	0.5437	50.73	0.5998	0.6229	0.6566
		Q2	0.4857	46.12	0.5305	0.4972	0.5549
		Q1+Q2	0.5147	48.43	0.5652	0.5600	0.6057
Yes	SSDR (SPRH, LLE=400D, SLPP=50D) [4]	Q1	0.6309	58.33	0.6514	0.6578	0.6879
		Q2	0.4081	36.26	0.4480	0.4612	0.4681
		Q1+Q2	0.5195	47.29	0.5497	0.5595	0.5780

AP-HR: Mean Average Precision (highly relevant)
FT-HR: Mean First Tier (Highly Relevant)
DAR: Mean Dynamic Average Recall
NCG @25: Mean Normalized Cumulated Gain @25
NDCG @25: Mean Normlized Discounted Cumulated Gain @25