Eurographics Workshop on 3D Object Retrieval (2015)
I. Pratikakis, M. Spagnuolo, T. Theoharis, L. Van Gool, and R. Veltkamp (Editors)

SHREC'15: Range Scans based 3D Shape Retrieval

A. Godil[1†], H. Dutagaci[2†], B. Bustos[5], S. Choi[3], S. Dong[7], T. Furuya[4], H. Li[7], N. Link[8], A. Moriyama[4], R. Meruane[5], R. Ohbuchi[4], D. Paulus[8], T. Schreck[6], V. Seib[8], I. Sipiran[6], H. Yin[7], C. Zhang[7]

[1]National Institute of Standards and Technology, USA, [2]Eskisehir Osmangazi University, Turkey
[3]Seoul National University, South Korea, [4]University of Yamanashi, Japan
[5]Universidad de Chile, Chile, [6] University of Konstanz, Germany
[7] Beijing Technology and Business University, China, [8] University of Koblenz-Landau, Germany

Abstract

The objective of the SHREC'15 Range Scans based 3D Shape Retrieval track is to evaluate algorithms that match range scans of real objects to complete 3D mesh models in a target dataset. The task is to retrieve a rank list of complete 3D models that are of the same category given the range scan of a query object. This capability is essential to many computer vision systems that involves recognition and classification of objects in the environment based on depth information. In this track, the target dataset consists of 1200 3D mesh models and the query set has 180 range scans of 60 physical objects. Six research groups participated in the contest with a total of 16 different runs. This paper presents the track datasets, participants' methods and the results of the contest.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Range data, H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—

1. Introduction

Shape retrieval and classification is a challenging and active area of research [TV07]. The interest in partial shape retrieval and classification [LBZ*13, SPS14], especially with range scan images, has increased with the wide availability of portable 3D depth sensors with increasing accuracy, speed, and lower cost (e.g. Microsoft Kinect). Use of depth information has considerably increased the accuracy of scene segmentation algorithms, leading to semantic interpretation of the environment. Such semantic mapping of the environment demands systems that are capable of categorizing generic objects from single depth views. Semantic scene interpretation is a key capability of service robots that can be used in domestic and industrial applications, elderly care, security, and search and rescue operations.

Shape matching of range scans to complete 3D models is also essential for model retrieval from a database when the available form of the query is a physical object. This scheme has applications in multimedia querying, inventory creation, advanced manufacturing, and reverse engineering.

The previous SHREC challenges in the range scan based retrieval are SHREC'09 [DGA*09], SHREC'10 [DGC*10], and the SHREC'13 [SMB*14], the last one, which was based on a large set of synthetic range scans. This year's contest is a continuation of the objective stated in [DGA*09, DGC*10] with a larger set of query range scans and a larger target dataset with more categories.

The task of categorization of a novel generic object from appearance or shape is more challenging than instance recognition since a model of the exact same object is not available to the system [RCSM03]. The within class variations are not limited to changes in pose, acquisition noise, and viewpoint. Each semantic category contains objects with widely varying geometric and textural properties; i.e. there is a semantic gap between shape-based features and the semantic category of the objects. In our case, the task is even more difficult, since only one view of the object will be

Disclaimer: Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by NIST nor does it imply that the products mentioned are necessarily the best available for the purpose.
† Organizers

DOI: 10.2312/3dor.20151069

used to match the query to a dataset of 3D triangular mesh models. In addition, the range scans are noisy and incomplete [TV07], and the triangular mesh models are obtained from the Web, hence many are not of high quality with uneven resolution and inconsistent surface normals. These issues of range scan matching to complete triangular meshes poses a challenge to many of the existing shape retrieval algorithms.

2. The Data Set

The dataset† used for the track consist of two parts: The target dataset and the range scan query dataset, which are described in the following subsections.

2.1. Target Set

The target dataset used for this track is the dataset used in SHREC'12 generic shape benchmark, which contains 1200 complete 3D models in 60 classes [LGA*12]. The target dataset is mainly based on SHREC'11 Generic benchmark [DGD*11], SHREC'10 Generic 3D Warehouse [VGD*10] and the Princeton Shape Benchmark [SMKF04a]. In the dataset there are many intra-class variations. For a few classes, articulation and deformation is also present. The classes are defined with respect to their semantic categories and are listed in Table 1. In each class there are 20 models. The file format to represent the 3D models is the ASCII Object File Format (*.off).

Bird	Fish	NonFlyingInsect
FlyingInsect	Biped	Quadruped
ApartmentHouse	Skyscraper	SingleHouse
Bottle	Cup	Glasses
HandGun	SubmachineGun	MusicalInstrument
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

Table 1: *60 classes of the target database.*

† http://www.itl.nist.gov/iad/vug/sharp/contest/2015/Range/

2.2. Query Set

The query set is composed of 180 range images, which are acquired by capturing 3 range scans of 60 physical objects, mainly toys and a few 3D printed items from arbitrary viewpoints. The classes of all the query objects are represented in the dataset.

The range images are captured using a Minolta Laser Scanner (Figure 1). We removed the background noise manually and used ASCII Object File Format (*.off) to represent the scans in triangular meshes. Figure 2 and Figure 3 show examples of query objects and their range scans, respectively.

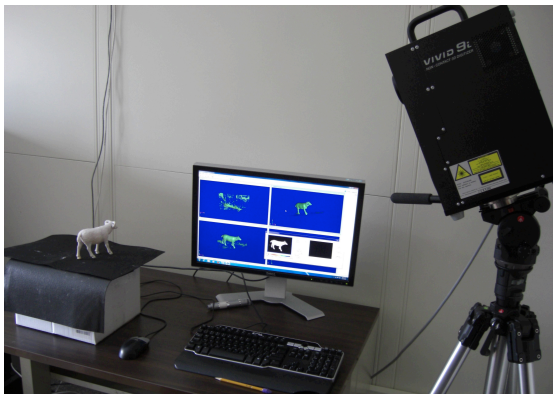

Figure 1: *Setup for range scanning.*

Figure 2: *Examples from the set of the objects that were scanned to obtain queries.*

3. Evaluation Measures

The participants have submitted rank lists for the query inputs. The length of each rank list is equal to the size of the target dataset. Using the rank lists the following evaluation measures were calculated: 1) Nearest Neighbor (NN), 2) First Tier (FT), 3) Second Tier (ST), 4) F-measure (F), 5) Average Precision (AP) 6) Discounted Cumulative Gain

Figure 3: *Examples from the query set.*

(DCG), and 7) the Precision-Recall Curves. The measures are described in [SMKF04b].

4. Submissions

Six research groups participated in the SHREC'15 Shape Retrieval Contest of Range Scans with 16 runs. The participants and their methods are listed as follows:

- Cross-Domain Manifold for Range Scan-based 3D Model Retrieval by T. Furuya, A. Moriyama, and R. Ohbuchi
- 2D-DCT Coefficients of Silhouettes for Range Scan Matching by H. Dutagaci
- Fast Point Feature Histograms, Bag of Feature and Modified Shape Distribution approach for Range Scan Matching by C. Zhang, S. Dong, H. Yin and H.Li
- Partial Shape Retrieval with Fast Point Feature Histograms (FPFH) and Signature Quadratic Form Distance (SQFD) by B. Bustos, R. Meruane, I. Sipiran and T. Schreck
- Depth Image Similarity based Method for Range Scan Matching by S. Choi
- Hough-Voting in a Continuous Voting Space for Range Scan Matching by V. Seib, N. Link and D. Paulus

5. Description of Methods

5.1. Ranking on Cross-Domain Manifold for Range Scan-based 3D Model Retrieval by T. Furuya, A. Moriyama, and R. Ohbuchi

To compare a range scan and a complete 3D model, most of existing methods compare the range scan with a set of multi-view rendered images of the 3D model. However, there is a gap between range scans and rendered images of 3D models. As range scans contain "high frequency noise" such as jagged perimeter edges and cracks, these range scans are often dissimilar to rendered images of 3D models.

We employ a similarity metric learning algorithm called Cross-Domain Manifold Ranking (CDMR) [FO14b] to partially overcome the gap between range scans and 3D models. While the CDMR can be performed in either unsupervised, supervised, or semi-supervised mode, we use unsupervised CDMR for this track.

5.1.1. Cross-Domain Manifold Ranking

The CDMR first creates a Cross-Domain Manifold (CDM) graph by connecting range scans and 3D models, which are represented as nodes, by using visual feature similarities among them. The CDM graph is represented as a matrix $\mathbf{W}$ having size $(N_R + N_M) \times (N_R + N_M)$, where N_R and N_M are the number of range scans and 3D models in a database respectively. For this track, $N_R = 180$ and $N_M = 1,200$. The element $\mathbf{W}_{ij}$ in $\mathbf{W}$ indicates similarity between the node (i.e., range scan or 3D model) i and the node j. $\mathbf{W}_{ij}$ is computed as $\mathbf{W}_{ij} = exp(-d_{ij}/\sigma)$ where d_{ij} is a distance between two nodes, whose computation algorithms will be described in Section 5.1.2. A parameter σ controls diffusion of relevance across the CDM. We use different values σ_{RM} and σ_{MM} to compute range scan-to-3D model similarities and 3D model-to-3D model similarities, respectively. Note that, for this track, we omit intra-domain connections among range scans to reduce the number of parameters to be tuned. That is, similarity $\mathbf{W}_{ij}$ between two range scans i and j is set to 0. This form of CDM is similar to a manifold graph used in the SCMR algorithm by Tatsuma et al. [ZWG*04]. After generating $\mathbf{W}$ representing the CDM, Manifold Ranking algorithm [ZWG*04] is applied on $\mathbf{W}$ to generate rankings of the 3D models for the given queries.

5.1.2. Visual Distances for Cross-Domain Manifold

Range scan-to-3D model distance: We use an algorithm called P-SV-DSIFT, which is based on P-BF-DSIFT algorithm [OF09], to compute a distance between a range scan and a 3D model. The range scan is first rendered from a

single viewpoint into a depth image having 256×256 pixels. To reduce high frequency noises, we blur the depth image and then apply closing operation to the blurred image with a circular structural elements of radius 3 pixels. On the other hand, each retrieval target 3D model is rendered from 42 viewpoints spaced uniformly in solid angle to generate 42 depth images with 256×256 pixels. We randomly and densely extract a set of about 1,200 SIFT [Low04] distinctive features from the image. To reduce influence of high frequency noise, we employ Lower Frequency Emphasis (LFE) [OF09], which is an importance sampling of larger scale features. We set the parameter W for LFE to 1.0, which means that the same number of SIFT features are sampled across an image pyramid of SIFT algorithm. The set of 1,200 SIFT features is aggregated into a feature vector per image, i.e., per view, by using Super Vector (SV) coding [ZYZH10]. We use soft-assignment variant of SV coding [FO14a]. We learn a codebook with 2,500 codewords. The aggregated feature, i.e., P-SV-DSIFT, is power-normalized and then L2-normalized as with [FO14a]. A distance d_{ij} between a range scan i and a 3D model j is computed as a minimum distance among a P-SV-DSIFT vector of the range scan and the set of 42 P-SV-DSIFT features of 42 rendered images of the 3D model. We use Cosine distance, which is computed as $-(c+1)/2$ where c is Cosine similarity between two P-SV-DSIFT features.

3D model-to-3D model distance: We use SV-DSIFT algorithm [FO14a] to compute a distance between two 3D models. The 3D model is rendered from 42 viewpoints to generate 42 depth images having 256×256 pixels. About 300 SIFT features are randomly and densely extracted from the image. We use LFE with $W = 1.0$. The set of about $42 \times 300 = 12,600$ SIFT features is aggregated into a feature vector per 3D model by using the soft-assignment variant of SV coding. SV-DSIFT feature is power-normalized and then L2-normalized as with P-SV-DSIFT. A distance d_{ij} between two 3D models i and j is computed as Cosine distance between two SV-DSIFT features for the two 3D models.

We submitted the following five runs for this track:
PSVDSIFT: A ranking of 3D models is generated without using the CDMR. That is, the query and the 3D models are compared by using P-SV-DSIFT.
CDMR: A ranking of 3D models is generated by relevance diffusion on the CDM graph. Since ground truth was not available for the dataset, we could not tune the parameters for CDMR, i.e., σ_{RM} and σ_{MM}. We tried the following four combinations of the parameters; $(\sigma_{RM}, \sigma_{MM})$ = (0.01, 0.01), (0.01, 0.05), (0.05, 0.01), (0.05, 0.05).

5.2. 2D-DCT Coefficients of Silhouettes for Range Scan Matching by H. Dutagaci

Observing that the silhouette of an object gives sufficient information about its category most of the time, we adopted a scheme where the silhouettes of both range queries and the silhouettes of 3D models obtained from various viewpoints are described by 2D-Discrete Cosine Transform (2D-DCT) coefficients. We rendered the 3D complete mesh models from 66 view points distributed over the view sphere to obtain 66 silhouette images of size 400×400. These silhouettes are scale normalized so that the binary image covers the entire silhouette with a predetermined margin of 4 pixels. Since 2D-DCT is dependent on the in-plane orientation of the shape, we have pose-normalized the silhouette in the image using second moments such that the principal orientation of the silhouette aligns with the horizontal axis of the image. We applied the same normalization scheme to the silhouettes of the range scans prior to extraction of 2D-DCT coefficients. $K \times K$ low frequency 2D-DCT coefficients are retained as descriptors for the pose-normalized silhouettes of both range scans and 3D models. In this track we set K to 10, hence each silhouette is represented with a feature vector of size 100. For matching, the L_2-norms of the distances between the feature vector of the range scan and the feature vectors of the 66 views of the 3D target model are computed. The smallest distance is assigned to be the distance between the range scan and the 3D target model.

5.3. Fast Point Feature Histograms, Bag of Feature and Modified Shape Distribution Approach for Range Scan Matching by C. Zhang, S. Dong, H. Yin and H.Li

In this approach, every model in the target dataset is converted to a point cloud so as to extract consistent descriptors with the input. First, we calculate the Fast Point Feature Histograms (FPFH) [RBB09] features of each point in the corresponding point cloud of a model, which is either from query or target dataset. And then the global FPFH descriptor (BoF-FPFH) of each point cloud is computed by using Bag of Feature (BoF) approach [OF08]. Modified Shape Distribution (MSD) [LLL*15] algorithm is adopted to obtain the distance information between each pair of points, which we sample from the surface of the 3D model randomly. At last, we adopt the IF-IDF model to implement shape matching.

5.3.1. Preprocessing

FPFH descriptor is intrinsically invariant to scale, orientation, and position. However, scale and position normalization is implemented so that the same parameters of the FPFH descriptor can be set for different 3D models. Then, we randomly sample 1,024 sample points for each model.

5.3.2. BoF-FPFH Descriptor Extraction

The BoF-FPFH feature extraction can be divided into two stages: FPFH feature extraction and encoding the BoF-FPFH descriptor by BoF approach:

FPFH feature extraction: FPFH is a robust multi-dimensional feature which is initially employed in the 3D

registration and describes the local geometry around a point p for 3D point cloud. More specifically, a point p's FPFH is based on the relationships between the points in the k-nearest-neighborhood and their estimated surface normals. Simply put, it attempts to capture as best as possible the sampled surface variations by taking into account all the interactions between the directions of the estimated normals. In our experiments, we calculate FPFH features of all random sample points. As an alternative to calculating FPFP features of all sample points, Harris keypoints of the 3D model are extracted from the 1,024 sample points to enhance the computation time. The mean numbers of extracted Harris keypoints are approximately 331 and 90 for the target dataset and query dataset, respectively.

Encoding the BoF-FPFH descriptor: To calculate BoF-FPFH, 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 FPFH features of the models in the target dataset. Note that the 3D models' point set is either represented by the random sample point set or the Harris keypoint set. K-means clustering is performed with $K = 512$. Then, the frequency histogram based on the visual words of the FPFH features is counted as the global feature vector (BoF-FPFH) of the 3D target model. Similarly, the frequency histogram of the visual words is obtained for a query model as its BoF-FPFH feature vector.

5.3.3. IF-IDF Model

We use the IF-IDF model (term frequency-inverse document frequency) [WMB99] to define importance of a visual word and use the following IF-IDF function to compute term weights:

$$\overline{h}_j = \left(h_j / \sum_i h_i \right) log\left(N/f_j\right) \tag{1}$$

where N denotes the total number of FPFH feature in this collection and f_j the frequency of visual word j in the whole collection. As a result, we can evaluate the similarity distance by the formula:

$$D(i,j)_{TF-IDF} = log_{1/2}\left(\left\langle h,\overline{h}\right\rangle / \|h\| \left\|\overline{h}\right\|\right). \tag{2}$$

At last, the similarity distances of BoF-FPFH and MSD are normalized and then summed with equal weight (BoF-FPFH-MSD) to implement feature fusion. We have submited two results based on the random sample point sets and Harris keypoint sets.

5.4. Partial Shape Retrieval with Fast Point Feature Histograms (FPFH) and Signature Quadratic Form Distance (SQFD) by B. Bustos, R. Meruane, I. Sipiran and T. Schreck

This method involves the application of a flexible distance used to compare two shapes which are represented by feature sets. The signature quadratic form distance [BUS10] is a context-free distance that has proven to be effective in the image retrieval domain. In addition, in this algorithm, we build a feature set composed of normalized local descriptors. The idea is to compute an intermediate representation for each shape using a set of local descriptors which are calculated around a set of representative surface points. This algorithm is a modified version of the method evaluated in [BBB*12].

All objects were scaled to fit in a $1 \times 1 \times 1$ box and then they were preprocessed in order to have uniform point density. We used two methods to select the sample points: In the first method we select interest points using Harris 3D [SB11]. We select 1 % of the number of vertices of a shape (with the highest Harris response) as keypoints. We called this method *fpfh1*. In the second method we do a dense interest point selection based on uniform sampling. We called this second method *fpfh2*. For each interest point we compute a local FPFH descriptor [RBB09].

The set of local descriptors of a shape forms the feature space of that shape. Next, a local clustering algorithm [LL04] is applied to obtain a set of representative descriptors. In brief, the clustering uses two thresholds to define the inter-cluster and intra-cluster properties of the space, so it does not depend on the number of clusters. Hence, the clustering only depends on the distribution of the descriptors in the feature space. Given a partitioning after the clustering, the intermediate representation S^P of an object P is defined as a set of tuples as follows:

$$S^P = (c_i^P, w_i^P), \quad i = 1,...,n \tag{3}$$

where c_i^P is the average local descriptor in the i-th cluster and w_i^P is the fraction of elements belonging to the i-th cluster. The representation of an object depends on the local clustering and two objects do not necessarily have the same number of clusters. We test this same approach using Spin Images [JH99] descriptor.

5.5. Depth Image Similarity based Method by S. Choi

This entry to the range scans track of SHREC'15 investigates the use similarities between depth images of a 3D model taken from various possible view points and the range scans. We exhaustively rotated each 3D target model with 30 degree intervals for x-axis and z-axis, so there are 12 x 12 = 144 total viewpoints. Then for each viewpoint, a depth image of size $d \times d$ is acquired. Similarily, for each 2.5D range query model we extracted 144 depth images. Then, one representative depth image with the widest surface is chosen for the query model. Finally, we calculated the distance score between all 144 depth images from a 3D model in the target set and 1 depth image from 2.5D range query model. Depth image distance is calculated as following steps:

1. Depth normalization: Pixel values in the depth images are normalized to be in the range of $[0,1]$. If a pixel is outside

the object shape boundary, a null value is assigned to that pixel.

2. Given two depth images A and B, the sum of pixelwise distanes is calculated:

$$\text{Distance}(A,B) = \sum_{p=1}^{d\times d} D(A(p),B(p)) \quad (4)$$

3. $D(a,b)$ function is a L1 distance measure based heuristic function. It is defined as follows:

- When a == null and b! = null, $D(a,b) = 1$
- When a! = null and b == null, $D(a,b) = 1$
- When a == null and b == null, $D(a,b) = 0$
- When a! = null and b! = null, $D(a,b) = -1 + |a-b|$

Per each 3D model in the target dataset, 144 depth image distances are calculated. 3D models in the target set are ranked by their respective least depth image distance. We submitted two runs: In *SNU_1*, d is set to 50. In *SNU_2*, d is set to 100.

5.6. Hough-Voting in a Continuous Voting Space by V. Seib, N. Link and D. Paulus

Our shape retrieval approach for the track is related to the Implicit Shape Model formulation by Leibe et al. [LLS04]. Recently, adaptations of this method to 3D data were proposed [KPW*10, STDS11, WZS13]. In contrast to the original formulation, the adaptations to 3D data all use a discrete Hough-space for voting. We use a continuous voting space and omit the vector quantization of features in order not to lose the feature's descriptiveness. To be able to generalize from learned shapes, we match each extracted feature with the k best matches in the learned dictionary.

5.6.1. Shape Retrieval Approach

In training, key points are extracted from full 3D models using a uniform voxel grid and a SHOT descriptor [TSDS10] is computed for each key point. For each feature, a vector pointing from the feature to the object's centroid is obtained, in the following referred to as *center vector*. To classify objects, features are detected on the input data in the same manner as in the training stage. Matching detected features with the previously trained data pool yields a list of feature correspondences. The distance between learned feature descriptor f_l and detected feature descriptor f_d is determined by the distance function $d(f_l,f_d) = \|f_l - f_d\|_2$. Since we can not expect to encounter the same objects during classification as were used in training, each detected feature is associated with the k best matching features from the learned data pool.

The center vectors of the created correspondences are used to create hypotheses on object center locations in a continuous voting space. A separate voting space for each class is used. Each vote in the voting space is weighted by its likelihood

$$\omega = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{d(f_l,f_d)^2}{2\sigma^2}\right). \quad (5)$$

The value σ^2 is class specific and is determined during training by the sample covariance.

Each voting space can be seen as a sparse representation of a probability density function. Maxima in the probability density function are detected using the Mean-shift algorithm. We use a modified Mean-shift vector as proposed by Cheng [Che95] to account for weighted votes. In a final step we merge found maxima positions from all voting spaces of individual classes. In case multiple maxima are found at the same position, i.e. if they are closer than half of the kernel bandwidth, only the maximum with the highest probability is retained.

5.6.2. Evaluation

We created point clouds from the provided normalized dataset and sampled the point clouds with a grid size of 1.0 cm. The models were trained from SHOT features calculated with a radius of 1 m.

The presented algorithm returns a list of results ranked by the common weight of the contributing votes. Since this challenge requires of report distances, we apply this simple transformation from weights to distances for each object i: $dist = \omega_{max} - \omega_i$ (where ω_{max} is the weight of most likely object hypothesis).

Results were obtained with the following parameters:

- continuous-hough_bw-3.0m_ls-0.25_k3: bandwidth for Mean-shift = 3.0 m, leaf-size for key point sampling = 0.25 (relative to the leaves size during training) and matching k = 3 nearest features.
- continuous-hough_bw-0.5m_ls-0.25_k5: bandwidth for Mean-shift = 0.5 m, leaf-size for key point sampling = 0.25 (relative to the leaves size during training) and matching k = 5 nearest features.

6. Results

The six participants of the SHREC'15 Shape Retrieval Contest of Partial Models submitted 16 sets of rank lists each. The results for the 16 submissions are given in Figure 4 as scalar performance measures. The precision-recall curves of the best runs of the participants are presented in Figure 5.

The best performance is achieved by Furuya et al.'s CDMR method, where SIFT descriptors are extracted from rendered views of 3D models. The 2D-DCT method of Dutagaci, which is again a view-based method, comes second in terms of retrieval performance. This method yielded moderate results as compared to Furuya et al.'s methods, however, the retrieval success is much higher than the other four sets

of runs. Choi's view-based method comes third, although with much lower performance figures. Zhang et al.'s, Bustos et al.'s, and Seib et al.'s algorithms are all based on point sampling from model surfaces. Their results are similar to each other, however much lower than those of the view-based methods.

We can conclude that view-based methods are more suitable to describe the models in the target dataset as compared to descriptors that rely on surface properties of the 3D models. The 3D triangular meshes in the target dataset are obtained from the Web, hence are not of high quality, with uneven resolution, internal triangles, incomplete parts, and inconsistent surface normals; therefore a reliable surface representation of the model cannot be obtained through point sampling of the polygons. In contrast, rendering depth images of the 3D models from various view-points allows a more coherent representation among the 3D target models and the range-scans.

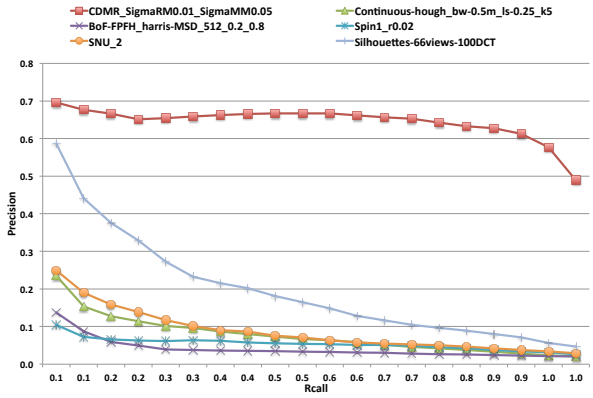

Figure 5: *Precision-recall curves.*

7. Conclusions

In this paper, we have described and compared six algorithms and their variants of six research groups that participated in the SHREC'15 - Range Scans based 3D Shape Retrieval. The algorithms accept a range scan as the input and retrieve similar models from a database of complete 3D models. The methods based on view-based methods yielded better overall performance as compared to the methods based on surface point descriptors.

8. Acknowledgement

Benjamin Bustos has been partial funded by FONDECYT(Chile) Project 1140783.

References

[BBB*12] BIASOTTI S., BAI X., BUSTOS B., CERRI A., GIORGI D., LI L., MORTARA M., SIPIRAN I., ZHANG S., SPAGNUOLO M.: Shrec'12 track: Stability on abstract shapes. 5

[BUS10] BEECKS C., UYSAL M. S., SEIDL T.: Signature quadratic form distance. In *Proceedings of the ACM International Conference on Image and Video Retrieval* (2010), ACM, pp. 438–445. 5

[Che95] CHENG Y.: Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 17*, 8 (1995), 790–799. 6

[DGA*09] DUTAGACI H., GODIL A., AXENOPOULOS A., DARAS P., FURUYA T., OHBUCHI R.: Shrec'09 track: Querying with partial models. In *3DOR* (2009). 1

[DGC*10] DUTAGACI H., GODIL A., CHEUNG C., FURUYA T., HILLENBRAND U., OHBUCHI R.: Shrec'10 track: Range scan retrieval. 1

[DGD*11] DUTAGACI H., GODIL A., DARAS P., AXENOPOULOS A., LITOS G., MANOLOPOULOU S., GOTO K., YANAGIMACHI T., KURITA Y., KAWAMURA S., ET AL.: Shrec'11 track: generic shape retrieval. In *Proceedings of the 4th Eurographics conference on 3D Object Retrieval* (2011), Eurographics Association, pp. 65–69. 2

[FO] FURUYA T., OHBUCHI R.: Hashing cross-modal manifold for scalable sketch-based 3d model retrieval.

[FO14a] FURUYA T., OHBUCHI R.: Fusing multiple features for shape-based 3d model retrieval. In *Proc. British Machine Vision Conference 2014 (BMVC 2014)* (2014). 4

[FO14b] FURUYA T., OHBUCHI R.: Similarity metric learning for sketch-based 3d object retrieval. *Multimedia Tools and Applications* (2014), 1–26. 3

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

[KPW*10] KNOPP J., PRASAD M., WILLEMS G., TIMOFTE R., VAN GOOL L.: Hough transform and 3d surf for robust three dimensional classification. In *Computer Vision–ECCV 2010*. Springer, 2010, pp. 589–602. 6

[LBZ*13] LIU Z.-B., BU S.-H., ZHOU K., GAO S.-M., HAN J.-W., WU J.: A survey on partial retrieval of 3d shapes. *Journal of Computer Science and Technology 28*, 5 (2013), 836–851. 1

[LGA*12] LI B., GODIL A., AONO M., BAI X., FURUYA T., LI L., LÓPEZ-SASTRE R., JOHAN H., OHBUCHI R., REDONDO-CABRERA C., ET AL.: Shrec'12 track: Generic 3d shape retrieval. 2

[LL04] LEOW W. K., LI R.: The analysis and applications of adaptive-binning color histograms. *Computer Vision and Image Understanding 94*, 1 (2004), 67–91. 5

[LLL*15] LI B., LU Y., LI C., GODIL A., SCHRECK T., AONO M., BURTSCHER M., CHEN Q., CHOWDHURY N. K., FANG B., ET AL.: A comparison of 3d shape retrieval methods based on a large-scale benchmark supporting multimodal queries. *Computer Vision and Image Understanding 131* (2015), 1–27. 4

[LLS04] LEIBE B., LEONARDIS A., SCHIELE B.: Combined object categorization and segmentation with an implicit shape model. In *Workshop on statistical learning in computer vision, ECCV* (2004), vol. 2, p. 7. 6

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision 60*, 2 (2004), 91–110. 4

Participants	Methods	NN	FT	ST	F	AP	DCG
Furuya et al.	P-SV-DSIFT	0.639	0.413	0.558	0.403	0.448	0.712
	CDMR_SigmaRM0.01_SigmaMM0.01	0.650	0.438	0.580	0.422	0.474	0.721
	CDMR_SigmaRM0.01_SigmaMM0.05	0.661	0.569	0.737	0.536	0.644	0.787
	CDMR_SigmaRM0.05_SigmaMM0.01	0.656	0.428	0.567	0.415	0.462	0.716
	CDMR_SigmaRM0.05_SigmaMM0.05	0.578	0.471	0.634	0.458	0.529	0.733
Dutagaci	Silhouettes-66views-100DCT	0.478	0.190	0.275	0.194	0.197	0.541
Zhang et al.	BoF-FPFH_harris-MSD	0.072	0.036	0.064	0.042	0.042	0.336
	BoF-FPFH-MSD	0.056	0.036	0.062	0.041	0.041	0.337
Bustos et al.	Fpfh1_r0.01	0.033	0.018	0.037	0.024	0.029	0.316
	Fpfh2_u0.02_r0.10	0.017	0.016	0.034	0.023	0.033	0.311
	Spin1_r0.02	0.033	0.041	0.087	0.053	0.054	0.348
	Spin2_u0.02_r0.05	0.022	0.022	0.043	0.028	0.041	0.322
Choi	SNU_1	0.167	0.076	0.119	0.081	0.084	0.388
	SNU_2	0.178	0.078	0.119	0.082	0.087	0.395
Seib et al.	Continuous-hough_bw-0.5m_ls-0.25_k5	0.139	0.075	0.124	0.085	0.077	0.389
	Continuous-hough_bw-3.0m_ls-0.25_k3	0.139	0.077	0.121	0.085	0.074	0.383

Figure 4: *Retrieval performances.*

[OF08] OHBUCHI R., FURUYA T.: Accelerating bag-of-features sift algorithm for 3d model retrieval. In *Proc. SAMT 2008 Workshop on Semantic 3D Media (S-3D)* (2008), pp. 23–30. 4

[OF09] OHBUCHI R., FURUYA T.: Scale-weighted dense bag of visual features for 3d model retrieval from a partial view 3d model. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on* (2009), IEEE, pp. 63–70. 3, 4

[RBB09] RUSU R. B., BLODOW N., BEETZ M.: Fast point feature histograms (fpfh) for 3d registration. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on* (2009), IEEE, pp. 3212–3217. 4, 5

[RCSM03] RUIZ-CORREA S., SHAPIRO L. G., MEILA M.: A new paradigm for recognizing 3-D object shapes from range data. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision* (2003), p. 1126. 1

[SB11] SIPIRAN I., BUSTOS B.: Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes. *The Visual Computer 27*, 11 (2011), 963–976. 5

[SMB*14] SIPIRAN I., MERUANE R., BUSTOS B., SCHRECK T., LI B., LU Y., JOHAN H.: A benchmark of simulated range images for partial shape retrieval. *The Visual Computer 30*, 11 (2014), 1293–1308. 1

[SMKF04a] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The princeton shape benchmark. In *Shape modeling applications, 2004. Proceedings* (2004), IEEE, pp. 167–178. 2

[SMKF04b] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The princeton shape benchmark. In *Shape Modeling International* (2004). 3

[SPS14] SAVELONAS M. A., PRATIKAKIS I., SFIKAS K.: An overview of partial 3d object retrieval methodologies. *Multimedia Tools and Applications* (2014), 1–26. 1

[STDS11] SALTI S., TOMBARI F., DI STEFANO L.: On the use of implicit shape models for recognition of object categories in 3d data. In *Computer Vision–ACCV 2010*. Springer, 2011, pp. 653–666. 6

[TSDS10] TOMBARI F., SALTI S., DI STEFANO L.: Unique signatures of histograms for local surface description. In *Computer Vision–ECCV 2010*. Springer, 2010, pp. 356–369. 6

[TV07] TANGELDER J., VELTKAMP R.: A survey of content based 3d shape retrieval methods. *Multimedia Tools and Applications* (2007). 1, 2

[VGD*10] VANAMALI T., GODIL A., DUTAGACI H., FURUYA T., LIAN Z., OHBUCHI R.: Shrec'10 track: Generic 3d warehouse. 2

[WMB99] WITTEN I. H., MOFFAT A., BELL T. C.: *Managing gigabytes: compressing and indexing documents and images*. Morgan Kaufmann, 1999. 5

[WZS13] WITTROWSKI J., ZIEGLER L., SWADZBA A.: 3d implicit shape models using ray based hough voting for furniture recognition. In *3D Vision-3DV 2013, 2013 International Conference on* (2013), IEEE, pp. 366–373. 6

[ZWG*04] ZHOU D., WESTON J., GRETTON A., BOUSQUET O., SCHÖLKOPF B.: Ranking on data manifolds. *Advances in neural information processing systems 16* (2004), 169–176. 3

[ZYZH10] ZHOU X., YU K., ZHANG T., HUANG T. S.: Image classification using super-vector coding of local image descriptors. In *Computer Vision–ECCV 2010*. Springer, 2010, pp. 141–154. 4