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

SHREC'12 Track: Generic 3D Shape Retrieval

B. Li[1,6], A. Godil[1], M. Aono[2], X. Bai[3], T. Furuya[4], L. Li[3], R. López-Sastre[5], H. Johan[6], R. Ohbuchi[7], C. Redondo-Cabrera[5], A. Tatsuma[2], T. Yanagimachi[7], S. Zhang[3]

[1] National Institute of Standards and Technology, Gaithersburg, USA [2] Toyohashi University of Technology, Japan
[3] Northwestern Polytechnical University, Xi'an, China [4] Nisca Corp., Yamanashi, Japan
[5] Department of Signal Theory and Communications, University of Alcalá, Spain
[6] School of Computer Engineering, Nanyang Technological University, Singapore [7] University of Yamanashi, Yamanashi, Japan

Abstract

Generic 3D shape retrieval is a fundamental research area in the field of content-based 3D model retrieval. The aim of this track is to measure and compare the performance of generic 3D shape retrieval methods implemented by different participants over the world. The track is based on a new generic 3D shape benchmark, which contains 1200 triangle meshes that are equally classified into 60 categories. In this track, 16 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

Figure 1: *Examples of generic 3D models.*

Generic 3D model retrieval is a fundamental research direction in the field of 3D model retrieval. Generic models include the 3D objects that we often see in our common life. Some examples are shown in Figure 1. Compared to professional models, generic models usually have more variations, for example, a chair can have diverse shapes and it may have a wheel or not; a table can be round or rectangular; an insect may be able to fly or does not have wings. Thus, it is non-trivial to classify different models within one class into a semantic group. On the other hand, generic models represent 3D objects that are utmost important to us and there are a lot of needs to retrieve such models.

Several generic 3D shape benchmarks have been built, such as Princeton Shape Benchmark (**PSB**) [SMKF04], National Taiwan University database (**NTU**) [CTSO03], Konstanze 3D Model Benchmark (**CCCC**) [Vra04], NIST Generic Shape Benchmark (**NSB**) [FGLW08], SHREC'10 Generic 3D Warehouse [PGD*10], and SHREC'11 Generic 3D Benchmark [DGD*11].

However, there are still some facets that can be further improved. For example, for some databases different classes have different number of models, which is a bias for retrieval; for other datasets some important types of models are not included. In fact, there are apparent overlapping in terms of classes among the aforementioned datasets and we can utilize this to merge and select the models in the same class to build a class for a new generic 3D model dataset. Hence, in order to create a more comprehensive benchmark and overcome some shortcomings of previous generic datasets, we built a new dataset named SHREC'12 Generic 3D Benchmark based on the above mentioned benchmarks. We also randomly assigned the index number of each model and developed evaluation code specially for the new benchmark. In this paper, we reported the results of five 3D retrieval algorithms tested in the generic shape retrieval track of SHREC 2012, held in conjunction with the fifth Eurographics Workshop on 3D Object Retrieval (EG3DOR'12).

2. Data Collection

The dataset comprises 1200 models, divided into 60 classes, with 20 models each. In order to build a more comprehen-

DOI: 10.2312/3DOR/3DOR12/119-126

Table 1: *60 classes of the SHREC'12 Generic 3D Shape Benchmark.*

Bird	Fish	NonFlyingInsect
FlyingInsect	Biped	Quadruped
ApartmentHouse	Skyscraper	SingleHouse
Bottle	Cup	Glasses
HandGun	SubmachineGun	Guitar
Mug	FloorLamp	DeskLamp
Sword	Cellphone	DeskPhone
Monitor	Bed	NonWheelChair
WheelChair	Sofa	RectangleTable
RoundTable	Bookshelf	HomePlant
Tree	Biplane	Helicopter
Monoplane	Rocket	Ship
Motorcycle	Car	MilitaryVehicle
Bicycle	Bus	ClassicPiano
Drum	HumanHead	ComputerKeyboard
TruckNonContainer	PianoBoard	Spoon
Truck	Violin	Bookset
Knife	Train	Plier
Chess	City	Computer
Door	Face	Hand

sive generic 3D benchmark with more diversity, we combine and merge models based on several previous SHREC generic 3D benchmarks. In detail, we create our SHREC'12 Generic 3D Benchmark mainly based on four datasets: SHREC'11 Generic 3D Benchmark [DGD*11], SHREC'10 Generic 3D Warehouse [PGD*10], Princeton Shape Benchmark [SMKF04] and SHREC'07 Watertight Shape Benchmark [VtH07].

Firstly, we utilize the 1000 models of SHREC'11 Generic 3D Benchmark [DGD*11], which contains the 800 models of the NIST Generic Shape Benchmark (**NSB**) [FGLW08] and the selected 200 models from the SHREC'10 Generic 3D Warehouse [PGD*10]. Secondly, we select three additional classes from the SHREC'10 Generic 3D Warehouse [PGD*10], which are "Bookset", "Knife" and "Train". Thirdly, we add six new classes from the Princeton Shape Benchmark (**PSB**) [SMKF04], and they are "chess", "city", "computer", "door", "face" and "hand". To make each class has 20 models, we also select some models from the Konstanze 3D Model Benchmark (**CCCC**) [Vra04] and National Taiwan University database (**NTU**) [CTSO03]. Finally, we also include the "plier" class of the SHREC'07 Watertight Shape Benchmark (**WSB**) [VtH07] into our dataset. The file format to represent the 3D models is the ASCII Object File Format (*.off). Table 1 lists all the 60 classes in the SHREC'12 Generic 3D Benchmark.

3. Evaluation

The participants submitted a 1200×1200 distance matrix per method/run. The matrix gives the pairwise dissimilarity values of all the possible model pairs in the dataset. Using the dissimilarity matrices provided by the participants, we performed our evaluations based on seven standard metrics which are widely used by 3D model retrieval community: Precision-Recall curve (*PR*), Nearest Neighbor (*NN*), First-Tier (*FT*), Second-Tier (*ST*), E-Measure (*E*), Discounted Cumulative Gain (*DCG*) [SMKF04] and Average Precision (*AP*).

4. Participants

There are 5 groups who have successfully participated in the SHREC'12 Generic 3D Shape Retrieval track. In total, they have submitted 16 dissimilarity matrices. The details about the participants and their runs are as follows.

- *LSD-r02*, *LSD-r03*, *LSD-r08* and *LSD-sum* submitted by Xiaoliang Bai, Liang Li and Shusheng Zhang from Northwestern Polytechnical University, China
- *ZFDR* submitted by Bo Li and Henry Johan from Nanyang Technological University, Singapore
- *3DSP_L2_200_hik*, *3DSP_L2_1000_hik*, *3DSP_L3_200_hik*, *3DSP_L2_200_chi2* and *3DSP_L2_1000_chi2* submitted by Carolina Redondo-Cabrera and Roberto Javier López-Sastre from University of Alcalá, Spain
- *DVD*, *DVD+DB* and *DVD+DB+GMR* submitted by Atsushi Tatsuma and Masaki Aono from Toyohashi University of Technology, Japan
- *DSIFT*, *DGSIFT* and *DG1SIFT* submitted by Tomohiro Yanagimachi, Takahiko Furuya and Ryutarou Ohbuchi from University of Yamanashi and Nisca Corp, Japan

5. Methods

5.1. 3D Model Retrieval Using Local Shape Distributions, by X. Bai, L. LI and S. Zhang

The idea of the proposed method is to perform 3D model retrieval using the local shape features. The method is inspired by the Bag of Geodesic Histograms (BOGH) algorithm [LGB*11]. However, the method follows a different strategy by suggesting the new Local Shape Distribution (LSD) descriptor as the shape representation for 3D objects.

5.1.1. Local Shape Distribution Descriptor

Let P denotes a surface point of a 3D object. Its r-neighborhood is defined as the spherical region centered at P with the radius r. The LSD descriptor associated to this region is a histogram-vector of the Euclidean-distances between P and other surface points within the region. Since all the points in the r-neighborhood of p have their own contributions to the local shape of the 3D object in this region, and such contributions are decreased with the increase of distances between the points and the center of the region, i.e. point P, each bin of the LSD histogram is Gaussian weighted ($\sigma = 0.3$), with an attempt to accurately indicate the shape distribution in the region.

5.1.2. Feature Extraction

The proposed method starts feature extraction by randomly sampling n points on the surface of a 3D object. It assumes that the scale normalization on the model has been conducted beforehand. For each sample point, the LSD descriptor of its r-neighborhood is computed, which is composed of d bins (d = 32). After that, the k-means algorithm is employed to carry out clustering on the resulting n LSD descriptors. The aim of this step is to select those characteristic descriptors, i.e. the centers of k clusters, in order to improve the speed of similarity matching. By this way, the 3D model is represented by a set of k LSD descriptors. In this track, n and k are set to 3000 and 200 respectively. The feature extraction process is shown in Figure 2.

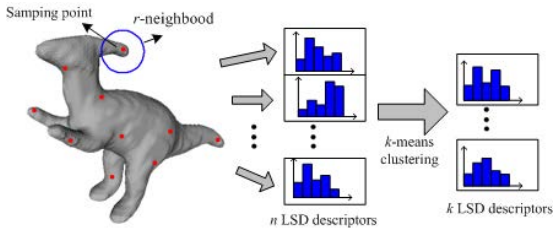


Figure 2: *Feature extraction process.*

5.1.3. Similarity Matching

The similarity matching step of the proposed method is analogous to that of BOGH. Let L_Q and L_C denote the LSD descriptor sets of a query and a target 3D model respectively. The Hungarian algorithm [Kuh10] is employed to establish the correspondence between L_Q and L_C. The dissimilarity between two descriptors is measured by the χ^2 distance.

5.1.4. Parameters Settings

Three configurations of the proposed method (r = 0.2, 0.3 and 0.8) were chosen to calculate the dissimilarity matrices (LSD-r02, LSD-r03 and LSD-r08). A version (LSD-sum) combining the above resulting matrices under the sum rule was also presented. Software to compute LSD descriptor is freely available at [BLZ12].

5.2. Hybrid Shape Descriptor ZFDR, by B. Li and H. Johan [LJ11]

The hybrid shape descriptor ZFDR [LJ11], containing both visual and geometric information of a 3D model, is composed of four parts: **Z**ernike moments feature, **F**ourier descriptor feature, **D**epth information feature and **R**ay-based feature. The shape descriptor computation process consists of two steps: 3D model normalization and feature extraction, as graphically shown in Figure 3. Continuous Principle Component Analysis (CPCA) [Vra04] alignment algorithm is utilized during the normalization step to align the 3D

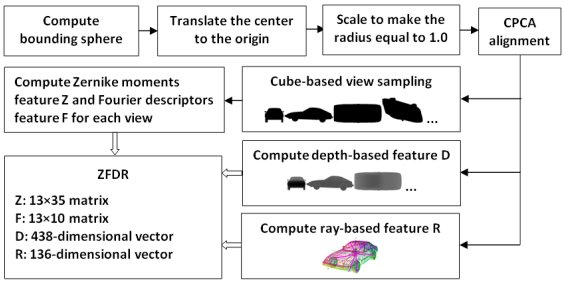


Figure 3: *ZFDR feature extraction process [LJ11].*

model. The details about the feature extraction are described as follows.

An image descriptor [ZL02], which comprises Zernike moments and Fourier descriptors, is adopted to represent the features of a silhouette view. These two features characterize the visual information of a 3D model. They are effective in representing certain types of models (e.g. "sea animal"), while not as effective as depth buffer-based features for some other classes (like "car") [BKS*04]. Therefore, we devise a hybrid shape descriptor by also integrating certain geometric information of a 3D model. Particularly, we integrate the depth buffer-based feature and ray-based with spherical harmonic representation feature developed by Vranic [Vra04] into our hybrid shape descriptor.

1) Cube-Based View Sampling: Considering the tradeoff between the feature extraction time and retrieval efficiency, the approach samples 13 silhouette views to represent a 3D model by setting cameras at the following locations on a cube: (1,0,0), (0,1,0), (0,0,1), (1,1,1), (-1,1,1), (-1,-1,1), (1,-1,1), (1,0,-1), (0,1,-1), (1,1,0), (0,1,1), (1,0,1), (1,-1,0).

2) Zernike Moments Feature (Z): Zernike moments feature is utilized to extract the region-based features of a silhouette view. For each sample view, the algorithm computes the Zernike moments [KH90] (up to the 10^{th} order, totally 35 moments). Then, the Zernike moments features are concatenated orderly according to the order of the view sequence, resulting a 13×35 matrix as the Zernike moments feature of a 3D model.

3) Fourier Descriptor Feature (F): Centroid distance-based Fourier descriptor [ZL01] is adopted to extract the contour feature of a silhouette view. The first 10 Fourier coefficients are used as the Fourier descriptor of a view. The combination of the Fourier descriptors of 13 views forms a 13×10 matrix as the Fourier descriptor feature of a 3D model.

4) Depth Information Feature (D): This feature is to extract the Fourier features of the six depth buffer images of a 3D model. First, the six depth views of a 3D model are rendered, then 2D Fourier Transform is applied on the depth

views, and finally 438 Fourier coefficients are employed as the depth features of a 3D model.

5) Ray-Based Feature (R): First, a set of rays emanating from the center of the model are shoot and based on the outmost intersections between the rays and model, the ray-based feature vector in the spatial domain is extracted. Then, Spherical Harmonics Transform [KFR03] is applied on the obtained radial distance feature vector to transform it from the spatial domain to the spectral domain. Finally, a 136-dimensional feature vector is obtained to depict the ray-based features of a 3D model.

6) Hybrid Shape Descriptor ZFDR Distance Computation: The hybrid shape descriptor ZFDR of a model is a combination of Zernike moments feature Z, Fourier descriptor F, Depth information feature D and Ray-based descriptor R. First, appropriate distance metrics are assigned to measure the component distances d_Z, d_F, d_D and d_R between two models, then the four component distances are linearly combined to form the hybrid descriptor distance d_{ZFDR}. For more details about the shape descriptor computation, please refer to [LJ11].

5.3. 3D Shape Recognition via 3D Spatial Pyramids, by C. Redondo-Cabrera and R. López-Sastre

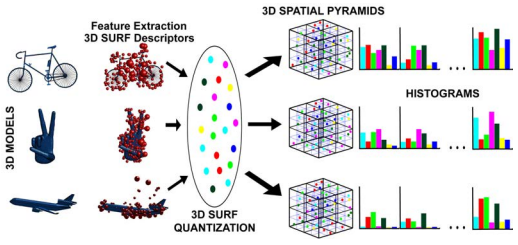


Figure 4: *3D Spatial Pyramid Model.*

The 3D Spatial Pyramid (3DSP) is a method for 3D shape recognition inspired by the work [LSP06, KPW*10]. The approach is shown in Figure 4. It starts from a 3D shape of the object of interest. Each shape is characterized by a set of 3D SURF local descriptors [KPW*10]. In contrast to a random or dense coverage of the shape with spin images [JH99], the 3D SURF is equipped with a 3D interest point detector, which picks out a repeatable and salient set of interest points in the shapes. The local 3D SURF descriptors are computed in these points. Then, by following a traditional *Bag of Words* approach, it quantizes these 3D descriptors, into 3D visual words. Finally, each 3D shape can be characterized by a histogram of its 3D visual words.

In the approach, it proposes to adapt the Spatial Pyramid Matching Kernel [LSP06] to work with 3D shapes. It models a 3D shape by an orderless set of 3D visual words. That is, if we define a visual codebook of size K, each 3D feature is associated to a codebook label $\{1,\ldots,K\}$. However, the 3DSP method should be able to capture the spatial distribution of such labels at different scales and locations in a working volume $\Omega^{(0)}$. Similar to [LSP06], but now in 3D, it defines a pyramid structure by partitioning $\Omega^{(0)}$ into fine sub-cubes (see Figure 4). For each level l of the pyramid, the volume of the previous level, $\Omega^{(l-1)}$, is decomposed into eight sub-cubes, hence a pyramid $P(L)$ of L levels contains $D = 8^L$ sub-cubes.

Once the pyramid $P(L)$ is composed, the method proposes to characterize each 3D shape $\mathcal{S}$ by a weighted ensemble of histograms $H(\mathcal{S}) = \left[\omega_0 H^0(\mathcal{S}), \omega_1 H^1(\mathcal{S}), ..., \omega_L H^L(\mathcal{S})\right]$, where $H^l(\mathcal{S})$ is the histogram of the features in the level l of the pyramid. Each $H^l(\mathcal{S})$ is obtained by concatenating 8^l histograms computed in all of the 8^l sub-cubes for level l. In order to penalize the future matches (between 3D shapes) found in larger volumes, it defines the weight ω_l as $\frac{1}{2^{L-l}}$.

Given two 3D shapes $\mathcal{S}_X$ and $\mathcal{S}_Y$, the algorithm first computes their corresponding 3DSP representations $H(\mathcal{S}_X)$ and $H(\mathcal{S}_Y)$. In order to compute the dissimilarity of these 3D shapes, it proposes two distance functions. First, it proposes the HIK^*, the histogram intersection kernel including a small modification for normalization. The *distance* between two histograms $\mathcal{D}(H(\mathcal{S}_X), H(\mathcal{S}_Y))_{HIK^*}$ is defined as,

$$\mathcal{D}(H(\mathcal{S}_X), H(\mathcal{S}_Y))_{HIK^*} = 1 - \frac{\sum_{i=1}^{N} \min(H(\mathcal{S}_X)_i, H(\mathcal{S}_Y)_i)}{\max(\sum_{i=1}^{N} H(\mathcal{S}_X)_i, \sum_{i=1}^{N} H(\mathcal{S}_Y)_i)}, \quad (1)$$

where N is the number of components of histograms $H(\mathcal{S}_X)$ and $H(\mathcal{S}_Y)$.

Second, it proposes to use the χ^2 distance. The distance measures the χ^2 dissimilarity between two histograms $\mathcal{D}(H(\mathcal{S}_X), H(\mathcal{S}_Y))_{\chi^2}$ as,

$$\mathcal{D}(H(\mathcal{S}_X), H(\mathcal{S}_Y))_{\chi^2} = \frac{1}{2} \sum_{i=1}^{N} \frac{(H(\mathcal{S}_X)_i - H(\mathcal{S}_Y)_i)^2}{H(\mathcal{S}_X)_i + H(\mathcal{S}_Y)_i}. \quad (2)$$

For the track, it submitted 5 different runs of the 3DSP methods, using the following parameters: visual vocabulary size $K = 200, 1000, 2000$; pyramid levels $L = 0, 1, 2, 3$.

5.4. Dense Voxel Spectrum Descriptor and Globally Enhanced Manifold Ranking, by A. Tatsuma and M. Aono

The approach proposes a novel 3D shape feature called Dense Voxel Spectrum Descriptor (DVD) that aims to capture 3D spatial information. 3D spatial information is the information that describes how each piece of a 3D geometric shape occupies which location of a 3D volumetric space. It also proposes a novel Manifold Ranking algorithm that grasps both local and global structures in feature space.

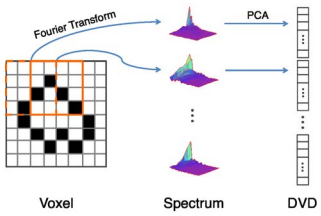


Figure 5: *Dense Voxel spectrum Descriptor (DVD).*

5.4.1. Dense Voxel Spectrum Descriptor

The overview of how the method defines the proposed DVD feature is illustrated in Figure 5. Voxels are represented as 2D pixels for simplicity in the figure. The essential idea of DVD is in that Fourier spectrum related to 3D voxels should not be computed directly from the entire voxel. Instead it cuts off voxels by a fixed block size, and applies Fourier transform to each block in order to obtain DVD feature.

DVD is, however, sensitive to the position, size, and orientation of the 3D object. To circumvent this, the method has employed a couple of pose normalizations developed by the authors previously: PointSVD and NormalSVD [TA09].

After pose normalization, it computes voxels. To do this, it generates random points on the surface and quantizes them into voxel space of size $64 \times 64 \times 64$. The non-empty voxel has a distance from the center of voxel space.

The voxel produced as above is cut off by a fixed block size with a fixed sliding (overlapping) window, and the Fourier spectrum is computed for each fixed block. The method adopts 32 as the size for a block, and 16 as the size for a sliding window. It should be noted that the method only uses lower frequencies of Fourier spectrum in the range $1 \leq x,y,z \leq 8$, because higher frequencies tend to have more noises. Finally, the net Fourier spectra are normalized with L1 norm.

DVD consists of concatenated Fourier spectra assembled from a block of voxels. This definition of DVD results in a high dimensional feature vector. Specifically, with the above-mentioned adopted sizes, the total dimension of DVD becomes $((64-32)/16+1)^3 \times 8^3 = 13,824$. The approach reduces the dimension from $8^3 = 512$ for lower frequencies down to 20 by applying Principal Component Analysis (PCA). The total reduced dimension amounts to $((64-32)/16+1)^3 \times 20 = 540$. Normalization with L1 norm is applied to this reduced dimensional feature vector.

The method employs a Manhattan distance for dissimilarity of DVD between two 3D objects. In addition to DVD, it uses a Depth-buffer shape descriptor [TA09] to make a composite feature vector.

5.4.2. Globally Enhanced Manifold Ranking

Recently, several methods, originated from Manifold Ranking (MR) [ZWG*04], have been reported to achieve high search performance by considering local structures in feature space, and having them reflected to ranking scores [DGD*11]. Considering only local structures in manifold learning, however, might cause some problems [BM05]. Thus, a new method has been developed for computing ranking scores by taking both local and global structures into account, which is called Globally enhanced Manifold Ranking (GMR).

In MR, a neighborhood graph is first generated together with all data including a given search query, then affinity matrix W is computed by using Gaussian kernel as their weights. Subsequently, affinity matrix W is normalized by diagonal matrix $D_{ii} = \sum_j W_{ij}$, to obtain a matrix $S = D^{-1/2}WD^{-1/2}$. Given a column vector $\mathbf{q}$ with 1s for query element and 0s otherwise, the ranking function f in MR is represented as follows:

$$f = (I - \alpha S)^{-1}\mathbf{q} \tag{3}$$

Zhou et al. [ZBS11] also generalized MR by using a graph Laplacian. They derived a ranking function f by using Green's function of an iterated graph Laplacian L as below.

$$f = \{(\beta I + L)^{-1}\}^m \mathbf{q} \tag{4}$$

Ranking scores with MR have relied only on local structures represented by a neighborhood graph. In the research on robust Laplacian Eigenmaps, Roychowdhury et al. [RG09] demonstrated that it was possible to capture global structure by using Minimum Spanning Tree (MST). Their proposed GMR also takes advantage of MST to capture global structures. In GMR, they first generate a neighborhood graph as well as an MST from all data including a given search query. They then compute affinity matrices W_{NN} for a neighborhood graph and W_{MST} for an MST, respectively. Since a composite of graphs can be done by adding affinity matrices, the graph Laplacian of the GMR can be expressed by a simple weighted addition.

$$L_{GMR} = L_{NN} + \lambda L_{MST} = (D_{NN} - W_{NN}) + \lambda(D_{MST} - W_{MST}) \tag{5}$$

From Equations (4) and (5), the ranking function of GMR is finally reduced to the following equation.

$$f = \{(\beta I + L_{GMR})^{-1}\}^m \mathbf{q} \tag{6}$$

5.5. Sum of Visual Distances for Generic 3D Model Retrieval, by T. Yanagimachi, T. Furuya, R. Ohbuchi [OF10]

The method is essentially the one described in [OF10] and an overview of the algorithm is shown in Figure 6. It involves multi-viewpoint rendering, Bag-of-Features (Bag-of-Words) integration of thousands of local visual features per

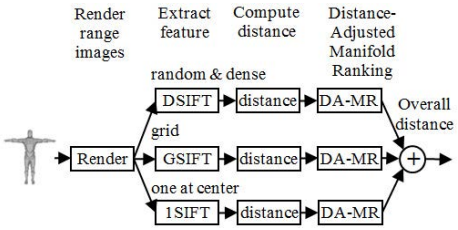
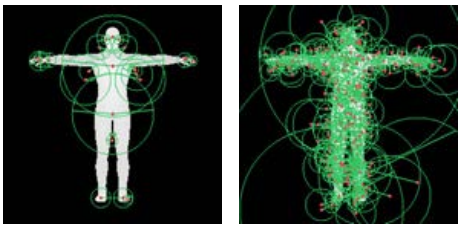


Figure 6: *Three distances computed using three visual features are summed, after normalization, to yield an overall distance.*

3D model, distance metric learning, and heterogeneous feature combination. A nice property of such an appearance-based approach is that 3D models in almost any shape representation can be compared.

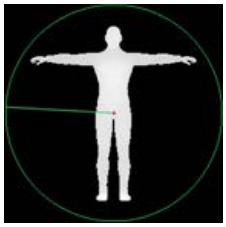
The method first renders a 3D model into range images from multiple (in this case 42) viewpoints. It then extracts visual features from the images (Figure 7) for comparison. The method extracts two sets of local features, Dense SIFT, or DSIFT [FO09], and Grid SIFT, or GSIFT [OF10] per range image. In addition, it extracts a global feature One SIFT, or 1SIFT, per range image. The three variations of the methods are named after the features involved. For example, the best performing DG1SIFT include all the three, while DGSIFT includes DSIFT and GSIFT. A global visual feature is added since this 3D Generic track database appeared to consist almost entirely of rigid models.



(a) Salient SIFT (original SIFT) [Low04])

(b) Dense SIFT (DSIFT)

(c) Grid SIFT (GSIFT)



(d) One SIFT (1SIFT)

Figure 7: *The method combines multiple visual features per view.*

Both DSIFT and GSIFT sample each range image with hundreds of SIFT features [Low04]. The DSIFT employs random and dense sampling pattern with prior to concentrate samples on or near 3D model. The GSIFT employs a simple grid sampling pattern. Thousands to tens of thousands of DSIFT (or GSIFT) features are integrated into a feature vector per 3D model by using BoF approach. DSIFT or GSIFT combined with BoF integration assume invariance against articulation and global deformation. Global feature 1SIFT samples a range image at its center with a SIFT feature. According to the experiments for rigid 3D objects [OF10], 1SIFT performed comparably to LFD [CTSO03].

For DSIFT and GSIFT, a distance between a pair of 3D models is given simply by applying a distance metric, e.g., L1-norm. For 1SIFT, distance computation is a little bit more involved. Assume that there is a pair of 3D models, each having k 1SIFT features extracted from k viewpoints. Using 1SIFT, a distance among the two models is the minimum of distances among all the (k-1)k/2 distances among k 1SIFT features each. Comparing 3D models using 1SIFT is slower than using DSIFT or GSIFT with BoF integration.

To gain extra retrieval accuracy, distance metric learning is applied based on Manifold Ranking (MR) [ZBL*03] to each of three distances computed from the three visual features. The distance adjustment described in [OF10] is applied prior to MR so that MR is more effective for high dimensional (e.g., 30k dimension) feature per 3D model of DSIFT and GSIFT. Affinities or rank values resulted from applying MR to distances of three features (in the case of DG1SIFT) are normalized and then combined into an overall affinity value by summation with equal weight.

6. Results

In this section, we perform a comparative evaluation of the results of the 16 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*.

Figure 8 shows the Precision-Recall performances of all the 16 runs while Figure 9 compares the best runs of each group. Table 2 lists the other 6 performance metrics of all the 16 runs. As can be seen from Figure 9 and Table 2, Yanagimachi's DG1SIFT performs the best, followed by Tatsuma's DVD+DB+GMR and Li's ZFDR, in terms of groups.

As can be seen from Figure 8, overall, Li's ZFDR method is comparable to Tatsuma's DVD+DB approach. However, after applying an enhanced Manifold Ranking learning method, which considers both the local and global structures in feature space, Tatsum et al. have achieved an apparent performance improvement which can be seen by the resulting DVD+DB+GMR method. Compared to DVD+DB, DVD+DB+GMR has a 6.1% and 15.7% gain in DCG and AP, respectively. In fact, all the three runs proposed by Yanagimachi et al. also have adopted Manifold Ranking method to improve the retrieval performance. This indicates

the advantage of employing machine learning approach in the 3D model retrieval research field.

However, we can see that all the runs except two of Yanagimachi's (DGSIFT and DG1SIFT) have the precision of 1 when *recall* <= 0.05. Based on the submitted results of the DG1SIFT method, we have found that for 18 classes, especially for "Door", "SingleHouse", "Train", "Tree", "Truck" and "Violin", it cannot achieve a precision of 1 at *recall* = 0.05. This should be because of the distance adaption and re-ranking during the Manifold Ranking process in the DG1SIFT algorithm, which can be regarded as one disadvantage of this technique. But overall, it apparently improves the retrieval performance.

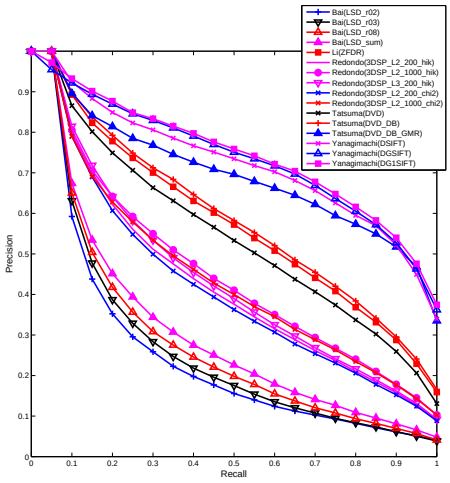


Figure 8: *Precision-Recall plot performance comparison of all the 16 runs of the 5 groups.*

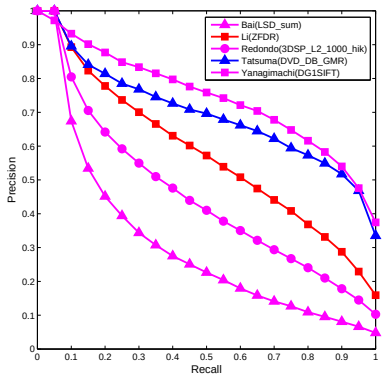


Figure 9: *Precision-Recall plot performance comparison of the best runs of each group.*

In conclusion, among the 5 methods submitted, 3 groups (Bai, Redondo and Yanagimachi) employ a local shape descriptor, 1 group (Tatsuma) adopts a global shape descriptor and the rest one (Li) uses both global and local features. Similarly, the three groups (Bai, Redondo and Yanagimachi) extracting local features have applied the Bag-of-Words framework or K-means clustering on the local features, which shows the popularity of the Bag-of-Words technique in dealing with local features. Finally, two groups utilize Manifold Ranking techniques.

7. Conclusions

In this paper, we first present the motivation of the organization of this generic 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 (16 runs) submitted by the 5 groups. Finally, a comprehensive evaluation has been conducted in terms of 7 different performance metrics. Based on the comparative evaluation, Yanagimachi's DG1SIFT method performs the best, followed by Tatsuma's DVD+DB+GMR method and Li's ZFDR approach in terms of groups. According to the track, Manifold Ranking learning method and Bag-of-Words approach are two popular and promising techniques in generic 3D shape retrieval, which shows current research trend in the field of generic 3D model retrieval.

References

[BKS*04] BUSTOS B., KEIM D. A., SAUPE D., SCHRECK T., VRANIC D. V.: Using entropy impurity for improved 3D object similarity search. In *ICME* (2004), pp. 1303–1306. 3

[BLZ12] BAI X., LI L., ZHANG S.: Software for 3D model retrieval using local shape distributions. http://code.google.com/p/shape-retrieval, Feb. 2012. 3

[BM05] BENGIO Y., MONPERRUS M.: Non-local manifold tangent learning. In *Advances in Neural Information Processing Systems 17* (2005), pp. 129–136. 5

[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. 1, 2, 6

[DGD*11] DUTAGACI H., GODIL A., DARAS P., AXENOPOULOS A., LITOS G. C., MANOLOPOULOU S., GOTO K., YANAGIMACHI T., KURITA Y., KAWAMURA S., FURUYA T., OHBUCHI R.: SHREC '11 track: Generic shape retrieval. In *3DOR* (2011), pp. 65–69. 1, 2, 5

[FGLW08] FANG R., GODIL A., LI X., WAGAN A.: A new shape benchmark for 3D object retrieval. In *ISVC (1)* (2008), pp. 381–392. 1, 2

[FO09] FURUYA T., OHBUCHI R.: Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features. In *CIVR* (2009). 6

[JH99] JOHNSON A., HEBERT M.: Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 5 (1999), 433–449. 4

[KFR03] KAZHDAN M. M., FUNKHOUSER T. A., RUSINKIEWICZ S.: Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Symposium on Geometry Processing* (2003), pp. 156–164. 4

Table 2: *Other Performance metrics for the performance comparison.*

Participant	Method	NN	FT	ST	E	DCG	AP
Bai	LSD-r02	0.431	0.179	0.257	0.175	0.510	0.326
Bai	LSD-r03	0.474	0.196	0.277	0.190	0.528	0.341
Bai	LSD-r08	0.489	0.212	0.301	0.207	0.545	0.359
Bai	LSD-sum	0.517	0.232	0.327	0.224	0.565	0.381
Li	ZFDR	0.818	0.491	0.621	0.442	0.776	0.650
Redondo	3DSP_L2_200_hik	0.671	0.349	0.470	0.329	0.669	0.501
Redondo	3DSP_L2_1000_hik	0.685	0.376	0.502	0.351	0.685	0.526
Redondo	3DSP_L3_200_hik	0.708	0.361	0.481	0.335	0.679	0.462
Redondo	3DSP_L2_200_chi2	0.667	0.341	0.462	0.322	0.662	0.442
Redondo	3DSP_L2_1000_chi2	0.662	0.367	0.496	0.346	0.678	0.468
Tatsuma	DVD	0.790	0.459	0.588	0.416	0.756	0.621
Tatsuma	DVD+DB	0.831	0.496	0.634	0.450	0.785	0.661
Tatsuma	DVD+DB+GMR	0.828	0.613	0.739	0.527	0.833	0.765
Yanagimachi	DSIFT	0.863	0.639	0.787	0.562	0.857	0.792
Yanagimachi	DGSIFT	0.860	0.654	0.794	0.572	0.866	0.803
Yanagimachi	DG1SIFT	0.879	0.661	0.799	0.576	0.871	0.811

[KH90] KHOTANZAD A., HONG Y.: Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence 12*, 5 (1990), 489–497. 3

[KPW*10] KNOPP J., PRASAD M., WILLEMS G., TIMOFTE R., VAN GOOL L.: Hough transform and 3D SURF for robust three dimensional classification. In *Proceedings of the 11th European Conference on Computer vision* (2010). 4

[Kuh10] KUHN H. W.: The Hungarian method for the assignment problem. In *50 Years of Integer Programming 1958-2008*. 2010, pp. 29–47. 3

[LGB*11] LIAN Z., GODIL A., BUSTOS B., DAOUDI M., HERMANS J., KAWAMURA S., KURITA Y., LAVOUÉ G., NGUYEN H. V., OHBUCHI R., OHKITA Y., OHISHI Y., PORIKLI F., REUTER M., SIPIRAN I., SMEETS D., SUETENS P., TABIA H., VANDERMEULEN D.: SHREC '11 track: Shape retrieval on non-rigid 3D watertight meshes. In *3DOR* (2011), pp. 79–88. 2

[LJ11] LI B., JOHAN H.: 3D model retrieval using hybrid features and class information. *Multimedia Tools and Applications* (2011), 1–26 (Online first version). 3, 4

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110. 6

[LSP06] LAZEBNIK S., SCHMID C., PONCE J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2006). 4

[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. 5, 6

[PGD*10] PORETHI V. T., GODIL A., DUTAGACI H., FURUYA T., LIAN Z., OHBUCHI R.: SHREC'10 track: Generic 3D warehouse. In *3DOR* (2010), pp. 93–100. 1, 2

[RG09] ROYCHOWDHURY S., GHOSH J.: Robust Laplacian eigenmaps using global information. In *Proc. of AAAI Fall Symp on Manifold Learning and Its Applications* (2009). 5

[SMKF04] SHILANE P., MIN P., KAZHDAN M. M., FUNKHOUSER T. A.: The Princeton shape benchmark. In *SMI* (2004), pp. 167–178. 1, 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. 5

[Vra04] VRANIC D.: *3D Model Retrieval*. PhD thesis, University of Leipzig, 2004. 1, 2, 3

[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

[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. 6

[ZBS11] ZHOU X., BELKIN M., SREBRO N.: An iterated graph laplacian approach for ranking on manifolds. In *Proc. of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (2011), ACM, pp. 877–885. 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. 3

[ZL02] ZHANG D., LUO G.: An integrated approach to shape based image retrieval. In *Proc. of the 5th Asian Conference on Computer Vision (ACCV 2002)* (2002), pp. 652–657. 3

[ZWG*04] ZHOU D., WESTON J., GRETTON A., BOUSQUET O., SCHÖLKOPF B.: Ranking on data manifolds. In *Advances in Neural Information Processing Systems 16* (2004), MIT Press. 5