SHREC'08 Entry: Local 2D Visual Features for CAD Model Retrieval

Kunio Osada, Takahiko Furuya, Ryutarou Ohbuchi

University of Yamanashi†

ABSTRACT

A local shape feature has an advantage in dealing with deformable or articulated 3D models. We evaluate the performance of our local, 2D visual features and their integration method based of the bag-of-features approach on the SHREC'08 CAD Model Track task. The evaluation showed that, it performed very well, winning the 2nd place in the contest, although it lost to a method that employs supervised learning of classes in the benchmark dataset.

KEYWORDS: Content-based retrieval, multi-scale feature, 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

To enter the last year's SHREC 2007 *CAD Model Track* (*CMT*), we used the shape feature for polygon soup models, and applied an unsupervised dimension reduction to find a subspace adapted to distribution features for the kind of models found in the database [6]. An interesting observation was that the feature extracted from convex hull of the models in the database still achieved quite good retrieval performance, about as good as the second place in the contest. However, we also observed that some of the models in a class shares local feature but not their overall shape, the kind extracted by convex hull.

In this entry to the SHREC 2008 CMT, we employ a method that compares shape by multi-scale, local, 2D visual features [4]. Local features are quite promising, especially for an articulated shape or for partial matching. However, distance computation using hundreds or thousands of local features per model can be costly. In the method we employed [4], we fuse local visual features into a feature vector for efficient model-to-model comparison.

2 METHOD

The method we used, whose detail is described in [4], compares shapes of 3D models visually by using a set of local features extracted from 2D depth images of the model (Figure 1). The depth images are generated by viewing the model placed at the coordinate origin from multiple orientations. Of required three axes of rotational invariance, two are achieved by the multiple-view rendering of the model.

The local feature set for each depth image is extracted by using the SIFT, which first selects (what the algorithm thinks) salient points and then computes a 128D feature that encodes scale and orientation of grayscale changes local to the salient points in the image. The rotational invariance of the SIFT feature handles the one remaining axis of rotational degrees-of-freedom. The SIFT algorithm extracts more than a dozen of features per depth image, and a model is rendered from 42 view orientations. The 3D model thus has a set of thousands of local visual features.

Computing a distance among a pair of 3D models having thousands of local features each can be expensive. A simple-minded algorithm would take $O(n^2)$ time assuming the average number of local feature per model to be n. With a large database, this cost of distance computation becomes dominant compared to the cost of feature extraction

To simplify distance computation, we combine these local visual features into a single feature vector by using so-called *bag-of-features* (*BoF*) approach [1, 5, 7]. The BoF approach is inspired originally by the *bag-of-words* approach in text retrieval, which characterizes a text document by a histogram of words' occurrences in the document. In our method, a vector-quantized local feature is treated as (visual) word, whose frequency per 3D model is accumulated into a histogram, which becomes the feature for the model.

The codebook for vector quantization is learned unsupervised, off-line, by using thousands of visual features generated from the models in the database to be queried. We used the well-known k-means clustering algorithm to learn the codebook. After the clustering using k-means algorithm, the barycenter of each cluster becomes the representative vector of the cluster. Once the codebook is learned, given a feature vector, its quantized vector can be found by searching through the codebook (i.e, the list of barycenters) for the vector closest to the feature vector. Given k cluster centers, that is, the vocabulary size of k, vector quantization takes $O(k)$ time. Vector quantization takes significant amount of time for a large number of local features, e.g., for preprocessing an entire database.

3 EXPERIMENTS AND RESULTS

We used the SHREC 2008 CAD Model track, which queries *Engineering Shape Benchmark* (ESB) database [2], for performance evaluation. There are 45 query models, and the ground truth classes in the database are "highly relevant" only. We varied a parameter of the method, the vocabulary size k to find a best performing result file. To compute distance among the histograms, we compared the *Kullback-Leibler Divergence* (*KLD*) used in [4] and *Cosine* measure for their retrieval performance.

Table 1 shows the performance of the proposed BoF-SIFT algorithm using two different distance measures and three different vocabulary sizes. The table also lists the performance figures for (1) our *Semi-Supervised Dimension Reduction* (*SSDR*) based method [8], and (2) our previous *Unsupervised Dimension Reduction* (*UDR*) based method that finished 1st in the SHREC 2007 CMT [7].

The BoF-SIFT using *Cosine* measure and vocabulary size k=1,200 achieved the best *Mean First Tier* (*FT*) of 44%, outperforming our previous UDR-based algorithm [5] with its FT=41%.

In the SHREC'08 CMT, the BoF-SIFT method with its FT=44% ranked at 2nd place. The method using morphological multiresolution feature set combined with SSDR [7] won the 1st

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

place with FT=78%, and the method by Napoleon et al finished the 3rd.

When we looked into each query, the SSDR-based method achieved FT=100% for the majority of the queries. However, some of the queries do favor the local features of BoF-SIFT instead of the SSDR-based method. For example, the BoF-SIFT works better than the SSDR-based method for the Query 19; the BoF-SIFT gave FT=69%, while the SSDR-based method gave FT=0%. In this query, a local feature, the teeth of a gear, is more important than the global shape of cylinder. The SSDR-based method retrieved cylindrical objects instead of gears. An appropriate fusion of these two methods would give us a stronger hybrid approach.

4 CONCLUSION

The Bag-of-Features SIFT method, which combines thousands of local visual features into a feature vector, is quite effective. It produced Mean First Tier of 44% to rank at 2nd place in the SHREC'08 CAD Model Track, trailing only the method based on Semi-Supervised Dimension Reduction [8].

In the future, we plan to explore methods to reduce computational costs for vector quantization and codebook learning. We also plan to apply some form of learning-based algorithm, supervised and unsupervised, to the BoF-SIFT method for a performance boost. We also would like to investigate a fusion of a global feature based shape matching method such as [8] and a local feature based method like this for better retrieval performance.

REFERENCES

[1] 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, pp.59-74, (2004)

[2] S. Jayanti, Y. Kalyanaraman, N. Iyer, K. Ramani, Developing An Engineering Shape Benchmark for CAD Models, CAD, 38(9), (2006)

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

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

[5] J. Sivic, A. Zisserman, Video Google: A text retrieval approach to object matching in Videos, Proc. ICCV 2003, Vol. 2, pp. 1470-1477, (2003).

[6] R.C. Veltkamp, F.B. ter Harr, SHREC 2007 3D Shape Retrieval Contest, Dept of Info and Comp. Sci., Utrecht Univ., Tech. Report UU-CS-2007-015, (2007).

[7] J. Winn, A. Criminisi, T. Minka, Object categorization by learned universal visual dictionary, Proc. ICCV05, Vol. II, pp.1800-1807, (2005).

[8] A. Yamamoto, M. Tezuka, T. Shimizu, and R. Ohbuchi, SHREC'08 Entry: Semi-Supervised Learning for 3D CAD Model Retrieval, Proc. SMI 2008 (2008)..

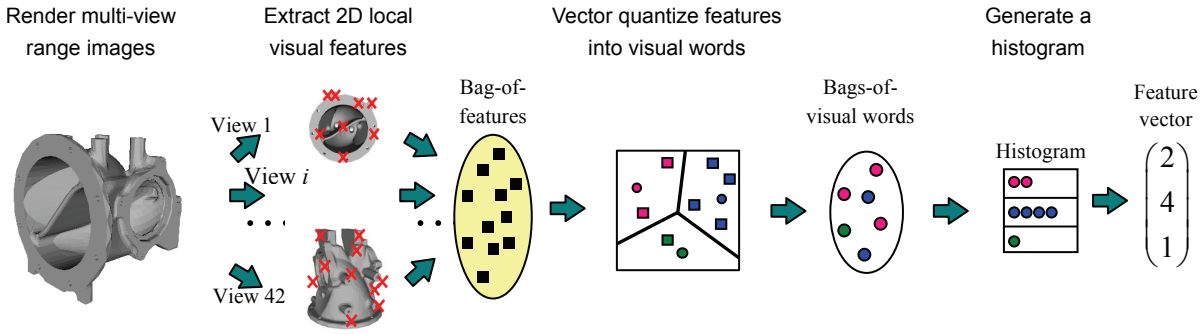

Figure 1. *Bag-of-Feature SIFT* algorithm integrates thousands of local visual features per model, extracted from multi-view depth images, into a feature vector for efficient feature-to-feature comparison.

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

Supervised	Method	Distance measure	Vocabulary size *k*	AP-HR	FT-HR [%]	DAR	NCG @25	NDCG @25
No	BoF-SIFT	Cosine	1000	0.4489	42.26	0.5324	0.4727	0.5237
			1200	0.4764	44.28	0.5676	0.5013	0.5603
			1500	0.4682	43.71	0.5555	0.4876	0.5418
		KLD	1000	0.4506	43.83	0.5441	0.4815	0.5309
			1200	0.4360	43.69	0.5367	0.4724	0.5190
			1500	0.4344	43.61	0.5300	0.4719	0.5186
No	UDR [6]			0.4337	41.23	0.5357	0.5023	0.5341
Yes	SSDR [8]			0.7997	78.17	0.7943	0.7887	0.7916

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