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

# SHREC'14 Track: Extended Large Scale Sketch-Based 3D Shape Retrieval

B. Li[†‡1], Y. Lu[†‡1], C. Li[†2], A. Godil[†2], T. Schreck[†3], M. Aono[‡4], M. Burtscher[‡1], H. Fu[‡5], T. Furuya[‡6], H. Johan[‡7], J. Liu[‡8], R. Ohbuchi[‡6], A. Tatsuma[‡4], C. Zou[‡9]

[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, Toyohashi, Japan
[5] School of Creative Media, City University of Hong Kong, Hong Kong, China
[6] Computer Science and Engineering Department, University of Yamanashi, Yamanashi, Japan
[7] Visual Computing, Fraunhofer IDM@NTU, Singapore
[8] Media Laboratory, Huawei Technologies Co. Ltd., Shenzhen, China
[9] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

**Abstract**

*Large scale sketch-based 3D shape retrieval has received more and more attentions in the community of content-based 3D object retrieval. The objective of this track is to evaluate the performance of different sketch-based 3D model retrieval algorithms using a large scale hand-drawn sketch query dataset on a comprehensive 3D model dataset. The benchmark contains 12,680 sketches and 8,987 3D models, divided into 171 distinct classes. In this track, 12 runs were submitted by 4 groups and their retrieval performance was evaluated using 7 commonly used retrieval performance metrics. We hope that this benchmark, the comparative evaluation results and the corresponding evaluation code will further promote the progress of this research direction for the 3D model retrieval community.*

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

## 1. Introduction

Sketch-based 3D model retrieval targets retrieving a list of 3D models based on sketch input. Compared to the schemes of Query-by-Model, it is more intuitive and convenient for even novice users to learn and search for relevant models. It also has many applications including sketch-based modeling and recognition, and sketch-based 3D animation [TWLB09].

In SHREC'12 [LSG*12] and SHREC'13 [LLG*13], two tracks have been successfully organized on the topic of sketch-based 3D model retrieval. They foster this research area by providing a small-scale and a large-scale sketch-based retrieval benchmark respectively and attracting state-of-the-art algorithms to participate and compete each other. However, even the large scale SHREC'13 Sketch Track Benchmark (**SHREC13STB**) [LLG*13] based on Eitz et al. [EHA12] and Princeton Shape Benchmark (PSB) [SMKF04] contains only 90 classes of 7,200 sketches and 1,258 models. Compared with the complete dataset of 250 classes built in Eitz et al. [EHA12], there is still much room left for further improvement to make it more comprehensive in terms of completeness of object classes existing in the real world. Thus, it is highly necessary to build up an even larger sketch-based 3D retrieval benchmark in terms of both sketches and models to help us to further evaluate the scalability of existing or newly developed sketch-based 3D model retrieval algorithms.

Considering this, we built a SHREC'14 Large Scale

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



DOI: 10.2312/3dor.20141058

Sketch Track Benchmark (**SHREC14LSSTB**) by extending the **SHREC13STB** [LLG*13] by means of identifying and consolidating relevant models for the 250 classes of sketches from the major previously proposed 3D object retrieval benchmarks. These previous benchmarks have been compiled with different goals in mind and to date, not been considered in their sum. Our work is the first to integrate them to form a new, larger benchmark corpus for sketch-based retrieval.

Specifically, besides the PSB used in **SHREC13STB**, the other available 3D model benchmark sources considered include 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], Bonn's Architecture Benchmark (**BAB**) [WBK09], and the Engineering Shape Benchmark (**ESB**) [JKIR06]. Fig. 1 shows some example models for the four specific benchmarks. Totally, this large-scale benchmark has 13,680 sketches and 8,987 models, classified into 171 classes.


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

Based on this new benchmark, we organize this track to further foster this challenging research area by soliciting retrieval results from current state-of-the-art retrieval methods for comparison, especially in terms of scalability. We also provide corresponding evaluation code for computing a set of performance metrics similar to those used in the Query-by-Model retrieval technique.

## 2. Data Collection

Our extended large scale sketch-based 3D model retrieval benchmark† is motivated by a latest large collection of human-drawn sketches built by Eitz et al. [EHA12] and SHREC'13 Sketch Track Benchmark (**SHREC13STB**) [LLG*13].

To explore how human draw sketches and human sketch recognition, Eitz et al. [EHA12] 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. More importantly, it avoids the bias issue since they collect the same number of sketches for every class and the number of sketches for one class is also adequate for a large scale retrieval benchmark. The sketch variation within one class is also adequate enough.

**SHREC13STB** [LLG*13] has found 1,258 relevant models for 90 of the total 250 classes from the PSB benchmark. However, it is not complete and large enough. A majority of 160 classes has not been included. Thus, we believe a new sketch-based 3D model retrieval benchmark based on the above two datasets, but extended by finding more models from other 3D data sources will be more comprehensive and appropriate for the evaluation of a sketch-based 3D model retrieval algorithm, especially for the property of scalability which is very important for practical applications.

Considering this and to build up a better and more comprehensive large-scale sketch-based 3D retrieval benchmark, we extend the search to other available benchmarks, as mentioned above. We have found 8,987 models for 171 classes (for the remaining 79 classes we did not find relevant models in the selected benchmarks), which substantially increase the scale of the benchmark and form the currently largest scale sketch-based retrieval benchmark. We (one undergraduate student, one master student, one researcher with a master degree and one with a PhD degree) adopted 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. This benchmark provides an important resource for the community of sketch-based 3D retrieval and will foster the development of practical sketch-based 3D retrieval applications. Fig. 2 shows several example sketches and their relevant models.

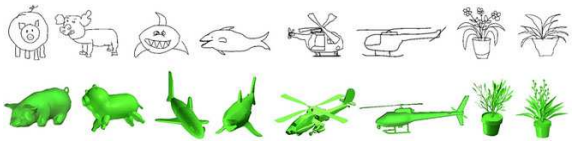

**Figure 2:** *Example 2D sketches and their relevant 3D models in the benchmark.*

We randomly select 50 sketches from each class for training and use the remaining 30 sketches per class for testing, while the relevant models as a whole remain as the target dataset. Participants need to submit results on the training and testing datasets, respectively. To provide a complete reference for the future users of our benchmark, we will eval-

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

uate the participating algorithms on both the testing dataset and the complete benchmark.

### 2.1. 2D Sketch Dataset

The 2D sketch query set contains 13,680 sketches (171 classes, each with 80 sketches) from Eitz et al.'s [EHA12] human sketch recognition dataset, each of which has relevant models in the selected 3D benchmarks.

### 2.2. 3D Model Dataset

In total, the 3D model dataset in this benchmark contains 8,987 models classified into 171 classes. On average, each class has around 53 models. Each model is saved in .OFF format as a text file. The same 3D dataset was also used in evaluating generic 3D shape retrieval algorithms in the SHREC'14 track on comprehensive 3D shape retrieval (`http://www.itl.nist.gov/iad/vug/sharp/contest/2014/Generic3D/`).

### 2.3. Ground Truth

All the sketches and models are categorized according to the classifications in Eitz et al. [EHA12] and the selected source benchmarks, respectively. In our classification and evaluation, we adopt the class names in Eitz et al. [EHA12].

## 3. Evaluation

To have a comprehensive evaluation of the retrieval algorithm, we employ seven commonly adopted performance metrics [SMKF04, SHR14] in Information Retrieval Evaluation that are also widely used in the 3D model retrieval field. They are Precision-Recall (PR) diagram, 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 [SHR14] to compute them.

## 4. Participants

Four groups have participated in the SHREC'14 track on Extended Large Scale Sketch-Based 3D Shape Retrieval. 12 rank list results (runs) for 6 different methods developed by 4 groups have been submitted. The participants and their runs are listed as follows:

- *BF-fGALIF*, *CDMR* ($\sigma_{SM}$=*0.1*, $\alpha$=*0.6*), *CDMR* ($\sigma_{SM}$=*0.1*, $\alpha$=*0.3*), *CDMR* ($\sigma_{SM}$=*0.05*, $\alpha$=*0.6*), and *CDMR* ($\sigma_{SM}$=*0.05*, $\alpha$=*0.3*) submitted by Takahiko Furuya and Ryutarou Ohbuchi from the University of Yamanashi, Yamanashi, Japan (Section 5.1)
- *SBR-VC* ($\alpha$=*1*) and *SBR-VC* ($\alpha = \frac{1}{2}$) submitted by Bo Li and Yijuan Lu from Texas State University, USA; Henry Johan from Fraunhofer IDM@NTU, Singapore; and Martin Burtscher from Texas State University, USA (Section 5.2)
- *OPHOG* and *SCMR-OPHOG* submitted by Atsushi Tatsuma and Masaki Aono from Toyohashi University of Technology, Japan (Section 5.3)
- *BOF-JESC (Words800_VQ)*, *BOF-JESC (Words1000_VQ)*, and *BOF-JESC (FV_PCA32_Words128)* submitted by Changqing Zou from Chinese Academy of Sciences, China; Hongbo Fu from the City University of Hong Kong, China; and Jianzhuang Liu from Huawei Technologies Co. Ltd., China (Section 5.4)

## 5. Methods

### 5.1. Ranking on Cross-Domain Manifold for Sketch-based 3D Model Retrieval, by T. Furuya and R. Ohbuchi

To compare a hand-drawn sketch to a 3D model, most of existing methods compare a sketch with a set of multi-view rendered images of a 3D model. However, there is a gap between sketches and rendered images of 3D models. As hand-drawn sketches contain "noise", such as shape abstraction, semantic influence, stylistic variation, and wobbly lines, these sketches are often dissimilar to rendered images of 3D models.

Our algorithm employs an unsupervised distance metric learning to partially overcome the gap between sketches and 3D models [LLG*14] [FO13]. Our algorithm called Cross-Domain Manifold Ranking, or CDMR [FO13], tries to bridge the gap between features extracted in two heterogeneous domains, i.e., domain of sketches and domain of rendered images of 3D models. While the CDMR algorithm could perform in either an unsupervised, semi-supervised, or supervised mode, we use unsupervised CDMR for this track.

Figure 3 shows an overview of the CDMR. It first creates two separate manifolds of features, i.e., a manifold of sketch features and a manifold of 3D model features. The feature manifolds are computed by using an algorithm best

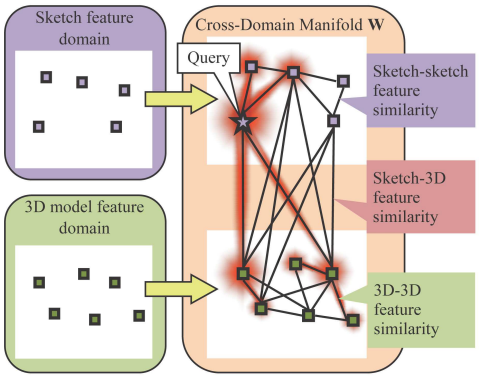


**Figure 3:** *Feature comparison using Unsupervised Cross-Domain Manifold Ranking (CDMR).*

suited for each of the domains; BF-fGALIF [FO13] (slightly modified BF-GALIF [ERB*12]) is used to compare sketches and BF-DSIFT [FO09] is used to compare 3D models. These two feature manifolds are then inter-linked to form a Cross-Domain Manifold (CDM) by using an algorithm capable of sketch-to-3D comparison, that is, the BF-fGALIF. Using the CDM, similarity values between a sketch query and 3D models are computed by diffusing relevance on the CDM. The relevance originates from the query, and it diffuses towards 3D models via edges of the CDM by using a process identical to Manifold Ranking [ZWG*04]. The higher the relevance value of a 3D model, the closer it is to the query.

Unlike previous sketch-to-3D model comparison algorithms, the CDMR tries to maintain manifolds of sketches and 3D models. This often positively contributes to ranking accuracy. Also, if a large enough number of sketches and their inter-similarity values are available, the CDMR performs a form of automatic query expansion on the manifold of sketches.

#### 5.1.1. Forming a Cross Domain Manifold

A CDM is a graph, whose vertices are either sketches or 3D models. The CDM graph $\mathbf{W}$ is represented by a matrix having size $(N_s + N_m) \times (N_s + N_m)$, where $N_s$ and $N_m$ are the number of sketches and 3D models in a database respectively. For this track, $N_s$ = 13,680 and $N_m$ = 8,987.

The element of the matrix $\mathbf{W}$, i.e., $\mathbf{W}_{ij}$, indicates similarity between a sketch (or a 3D model) $i$ and a sketch (or a 3D model) $j$. (For details, please refer to [FO13].) Distances are computed for each pair of vertices $i$ and $j$ by using the feature comparison methods i.e., BF-fGALIF and BF-DSIFT. The distances are then converted into similarities by using the following equation where $d(i, j)$ is distance between vertices $i$ and $j$.

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

The parameter $\sigma$ controls diffusion of relevance value across the CDM. We use different values $\sigma_{SS}$, $\sigma_{MM}$, and $\sigma_{SM}$ to compute sketch-to-sketch similarity, 3D model-to-3D model similarity, and sketch-to-3D model similarity, respectively. These similarity values must be computed either by feature similarity or semantic similarity (if available.)

As mentioned above, sketch-to-3D model comparison uses BF-fGALIF algorithm [LLG*14] [FO13], which is a slightly modified version of BF-GALIF [ERB*12]. BF-fGALIF compare a sketch and multi-view rendered images of a 3D model by using sets of Gabor filter-based local features. A 3D model is rendered into Suggestive Contour (SC) [DFRS03] images from multiple viewpoints. The sketch image and the SC images of the 3D model are rotation-normalized by using responses of multi-orientation Gabor filters computed of the image. After normalizing for rotation, fGALIF features are densely extracted from the image. The set of fGALIF features are integrated into a feature vector per image by using Bag-of-Features (BF) approach. A BF feature of the sketch is compared against a set of per-view BF features of the 3D model to find a distance between the sketch and the 3D model.

For sketch-to-sketch comparison, BF-fGALIF features are extracted from the sketches. Unlike the BF-fGALIF for sketch-to-3D model comparison, the BF-fGALIF for sketch-to-sketch comparison does not perform rotation normalization. This is because most of the sketches drawn by human beings are aligned to a canonical orientation.

To compare 3D models, we use the BF-DSIFT [FO09] algorithm. It is also a view-based algorithm. A set of multi-scale, rotation-invariant local visual features is densely extracted from multi-view rendered range images of a 3D model. The set of local visual features is then BF-integrated per 3D model for comparison.

#### 5.1.2. Ranking on the Cross Domain Manifold

After generating $\mathbf{W}$ representing a CDM, Manifold Ranking (MR) algorithm [ZWG*04] is applied on $\mathbf{W}$ to diffuse relevance value over the CDM from a query. We use the closed form of the MR (equation (1)) to find relevance values in $\mathbf{F}$ given "source" matrix $\mathbf{Y}$. In equation (1), $\mathbf{I}$ is an identity matrix and $\mathbf{S}$ is a symmetrically normalized matrix of $\mathbf{W}$ and $\alpha$ is a parameter. $\mathbf{F}_{ij}$ is the relevance value of the 3D model $j$ given the sketch $i$. A higher relevance means a smaller distance.

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

Among the parameters for the CDMR (i.e., $\sigma_{SS}$, $\sigma_{MM}$, $\sigma_{SM}$ and $\alpha$), we fixed $\sigma_{SS}$ to 0.02 and $\sigma_{MM}$ to 0.005 through preliminary experiments. However, for $\sigma_{SM}$ and $\alpha$), we could not tune these values since ground truth was not available for the experiments. We tried the following combinations of the parameters; ($\sigma_{SM}$, $\alpha$) = (0.1, 0.6), (0.1, 0.3), (0.05, 0.6), (0.05, 0.3), which may not be optimal and might decrease retrieval accuracy compared to the method without CDMR, i.e, BF-fGALIF.

### 5.2. Efficient Sketch-Based 3D Model Retrieval Based on View Clustering and Parallel Shape Context Matching (SBR-VC) [LLJ13] [LLG*13] [LLG*14], by B. Li, Y. Lu, H. Johan, and M. Burtscher

The SBR-VC algorithm first clusters a set of sample views of each model into an appropriate number of representative views, according to its visual complexity defined as the viewpoint entropy distribution of its sample views. Next, an efficient parallel relative shape context matching [BMP02]

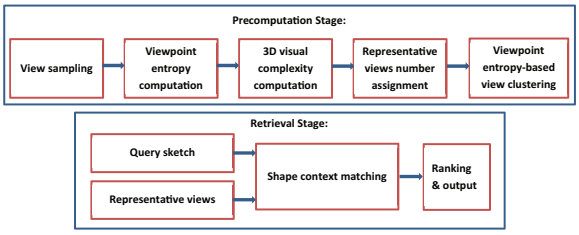


**Figure 4:** *Overview of the SBR-VC algorithm: the first row is for the precomputation whereas the second row is for the retrieval stage [LLG*13] [LLG*14].*

algorithm is employed to compute the distances between a 2D sketch and the representative silhouette views of a 3D model. Before retrieval, the relative shape context features of the representative views of all 3D target models are precomputed. Figure 4 presents an overview of the algorithm, which is described in more detail below.

#### 5.2.1. Precomputation

**(1) Viewpoint entropy-based adaptive view clustering**. This clustering is performed in four steps. For each 3D model, the first step computes the viewpoint entropy of 81 views that are sampled by subdividing a regular icosahedron using the Loop subdivision rule. The second step calculates the viewpoint entropy-based 3D visual complexity for each model. The mean and standard deviation entropies $m$ and $s$ of all sample views of each 3D model are computed first. The 3D visual complexity of each model is defined as

$$C = \sqrt{\frac{\hat{s}^2 + \hat{m}^2}{2}}, \tag{2}$$

where $\hat{s}$ and $\hat{m}$ are the entropies $s$ and $m$ normalized relative to their maximum and minimum over all the models. Hence, $C \in [0,1]$. This metric has the ability to quantitatively measure the semantic difference between models belonging to different categories. In the third step, the visual complexity $C$ of a 3D model is utilized to determine the number of representative views

$$N_c = \lceil \alpha \cdot C \cdot N_0 \rceil, \tag{3}$$

where $\alpha$ is a constant and $N_0$ is the number of sample views for each 3D model. $N_0$ is 81 in the presented SBR-VC algorithm. For large-scale retrieval, $\alpha$ is chosen as 1 or $\frac{1}{2}$, which corresponds to an average of 18.5 or 9.5 representative views, respectively, for each model in the dataset. The fourth step applies Fuzzy C-Means view clustering to the viewpoint entropy values of the 81 sample views, together with their viewpoint locations, to generate the representative views for each model.

**(2) Feature view generation**. Outline feature views for the 2D sketches and the 3D models are generated. In the 3D case, silhouette views are first rendered followed by outline feature extraction. In the 2D case, silhouette views are generated based on binarization, Canny edge detection, closing, dilation, and hole filling.

**(3) Relative shape context computation**. Rotation-invariant relative shape context features [BMP02] are extracted to represent both sketches and sample views. 50 feature points are uniformly sampled for each outline feature view based on cubic B-Spline interpolation.

#### 5.2.2. Online retrieval

With a 2D query sketch, a target 3D database, and the precomputed relative shape context features of the representative views of each model, the online retrieval algorithm works as follows.

**(1) Sketch feature extraction**. First, an outline feature view of the 2D sketch is generated. Then, its relative shape context features are computed.

**(2) 2D-3D distance computation**. The relative shape context matching between the sketch and each representative view of a model is performed in parallel. The minimum 2D-3D matching cost is chosen as the sketch-model distance.

**(3) 2D-3D distance ranking**. The sketch-model distances are sorted in ascending order and the models are ranked accordingly.

SBR-VC ($\alpha = 1$) and SBR-VC ($\alpha = \frac{1}{2}$) represent two runs of the SBR-VC algorithm with corresponding $\alpha$ values. The 70x performance speedup achieved over the serial code [LLG*13] is mainly due to the parallelization and code optimization of the relative shape context matching algorithm.

### 5.3. Overlapped Pyramid of HOG and Similarity Constrained Manifold Ranking, by A. Tatsuma and M. Aono

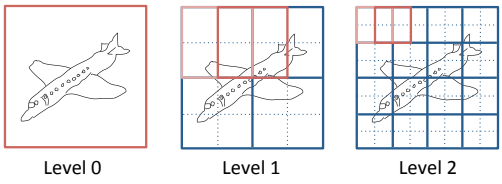


**Figure 5:** *Overview of the Overlapped Pyramid of HOG.*

We propose a new feature vector known as Overlapped Pyramid of Histograms of Orientation Gradients (OPHOG) which is an extended version of pyramid of histograms of orientation gradients [BZM07] proposed in the field of image classification. An overview of the proposed OPHOG is illustrated in Figure 5. OPHOG divides an image into overlapped cells by stages, and extracts an orientation histogram from each cell.

We perform preprocessing to a 3D model and a sketch image before extracting OPHOG features. In the preprocessing of the 3D model, we generate depth buffer images with $300 \times 300$ resolution from the 102 viewpoints that are composed of the vertices of a unit geodesic sphere. To obtain a sketch-like image, we apply Laplacian filtering, thinning transformation and Gaussian filtering to the depth buffer image. Similarly, in the preprocessing of the sketch image, we resize it to $300 \times 300$ resolution, and employ thinning transformation and Gaussian filtering.

After preprocessing, OPHOG divides a given image into cells using a regular sliding window determined by the spatial level. The window size $w$ and stride size $s$ are defined by the image size $n$ and spatial level $l$ as follows:

$$w = n/2^l, \quad s = w/2. \tag{4}$$

Moreover, the number of cells at level $l$ becomes $(2^{l+1}-1)^2$.

The OPHOG feature is obtained by concatenating all of the orientation histograms calculated for each cell. The orientation histogram is constructed by voting gradient magnitude to the corresponding orientation bin. The gradient magnitude $g$ and orientation $\theta$ are defined as follows:

$$g(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2}, \tag{5}$$

$$\theta(x,y) = \tan^{-1} \frac{f_x(x,y)}{f_y(x,y)}, \tag{6}$$

where

$$f_x(x,y) = L(x+1,y) - L(x-1,y),$$
$$f_y(x,y) = L(x,y+1) - L(x,y-1),$$

and $L(x,y)$ denotes the image value at pixel $(x,y)$.

Finally, to decrease the influence of the noise in a sketch image, we transform the OPHOG feature vector into its rank order vector and normalize the rank order vector using $\ell_2$ normalization.

During implementation, we set the number of histogram bins to 40 and limit the number of levels to 3.

For comparing a sketch image to a 3D model, we calculate the minimum Euclidean distance, which is denoted by the following equation:

$$d(s,m) = \min_{i=1,\cdots,102} ||\mathbf{v}^{(s)} - \mathbf{v}_i^{(m)}||, \tag{7}$$

where $\mathbf{v}^{(s)}$ is the feature vector of sketch image $s$, and $\mathbf{v}_i^{(m)}$ denotes the feature vector of the $i$th depth buffer image rendered from 3D model $m$.

We also propose the extended manifold ranking method [ZWG*04] constrained by the similarity between a sketch image and a 3D model. In the following, we call this method Similarity Constrained Manifold Ranking (SCMR).

Suppose we have feature vectors of 3D model $\mathbf{x}_1, \ldots, \mathbf{x}_n$. SCMR aims to assign to each feature vector $\mathbf{x}_i$ a ranking score $r_i$ which reflects the non linear structure of the data manifold. To reflect the data relations represented with the affinity matrix $W$ within the ranking scores, we defined the following cost function:

$$\frac{1}{2} \sum_{i,j=1}^{n} \left( \frac{r_i}{\sqrt{D_{ii}}} - \frac{r_j}{\sqrt{D_{jj}}} \right)^2 W_{ij}, \tag{8}$$

where $D_{ii} = \sum_j W_{ij}$.

To preserve the similarity between a query sketch-image and a target 3D model in the ranking score, we add the following fitting constraint term:

$$\sum_{i=1}^{n} (r_i - p_i)^2, \tag{9}$$

where $p_i = \exp(-d(s,m_i)^2/\sigma^2)$ is the similarity between the query sketch-image and $i$th target 3D model.

The optimal ranking score is obtained by minimizing following cost function:

$$E(r) = \frac{1}{2} \sum_{i,j=1}^{n} \left( \frac{r_i}{\sqrt{D_{ii}}} - \frac{r_j}{\sqrt{D_{jj}}} \right)^2 W_{ij} + \mu \sum_{i=1}^{n} (r_i - p_i)^2, \tag{10}$$

where $\mu > 0$ is a regularization parameter.

Differentiating $E(r)$ with respect to $r$ and rearranging, we obtain

$$\mathbf{r} = (I - \alpha M)^{-1} \mathbf{p}, \tag{11}$$

where $M = D^{-1/2} W D^{-1/2}$, $\mathbf{r} = [r_1, \ldots, r_n]^{\mathrm{T}}$, $\mathbf{p} = [p_1, \ldots, p_n]^{\mathrm{T}}$, and $\alpha \in [0,1)$ is a tuning parameter.

Clearly, the matrix $(I - \alpha M)^{-1}$ can be calculated off-line. The ranking score can be obtained by simple matrix-vector multiplication.

In SCMR, we use Depth Buffered Super-Vector Coding, which we propose for the large scale comprehensive 3D shape retrieval track as the feature vector for a 3D model. Furthermore, we calculate the affinity matrix using a locally constrained diffusion process [YKTL09].

### 5.4. BOF-JESC Based Descriptor, by C. Zou , H. Fu, and J. Liu

The proposed mid-level feature descriptor BOF-JESC follows the bag-of-features framework and employs a junction-based extended shape context to characterize the local details within the four concentric circles centered at key-points. BOF-JESC extracts a global histogram for each image *im* (*im* denotes a binary image obtained from a query sketch/model view in this work). Edge point location in a local patch of BOF-JESC is quantized into 40 bins of a log-polar coordinate system with the radius set to 0.075, 0.15,

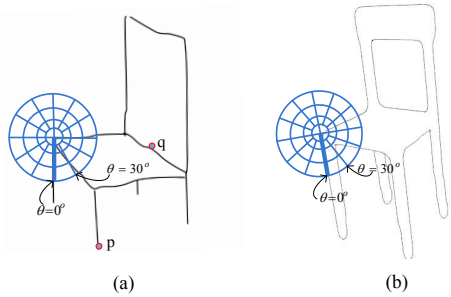


**Figure 6:** *Illustration for the junction-based extended shape context feature descriptor. Two local patches on a junction of a query sketch and a model view are shown in (a) and (b), respectively.*

0.25 and 0.35 of $R_{im}$ ($R_{im} = \sqrt{W * H}$ where $W$ and $H$ is the width and height of the bounding box of *im*). In BOF-JESC, the circle with the shortest radius within a local patch is divided into four bins (as shown in Fig. 6), which comes from the facts that the bins with small areas are more sensitive to the statistics of the edge points.

The 40 dimensional local feature of BOF-JESC has the following characteristics:

- BOF-JESC selects all the junctions (we uses the method in [MAFM08] to extract the junctions in *im*, and the points with degree one, e.g. the point *p* in Fig. 6a, are also treated as junctions), and the mid-points in the lines connecting two adjacent junctions (e.g. the point *q* in Fig. 6a) into the key-point set to generate local features;
- BOF-JESC aligns the reference axis with $\theta = 0$ of the log-polar coordinate system to the average direction of the tangent lines of the ten nearest points in the longest edge connecting the corresponding key-point, this step obtains a rotation invariance;
- BOF-JESC quantizes the edge points on the boundary of two neighboring bins into the bin with a greater angle (relative to the the reference axis in the anti-clockwise direction);
- BOF-JESC normalizes a 40 dimensional local feature with $\ell_1$-norm regularization.

After the local features based on key-points are extracted from all the model views in a database, BOF-JESC employs K-means to obtain *d* "visual words" and finally builds a global $\ell_2$-normalized histogram (i.e. a *d* dimensional feature vector) for each model view in the off-line stage.

#### 5.4.1. Implementation

We sample 42 views for each 3D model uniformly on the unit viewpoint sphere. The vocabulary is obtained by the following steps: 1) concentrating the local features of all the model views in the database, 2) sampling 1 million local features from concentrated features, 3) utilizing KNN to obtain *N* words. The query-to-model distance metric is based on the nearest neighbor (NN) strategy, which finds the closest view to the query in the feature space, and treats such a minimum query-to-view distance as the query-to-model distance. The vocabulary sizes are set to 600, 800, 1000, and 1200. Besides the standard framework of the bag-of-feature method using k-means, we also evaluate the performance of the Fisher Vector [PLSP10] combined with JESC features.

## 6. Results

In this section, we perform a comparative evaluation of the 12 runs of the 6 methods submitted by 4 groups. We measure retrieval performance based on the 7 metrics mentioned in Section 3: PR, NN, FT, ST, E, DCG and AP.

As described in Section 2, the complete query sketch dataset is divided into "Training" and "Testing" datasets, which is to accustom to machine learning-based retrieval algorithms. To provide complete reference performance data for both learning-based methods and non-learning based approaches (like all the 6 participating methods), we evaluate the submitted results on both "Training" and "Testing" datasets, as well as the complete sketch dataset. Figure 7 and Table 1 compare the participating methods in terms of the 7 performance metrics on the above three datasets, respectively.

As shown in the aforementioned figure and table, Tatsuma's SCMR-OPHOG performs best, followed by their OPHOG, while the overall performance of the top methods from other groups are very close. We can see that other groups could catch up with OPHOG in terms of overall performance; but after employing a manifold ranking method SCMR, Tatsuma's group achieved much better performance. For example, compared with OPHOG, SCMR-OPHOG has a gain of 77.3%, 74.5%, 52.94%, 10.3%, and 150.0% in FT, ST, E, DCG and AP, respectively. However, compared to the performance obtained in the SHREC'12 and SHREC'13 sketch-based 3D model retrieval tracks [LSG*12] [LLG*13], the performance of all the participants has decreased sharply due to much more challenging data in the new benchmark. It is also worth to pay more attention to the scalability issues when developing sketch-based 3D retrieval algorithms, especially for large-scale retrieval applications. More details about the retrieval performance with respect to different classes for each participating method can be found in the track homepage [SHR14].

In addition, we have an approximate efficiency performance comparison by asking participants to provide timing information. The average response time per query on the "Testing" dataset based on a modern computer is compared in Table 2. Obviously, BF-fDSIFT is the most efficient, followed by BOF-JESC and SBR-VC ($\alpha = \frac{1}{2}$). OPHOG, SCMR-OPHOG and SBR-VC ($\alpha = 1$) are comparable in terms of speed, while CDMR has the by order of magnitude highest time consumption, thus it has more space for

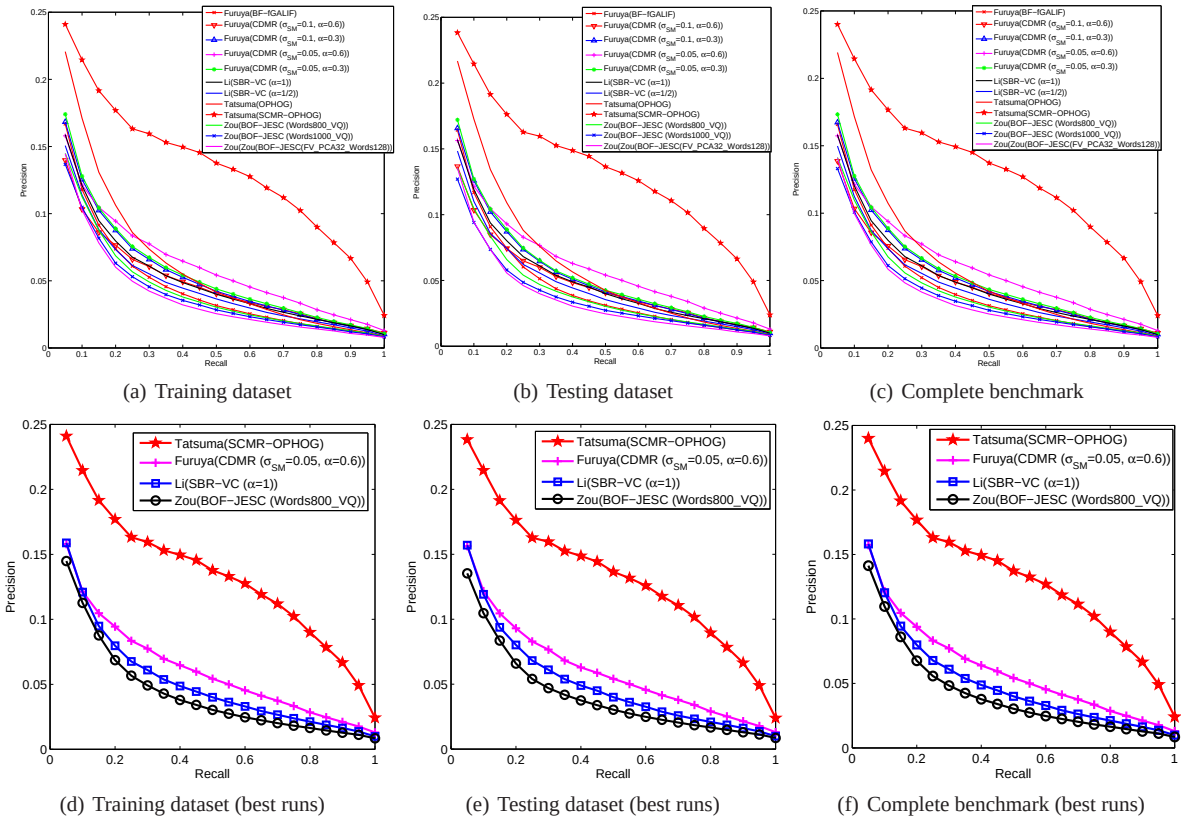


**Figure 7:** *Precision-Recall diagram performance comparisons on different datasets of the SHREC'14 Sketch Track Benchmark for the 12 runs of the 4 participating groups.*

further improvement in this regard. We believe that the timing information is useful for an approximate comparison of the runtime requirements of the algorithms.

Finally, we classify all participating methods with respect to the techniques employed. Three groups (Furuya, Tatsuma and Zou) utilize local features while one group (Li) employs a global feature. Two (Furuya and Zou) of the three methods based on local features apply the Bag-of-Features framework while Manifold Ranking is also used in two (Furuya and Tatsuma) of the three local feature-based algorithms.

## 7. Conclusions

In conclusion, this large scale sketch-based retrieval track is to further foster this challenging and interesting research direction encouraged by the success of SHREC'12 and SHREC'13 sketch-based 3D shape retrieval tracks. Though the benchmark is even more challenging, we still have 4 groups who have successfully participated in the track and contributed 12 runs of 6 methods. This track provides a common platform to solicit current sketch-based 3D model retrieval approaches in terms of this large scale retrieval scenario. It also helps us identify state-of-the-art methods as well as future research directions for this research area.

We have noticed that the obtained retrieval performance is far from satisfactory and existing sketch-based retrieval methods drop apparently when scaled to a very large collection. Therefore, we identify the future direction of this research area is developing more robust algorithms which are scalable to different sizes and diverse types of sketch queries and models. To achieve this, we recommend utilizing techniques from other related disciplines, such as machine learning and pattern recognition, as well as developing more powerful local features to improve the retrieval performance. We also hope that the large scale sketch retrieval benchmark will become a useful reference for researchers in this community.

## 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,

**Table 1:** *Performance metrics comparison on the SHREC'14 Sketch Track Benchmark.*

| Participant | Method | NN | FT | ST | E | DCG | AP |
|---|---|---|---|---|---|---|---|
| **Training dataset** | | | | | | | |
| Furuya | BF-fGALIF | 0.113 | 0.050 | 0.079 | 0.036 | 0.321 | 0.045 |
| | CDMR ($\sigma_{SM}$=0.1, $\alpha$=0.6) | 0.069 | 0.046 | 0.074 | 0.031 | 0.308 | 0.048 |
| | CDMR ($\sigma_{SM}$=0.1, $\alpha$=0.3) | 0.104 | 0.055 | 0.087 | 0.039 | 0.324 | 0.053 |
| | CDMR ($\sigma_{SM}$=0.05, $\alpha$=0.6) | 0.085 | 0.058 | 0.094 | 0.040 | 0.325 | 0.060 |
| | CDMR ($\sigma_{SM}$=0.05, $\alpha$=0.3) | 0.109 | 0.057 | 0.090 | 0.041 | 0.329 | 0.055 |
| Li | SBR-VC ($\alpha$=1) | 0.097 | 0.050 | 0.081 | 0.038 | 0.320 | 0.050 |
| | SBR-VC ($\alpha = \frac{1}{2}$) | 0.094 | 0.047 | 0.077 | 0.035 | 0.316 | 0.046 |
| Tatsuma | OPHOG | **0.158** | 0.066 | 0.097 | 0.051 | 0.340 | 0.060 |
| | SCMR-OPHOG | **0.158** | **0.118** | **0.172** | **0.078** | **0.375** | **0.132** |
| Zou | BOF-JESC (Words800_VQ) | 0.107 | 0.043 | 0.068 | 0.031 | 0.312 | 0.042 |
| | BOF-JESC (Words1000_VQ) | 0.101 | 0.040 | 0.064 | 0.028 | 0.307 | 0.039 |
| | BOF-JESC (FV_PCA32_Words128) | 0.099 | 0.040 | 0.062 | 0.027 | 0.304 | 0.038 |
| **Testing dataset** | | | | | | | |
| Furuya | BF-fGALIF | 0.115 | 0.051 | 0.078 | 0.036 | 0.321 | 0.044 |
| | CDMR ($\sigma_{SM}$=0.1, $\alpha$=0.6) | 0.065 | 0.046 | 0.075 | 0.031 | 0.308 | 0.047 |
| | CDMR ($\sigma_{SM}$=0.1, $\alpha$=0.3) | 0.100 | 0.056 | 0.087 | 0.039 | 0.325 | 0.052 |
| | CDMR ($\sigma_{SM}$=0.05, $\alpha$=0.6) | 0.081 | 0.058 | 0.094 | 0.040 | 0.326 | 0.060 |
| | CDMR ($\sigma_{SM}$=0.05, $\alpha$=0.3) | 0.109 | 0.057 | 0.089 | 0.041 | 0.328 | 0.054 |
| Li | SBR-VC ($\alpha$=1) | 0.095 | 0.050 | 0.081 | 0.037 | 0.319 | 0.050 |
| | SBR-VC ($\alpha = \frac{1}{2}$) | 0.083 | 0.047 | 0.075 | 0.035 | 0.315 | 0.046 |
| Tatsuma | OPHOG | **0.160** | 0.067 | 0.099 | 0.052 | 0.341 | 0.061 |
| | SCMR-OPHOG | **0.160** | **0.115** | **0.170** | **0.079** | **0.376** | **0.131** |
| Zou | BOF-JESC (Words800_VQ) | 0.086 | 0.043 | 0.068 | 0.030 | 0.310 | 0.041 |
| | BOF-JESC (Words1000_VQ) | 0.082 | 0.038 | 0.062 | 0.027 | 0.304 | 0.037 |
| | BOF-JESC (FV_PCA32_Words128) | 0.089 | 0.038 | 0.060 | 0.026 | 0.302 | 0.036 |
| **Complete benchmark** | | | | | | | |
| Furuya | BF-fGALIF | 0.114 | 0.050 | 0.079 | 0.036 | 0.321 | 0.045 |
| | CDMR ($\sigma_{SM}$=0.1, $\alpha$=0.6) | 0.068 | 0.046 | 0.074 | 0.031 | 0.308 | 0.048 |
| | CDMR ($\sigma_{SM}$=0.1, $\alpha$=0.3) | 0.102 | 0.055 | 0.087 | 0.039 | 0.324 | 0.053 |
| | CDMR ($\sigma_{SM}$=0.05, $\alpha$=0.6) | 0.084 | 0.058 | 0.094 | 0.040 | 0.325 | 0.060 |
| | CDMR ($\sigma_{SM}$=0.05, $\alpha$=0.3) | 0.109 | 0.057 | 0.090 | 0.041 | 0.329 | 0.054 |
| Li | SBR-VC ($\alpha$=1) | 0.096 | 0.050 | 0.081 | 0.038 | 0.319 | 0.050 |
| | SBR-VC ($\alpha = \frac{1}{2}$) | 0.090 | 0.047 | 0.077 | 0.035 | 0.316 | 0.046 |
| Tatsuma | OPHOG | **0.159** | 0.066 | 0.098 | 0.051 | 0.341 | 0.061 |
| | SCMR-OPHOG | 0.158 | **0.117** | **0.171** | **0.078** | **0.376** | **0.132** |
| Zou | BOF-JESC (Words800_VQ) | 0.099 | 0.043 | 0.068 | 0.031 | 0.311 | 0.042 |
| | BOF-JESC (Words1000_VQ) | 0.094 | 0.039 | 0.063 | 0.028 | 0.306 | 0.039 |
| | BOF-JESC (FV_PCA32_Words128) | 0.095 | 0.039 | 0.061 | 0.027 | 0.303 | 0.037 |

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.



## References

[BMP02] BELONGIE S., MALIK J., PUZICHA J.: Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell. 24*, 4 (2002), 509–522. 4, 5

[BZM07] BOSCH A., ZISSERMAN A., MUNOZ X.: Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval* (New York, NY, USA, 2007), CIVR '07, ACM, pp. 401–408. 5

[DFRS03] DECARLO D., FINKELSTEIN A., RUSINKIEWICZ S., SANTELLA A.: Suggestive contours for conveying shape. *ACM Trans. Graph. 22*, 3 (2003), 848–855. 4

**Table 2:** *Timing information comparison: T is the average response time (in seconds) per query.*

| **Contributor** (with computer configuration) | **Method** | **Language** | *T* |
|---|---|---|---|
| **Furuya** (CPU: Intel(R) Core i7 3930K @3.20 GHz (the programs ran on a single thread); Memory: 64 GB; OS: Ubuntu 12.04) | BF-fDSIFT | C++ | 1.82 |
| | CDMR | C++ | 3742.15 |
| **Li** (CPU: Intel(R) Xeon(R) CPU X5675 @3.07 GHz (2 processors, 12 cores); Memory: 20 GB; OS: Windows 7 64-bit) | SBR-VC ($\alpha$=1) | C/C++ | 27.49 |
| | SBR-VC ($\alpha = \frac{1}{2}$) | C/C++ | 15.16 |
| **Tatsuma** (CPU: Intel(R) Xeon(R) CPU E5-2630 @2.30GHz (2 processors, 12 cores); Memory: 64 GB; OS: Debian Linux 7.3) | OPHOG | C++, Python | 23.85 |
| | SCMR-OPHOG | C++, Python | 25.67 |
| **Zou** (CPU: Intel(R) Xeon(R) CPUW3550@3.07GHz (the programs ran on a single thread); Memory: 24 GB; OS: Windows 7 64-bit) | BOF-JESC | Matlab | 6.10 |


[EHA12] EITZ M., HAYS J., ALEXA M.: How do humans sketch objects? *ACM Trans. Graph. 31*, 4 (2012), 44:1–44:10. 1, 2, 3

[ERB*12] EITZ M., RICHTER R., BOUBEKEUR T., HILDEBRAND K., ALEXA M.: Sketch-based shape retrieval. *ACM Trans. Graph. 31*, 4 (2012), 31:1–31:10. 4

[FO09] FURUYA T., OHBUCHI R.: Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features. In *CIVR* (2009). 4

[FO13] FURUYA T., OHBUCHI R.: Ranking on cross-domain manifold for sketch-based 3D model retrieval. In *Cyberworlds 2013* (2013), pp. 1–8. 3, 4

[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

[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

[LLG*13] LI B., LU Y., GODIL A., SCHRECK T., AONO M., JOHAN H., SAAVEDRA J. M., TASHIRO S.: SHREC'13 track: Large scale sketch-based 3D shape retrieval. In *3DOR* (2013), Castellani U., Schreck T., Biasotti S., Pratikakis I., Godil A., Veltkamp R. C., (Eds.), Eurographics Association, pp. 89–96. 1, 2, 4, 5, 7

[LLG*14] LI B., LU Y., GODIL A., SCHRECK T., BUSTOS B., FERREIRA A., FURUYA T., FONSECA M. J., JOHAN H., MATSUDA T., OHBUCHI R., PASCOAL P. B., SAAVEDRA J. M.: A comparison of methods for sketch-based 3D shape retrieval. *Computer Vision and Image Understanding 119* (2014), 57–80. 3, 4, 5

[LLJ13] LI B., LU Y., JOHAN H.: Sketch-based 3D model retrieval by viewpoint entropy-based adaptive view clustering. In *Eurographics Workshop on 3D Object Retrieval (3DOR), 2013* (2013), pp. 49–56. 4

[LSG*12] LI B., SCHRECK T., GODIL A., ALEXA M., BOUBEKEUR T., BUSTOS B., CHEN J., EITZ M., FURUYA T., HILDEBRAND K., HUANG S., JOHAN H., KUIJPER A., OHBUCHI R., RICHTER R., SAAVEDRA J. M., SCHERER M., YANAGIMACHI T., YOON G.-J., YOON S. M.: SHREC'12 track: Sketch-based 3D shape retrieval. In *Eurographics Workshop on 3D Object Retrieval (3DOR), 2012* (2012), pp. 109–118. 1, 7

[MAFM08] MAIRE M., ARBELAEZ P., FOWLKES C., MALIK J.: Using contours to detect and localize junctions in natural images. In *CVPR* (2008), IEEE Computer Society. 7

[PLSP10] PERRONNIN F., LIU Y., SÁNCHEZ J., POIRIER H.: Large-scale image retrieval with compressed fisher vectors. In *CVPR* (2010), IEEE, pp. 3384–3391. 7

[SHR14] http://www.itl.nist.gov/iad/vug/sharp/contest/2014/SBR/, 2014. 3, 7

[SMKF04] SHILANE P., MIN P., KAZHDAN M. M., FUNKHOUSER T. A.: The Princeton shape benchmark. In *SMI* (2004), pp. 167–178. 1, 3

[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

[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

[TWLB09] TA A.-P., WOLF C., LAVOUÉ G., BASKURT A.: 3D object detection and viewpoint selection in sketch images using local patch-based Zernike moments. In *CBMI* (2009), Kollias S. D., Avrithis Y. S., (Eds.), IEEE Computer Society, pp. 189–194. 1

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

[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. 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. 4, 6