Viewpoint-Aware Representation for Sketch-Based 3D Model Retrieval

Changqing Zou, Changhu Wang, *Member, IEEE*, Yafei Wen, Lei Zhang, *Senior Member, IEEE*, and Jianzhuang Liu, *Senior Member, IEEE*

***Abstract*—We study the problem of sketch-based 3D model retrieval, and propose a solution powered by a new query-to-model distance metric and a powerful feature descriptor based on the bag-of-features framework. The main idea of the proposed query-to-model distance metric is to represent a query sketch using a compact set of sample views (called *basic views*) of each model, and to rank the models in ascending order of the representation errors. To better differentiate between relevant and irrelevant models, the representation is constrained to be essentially a combination of basic views with similar viewpoints. In another aspect, we propose a mid-level descriptor (called BOF-JESC) which robustly characterizes the edge information within junction-centered patches, to extract the salient shape features from sketches or model views. The combination of the query-to-model distance metric and the BOF-JESC descriptor achieves effective results on two latest benchmark datasets.**

***Index Terms*—3D model retrieval, junction-based local feature, query-to-model distance, viewpoint-aware representation.**

I. Introduction

SKETCH-BASED 3D model retrieval, which takes a hand-drawn sketch as a query, provides a more natural way for end users to obtain their desired 3D objects. In recent years, this topic has attracted increasingly more and more attention (e.g., [4], [6], [8], [13], [14], [17]). For instance, Eitz *et al.* [4] developed a 3D model retrieval system based on the Bag-of-Features (BoF) framework. In this system, the matching between a query and the sampled views of a 3D model is treated as the comparison of visual word histograms. Eitz's method is efficient and obtains a good performance when the sampled views of a 3D model are relatively dense. The method in [6] achieves a good performance on a watertight model benchmark by a two-stage method: 1) obtaining a set of candidate views associated with view-context features, and 2) performing 2D-3D matching based on the relative shape context distance. However, the method is inefficient and the query response time is relatively long due to its shape matching algorithm with high computational complexity. Yoon *et al.* [17] proposed an algorithm based on the diffusion tensor field feature representation to match a sketch with 13 views of a model. In [13], Saavedra *et al.* proposed to extract global features to characterize sample views. The precision of the methods in [13] and [17] is partially limited by the sampling density of the model views (they utilize a small number of sampling views for a 3D model).

In this paper, we propose a new approach to solve the sketch-based 3D model retrieval problem (the flow chart is illustrated in Fig. 1). In general, the approach is powered by two parts. The first part is the query-to-model distance metric based on the *viewpoint-aware representation* (abbreviated as VAR). Briefly, in this distance metric, each 3D model is represented by a compact set of sample views, called *basic views*, based on which a query sketch is approximately represented under a *viewpoint consistency constraint* in the feature space, and the representation error is used to measure the query-to-model distance. Our experimental results indicate that the proposed query-to-model distance metric achieves a higher sketch-model matching precision than the one based on the nearest neighbor (NN) strategy. In addition, we propose an effective mid-level descriptor, called BOF-JESC, to capture the salient shape features in a query sketch or a model view. The comparison results show that the proposed approach outperforms the state-of-the-art methods in both accuracy and speed on the latest large dataset SHREC13 [7].

II. BOF-JESC Descriptor

Our proposed mid-level feature descriptor BOF-JESC follows the bag-of-features framework. It employs a junction-based extended shape context to characterize the local details within the four concentric circles centered at the key points. The motivation of the BOF-JESC descriptor comes from two aspects: 1) the local patch centered at a junction takes into account contour salience, hence can capture important cues for perceptual organization and shape discrimination, as discussed in [10], and 2) the local descriptor shape context [2] is tailored for the images in this work (i.e., the sketches or model views)

Manuscript received January 30, 2014; revised April 17, 2014; accepted April 29, 2014. Date of publication May 05, 2014; date of current version May 12, 2014. This work was supported by Scientific Research Fund of Hunan Provincial Education Department (Grant 13C073), and by the Construct Program of the Key Discipline in Hunan Province. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiao-Ping Zhang.

C. Zou is with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, the Department of Physics and Electronic Information Science, Hengyang Normal University, Hengyang 421008, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: aaronzou1125@gmail.com).

C. Wang and L. Zhang are with Microsoft Research Asia, Beijing, China. (e-mail: chw@microsoft.com, leizhang@microsoft.com).

Y. Wen is with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: wenyf@siat.ac.cn).

J. Liu is with the Media Laboratory, Huawei Technologies Co., Ltd., Shenzhen 518129, China, and also with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: liu.jianzhuang@huawei.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2321764

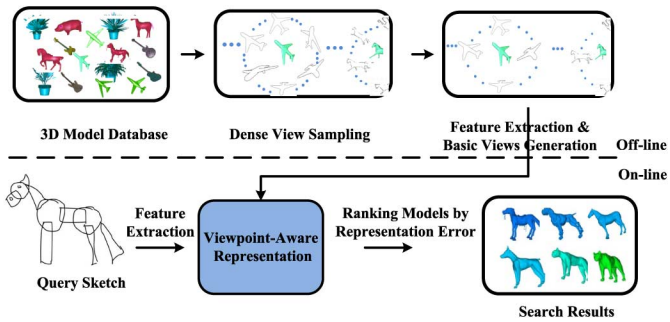

Fig. 1. Flowchart of the proposed sketch-based 3D model retrieval solution.

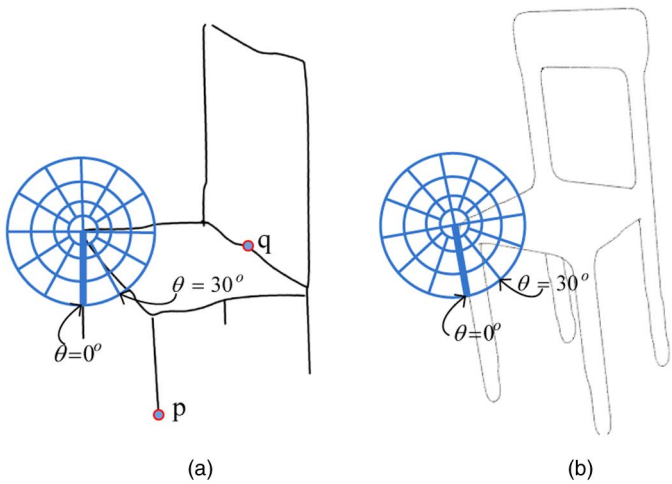

Fig. 2. Illustration for the junction-based extended shape context feature descriptor. Two local patches on a junction of a query sketch and a model view are shown in (a) and (b), respectively.

since they only contain contours. It has been evaluated by [11] to have a high discrimination performance.

In this work, 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 as shown in Fig. 2 (i.e. the number of points is recorded in each bin). In our experiments, the best performance is achieved by setting the radius of the log-polar coordinate to 0.075, 0.15, 0.25 and 0.35 of R_{im} ($R_{im} = \sqrt{W * H}$ where W and H is the width and hight of the bounding box of im). The circle with the shortest radius is divided into four bins, as shown in Fig. 2, which is based on the fact that the bins with small areas are more sensitive to the statistics of the edge points.

In general, the proposed 40 dimensional BOF-JESC has the following characteristics:

- BOF-JESC selects all the junctions (we use the method in [10] to extract the junctions in im, and the points with degree one, e.g. the point p in Fig. 2(a), are also treated as junctions), and the mid-points in the lines connecting two adjacent junctions (e.g. the point q in Fig. 2(a)) 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 L1-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 L_2-normalized histogram (i.e. a d dimensional feature vector) for each model view in the off-line stage.

III. VAR-Based Query-to-Model Distance Metric

A. Viewpoint-Aware Representation (VAR)

1) Representation-based Distance Metric: The primary task of sketch-based 3D model retrieval is to measure the distance between 1) a query sketch s represented by a d-dimensional feature vector $\mathbf{s}$ and 2) a 3D model m characterized by a set $\mathcal{V}_m$ of 2D sample views, where each view v_i is represented by a d-dimensional feature vector $\mathbf{v}_i$.

The most commonly used query-to-model distance metric is based on the nearest neighbor (NN) strategy, which can be formulated as

$$Dist_{\mathrm{NN}}(s, m|\mathcal{V}_m) = \min_{v_i \in \mathcal{V}_m} \|\mathbf{s} - \mathbf{v}_i\|_2, \tag{1}$$

where the distance between query sketch s and 3D model m is calculated as the Euclidean distance between feature $\mathbf{s}$ and the feature of the view closest to $\mathbf{s}$.

NN-based distance metric utilizes individual views separately. Therefore, its reliability is sensitive to the sample views. Particularly, it needs sufficiently dense views to guarantee that a relevant query sketch can be close enough to one of these views. With fewer views available, the minimum query-to-view distance becomes an unreliable estimate of the query-to-model distance.

To obtain a more reliable (robust) distance metric, we utilize the observation that a view can usually be approximately represented by a linear combination of several other views from the neighboring viewpoints in the feature space (an illustrative comparison with traditional NN distance metric is shown in Fig. 3). This inspires us to simultaneously measure the query's distance to multiple views, instead of individual views. It's worth to note that this idea is similar to the one used in the locality-constrained linear coding for image classification [15].

Therefore, we propose to approximately *represent* the query sketch based on a set $\mathcal{V}_m$ of n basic views of 3D model m, and to utilize the representation error as a measure of query-to-model distance, which is formulated as

$$Dist_{\mathrm{VAR}}(s, m|\mathcal{V}_m) = \|\mathbf{s} - \mathbf{B}_m\mathbf{x}^*\|_2, \tag{2}$$

where $\mathbf{B}_m \in \mathbb{R}^{d \times n}$ is the basic view feature matrix, with i-th column representing the feature vector of i-th basic view; $\mathbf{x}^* \in \mathbb{R}^{n \times 1}$ is the representation coefficient vector that represents the optimal projection of query feature $\mathbf{s}$ on the basic view features.

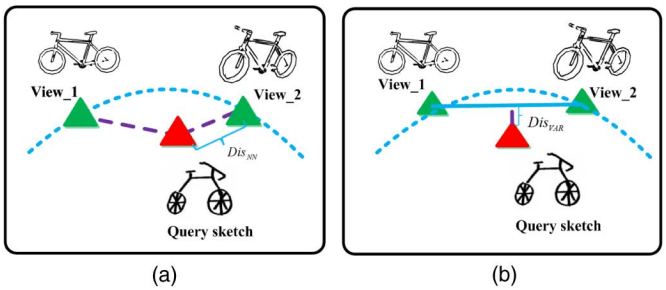

Fig. 3. An illustrative comparison of two query-to-model distance metrics, based on (a) nearest neighbor (NN) and (b) viewpoint-aware representation (VAR). The features of the query sketch (shown in red) and two sample views (shown in green) are plotted in 2D, and the dotted curve lines represent other views with similar viewpoints. It is shown that the NNbased metric suffers from the sparsely sampled views and thus overestimates the distance, whereas the VAR-based metric successfully captures the relevance by approximating the query sketch by a linear combination of the two basic views.

2) Viewpoint Constraint for VAR: The computation of the representation coefficient vector $\mathbf{x}^*$ (in Eq. (2)) for query sketch s on a set $\mathcal{V}_m$ of basic views can be formulated as an optimization problem minimizing the representation error of query feature $\mathbf{s}$. In addition to this criterion, the solution space should be constrained avoid geometrically unreasonable solutions, i.e. the combinations of basic views from quite different viewpoints. Therefore, we constrain that the query sketch should be represented as a combination of basic views observed from similar (neighboring) viewpoints. Hence the problem can be formulated as

$$\mathbf{x}^* = \arg\min_{\mathbf{x}}(\|\mathbf{s} - \mathbf{B}_m\mathbf{x}\|_2^2 + \lambda\mathbf{x}^T\mathbf{L}_m\mathbf{x}) \text{s.t.} \mathbf{1}^{\mathrm{T}}\mathbf{x} = 1, \quad (3)$$

where the positive regularization parameter λ balances the representation error and the viewpoint consistency of the involved basic views; the coefficient x_i is the i-th component of $\mathbf{x}$; the symmetric matrix $\mathbf{L}_m \in \mathbb{R}^{n\times n}$ encodes the pairwise viewpoint distances between n basic views in $\mathcal{V}_m$. Formally, $\mathbf{L}_m$ is defined as

$$\mathbf{L}_m = \begin{pmatrix} \sum_{j=1}^{n}\widetilde{d}(m)_{1,j} & \cdots & -\widetilde{d}(m)_{1,n} \\ \vdots & \vdots & \vdots \\ -\widetilde{d}(m)_{i,1} & \cdots & -\widetilde{d}(m)_{i,n} \\ \vdots & \vdots & \vdots \\ -\widetilde{d}(m)_{n,1} & \cdots & \sum_{j=1}^{n}\widetilde{d}(m)_{n,j} \end{pmatrix}, \quad (4)$$

where $\widetilde{d}(m)_{i,j} = \max[\exp(-(\frac{d(m)_{i,j}}{\gamma})^2) - 0.1, 0]$, $d(m)_{i,j}$ is the viewpoint distance (i.e., the spherical distance on the unit view sphere) between i-th and j-th basic views of model m. The parameter γ controls the range (number) of the influential neighboring views.

In Eq. (3), the first term requires the represented feature $\mathbf{B}_m\mathbf{x}$ to be close enough to the query feature $\mathbf{s}$. The second term gives preference to combinations of geometrically consistent basic views (i.e., which cause small value of $\mathbf{x}^\top\mathbf{L}_m\mathbf{x}$), and accordingly penalizes the combinations of basic views with large variations in viewpoints. Both of the terms in Eq. (3) are convex and the Eq. (3) has an analytical solution (the detail solution procedure for Eq. (3) can be referred to a similar problem in [15]).

TABLE I
THE AVERAGE BASIC VIEW NUMBERS OF SOME CATEGORIES IN THE DATASET USED IN [9]

human	cup	glasses	plane	ant	chair
34	38	39	33	40	41
table	teddy	hand	plier	fish	bird
42	36	41	37	25	35

B. Basic View Generation

In this subsection, we describe how to generate a compact set of basic views for a 3D model.

1) Dense View Sampling: For each model m, we employ the algorithm in [16] to densely sample a set $\mathcal{D}_m$ of N (set empirically to 400) views, which are distributed uniformly on the unit viewpoint sphere. To obtain these views, suggestive contours [3] are adopted to render the 3D models.

2) Basic View Selection: The densely sampled views are inevitably redundant and usually cause high computational cost and memory requirements of retrieval system. Thus, we select a compact set $\mathcal{V}_m \subset \mathcal{D}_m$ of basic views for each model m, according to two criteria: 1) the basic views should approximately represent all views in $\mathcal{D}_m$, and 2) the number of basic views should be as small as possible and determined adaptively by 3D models with various complexity.

As the optimal selection is NP-hard, we propose a greedy algorithm to select basic views sequentially. Given a current set $\mathcal{V}_m^{(k)} = \{v_i\}_{i=1}^k$ of k selected views, the next view v_{k+1} is selected to minimize the accumulated representation error of all the views in $\mathcal{D}_m$ on the basic views $\mathcal{V}_m^{(k)} \cup \{v_{k+1}\}$, formulated as

$$v_{k+1} = \arg\min_{v\in\mathcal{D}_m\setminus\mathcal{V}_m^{(k)}} \sum_{u\in\mathcal{D}_m} Dist_{\mathrm{VAR}}(u, m|\mathcal{V}_m^{(k)} \cup \{v\}). \quad (5)$$

The selection terminates adaptively at the n-th view when the relative reduction of the representation error from $\mathcal{V}_m^{(n)}$ to $\mathcal{V}_m^{(n+1)}$ is smaller than 1%. In the implementation, $\mathcal{V}_m$ is initialized as the set of exemplar views identified by affinity propagation [5] over $\mathcal{D}_m$ based on the Euclidean distance of the feature vectors. Table I lists the average basic view numbers of some categories in the dataset used in [9], which indicates that the basic view selection algorithm can obtain an adaptive number of basic views of a 3D model. Note that the basic view numbers of the listed categories also consist with the results of the entropy-based algorithm in [8].

C. VAR-based 3D Model Retrieval

Given a set $\mathcal{M}$ of 3D models in the database, the on-line retrieval system first computes the query-to-model distance for each model $m \in \mathcal{M}$ using Eq. (2), and then ranks the models in ascending order according to the query-to-model distances. Finally, the top-ranked models are returned as the search results. This framework can be easily integrated with efficient indexing structures of 3D models to provide sub-linear complexity to the number of 3D models in the database.

TABLE II
EXPERIMENTAL SETTINGS

dataset	model	category	sketch
SHREC12	260	13	250
SHREC13	1258	90	7200

IV. EXPERIMENTS

A. Experimental Settings

To evaluate the performance of the proposed solution, we conduct experiments on the two benchmark datasets (SHREC12 and SHREC13) respectively used in [9] and [7]. Our experiments utilize the same settings with the SHREC'12 Track and SHREC'13 Track. The detailed experimental settings about datasets are listed in Table II. In Eq. (3), the constant λ is empirically chosen from $[0.001 : 0.002 : 0.5]$, and the best experiment result comes from $\lambda = 0.05$ in our test. The parameter γ in Eq. (4) is set to 0.8, which is selected from nine parameters $[0.2 : 0.1 : 1.0]$ because of its best precision.

The experiments achieve a relatively better performance when the vocabulary sizes (i.e. d in Section II-A) for SHREC12 and SHREC13 are set to 150 and 200. The proposed method is implemented with MATLAB, running on a PC with Intel(R) dual Core(TM) i5 CPU M540@2.5 GHz (only one CPU is used).

To qualitatively compare with other works, three performance metrics are adopted: 1) Top One (TO), which measures the precision of top-one results, averaged over all queries, 2) First Tier (FT), which measures the precision of top-C results (where C is the number of ground-truth models relevant to the query), averaged over all queries; and 3) Mean Average Precision (mAP), which summarizes the Average Precision of ranking lists for all queries.

B. Results

The evaluation results on the two datasets are listed in Table III, where VAR-BOF-JESC and NN-BOF-JESC denote the proposed solution and the method combining NN based metric distance and the BOF-JESC descriptor, respectively. Densely sampled views are used to compute the performance of NN-BOF-JESC (i.e., the set $\mathcal{V}_m$ in eq. (1) includes 400 views). From Table III, we can see that both VAR-BOF-JESC and NN-BOF-JESC outperform the related methods (see [7] for the details of the algorithms VS-SC, SBR-2D-3D, Saavedra-FDC, and Aono-EFSD) on the larger dataset SHREC13. On SHREC12, the query precision of VAR-BOF-JESC and NN-BOF-JESC is higher than that of BOF-SBR, HKO-KASD, HOG-SC, and Dilated_DG1SIFT (see [9] for the details of the algorithms BOF-SBR, HKO-KASD, HOG-SC, and Dilated_DG1SIFT).

The mAP of the proposed VAR-BOF-JESC is about one percent lower than that of SBR-2D-3D on SHREC12, while on SHREC13 the mAp of VAR-BOF-JESC is about two percent over that of SBR-2D-3D. The different performances on the two datasets are due to the different styles of the query sketches in SHREC12 and SHREC13. The query sketches in SHREC13 have more junctions and it is more likely to extract abundant of junctions from these sketches. It should be noted that the computational-complexity of the SBR-2D-3D [6] is much more than VAR-BOF-JESC (SBR-2D-3D takes more than 19 seconds on SHREC12 (260 models) for each query, while VAR-BOF-JESC finish a query within 1 second). The results of the VAR-BOF-JESC and NN-BOF-JESC on the two test datasets also indicate that the VAR-BOF-JESC based distance metric is superior to the NN-based distance metric when combined with the descriptor BOF-JESC.

TABLE III
PERFORMANCE COMPARISON USING TO [%], FT [%], AND MAP [%]

SHREC12				SHREC13			
Method	TO	FT	mAP	Method	TO	FT	mAP
SBR-2D-3D	68.8	41.5	55.6	**VAR-BOF-JESC**	**18.5**	**11.0**	**11.7**
VAR-BOF-JESC	**65.7**	**40.5**	**54.7**	NN-BOF-JESC	18.5	10.7	11.5
NN-BOF-JESC	65.3	38.4	52.6	VS-SC	16.1	9.7	11.3
VAR-GIST	56.2	36.6	49.6	SBR-2D-3D	13.3	7.9	9.6
NN-GIST	55.3	34.2	47.4	VAR-GIST	11	7.6	8.1
BOF-SBR	53.2	33.9	45.0	NN-GIST	11	7.4	7.9
HOG-SC	31.2	21.5	33.1	Saavedra-FDC	5.2	3.9	5.1
Dilated_DG1SIFT	21.2	16.8	30.2	Aono-EFSD	5.2	3.9	5.1

In terms of efficiency, the average query time is around 3.3 seconds for VAR-BOF-JESC, and around 5.4 seconds for NN-BOF-JESC on the SHREC13 dataset, both using kd-tree [1] acceleration (by conducting each algorithm 20 times using the same query sketch). It indicates that the computation of the VAR-based distance is as efficient as that of NN-based distance (considering that the number of the views for NN-BOF-JESC is greater than that for VAR-BOF-JESC).

To further study the adaptability of the VAR-based distance metric, we compare the performance of the method VAR-GIST which employs VAR-based distance metric and the global feature descriptor GIST [12] to the method NN-GIST which combines the NN-based distance metric and GIST (we extract a 512 dimensional global feature vector, following the implementation in [12], for each model view or query sketch whose orientation is normalized according to the symmetry axis of the bounding box). The results of VAR-GIST and VAR-GIST on the two datasets demonstrate that VAR-based distance metric is superior to the NN-based one in terms of effectiveness.

V. CONCLUSION

We have presented an effective sketch-based 3D model retrieval approach which benefits from two aspects: 1) a viewpoint-aware representation based query-to-model distance metric and 2) a powerful junction based mid-level feature descriptor named BOF-JESC. The basic views of a 3D model are generated adaptively and used to represent the query sketch under the viewpoint consistency constraint, resulting in a reliable query-to-model matching. In contrast to previous descriptors proposed in this related scenario, the BOF-JESC descriptor is more powerful to capture the salient features in a model view or a query sketch. Extensive experimental results demonstrate the superiority of our approach over the state-of-the-art methods.

REFERENCES

[1] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[2] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching, and object recognition using low distortion correspondences," in *Proc. CVPR*, 2005, pp. 26–33.

[3] D. DeCarlo, A. Finkelstein, S. Rusinkiewicz, and A. Santella, "Suggestive contours for conveying shape," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 848–855, 2003.

[4] M. Eitz, R. Richter, T. Boubekeur, K. Hildebrand, and M. Alexa, "Sketch-based shape retrieval," *ACM Trans. Graph.*, vol. 31, no. 4, p. 31, 2012.

[5] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.

[6] B. Li and H. Johan, "Sketch-based 3D model retrieval by incorporating 2D-3D alignment," *Multimedia Tools Applicat.*, pp. 1–23, 2012.

[7] B. Li, Y. Lu, A. Godil, and T. Schreck *et al.*, "Shrec'13 track: Large scale sketch-based 3d shape retrieval," in *Proc. Eurographics Workshop on 3D Object Retrieval*, 2013, pp. 89–96.

[8] B. Li, Y. Lu, and H. Johan, "Sketch-based 3d model retrieval by viewpoint entropy-based adaptive view clustering," in *Proc. Eurographics Workshop on 3D Object Retrieval*, 2013, pp. 49–56.

[9] B. Li, T. Schreck, and A. Godil *et al.*, "Shrec'12 track: Sketch-based 3D shape retrieval," in *Proc. Eurographics Workshop on 3D Object Retrieval*, 2012, pp. 109–118.

[10] M. Maire, P. Arbeláez, C. Fowlkes, and J. Malik, "Using contours to detect, and localize junctions in natural images," in *Proc. CVPR*, 2008, pp. 1–8.

[11] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, 2005.

[12] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[13] J. M. Saavedra, B. Bustos, T. Schreck, S. M. Yoon, and M. Scherer, "Sketch-based 3D model retrieval using keyshapes for global, and local representation," in *Proc. Eurographics Workshop on 3D Object Retrieval*, 2012, pp. 47–50.

[14] T. Shao, W. Xu, K. Yin, J. Wang, K. Zhou, and B. Guo, "Discriminative sketch-based 3D model retrieval via robust shape matching," *Comput Graph. Forum*, vol. 30, no. 7, pp. 2011–2020, 2011.

[15] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. CVPR*, 2010, pp. 3360–3367.

[16] A. Yershova and S. M. LaValle, "Deterministic sampling methods for spheres, and SO (3)," in *Proc. ICRA*, 2004, vol. 4, pp. 3974–3980.

[17] S. M. Yoon, M. Scherer, T. Schreck, and A. Kuijper, "Sketch-based 3D model retrieval using diffusion tensor fields of suggestive contours," in *Proc. ACM Multimedia*, 2010, pp. 193–200.