Eurographics Workshop on 3D Object Retrieval (2014)
B. Bustos, H. Tabia, J.P. Vandeborre, and R. Veltkamp (Editors)

SHREC'14 Track: Large Scale Comprehensive 3D Shape Retrieval

B. Li[†‡1], Y. Lu[†‡1], C. Li[†2], A. Godil[†2], T. Schreck[†3], M. Aono[‡4], Q. Chen[‡5], N.K. Chowdhury[‡4], B. Fang[‡5], T. Furuya[‡6], H. Johan[‡7], R. Kosaka[‡4], H. Koyanagi[‡4], R. Ohbuchi[‡6], A. Tatsuma[‡4]

[1] Department of Computer Science, Texas State University, San Marcos, USA
[2] Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, USA
[3] Computer and Information Science, University of Konstanz, Konstanz, Germany
[4] Department of Computer Science and Engineering, Toyohashi University of Technology, Japan
[5] College of Computer Science, Chongqing University, China [6] University of Yamanashi, Yamanashi, Japan
[7] Fraunhofer IDM@NTU, Singapore

Abstract

The objective of this track is to evaluate the performance of 3D shape retrieval approaches on a large-sale comprehensive 3D shape database that contains different types of models, such as generic, articulated, CAD and architecture models. The track is based on a new comprehensive 3D shape benchmark, which contains 8,987 triangle meshes that are classified into 171 categories. The benchmark was compiled as a superset of existing benchmarks and presents a new challenge to retrieval methods as it comprises generic models as well as domain-specific model types. In this track, 14 runs have been submitted by 5 groups and their retrieval accuracies were evaluated using 7 commonly used performance metrics.

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [Computer Graphics]: Information Systems—Information Search and Retrieval

1. Introduction

With the increasing number of 3D models created every day and stored in databases, the development of effective and scalable 3D search algorithms has become an important research area. In this contest, the task was to retrieve 3D models similar to a complete 3D model query from a new integrated large-scale comprehensive 3D shape benchmark including various types of models. Owing to the integration of the most important existing benchmarks to date, the newly created benchmark is the most exhaustive to date in terms of the number of semantic query categories covered, as well as the variations of model types. In particular, it combines generic and domain-dependent model types and therefore rates the retrieval performance with respect to cross-domain retrieval tasks. The shape retrieval contest will allow researchers to evaluate results of different 3D shape retrieval approaches when applied on a large scale comprehensive 3D database.

The benchmark is motivated by the latest large collection of human-drawn sketches built by Eitz et al. [EHA12]. To explore how human draw sketches and human sketch recognition work, they collected 20,000 human-drawn sketches, categorized into 250 classes, each with 80 sketches. This sketch dataset is exhaustive in terms of the number of object categories. Thus, we believe that a 3D model retrieval benchmark based on their object categorizations will be more comprehensive and appropriate than currently available 3D retrieval benchmarks to more objectively and accurately evaluate the real practical performance of a comprehensive 3D model retrieval algorithm if implemented and used in the real world.

Considering this, we built a SHREC'14 Large Scale Comprehensive Track Benchmark (**SHREC14LSGTB**) by collecting relevant models in the major previously proposed 3D

† Track organizers. For any questions related to the track, please contact Generic3D@nist.gov or li.bo.ntu0@gmail.com.
‡ Track participants.

Table 1: *Classification information of the 8 selected source benchmarks.*

Benchmarks	Types	Number of models	Number of classes	Average number of models per class
PSB	Generic	907 (train) 907 (test)	90 (train) 92 (test)	10 (train) 10 (test)
SHREC12GTB	Generic	1200	60	20
TSB	Generic	10,000	352	28
CCCC	Generic	473	55	9
WMB	Watertight (articulated)	400	20	20
MSB	Articulated	457	19	24
BAB	Architecture	2257	183 (function-based) 180 (form-based)	12 (function-based) 13 (form-based)
ESB	CAD	867	45	19

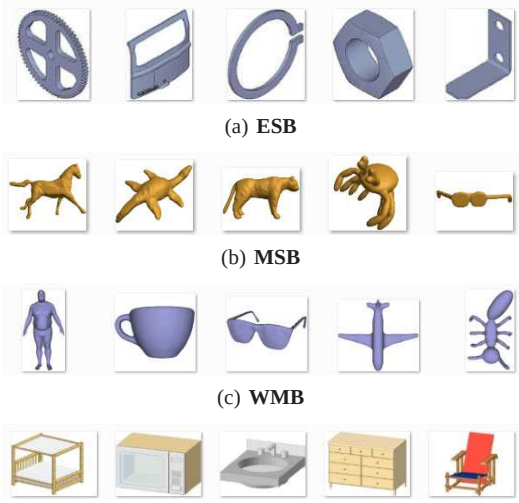
object retrieval benchmarks. When creating the benchmark, our target was to find models for as many of the 250 classes as possible, and for each class find as many models as possible. These previous benchmarks have been compiled with different goals in mind and to date, have not been considered in their union. Our work is the first to integrate them to form a new, larger benchmark corpus for comprehensive 3D shape retrieval.

To avoid adding replicate models, we selected the following 8 benchmarks: the Princeton Shape Benchmark (**PSB**) [SMKF04], the SHREC'12 Generic Track Benchmark (**SHREC12GTB**) [LGA*12], the Toyohashi Shape Benchmark (**TSB**) [TKA12], the Konstanz 3D Model Benchmark (**CCCC**) [Vra04], the Watertight Model Benchmark (**WMB**) [VtH07], the McGill 3D Shape Benchmark (**MSB**) [SZM*08], the Bonn Architecture Benchmark (**BAB**) [WBK09], and the Engineering Shape Benchmark (**ESB**) [JKIR06]. Table 1 lists their basic classification information while Fig. 1 shows some example models for the four specific benchmarks. Totally, the extended large-scale benchmark has 8,987 models, classified into 171 classes. The average number of models in each class is 53, which is much more than any of the benchmark in Table 1.

Based on this new challenging benchmark, we organize this track to foster this research area by soliciting retrieval results from current state-of-the-art 3D model retrieval methods for comparison, especially for practical retrieval performance. We will also provide evaluation code to compute a set of performance metrics, including those commonly used for evaluating query by example retrieval techniques.

2. Data Collection

The SHREC'14 Large Scale Comprehensive Retrieval Track Benchmark[†] has 8,987 models, categorized into 171 classes. We (one undergraduate student, one master student, one researcher with a master degree and one with a PhD degree) adopt a voting scheme to classify models. For each classification, we have at least two votes. If these two votes agree each other, we confirm that the classification is correct, otherwise, we perform a third vote to finalize the classification. All the models are categorized according to the classifications in Eitz et al. [EHA12], based on visual similarity. This 3D dataset was also used as the target 3D model dataset in evaluating sketch-based 3D shape retrieval algorithms in the SHREC'14 track on extended large scale sketch-based 3D shape retrieval (http://www.itl.nist.gov/iad/vug/sharp/contest/2014/SBR/).



(a) **ESB**

(b) **MSB**

(c) **WMB**

(d) **BAB**

Figure 1: *Example 3D models in **ESB**, **MSB**, **WMB** and **BAB** datasets.*

3. Evaluation

To have a comprehensive evaluation of the retrieval algorithm, we employ seven commonly adopted performance

[†] Available on http://www.itl.nist.gov/iad/vug/sharp/contest/2014/Generic3D/.

metrics in 3D model retrieval. They are Precision-Recall (PR) graph, Nearest Neighbor (NN), First Tier (FT), Second Tier (ST), E-Measures (E), Discounted Cumulated Gain (DCG) and Average Precision (AP). We also have developed the code to compute them. Besides the common definitions, we also develop two weighted variations for each benchmark by incorporating the model variations in each class. Basically, we use the number of available models to define the model variations. We assume there is a linear correlation between the number of available models in one class and the degree of variations of the class. Therefore, we adopt a weight based on the number of models or its reciprocal to define each weighted performance metric.

Proportionally weighted metrics m_p and reciprocally weighted metrics m_r (m=NN/FT/ST/E/DCG) are defined as follows.

$$m_p = \sum_{i=1}^{N} \frac{n_i}{N} \cdot m_i,$$
$$m_r = \sum_{i=1}^{N} \frac{\frac{1}{n_i}}{\sum_{j=1}^{N} \frac{1}{n_j}} \cdot m_i,$$

where N is the total number of models, n_i is the size of class to which the i^{th} model belongs, m_i is the non-weighted NN/FT/ST/E/DCG metric value for the i^{th} model. m_p assigns bigger weights to the classes with more variations. On the contrary, m_r highlights the performance in retrieving classes with few models/variations.

4. Participants

There are 5 groups who have successfully participated in the SHREC'14 Comprehensive 3D Shape Retrieval track. In total, they have submitted 14 dissimilarity matrices. The details about the participants and their runs are as follows.

- *CSLBP-Run-1*, *CSLBP-Run-2*, *CSLBP-Run-3*, *HSR-DE* and *KVLAD* submitted by Masaki Aono, Nihad Karim Chowdhury, Hitoshi Koyanagi, and Ryuichi Kosaka from Toyohashi University of Technology, Toyohashi, Japan (Section 5.1)
- *DBNAA_DERE* submitted by Qiang Chen and Bin Fang from Chongqing University, China (Section 5.2)
- *BF-DSIFT*, *VM-1SIFT*, *MR-BF-DSIFT*, *MR-D1SIFT* and *MR-VM-1SIFT* submitted by Takahiko Furuya and Ryutarou Ohbuchi from University of Yamanashi, Japan (Section 5.3)
- *ZFDR* submitted by by Bo Li and Yijuan Lu from Texas State University, USA; and Henry Johan from Fraunhofer IDM@NTU, Singapore (Section 5.4)
- *DBSVC* and *LCDR-DBSVC* submitted by Atsushi Tatsuma and Masaki Aono from Toyohashi University of Technology, Japan (Section 5.5)

5. Methods

5.1. Hybrid Shape Descriptors CSLBP*, HSR-DE, and KVLAD, by M. Aono, N.K., Chowdhury, H. Koyanagi, and R. Kosaka

We have investigated accurate 3D shape descriptors over the years for massive 3D shape datasets. In the Large Scale Comprehensive 3D Shape Retrieval track, we have attempted to apply three different methods with five runs. Note that all the five runs, we apply pose normalization [TA09] as pre-processing.

For the first three runs, we applied CSLBP*, a hybrid shape descriptor, composed of **C**enter-**S**ymmetric **L**ocal **B**inary **P**attern (CSLBP) feature [HPS06], **E**ntropy descriptor [CPW98], and optional Chain Code (CC). The difference between the three runs comes from the number of view projections and the existence of the optional CC: 16 views for CSLBP in Run-1, 24 views for CSLBP in Run-2 and Run-3, while no CC for Run-1 and Run-2 and CC addition in Run-3. CSLBP* is computed by first generating depth buffer images from multiple viewpoints for a given 3D shape object, then by analyzing gray-scale intensities to produce three-resolution level histograms (in our implementation, 256×256, 128×128, and 64×64), having 16 bins each, after segmenting each depth-buffer image into sub-images (16, 8, 4, respectively). In addition to CSLBP, we have augmented it with "Entropy", trying to capture the randomness of surface shapes, resulting in CSLBP*.

For the fourth run, we applied HSR-DE, another hybrid shape descriptor, composed of multiple Fourier spectra obtained by Hole, Surface-Roughness, Depth-buffer, Contour, Line, Circle, and Edge images, an extension to the method we published in [AKT13]. Figure 2 illustrates the method adopted in Run-4.

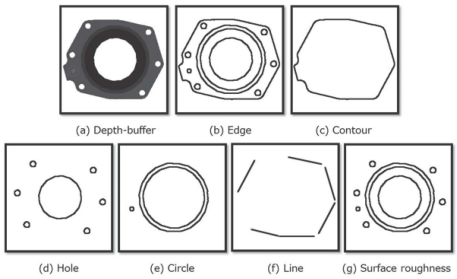


Figure 2: *An example of HSR-DE (Hole and Surface-Roughness descriptors with Depth-buffer and Edge features augmented) before conversion to Fourier spectra.*

For the fifth run, we applied KVLAD, a supervised learning method we developed by combining non-linear scale space [ABD12] with VLAD [JDSP10]. For the training stage, we employ SHREC2011 data and generate a code book of size 500, which is used for distance computation during the testing stage.

5.2. 3D Model Retrieval Descriptor DBNAA_DERE, by Q. Chen and B. Fang [CFYT14]

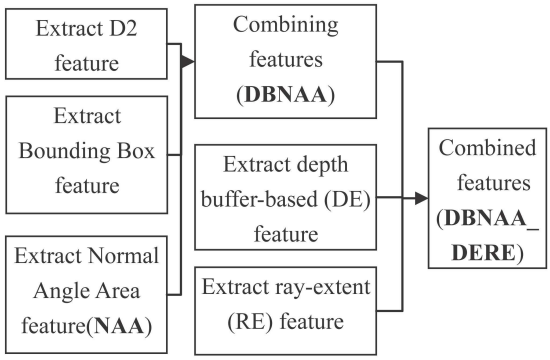


Figure 3: *DBNAA_DERE feature extraction procedure.*

First, we extract a Normal Angle Area (NAA), and then combined NAA with D2 [OFCD02] and Bounding Box feature to form DBNAA. Second, we combine DBNAA with Depth buffer images (DE), Ray extent (RE) [Vra04] to form DBNAA_DERE [CFYT14] descriptor. Figure 3 shows the feature extraction procedure.

(1) DBNAA feature extraction. DBNAA comprises three components: D2 feature, Bounding Box feature and Normal Angle Area feature. D2 feature [OFCD02] is extracted by using N=1024 samples, B=1024 bins, and Bounding Box feature is extracted after using Continuous Principle Component Analysis (CPCA) [Vra04] for model pose normalization.

$$L = \{Z_{max} - Z_{min}, Y_{max} - Y_{min}, X_{max} - X_{min}\},$$
$$F_{BB} = \{\frac{rank(L,1)}{rank(L,2)}, \frac{rank(L,2)}{rank(L,3)}\},$$

where $rank()$ is to sort the vector in ascending order.

NAA feature is based on mean angle α and average area S of each vertex,

$$\alpha = \frac{1}{N_{vj}} \sum_{\{n_i,n_j\} \subset F_{vj}} n_i \cdot n_j,$$
$$S = \frac{1}{N_{vj}} \sum_{i=1}^{N_{vj}} S_i,$$

where N_{vj} is the number of adjacent faces of each vertex, n_i is the normal of face i, F_{vj} is the set of normals of the adjacent faces.

After getting the mean angle α and average area S, they can be formed into a joint 2D histogram, where α is divided into N bins, S is also divided into N bins. Here N is set to 16 empirically. NAA is a N*N feature matrix.

1) DBNAA feature combination. Feature combination is carried out as follows,

$$d_{DBNAA} = \alpha * d_D + \beta * d_B + (1 - \alpha - \beta) * d_{NAA}$$

where the α and β is set to α=0.65, β=0.15 according to our experiments on SHREC'12 Track: Generic 3D Shape Retrieval [LGA*12] dataset.

2) DBNAA_DERE feature combination. Inspired by Li and Johan [LJ13], the Depth Buffer-based (DE) and Ray-Extent (RE) [Vra04] features are combined to the DBNAA framework as follows:

$$d_{DBNAA\_DERE} = \alpha * d_{DBNAA} + \beta * d_{DE} + (1 - \alpha - \beta) * d_{RE}$$

Here α=0.3, β=0.35 also according to the experiment on SHREC'12 Track: Generic 3D Shape Retrieval [LGA*12] dataset.

Since it could not get the correct class label of the test set, so the class information based retrieval method is not available here. For more details about the shape descriptor computation, please refer to [CFYT14].

5.3. Visual Feature Combination for Generic 3D Model Retrieval, by T. Furuya and R. Ohbuchi

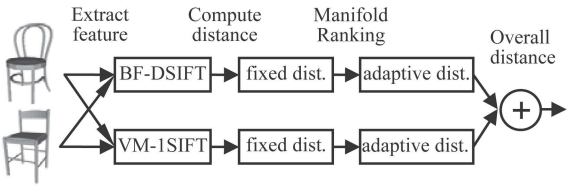


Figure 4: *Two feature-adaptive distances computed using two visual features (BF-DSIFT and VM-1SIFT) are combined.*

Our method is essentially the one described in [LGA*12] and [OF10]. An overview of our algorithm is shown in Figure 4. It involves multi-viewpoint rendering of 3D models, extraction of local and global visual features, and combination of feature-adaptive distances learned on both local and global visual features.

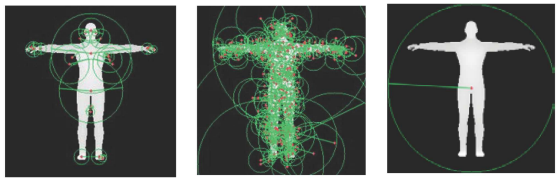
A nice property of such a view-based approach is that 3D models in almost any shape representations (polygon soup, uniform mesh, point cloud etc.) can be compared. And combination of local and global features would yield more robust retrieval results due to complementary features extracted by local and global approaches. Local feature is robust against articulation of 3D shapes, e.g., bending of joints. However, it has difficulty distinguishing some classes of rigid shapes, e.g, pipes bent in U shape and in S shape. On the other hand, global feature may distinguish these rigid shapes.

5.3.1. Visual feature extraction

Our method first renders a 3D model into range images from multiple (in this case 42) viewpoints spaced uniformly in

solid angle and image resolution of 256×256 pixels. Then it extracts a set of local visual features, Dense SIFT (DSIFT) [FO09] and a set of global visual features, One SIFT (1SIFT) [OF10] from the range images.

For DSIFT extraction, we randomly and densely sample feature points on the range image with prior to concentrate feature points on or near 3D model in the image (see Figure 5 (b)). From each feature point sampled on the image, we extract SIFT [Low04], which is multi-scale, rotation-invariant local visual feature. The number of feature points per image is set to 300 as in [FO09], resulting in about 13k DSIFT features per 3D model. The set of dense local features are integrated into a single feature vector per 3D model by using Bag-of-Features (BF) approach. We use ERC-Tree algorithm [GEW06] to accelerate both codebook learning (clustering of local features) and vector quantization of local features into visual words. The frequency histogram of vector-quantized DSIFT features becomes a BF-DSIFT feature vector for the 3D model.



(a) Original SIFT [Low04] (b) DSIFT (c) 1SIFT

Figure 5: *Our method combines dense local visual features (DSIFT) and global visual features (1SIFT).*

For 1SIFT extraction, we sample a feature point at the center of the range image and extract a SIFT feature from a large region covering the entire 3D model (see Figure 5 (c)). The number of 1SIFT per model is equal to the number of rendering viewpoints (42 1SIFT features in this case). Note that the set of 1SIFT features is not BF-integrated but is directly compared by using per-View Matching of 1SIFT features (VM-1SIFT).

5.3.2. Distance computation

For the experiments, ranking of retrieval results are performed by using two different distance metrics; fixed distance and feature-adaptive distance learned by using Manifold Ranking (MR) algorithm [ZBL*03].

(1) Fixed distance. Symmetric version of Kullback-Leibler Divergence (KLD) is used as fixed distance metric. KLD performs well when comparing a pair of probability distributions, i.e., histograms. For the BF-DSIFT, distance between a pair of 3D models $\mathbf{x}_i$, $\mathbf{x}_j$ is equivalent to KLD between BF-DSIFT feature vectors of the models (Equation (1)). For the VM-1SIFT, distance between a pair of 3D models is calculated by using Equation (2) where N_v is the number of 1SIFT features per model and x_{ip} is 1SIFT feature vector extracted from the view p of 3D model x_i.

$$d_{BF-DSIFT}(\mathbf{x}_i,\mathbf{x}_j) = d_{KLD(\mathbf{x}_i,\mathbf{x}_j)} \tag{1}$$

$$d_{VM-1SIFT}(\mathbf{x}_i,\mathbf{x}_j) = \sum_{p=1}^{N_v} \min_{1\le q\le N_v} d_{KLD}(\mathbf{x}_{ip},\mathbf{x}_{jq}) \tag{2}$$

(2) Feature-adaptive distance. To improve distance metric among 3D models, we apply MR to each of the BF-DSIFT and the VM-1SIFT to compute diffusion distance on a feature manifold. We first generate a $N_m \times N_m$ affinity matrix $\mathbf{W}$ where N_m is the number of 3D models (N_m=8,987 for this track) and $\mathbf{W}_{ij}$ indicates similarity between a pair of 3D models x_i, x_j. $\mathbf{W}_{ij}$ is computed by using Equation (3) where d is fixed distance of either BF-DSIFT (Equation (1)) or VM-1SIFT (Equation (2)).

$$\mathbf{W}_{ij} = \begin{cases} \exp(-\frac{d(\mathbf{x}_i,\mathbf{x}_j)}{\sigma}) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

We normalize $\mathbf{W}$ for $\mathbf{S}$;

$$\mathbf{S} = \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}} \tag{3}$$

where $\mathbf{D}$ is a diagonal matrix whose diagonal element is $\mathbf{D}_{ii} = \sum_j W_{ij}$.

We use the following closed form of the MR to find relevance values in $\mathbf{F}$ given "source" matrix $\mathbf{Y}$. $\mathbf{F}_{ij}$ is the relevance value of the 3D model i and j. A higher relevance means a smaller diffusion distance.

$$\mathbf{F} = (\mathbf{I} - \alpha\mathbf{S})^{-1}\mathbf{Y} \tag{4}$$

We add prefix "MR-" before the feature comparison method to indicate MR-processed algorithms (MR-BF-DSIFT and MR-VM-1SIFT). Table 2 summarizes the parameters used for MR-BF-DSIFT and MR-VM-1SIFT. To further improve retrieval accuracy, we combine diffusion distances of local features and global features. The diffusion distances of MR-BF-DSIFT and MR-VM-1SIFT are normalized and then summed with equal weight (MR-D1SIFT).

Table 2: *Parameters for the MR.*

Method	σ	α
MR-BF-DSIFT	0.0050	0.975
MR-VM-1SIFT	0.0025	0.900

5.4. Hybrid Shape Descriptor ZFDR, by B. Li, Y. Lu and H. Johan [LJ13]

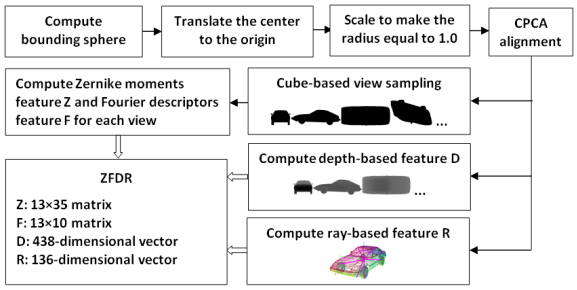


Figure 6: *ZFDR feature extraction process [LJ13].*

The comprehensive 3D model dataset contains both generic and professional (e.g. CAD and architecture models), rigid and non-rigid, articulated and non-articulated, watertight and non-watertight models. Due to the variations in the types and robustness considerations in retrieval performance, we employ the hybrid shape descriptor ZFDR devised in [LJ13] which integrates both visual and geometric information of a 3D model: **Z**ernike moments and **F**ourier descriptor features of 13 cube-based sample views, and **D**epth information feature of 6 depth buffer views and **R**ay-based features based on ray shooting from the center of the model to its farthest surface intersection points. Figure 6 illustrates the overview of the feature extraction process: 3D model normalization mainly utilizing Continuous Principle Component Analysis (CPCA) [Vra04] and extraction of four component features **Z**, **F**, **D** and **R**. The details are described as follows.

(1) View sampling. As a tradeoff between efficiency and accuracy, the approach sets cameras on the 4 top corners, 3 adjacent face centers and 6 middle edge points of a cube to generate 13 silhouette views to represent a 3D model.

(2) Zernike moments and Fourier descriptors features (ZF). For each silhouette view, up to 10^{th} order Zernike moments [KH90] (totally 35 moments) and first 10 centroid distance-based Fourier descriptors [ZL01] are computed to respectively represent the region-based and contour-based visual features of the 3D model.

(3) Depth information and Ray-based features (DR). To improve the versatility of the descriptor in characterizing diverse types of models, the depth buffer-based feature and ray-based with spherical harmonic representation feature developed by Vranic [Vra04] are integrated into the hybrid shape descriptor. The executable files [Vra04] are utilized to extract the 438-dimensional **D** and 136-dimensional **R** features.

(4) ZFDR hybrid shape descriptor distance. Scaled-ℓ_1 [Vra04] or Canberra distance metric is first applied to measure the component distances d_Z, d_F, d_D, and d_R between two models. Then, the hybrid descriptor distance d_{ZFDR} is generated by linearly combining the four component distances.

Please refer to the original paper [LJ13] for more details about the feature extraction and retrieval process.

5.5. Depth Buffered Super-Vector Coding, by A. Tatsuma and M. Aono

We propose a new 3D model feature known as Depth Buffered Super-Vector Coding (DBSVC), an approach categorized as a bag-of-visual words method [BBGO11, FO09]. DBSVC extracts 3D model features from rendered depth buffer images using a super-vector coding method [ZYZH10].

Figure 7 illustrates the generation of our proposed DBSVC feature. We first apply pose normalization. 3D models are usually defined by different authors with distinct authoring tools, which make the position, size, and orientation of 3D models diverse from each other. To solve this problem, we use Point SVD, a normalization method developed previously by the authors [TA09].

Post pose normalization, we enclose the 3D model with a unit geodesic sphere. From each vertex of the unit geodesic sphere, we render depth buffer images with 300×300 resolution, and a total of 38 viewpoints are defined.

After image rendering, we extract local features from each depth buffer image. Here, we propose a new local feature called Power SURF descriptor. SURF-128 descriptor extracts the sums of the wavelet responses, which are split up according to their signs [BETVG08]. The SURF-128 descriptor outperforms the regular SURF descriptor, but

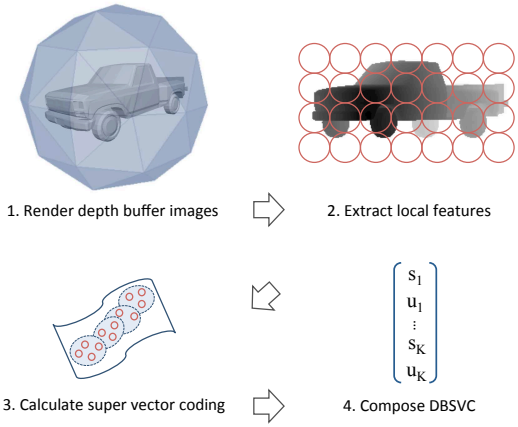


Figure 7: *Overview of the Depth Buffered Super Vector Coding (DBSVC)*

it turns more sparse. In the field of computer vision, Euclidean distance is known as a poor measure of similarity on sparse vectors [NS06]. The power normalization diminishes the sparseness of a vector [PSM10]. Thus, we apply the power and the ℓ_2 normalization to the SURF-128 descriptor, and call it the Power SURF descriptor. Moreover, we employ feature augmentation with patch coordinates [SPdC12]. The Power SURF descriptors are extracted from 98×98 pixel patches arranged every 5 pixels.

To calculate DBSVC, we generate a codebook of visual words in advance. The visual word is thus defined as the center of a cluster obtained by applying K-means clustering to the Power SURF descriptors, which are extracted from 3D models in the training dataset. For the training dataset, we use the NTU 3D Model Dataset (NMD) [CTSO03], consisting of 10910 unclassified models. We remove the decimated and the duplicated models from NMD, and use the remaining 4500 models as the training dataset. K-means clustering is performed with $K = 2048$.

We calculate DBSVC with the codebook of k visual words $\mathbf{v}_1, \ldots, \mathbf{v}_K$. Given a set of local features $\mathbf{x}_1, \ldots, \mathbf{x}_N$ extracted from a 3D model, let $q_{ki} = 1$ if $\mathbf{x}_i$ is assigned to $\mathbf{v}_k$ and 0 otherwise. For each $k = 1, \ldots, K$, we define

$$\begin{aligned} p_k &= \frac{1}{N}\sum_{i=1}^{N} q_{ki}, \\ s_k &= s\sqrt{p_k}, \\ \mathbf{u}_k &= \frac{1}{\sqrt{p_k}}\sum_{i=1}^{N} q_{ki}(\mathbf{x}_i - \mathbf{v}_k), \end{aligned}$$

where s is a constant chosen to balance s_k with $\mathbf{u}_k$ numerically. Then the DBSVC feature is obtained by

$$\mathbf{f}_{DBSVC} = [s_1, \mathbf{u}_1^{\mathrm{T}}, \ldots, s_K, \mathbf{u}_K^{\mathrm{T}}]^{\mathrm{T}}.$$

To diminish the sparseness, the DBSVC feature is normalized using the power and the ℓ_2 normalization.

We simply calculate the Euclidean distance for comparing DBSVC features between two 3D models.

In addition, ranking scores are calculated using our modified manifold ranking algorithm. We use a locally constrained diffusion process [YKTL09] for calculating the affinity matrix in the manifold ranking algorithm [ZWG*04], and call this method Locally Constrained Diffusion Ranking (LCDR).

6. Results

In conclusion, among the 5 participating groups, 2 groups (Aono and Tatsuma) employ a local shape descriptor, 2 groups (Chen and Li) adopt a global feature, and 1 group (Furuya) tests both local and local features. Two groups (Tatsuma and Furuya) that extract local features have applied the Bag-of-Words framework and K-means clustering on the local features, which shows the popularity of the Bag-of-Words technique in dealing with local features.

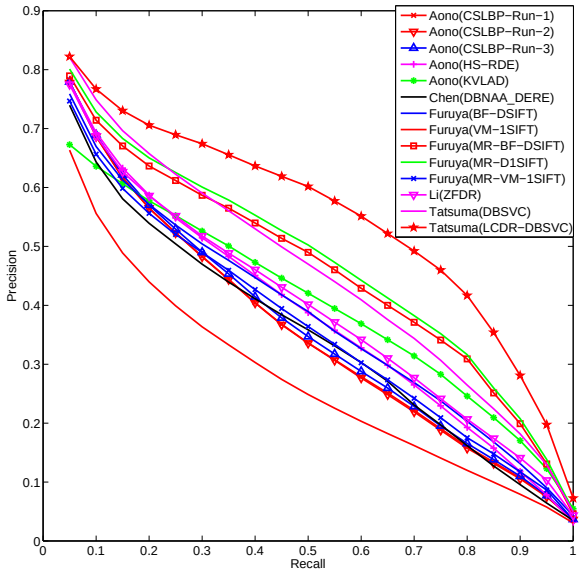


Figure 8: *Precision-Recall plot performance comparison of all the 14 runs of the 5 groups.*

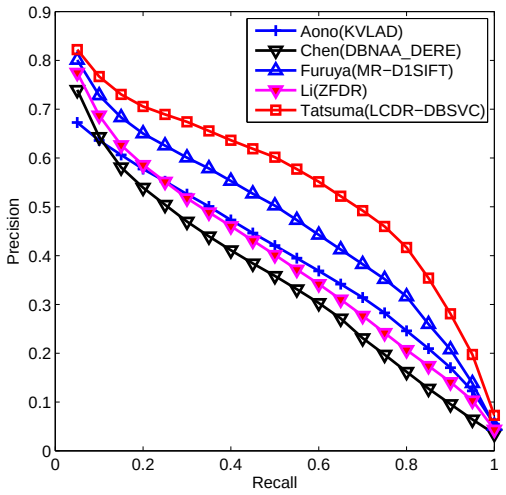


Figure 9: *Precision-Recall plot performance comparison of the best runs of each group.*

In this section, we perform a comparative evaluation of the results of the 14 runs submitted by the 5 groups. To have a comprehensive comparison, we measure the retrieval performance based on the 7 metrics mentioned in Section 3: PR, NN, FT, ST, E, DCG and AP, and the proportionally and reciprocally weighted NN, FT, ST, E, DCG.

Figure 8 shows the Precision-Recall performances of

Table 3: *Performance metrics for the performance comparison.*

Participant	Method	NN	FT	ST	E	DCG	AP
Aono	CSLBP-Run-1	0.840	0.353	0.452	0.197	0.736	0.349
	CSLBP-Run-2	0.842	0.352	0.450	0.197	0.735	0.347
	CSLBP-Run-3	0.840	0.359	0.459	0.200	0.740	0.355
	HSR-DE	0.837	0.381	0.490	0.203	0.752	0.378
	KVLAD	0.605	0.413	0.546	0.214	0.746	0.396
Chen	DBNAA_DERE	0.817	0.355	0.464	0.188	0.731	0.344
Furuya	BF-DSIFT	0.824	0.378	0.492	0.201	0.756	0.375
	VM-1SIFT	0.732	0.282	0.380	0.158	0.688	0.269
	MR-BF-DSIFT	0.845	0.455	0.567	0.229	0.784	0.453
	MR-D1SIFT	0.856	0.465	0.578	0.234	0.792	0.464
	MR-VM-1SIFT	0.812	0.368	0.467	0.194	0.737	0.357
Li	ZFDR	0.838	0.386	0.501	0.209	0.757	0.387
Tatsuma	DBSVC	**0.868**	0.438	0.563	0.234	0.790	0.446
	LCDR-DBSVC	0.864	**0.528**	**0.661**	**0.255**	**0.823**	**0.541**

Table 4: *Proportionally weighted performance metrics for the performance comparison.*

Participant	Method	NN	FT	ST	E	DCG
Aono	CSLBP-Run-1	175.324	75.507	100.098	28.904	159.435
	CSLBP-Run-2	175.631	74.830	98.638	28.868	159.036
	CSLBP-Run-3	175.073	75.915	100.682	29.115	159.862
	HSR-DE	175.864	80.634	107.474	29.509	161.943
	KVLAD	123.059	83.382	114.400	28.756	160.724
Chen	DBNAA_DERE	171.149	79.380	108.438	27.193	159.316
Furuya	BF-DSIFT	173.028	78.158	105.412	28.547	161.179
	VM-1SIFT	158.938	57.790	80.962	23.973	150.085
	MR-BF-DSIFT	174.762	92.451	120.921	31.160	166.318
	MR-D1SIFT	178.497	94.309	121.762	31.804	167.318
	MR-VM-1SIFT	172.998	77.332	99.818	28.245	158.999
Li	ZFDR	175.142	79.407	106.578	29.422	161.351
Tatsuma	DBSVC	**178.981**	88.434	120.341	32.321	167.176
	LCDR-DBSVC	177.863	**107.851**	**144.179**	**33.691**	**173.773**

all the 14 runs while Figure 9 compares the best runs of each group. Tables 3∼5 list the other 6 non-weighted and weighted performance metrics of all the 14 runs. As can be seen from Figure 9 and Tables 3∼5, Tatsuma's LCDR-DBSVC performs the best, followed by Furuya's MR-D1SIFT. From this result, we can say that using view-based features in combination with advanced feature coding and adaptive ranking provides best performance among the set of submitted methods. Usage of local visual features (as in LCDR-DBSVC) seems to work better, on average than coarser or global visual features (as in MR-D1SIFT).

As can be seen from Figure 8, if we compare approaches without employing a machine learning approach including Manifold Ranking, overall Li's ZFDR, Furuya's BF-DSIFT and Aono's HSR-DF are comparable to Tatsuma's DB-SVC approach. However, after applying a Manifold Ranking learning method, Tatsuma et al. have achieved an apparent performance improvement which can be seen by the resulting LCDR-DBSVC method. Compared to DBSVC, LCDR-DBSVC has a 20.6%, 17.4%, 9.0%, 4.2% and 21.3% gain in terms of non-weighted FT, ST, E, DCG and AP, respectively. In fact, Furuya et al.'s three "MR-" runs also have adopted

Table 5: *Reciprocally weighted performance metrics for the performance comparison.*

Participant	Method	NN	FT	ST	E	DCG
Aono	CSLBP-Run-1	4.504	2.060	2.437	1.224	3.877
	CSLBP-Run-2	4.538	2.065	2.442	1.226	3.879
	CSLBP-Run-3	4.471	2.103	2.480	1.245	3.894
	HSR-DE	4.459	2.160	2.584	1.281	3.956
	KVLAD	3.261	2.196	2.945	1.445	3.830
Chen	DBNAA_DERE	4.252	1.911	2.306	1.146	3.747
Furuya	BF-DSIFT	4.379	2.179	2.643	1.306	3.992
	VM-1SIFT	3.716	1.596	1.972	0.962	3.467
	MR-BF-DSIFT	4.622	2.551	3.016	1.504	4.206
	MR-D1SIFT	4.678	2.604	3.090	1.542	4.263
	MR-VM-1SIFT	4.256	2.037	2.442	1.215	3.829
Li	ZFDR	4.476	2.216	2.661	1.321	3.994
Tatsuma	DBSVC	4.803	2.523	3.022	1.523	4.269
	LCDR-DBSVC	**4.881**	**2.905**	**3.435**	**1.731**	**4.470**

Manifold Ranking method to improve the retrieval performance. This indicates the advantage of employing machine learning approach in the 3D model retrieval research field. We need to mention that the above finding is consistent with three types of metrics, either standard, or proportionally or reciprocally weighted ones.

7. Conclusions

In this paper, we first present the motivation of the organization of this comprehensive 3D shape retrieval track and then introduce the data collection process. Next, we briefly introduce our evaluation method, followed by the short descriptions of the 5 methods (14 runs) submitted by the 5 groups. Finally, a comprehensive evaluation has been conducted in terms of 7 different performance metrics, with or without proportional or reciprocal weights based on model variations in each class. According to the track, Manifold Ranking learning method and Bag-of-Words approach are two popular and promising techniques in comprehensive 3D shape retrieval, which shows a current research trend in the field of comprehensive 3D model retrieval.

Acknowledgments

This work of Bo Li and Yijuan Lu is supported by the Texas State University Research Enhancement Program (REP), Army Research Office grant W911NF-12-1-0057, and NSF CRI 1058724 to Dr. Yijuan Lu.

Henry Johan is supported by Fraunhofer IDM@NTU, which is funded by the National Research Foundation (NRF) and managed through the multi-agency Interactive & Digital Media Programme Office (IDMPO) hosted by the Media Development Authority of Singapore (MDA).

We would like to thank Yuxiang Ye and Natacha Feola who helped us to build the benchmark.

We would like to thank Mathias Eitz, James Hays and Marc Alexa who collected the 250 classes of sketches. We would also like to thank following authors for building the 3D benchmarks:

- Philip Shilane, Patrick Min, Michael M. Kazhdan, Thomas A. Funkhouser who built the Princeton Shape Benchmark (**PSB**);
- Atsushi Tatsuma, Hitoshi Koyanagi, Masaki Aono who built the Toyohashi Shape Benchmark (**TSB**);
- Dejan Vranic and colleagues who built the Konstanz 3D Model Benchmark (**CCCC**);
- Daniela Giorgi who built the Watertight Shape Benchmark (**WMB**);
- Kaleem Siddiqi, Juan Zhang, Diego Macrini, Ali Shokoufandeh, Sylvain Bouix, Sven Dickinson who built the McGill 3D Shape Benchmark (**MSB**);
- Raoul Wessel, Ina Blümel, Reinhard Klein, University of Bonn, who built the Bonn Architecture Benchmark (**BAB**);
- Subramaniam Jayanti, Yagnanarayanan Kalyanaraman, Natraj Iyer, Karthik Ramani who built the Engineering Shape Benchmark (**ESB**).

References

[ABD12] ALCANTARILLA P. F., BARTOLI A., DAVISON A. J.: KAZE features. In *ECCV (6)* (2012), Fitzgibbon A. W., Lazebnik S., Perona P., Sato Y., Schmid C., (Eds.), vol. 7577 of *Lecture Notes in Computer Science*, Springer, pp. 214–227. 3

[AKT13] AONO M., KOYANAGI H., TATSUMA A.: 3D shape retrieval focused on holes and surface roughness. In *Proc. of 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)* (2013), pp. 1–8. 3

[BBGO11] BRONSTEIN A. M., BRONSTEIN M. M., GUIBAS L. J., OVSJANIKOV M.: Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics 30*, 1 (Feb. 2011), 1–20. 6

[BETVG08] BAY H., ESS A., TUYTELAARS T., VAN GOOL L.: Speeded-up robust features (SURF). *Computer Vision Image Understanding 110*, 3 (June 2008), 346–359. 6

[CFYT14] CHEN Q., FANG B., YU Y.-M., TANG Y.: 3D cad model retrieval based on the combination of features. *Multimedia Tools and Applications* (2014), 1–19. 4

[CPW98] CHEN C. H., PAU L. F., WANG P. S. P.: *Handbook of Pattern Recognition and Computer Vision (2nd Edition)*. World Scientific Publishing Co., Inc., 1998. 3

[CTSO03] CHEN D.-Y., TIAN X.-P., SHEN Y.-T., OUHYOUNG M.: On visual similarity based 3D model retrieval. *Computer Graphics Forum 22*, 3 (2003), 223–232. 7

[EHA12] EITZ M., HAYS J., ALEXA M.: How do humans sketch objects? *ACM Trans. Graph. 31*, 4 (2012), 44:1–44:10. 1, 2

[FO09] FURUYA T., OHBUCHI R.: Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features. In *Proc. of the ACM International Conference on Image and Video Retrieval* (2009), CIVR '09, ACM, pp. 26:1–26:8. 5, 6

[GEW06] GEURTS P., ERNST D., WEHENKEL L.: Extremely randomized trees. *Machine Learning 63*, 1 (2006), 3–42. 5

[HPS06] HEIKKILÄ M., PIETIKÄINEN M., SCHMID C.: Description of interest regions with center-symmetric local binary patterns. In *ICVGIP* (2006), Kalra P. K., Peleg S., (Eds.), vol. 4338 of *Lecture Notes in Computer Science*, Springer, pp. 58–69. 3

[JDSP10] JEGOU H., DOUZE M., SCHMID C., PÉREZ P.: Aggregating local descriptors into a compact image representation. In *CVPR* (2010), IEEE, pp. 3304–3311. 3

[JKIR06] JAYANTI S., KALYANARAMAN Y., IYER N., RAMANI K.: Developing an engineering shape benchmark for cad models. *Computer-Aided Design 38*, 9 (2006), 939–953. 2

[KH90] KHOTANZAD A., HONG Y.: Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence 12*, 5 (1990), 489–497. 6

[LGA*12] LI B., GODIL A., AONO M., BAI X., FURUYA T., LI L., LÓPEZ-SASTRE R. J., JOHAN H., OHBUCHI R., REDONDO-CABRERA C., TATSUMA A., YANAGIMACHI T., ZHANG S.: SHREC'12 track: Generic 3D shape retrieval. In *3DOR* (2012), Spagnuolo M., Bronstein M. M., Bronstein A. M., Ferreira A., (Eds.), Eurographics Association, pp. 119–126. 2, 4

[LJ13] LI B., JOHAN H.: 3D model retrieval using hybrid features and class information. *Multimedia Tools Appl. 62*, 3 (2013), 821–846. 4, 6

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110. 5

[NS06] NISTER D., STEWENIUS H.: Scalable recognition with a vocabulary tree. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Washington, DC, USA, 2006), vol. 2 of *CVPR '06*, IEEE Computer Society, pp. 2161–2168. 7

[OF10] OHBUCHI R., FURUYA T.: Distance metric learning and feature combination for shape-based 3D model retrieval. In *Proceedings of the ACM workshop on 3D object retrieval* (2010), 3DOR '10, ACM, pp. 63–68. 4, 5

[OFCD02] OSADA R., FUNKHOUSER T. A., CHAZELLE B., DOBKIN D. P.: Shape distributions. *ACM Trans. Graph. 21*, 4 (2002), 807–832. 4

[PSM10] PERRONNIN F., SÁNCHEZ J., MENSINK T.: Improving the fisher kernel for large-scale image classification. In *Proceedings of the 11th European Conference on Computer Vision: Part IV* (Berlin, Heidelberg, 2010), ECCV '10, Springer-Verlag, pp. 143–156. 7

[SMKF04] SHILANE P., MIN P., KAZHDAN M. M., FUNKHOUSER T. A.: The Princeton shape benchmark. In *SMI* (2004), pp. 167–178. 2

[SPdC12] SÁNCHEZ J., PERRONNIN F., DE CAMPOS T.: Modeling the spatial layout of images beyond spatial pyramids. *Pattern Recognition Letters 33*, 16 (2012), 2216–2223. 7

[SZM*08] SIDDIQI K., ZHANG J., MACRINI D., SHOKOUFANDEH A., BOUIX S., DICKINSON S. J.: Retrieving articulated 3-D models using medial surfaces. *Mach. Vis. Appl. 19*, 4 (2008), 261–275. 2

[TA09] TATSUMA A., AONO M.: Multi-Fourier spectra descriptor and augmentation with spectral clustering for 3D shape retrieval. *Vis. Comput. 25* (2009), 785–804. 3, 6

[TKA12] TATSUMA A., KOYANAGI H., AONO M.: A large-scale shape benchmark for 3D object retrieval: Toyohashi Shape Benchmark. In *Proc. of 2012 Asia Pacific Signal and Information Processing Association (APSIPA2012)* (2012). 2

[Vra04] VRANIC D.: *3D Model Retrieval*. PhD thesis, University of Leipzig, 2004. 2, 4, 6

[VtH07] VELTKAMP R. C., TER HAAR F. B.: *SHREC 2007 3D Retrieval Contest*. Technical Report UU-CS-2007-015, Department of Information and Computing Sciences, Utrecht University, 2007. 2

[WBK09] WESSEL R., BLÜMEL I., KLEIN R.: A 3D shape benchmark for retrieval and automatic classification of architectural data. In *3DOR* (2009), Spagnuolo M., Pratikakis I., Veltkamp R. C., Theoharis T., (Eds.), Eurographics Association, pp. 53–56. 2

[YKTL09] YANG X., KÖKNAR-TEZEL S., LATECKI L. J.: Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition* (June 2009), pp. 357–364. 7

[ZBL*03] ZHOU D., BOUSQUET O., LAL T. N., WESTON J., SCHÖLKOPF B.: Learning with local and global consistency. In *NIPS* (2003), Thrun S., Saul L. K., Schölkopf B., (Eds.), MIT Press. 5

[ZL01] ZHANG D., LUO G.: A comparative study on shape retrieval using Fourier Descriptors with different shape signatures. In *Proc. of International Conference on Intelligent Multimedia and Distance Education (ICIMADE01)* (2001), pp. 1–9. 6

[ZWG*04] ZHOU D., WESTON J., GRETTON A., BOUSQUET O., SCHÖLKOPF B.: Ranking on data manifolds. In *Advances in Neural Information Processing Systems 16*, Thrun S., Saul L., Schölkopf B., (Eds.). MIT Press, Cambridge, MA, 2004. 7

[ZYZH10] ZHOU X., YU K., ZHANG T., HUANG T.: Image classification using super-vector coding of local image descriptors. In *Computer Vision - ECCV 2010*, Daniilidis K., Maragos P., Paragios N., (Eds.), vol. 6315 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2010, pp. 141–154. 6