Erschienen in: Multimedia Tools and Applications ; 74 (2015), 13. - S. 4707-4722

User-drawn sketch-based 3D object retrieval using sparse coding

Sang Min Yoon · Gang-Joon Yoon · Tobias Schreck

Abstract 3D object retrieval from user-drawn (sketch) queries is one of the important research issues in the areas of pattern recognition and computer graphics for simulation, visualization, and Computer Aided Design. The performance of any content-based 3D object retrieval system crucially depends on the availability of effective descriptors and similarity measures for this kind of data. We present a sketch-based approach for improving 3D object retrieval effectiveness by optimizing the representation of one particular type of features (oriented gradients) using a sparse coding approach. We perform experiments, the results of which show that the retrieval quality improves over alternative features and codings. Based our findings, the coding can be proposed for sketch-based 3D object retrieval systems relying on oriented gradient features.

Keywords 3D object retrieval · Sketch-based querying · Gradient descriptor · Sparse coding

1 Introduction

3D object data is becoming ubiquitous in many application areas. It serves as the basis for computer aided design, where machining parts, buildings, or other types of objects are being composed of 3D object parts, to form a blueprint for real-world construction. 3D objects are also indispensable elements for scientific simulation, visualization, serious and

S. M. Yoon (✉)
School of Computer Science, Kookmin University, Seoul, Korea
e-mail: sangmin.yoon@gmail.com

G.-J. Yoon
Department of Mathematics, Ewha W. University, Seoul, Korea
e-mail: gangjoon@gmail.com

T. Schreck
Computer and Information Science, University of Konstanz, Konstanz, Germany
e-mail: Tobias.Schreck@uni-konstanz.de

entertainment gaming, and so on. To re-use existing 3D object content from large 3D object repositories, methods are needed to retrieve the 3D objects requested by users. The area of 3D object retrieval includes methods for implementing similarity functions that allow retrieval and clustering in large 3D object databases.

To date, many retrieval methods have been proposed to compute the similarity between 3D objects for usage in retrieval algorithms. One of the most popular approaches is to extract descriptors (e.g., feature vectors) for each 3D object, and determine the similarity between a pair of objects based on the distance calculated with a metric defined in the descriptors. Regarding the execution of queries, most methods assume that the user query is carried out by the query-by-example paradigm. Thereby, the user provides one reference 3D object, to which the retrieved objects should be similar. The advantage of this paradigm is that both the query (the example) and the candidate objects from the databases are of the *same structure* and therefore, a single descriptor type is applicable. A drawback is that the user needs to provide a query object, which is obviously difficult if the user does not have a prior object, or if selection from the database is not possible because of its size.

An alternative approach to query-by-example is to allow the user to provide free-form sketches that approximate the shape of the 3D objects to be retrieved. Sketch-based query formulation has been successfully established in the domain of image retrieval [1, 22, 45]. Recently, also sketch-based retrieval has come into focus for the task of 3D object retrieval [23, 30, 32, 43]. The advantages of sketch-based 3D object retrieval are that no existing 3D object is required and that the user is not constrained in formulating the visual query. A disadvantage of this approach, however, is that the query and the candidate objects are *not of the same structure* anymore: The query is a free-form sketch, possibly with, variance in the level of detail and abstraction as sketched by the user; the 3D objects in the database are precisely modeled and defined according to the given 3D file format (e.g, polygon meshes).

Recent research has therefore addressed the development of approaches that compensate for the structural differences between the sketch and the target 3D object, to allow for effective retrieval. Two basic methods can be distinguished: 1) One method can transform, by non-photorealistic rendering methods, the target 3D object to a visual representation that is structurally more similar to hand-drawn sketches, prior to descriptor extraction. And 2) the other method can define descriptors that encode only those object properties which are common to both the user sketches and the 3D objects. An example of the latter is gradient descriptors, which consider the dominant orientations of the lines in images.

In this paper, we present an improved approach for supporting sketch-based 3D object retrieval based on oriented gradient features. Our general approach follows the architecture of a typical feature-based retrieval system: Feature vectors are extracted from query and candidate objects, and rankings are produced by a distance function defined on the feature vector representations. In case of sketch-based retrieval, the challenge is to find features that are structurally similar with respect to object content, when extracted from the different modalities 3D object (mesh) and user free-hand sketch. Features extracted from suggestive contours have shown to work well for sketch-based 3D object retrieval [43]. Our contribution in this paper is to define an appropriate representation for coding of oriented gradient features, that manage to improve the basic retrieval effectiveness. Figure 1 illustrates our retrieval workflow, explicitly considering the sparse coding scheme introduced.

Our method is based on combining both described approaches and optimizing the coding of an appropriate gradient descriptor. We apply the approach on an encompassing benchmark data set representing a variety of sketches, both with respect to user drawing styles and 3D object types represented. We experimentally show that the retrieval quality outperforms

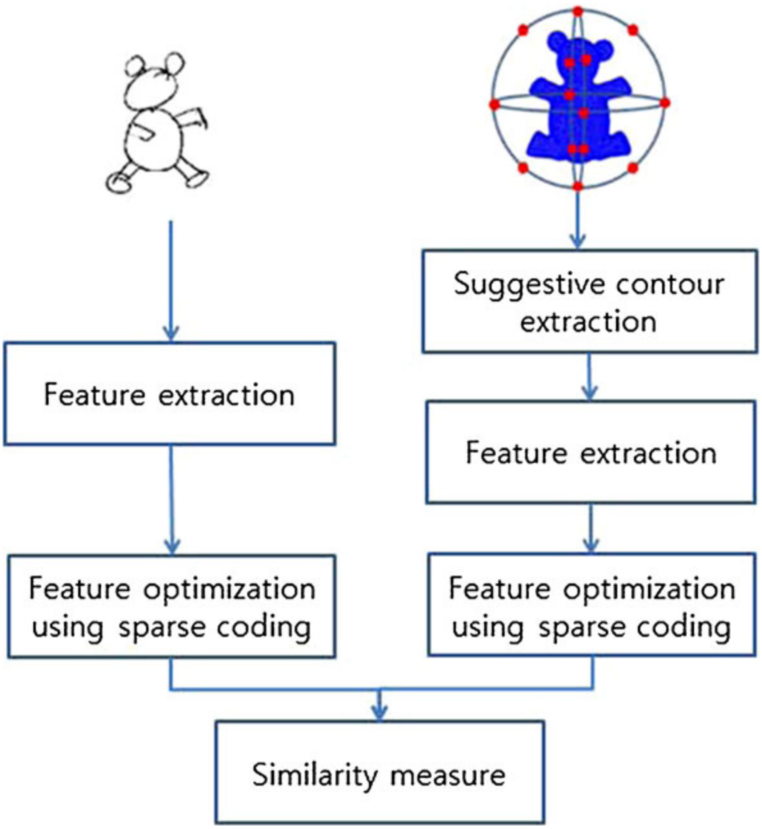


Fig. 1 Flowchart for user sketch-based 3D object retrieval using sparse coding

an existing state-of-the-art method and is therefore applicable for sketch-based 3D object retrieval.

The remainder of this paper is structured as follows. In Section 2 we recall related work about 3D object retrieval, local methods, and sketch-based methods. In Section 3 we explain in detail, how we extract features from 3D models using suggestive contours and feature optimization using sparse coding for robust 3D object retrieval. In Section 4 we present the experimental results showing the effectiveness of our proposed approach. We summarize our contribution and outline future work in the area in Section 5.

2 Related work

In this section, we review methods from global and partial 3D object retrieval, as well as approaches for retrieval based on user sketches.

2.1 3D object retrieval

The field of 3D object retrieval research is concerned with defining a similarity function operating based on 3D object data with the goal of implementing 3D search systems and other analytic applications. Over the last decade, many methods have been proposed for the query-by-sketch paradigm of 3D object retrieval. Surveys indicate [4, 40] that to date, many types of descriptors have been proposed for global and partial 3D object retrieval. Many methods form descriptors from the computed low-level features, e.g., from the surface, volumetric, or structural properties of the 3D objects. Among the most popular and robust descriptors for global retrieval are image-based descriptors. These extract features from 2D renderings of the models for the coding of feature vectors. In addition, descriptors based on heat kernel signatures [3] have recently become popular, as they can support specific desirable properties of the retrieval, such as invariance with respect to rigid transformations.

The effectiveness of shape retrieval methods is typically evaluated based on various kinds of benchmark data [19, 27, 38]. While on the macro (average) benchmarking levels best performing methods can be identified, typically, on the micro (class-specific) level different methods may provide best retrieval performance. This to date motivates the consideration of new descriptor models, suiting specific application needs.

2.2 Local methods

In addition to global descriptors, local descriptors have also been researched. These describe 3D models by using a set of local properties and can be applied to provide both global and partial 3D object retrieval capabilities. Example approaches are based on the detection of different types of local interest points [18, 39], or on the segmentation of a whole model into meaningful parts [13]. In [5], global descriptors were applied to object fragments for improved retrieval performance. The coding of local descriptors from sets of interest points is another challenge, with current approaches often resorting to the bag-of-features method [2].

2.3 Sketch-based methods

Recently, sketch-based methods have been researched for application in 3D object and image retrieval. The idea is to have the user provide a sketch of the visual properties of the shape to be searched. One of the first 3D object retrieval systems, the Princeton Shape Search Engine [25], provided a sketch interface with which users could upload shape projections, and spherical harmonics features were extracted from sketches and 3D objects for ranking of answer objects. However, specific customization of the features to match the properties of the human sketches was not a component on the approach. In another work, the idea of modeling by example [24] was introduced in which by means of object segmentation and 3D retrieval, the modeling process was accelerated by re-usage of existing object parts. Building on this idea, the system in [29] was extended by a sketch interface.

In [43], we proposed an approach for sketch-based 3D retrieval that extracts features from a non-photorealistic rendering of the 3D objects, thereby making the sketch and the target 3D views structurally more compatible with each other. Specifically, the suggestive contour rendering method [15] was used. The retrieval performance was improved compared to when features were extracted from photorealistic renderings, based on objects from [38] and a set of user-provided sketches. Sketch-based retrieval has also been considered in 2D image retrieval applications, and in conjunction with appropriate indexing techniques can provide efficient and effective results also for very large data sizes [1, 21, 22, 45]. Gradient features such as HOG [14] have shown effective in certain Computer Vision applications and also work well for sketch-based 3D retrieval. An alternative to gradient features are key shapes, which have been considered for 3D retrieval for example in [23, 31, 37] as a basis to form a descriptor. Currently, sketch-based 3D retrieval considers all user sketches uniformly. Potential improvements could be achieved classifying user sketches into different types during preprocessing. A recent work demonstrated that the classification of user sketches is possible [20].

3 Our proposed approach

We next describe a novel approach based on a sparse coding representation of gradient features. The method starts by extracting a set of 2D images from the original 3D objects from the target database. To this end, a set of 2D projections are obtained from a virtual camera system assumed to be distributed equally on a sphere around the object. From each camera, a suggestive contour image is generated like initially proposed in [43]. Suggestive contour rendering produces non-photorealistic views of a 3D object which are structurally comparable to a sketched version of the 3D model.

According to Fig. 2, then a histogram of oriented gradients is computed from each suggestive contour image by analyzing the image in the topological space of diffusion tensor fields (HOG-DTF) [2,3]. While this representation has shown to be basically effective, it can be improved by a sparse coding representation. Sparse coding [4, 5] is applied on the set of all HOG-DFT feature vectors of all suggestive contour images. The coding represents every original feature vector as a linear combination of a small number of codebook vectors which have been obtained during the sparse coding training phase.

We retrieve a query object given as a sketch image by measuring the similarity between the HOG descriptors obtained from the sketch image and any linear combination of the codebook vectors, after normalizing all the descriptors to the unit vectors.

As will be demonstrated in the experimental section, this representation is sufficient for preserving the most important characteristics of the space of feature vectors and provides noise removal which helps the retrieval process. We next describe the proposed method formally.

Let X_0 be a 3D object to be retrieved by evaluating its similarity to objects $\{Y_k\}_{k=1}^{M}$ in the database. To retrieve the 3D objects based only on a user-drawn 2D sketched query image($\mathbf{x}_s$), we project the 3D objects $\{Y_k\}_{k=1}^{M}$ in database from K viewpoints to 2D suggestive contour images, and then extract the HOG-DTF descriptors $\mathbf{x}$ and $\{\mathbf{y}_{i,j}\}_{i=1,j=1}^{M,K} \in \mathbb{R}^T$ from the 2D suggestive contour images of the $K \cdot M$ projections.

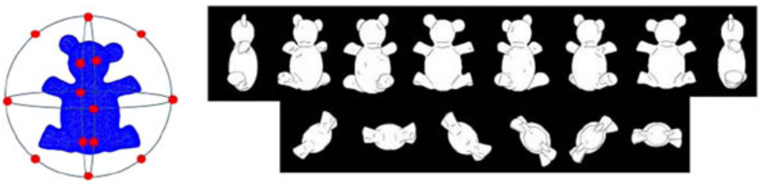


Fig. 2 Suggestive contours which are rendered from regularly distributed viewpoints

The Euclidean (ℓ_2) norm $\|\cdot\|_2$ is defined by $\|\mathbf{a}\|_2^2 = a_1^2 + \cdots + a_T^2$ for $\mathbf{a} = (a_1, \ldots, a_T) \in \mathbb{R}^T$. By $\|\cdot\|_1$, we denote the ℓ_1 norm defined by $\|a\|_1 = \sum_{i=1}^{T} |a_i|$. For our argument, we normalize the HOG-DTF vectors to be unit ones: $\|\mathbf{x}\|_2 = \|\mathbf{y}_{i,j}\|_2 = 1$ for $i = 1, \ldots, M$ and $j = 1, \ldots, K$. For the sake of simplicity, we number the descriptor vectors by grouping the vectors stemming from the same object as

$$\psi_{(i-1)K+j} = \mathbf{y}_{i,j}, \;\; i = 1, \ldots, M \quad \text{and} \quad j = 1, \ldots, K. \tag{1}$$

To retrieve the 3D object by measuring the similarity, there is a way to represent $\mathbf{x}$ in terms of $\psi_j's$ and find the largest coefficient from the representation because the larger the size of the coefficient is, the higher the similarity becomes. In general, the vectors ψ_j are not orthogonal to each other. Thus, our strategy for retrieval is to project $\mathbf{x}$ into the subspace $\mathbf{V}$ of $\mathbb{R}^T$ which is defined as the space of all possible spanned by the descriptors $\{\psi_\ell\}_{\ell=1}^{K\cdot M}$, $\mathbf{V} = \{w_1\psi + \cdots + w_{K\cdot M}\psi_{K\cdot M} : w_i \in \mathbb{R}, \;\; i = 1, \ldots, K \cdot M\}$,

$$\mathbf{w}_0 = \underset{\mathbf{w}\in\mathbb{R}^{K\cdot M}}{\operatorname{argmin}} \|w_1\psi_1 + w_2\psi_2 + \cdots + w_{K\cdot M}\psi_{K\cdot M} - \mathbf{x}\|_2 \tag{2}$$

and then rank the $3D$ objects $\{Y_i\}_{i=1}^{M}$ in the database according to the similarity evaluated from the size of the coefficients w_i. User-drawn sketch query images $\mathbf{x_s}$ are different according to their intention to retrieve the deformable 3D objects, but the projected suggestive contour image provides the salient boundary and inner features comparing to boundary image and ridges and valley images [43]. Thus, we may assume that there exists a linear combination $w_1\psi_1 + w_2\psi_2 + \cdots + w_{K\cdot M}\psi_{K\cdot M}$ with minimum norm error such that the coefficient vector $\mathbf{w}_0 = (w_1, \ldots, w_{K\cdot M})$ has sparse or there exists a constant $\epsilon << 1$ and an integer $S > 0$ satisfy the following inequality:

$$\|\mathbf{w}_0 - \mathbf{w}_S\|_1 \le \epsilon \tag{3}$$

where $\mathbf{w}_S \in \mathbb{R}^{K\cdot M}$ is the vector obtained from $\mathbf{w}_0$ by setting all but the largest S coefficients to be zero. After finding such a linear combination $\tilde{\psi} = w_1\psi_1 + w_2\psi_2 + \cdots + w_{K\cdot M}\psi_{K\cdot M}$, we choose the S largest values from the coefficients $\{\mathbf{w}_\ell\}_{\ell=1}^{K\cdot M}$ and we will rank the objects Y_k from which the S largest coefficients stem. To do this, however, we need to compute the amplitudes of all the coefficients even though all but a few amount of them will be discarded. Also we must locate the largest coefficients too. In the literature, fortunately, there is a technique called Compressed Sensing(CS) which enables us to directly find the significant coefficients and their locations in the sensing step. Mathematically speaking, let $\mathbf{w}_0$ be the coefficient vector solution that is sparse or satisfies the condition (3) for sufficiently small $\epsilon > 0$. Then, under the specific conditions on the matrix Ψ with $\psi_1, \ldots, \psi_{K\cdot M}$ as columns, the solution $\tilde{\mathbf{w}}$ to the problem:

$$\tilde{\mathbf{w}} = \underset{\mathbf{w}=(w_1,\ldots,w_{K\cdot M})^t\in\mathbb{R}^{K\cdot M}}{\operatorname{argmin}} \|\mathbf{w}\|_1 := \sum_{\ell=1}^{K\cdot M} |w_\ell| \quad \text{subject to } \Psi\mathbf{w} = \Psi\mathbf{w}_0 \tag{4}$$

satisfies the following inequality:

$$\|\tilde{\mathbf{w}} - \mathbf{w}_0\|_2 \le C\|\mathbf{w}_0 - \mathbf{w}_S\|_1 \tag{5}$$

for a constant C and an integer $S >$ depending only on Ψ ([9]). As shown in Eq. 5, CS is a data acquisition technique for finding a sparse representation. It turns out to be strongly efficient for acquiring and reconstructing sparse or compressible data. For details in compressive sensing and related algorithms, see [6–10, 16] and the references therein. Motivated by CS, we try to find the vector which has a good fitting to the input data (smaller l_2 norm error), and it has a grouped sparse representation (smaller l_1 norm) by adding the ℓ_1 sparsity term to the least squares problem (2) and solving the unconstrained problem called LASSO(Least Absolute Shrinkage and Selection Operator):

$$\tilde{\mathbf{w}} = \underset{\mathbf{w} \in \mathbb{R}^{K \cdot M}}{\operatorname{argmin}} \frac{1}{2} \|\Psi \mathbf{w} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{w}\|_1 \tag{6}$$

Even though the LASSO optimization which yields good results for structure detection an over-parameterized polynomial model in the presence of additive output noise meets the requirements as shown in [44], it tends to select only one variable from the group when there is a group of variables in which the pairwise correlations are very high ([46]). However, the significant information about $\mathbf{x}$ is concentrated on relatively few largest components of the retrieved vector from the 3D object as given in (6), so the sparse (or relatively few largest) components are grouped necessarily. For this reason, we add the least square regularization term to the problem (6) for small l_2 norm error by consider the following optimization problem:

$$\tilde{\mathbf{w}} = \underset{\mathbf{w} \in \mathbb{R}^{K \cdot M}}{\operatorname{argmin}} \frac{1}{2} \|\Psi \mathbf{w} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{w}\|_1 + \gamma \|\mathbf{w}\|_2^2 \tag{7}$$

This model (7) is a combination of the LASSO ($\gamma = 0$) and the Ridge regression ($\lambda = 0$), and it was proposed by [46] to overcome the limitations of the LASSO optimization. In Eq. 7, the first term measures the fitting and the second regularization term is added to recover a sparse data where λ controls the trade-off between sparsity and reconstruction fidelity. The third regularization term gives rise to grouping effect of highly correlated variables and removes the limitation on the number of selected variables.

As mentioned above, CS is well known to recover a sparse signal. Furthermore, it has a desirable property that if $\tilde{\mathbf{w}}$ is not sparse, the recovered vector $\tilde{\psi}$ given by

$$\tilde{\psi} = \tilde{w}_1 \psi_1 + \tilde{w}_2 \psi_2 + \cdots + \tilde{w}_{K \cdot M} \psi_{K \cdot M}$$

for the solution $\tilde{\mathbf{w}}$ to the problem (7) is as good as if we were to choose the largest values after calculating the values and detecting the corresponding locations. (That is, $\tilde{\mathbf{w}}$ satisfies the inequality as given (3)). Showing the trade-off between the reconstruction error and the sparsity, the formulation (7) allows us to attain a reliable approximation.

To solve the problem (7), we use the sparse coding algorithm proposed by Lee et al. [28]. Sparse coding [12, 17, 26, 35] is regarded as a suitable technique to optimally represent an input HOG-DTF in terms of a linear combination of atoms in an over-complete trained dictionary of basis vectors, with sparse coefficients that are sufficient for preserving specific features. Figure 3 shows the details of the feature extraction and similarity measure methodology using sparse coding.

In the 3D object retrieval process, we want the process to retain the following properties: First, it must be efficient at categorizing the descriptor in a higher dimension. Second, it should achieve a much lower retrieval error. Also, it should be able to capture the salient properties of images. The optimization retrieval by solving problem Eq. 6 meets our goal. For this reason, we have proposed a retrieval method via the optimization problem in Eq. 6.

From the vector $\tilde{\mathbf{w}} = (\tilde{w}_1, \ldots, \tilde{w}_{K \cdot M})$ solving the problem (6), we sort the components of $\tilde{\mathbf{w}}$ according to their sizes in $\{|\tilde{w}_1|, \ldots, |\tilde{w}_{K \cdot M}|\}$ as

$$|\tilde{w}_{\sigma(1)}| \geq \cdots \geq |\tilde{w}_{\sigma(K \cdot M)}|$$

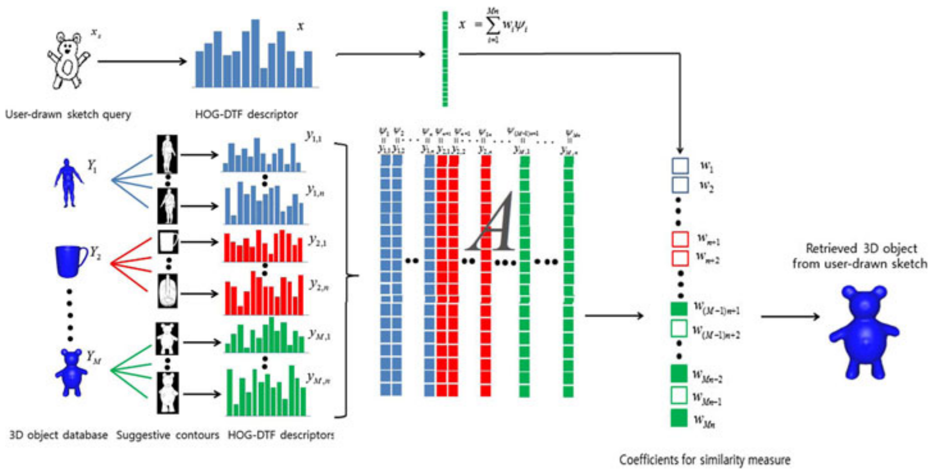


Fig. 3 The details of our proposed feature extraction and optimization procedure using sparse coding

For $\sigma(1)$, there exist unique $1 \leq i \leq M$ and $1 \leq j \leq K$ such that $\sigma(1) = (i-1)K + j$. Now, we retrieve X_0 as the 3D model Y_k in $\{Y_i\}_{i=1}^{M}$ from which the HOG-DTF vector $\psi_{\sigma(1)} = \mathbf{y}_{i,j}$ stems. Also, we rank the objects $\{Y_i\}_{i=1}^{M}$ according to the order $\{|\tilde{w}_{\sigma(1)}| \geq \cdots \geq |\tilde{w}_{\sigma(K \cdot M)}|\}$.

4 Experiments

We conducted experiments to evaluate the effectiveness of our proposed method as compared to two alternative retrieval methods [36, 37] and one alternative coding approach [42]. We discuss the obtained results in three parts: the experimental setup in Section 4.1; the retrieval results from representative sketched query images by numerous users in Section 4.2; and comparison of our method against state-of-the-art previous approaches in Section 4.3.

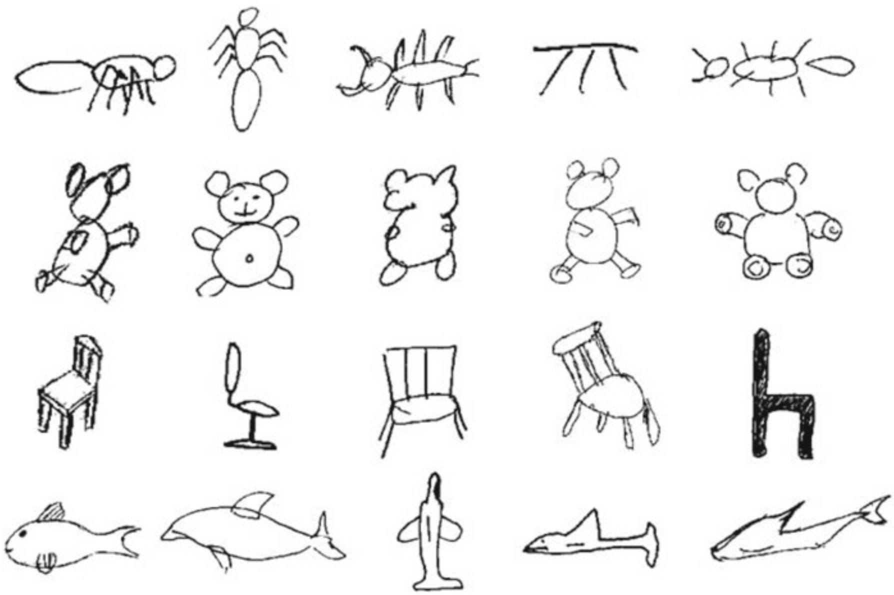
4.1 Experimental setup

We conducted experiments to evaluate the retrieval performance of our approach. For our experiments, we used several 3D mesh models from the Princeton 3D mesh segmentation Benchmark[1] [13]. We used 260(M) models from 13 classes, i.e. humans, cups, airplane, ant, chair, or sunglasses. Accordingly we rendered 260(M)×14(n) suggestive contour images, which we used for our evaluation. We also collected 250 user drawn sketch images from numerous users. We used the 14(n) different viewpoints of a 3D model to retrieve the 3D models, as Funkhouser et al. [25] used 13 orthographic viewpoints, Chen et al. [11] used 10 shaded boundary images from 20 viewpoints, and Macrini et al. [33] used 128 projected images for 3D model retrieval.

Figure 4a illustrates the represented 3D object classes, and Fig. 4b illustrates a selection of the sketched query images which we collected. For the query images, we asked five

[1] http://segeval.cs.princeton.edu/

(a)



(b)

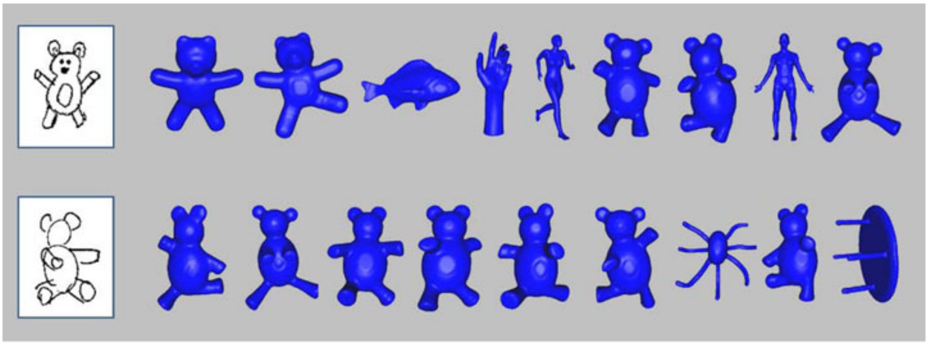
Fig. 4 3D model database which is used in our experiments. (**a**) Examples for the used 13 3D object query classes in our experiments, selected from the Princeton Shape Segmentation Benchmark. (**b**) Sketch query images hand-drawn for each query class by five different users colleagues to draw sketches, just based on the class names. As expected and desired for our experiment, the provided sketch images differ according to user interest and viewpoint. Some users draw the outer shape of the target object (either with or without perspective information), others draw skeletal features of the target object.

To measure the similarity, we first crop the user drawn sketched query image and normalize its size to 200×200 pixels and then extract the gradient feature vectors, followed by the sparse coding step.

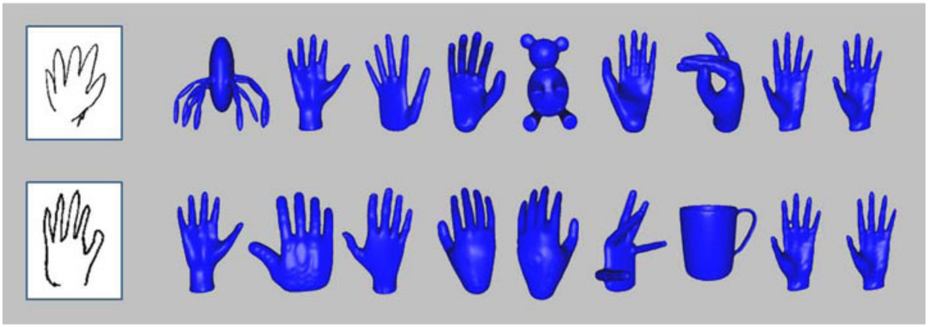
4.2 Evaluation of sketch-based 3D model retrieval

In this section, we show the results of the retrieved 3D objects from various user-drawn sketches. We first analyze the retrieved 3D objects by one sketched query image and the variation of its similarity measure. Figure 5a and b show the top ranked 3D objects, retrieved for user sketches for the concepts "bear" and "hand", respectively. In the top ranked 3D models as shown, most nearest neighbors retrieved with the sparsely coded HOG-DTF features are relevant to the query images. Some 3D objects are wrongly retrieved, however the projected images of these 3D models are still quite similar to the user's sketches.

We can also approximately recover the view the user had in mind, when drawing the sketch query image as shown in Fig. 6a and b. The example query image is sketched from

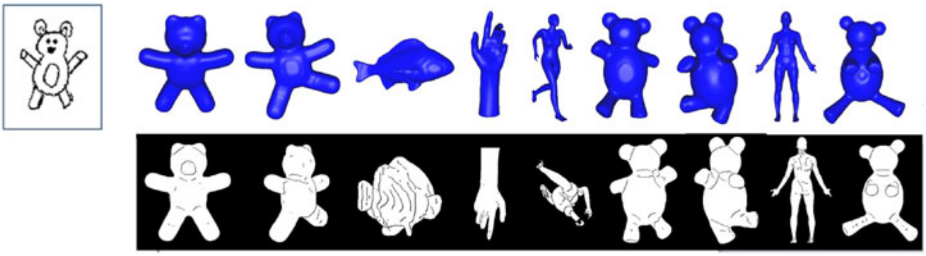


(a)

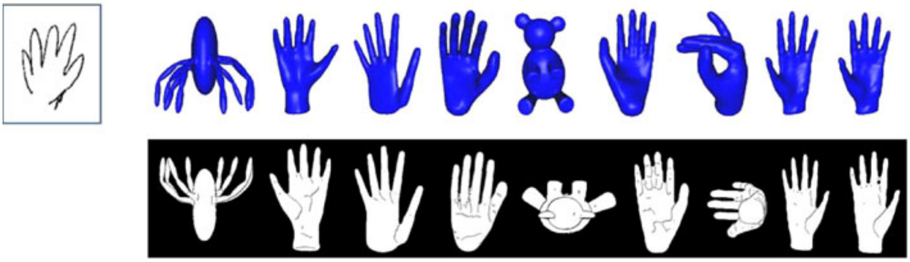


(b)

Fig. 5 High ranked 3D objects using our proposed approach. (**a**) Retrieved 3D objects from user-drawn "bear" query images. (**b**) Retrieved 3D objects from user-drawn "hand" query images



(a)



(b)

Fig. 6 High ranked 3D objects using our proposed approach. (**a**) Retrieved 3D objects and their estimated pose from user-drawn "bear" sketch query. (**b**) Retrieved 3D objects and their estimated pose from user-drawn "hand" sketch query

the front viewpoint of the target object. The closest views of the 3D models are also full-frontal or slightly in profile. One useful side application of our proposed method, is that we can also estimate the users viewpoint, given that database models are given in a consistent 3D reference frame. The wrongly retrieved 3D objects in Fig. 5 are highly ranked because of their projected and size-normalized images are very similar to user-drawn sketches as shown in Fig. 6.

4.3 Comparison of retrieval performance

To systematically evaluate and compare the performance of our proposed methodology we present the first tier precision of the different query benchmark classes. The first-tier measure is a well-known measure defined as the percentage of correctly retrieved objects within the top k retrieved objects in the respective query class, where k is the total number of relevant objects (i.e., the size of the query class). Figure 7 shows a comparison of the first-tier precision of our approach against the performance of the HELO in [36] and STELA descriptor as reported in [37], as well as the suggestive contour features without sparse coding representation and compressive sensing based classification as reported in [42]. Sparse coding-based feature optimization provides for improved first-tier precision in 8 of the 13 total classes, with significant improvements in some classes yielding up to 20 percentage points. In all but one of the remaining classes, the sparse coding representation performs at least comparable as the other methods. In particular, the first tier precision of "chair" is more than double comparing to previous approaches because our proposed feature optimization

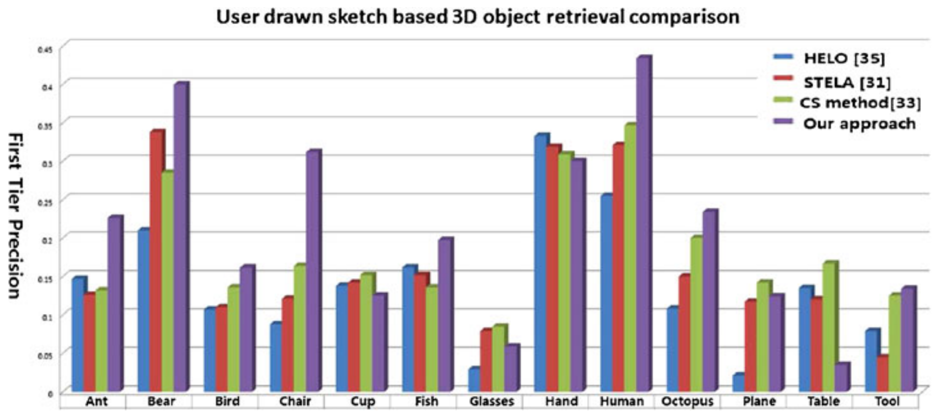


Fig. 7 Comparison of user-drawn sketch based 3D model retrieval between our proposed approach and previous approaches HELO[36], STELA[37], and compressive sensing based classification [42]

using sparse coding emphasize their own characteristics. However, the first tier precision of table is lower than previous approaches due to similar features of ant and octopus classes.

We note that there also exist few outlier classes which show unexpected low performance for a single technique. Specifically, in the class "plane" the HELO descriptor [36] is a low-performing outlier, which our approach obviously fails in the "tool" class. While we at present do not have a sound explanation for these findings, we attribute them as not overly significant for the overall picture, as they address only a smaller number of queries in the experiment.

At this point, we also note that an improvement of retrieval precision is a difficult to obtain goal, as it related to optimization of dedicated feature extraction or coding procedures, and cannot be attained by simple upscaling of hardware backend, etc.

We also took a detailed look at the sensitivity of the performance of our method in comparison to variation of it, by Nearest Neighbor(NN), Second Tier (ST), E-Measures (E), Discounted Cumulated Gain (DCG), and Average Precision (AP) [38]. These are standard, widely-used methods to evaluate the performance of information retrieval systems. Table 1 compares the sparse coding representation with the HOG-DTF features without sparse coding [43] and in the alternative compressive sensing encoding [42]. In particular, we see that the sparse coding based feature optimization is more effective than HOG-DTF based Euclidean distance approach for various evaluation methods. Figure 7 and Table 1 show that our proposed feature optimization increases the performance on the whole and keeps the balance between different classes comparing to previous approaches.

Table 1 Quantitative comparison of sketch-based 3D object retrieval between our proposed method and alternative approaches

Method	NN	ST	E	DCG	AP
Our approach	0.312	0.335	0.225	0.554	0.331
HOG with Euclidean distance [43]	0.220	0.286	0182	0.513	0.292
Compressive sensing based approach [42]	0.283	0.317	0.195	0.527	0.316

5 Discussion

In this paper we propose a user-drawn sketch based 3D object retrieval approach based on gradient feature vectors and its optimization using sparse coding. To extract meaningful features from a 3D object we rendered the suggestive contour from various viewpoints. The suggestive contour was then analyzed in the space of the diffusion tensor fields, and each pixel was represented as an ellipsoidal model, the direction and scale of which are determined by its eigenvalues and eigenvectors.

Based on our experimental evaluation, we argue that the proposed sparse coding-based feature optimization scheme provides better first-tier precision and precision-recall results than previous approaches based on oriented gradient features. Our method uses selective sparse coding instead of traditional vector quantization to extract the salient properties of the appearance descriptors of suggestive contours. This study demonstrates the success of sparse coding for complex feature optimization, providing highly efficient categorization of higher-dimension descriptors.

We note that our experiment indicates the improvement of sparse coding for the specific type of features considered (oriented gradients [43] and in compressed sensing coding [42]). The experiment also indicates the improvement over two further sketch-based 3D retrieval methods ([36, 37]). In the mean time, a larger number of sketch-based retrieval methods have been proposed and the comparison of our proposed scheme against them remains a study to be done. To this end, our future evaluation efforts need to build on benchmarks and method surveys as presented e.g., in [23, 30, 32].

Another interesting problem is to find methods to adapt the feature coding to the sketch in style of different users. It could be the case that different codings (or different features) would provide further improvements, when selectively applied to different sketching styles, such as planar or perspective, abstract, volumetric etc. as depending of the sketching user.

Acknowledgments S.M. Yoon was funded by the Korea Meteorological Administration Research and Development Program under Grant Weather Information Service Engine(WISE) project, 153-3100-3133-302-350. G.-J. Yoon was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No.2013024792).

References

1. Bozas K, Izquierdo E (2012) Large scale sketch based image retrieval using patch hashing. In: Lecture notes in CS, vol 7431, pp 210–219
2. Bronstein AM, Bronstein MM, Guibas LJ, Ovsjanikov M (2011) Shape google: geometric words and expressions for invariant shape retrieval. ACM Trans Graph 30:1:1–1:20
3. Bronstein AM, Bronstein MM, Kimmel R, Mahmoudi M, Sapiro G (2010) A gromov-hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. Int J Comput Vision 89:266–286
4. Bustos B, Keim D, Saupe D, Schreck T (2007) Content-based 3D object retrieval. IEEE Comput Graph Appl 27(4):22–27. Special Issue on 3D Documents
5. Bustos B, Schreck T, Walter M, Barrios J, Schaefer M, Keim D (2012) Improving 3D similarity search by enhancing and combining 3D descriptors. Springer Multimedia Tools Appl 58(1):81–108
6. Candès E (2006) Compressive sampling. In: Proceedings international congress of mathematics. Madrid pp 1433–1452
7. Candès E, Romberg J (2006) Quantitive robust uncertainty principles and otptimally sparse decompositions. Found Comput Math 6(2):227–254

8. Candès E, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency imformation. IEEE Trans Inform Theory 52(2):489–509
9. Candès E, Romberg J, Tao T (2006) Stable signal recovery from incomplete and inaccurate measurements. Commun Pur Appl Math 59(8):1207–1223
10. Candès E, Tao T (2006) Near optimal signal recovery from random projections: universal encoding strategies? IEEE Trans Inf Theory 52(12):5406–5425
11. Chen D-Y, Tian X-P, Shen Y-T, Ouhyoung M (2003) On visual similarity based 3D model retrieval. Comput Graph Forum 22(3):223–232
12. Chen SS, Donoho DL, Sauders MA (2001) Atomic decomposition by basis pursuit. SIAM Rev 43(1):129–159
13. Chen X, Golovinskiy A, Funkhouser T (2009) A benchmark for 3D mesh segmentation. ACM Trans Graph (Proc SIGGRAPH) 28(3):1–12
14. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: CVPR '05: Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 1. IEEE Computer Society, Washington, DC, pp 886–893
15. DeCarlo D, Finkelstein A, Rusinkiewicz S, Santella A (2003) Suggestive contours for conveying shape. ACM Trans Graph (Proc SIGGRAPH) 22(3):848–855
16. Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52(4):1289–1306
17. Donoho DL, Elad M (2003) Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. PNAS 100(5):2197–2202
18. Dutagaci H, Cheung CP, Godil A (2011) Evaluation of 3d interest point detection techniques. In: Proceedings EG workshop on 3D object retrieval, pp 57–64
19. Dutagaci H, Godil A, Daras P, Axenopoulos A, Litos GC, Manolopoulou S, Goto K, Yanagimachi T, Kurita Y, Kawamura S, Furuya T, Ohbuchi R (2011) Shrec '11 track: generic shape retrieval. In: Proceedings EG workshop on 3D object retrieval, pp 65–69
20. Eitz M, Hays J (2011) Learning to classify human object sketches. In: SIGGRAPH 2011: Talks
21. Eitz M, Hildebrand K, Boubekeur T, Alexa M (2009) A descriptor for large scale image retrieval based on sketched feature lines. In: Eurographics symposium on sketch-based interfaces and modeling, pp 29–36
22. Eitz M, Hildebrand K, Boubekeur T, Alexa M (2011) Sketch-based image retrieval: benchmark and bag-of-features descriptors. IEEE Trans Vis Comput Graph 17(11):1624–1636
23. Eitz M, Richter R, Boubekeur T, Hildebrand K, Alexa M (2012) Sketch-based shape retrieval. ACM Trans Graph (Proc SIGGRAPH) 31(4):31:1–31:10
24. Funkhouser T, Kazhdan M, Shilane P, Min P, Kiefer W, Tal A, Rusinkiewicz S, Dobkin D (2004) Modeling by example. ACM Trans Graph (Proc SIGGRAPH) 24(3):652–663
25. Funkhouser T, Min P, Kazhdan M, Chen J, Halderman A, Dobkin D, Jacobs D (2003) A search engine for 3D models. ACM Trans Graph 22(1):83–105
26. Gregor K, LeCun Y (2010) Learning fast approximations of sparse coding. Proceedings international conference on machine learning (ICML'10), Haifa
27. Jayanti S, Kalyanaraman Y, Iyer N, Ramani K (2006) Developing an engineering shape benchmark for CAD models. Comput Aided Des 38(9):939–953
28. Lee H, Battle A, Raina R, Ng AY (2007) Efficient sparse coding algorithms. Adv Neural Inf Process Syst 19:801–808
29. Lee J, Funkhouser T (2008) Sketch-based search and composition of 3D models. In: EUROGRAPHICS workshop on sketch-based interfaces and modeling
30. Li B, Lu Y, Godil A, Schreck T, Aono M, Johan H, Saavedra J, Tashiro S (2013) SHREC13 track: large scale sketch-based 3d shape retrieval. In: Eurographics workshop on 3D object retrieval
31. Li B, Lu Y, Johan H (2013) Sketch-based 3D model retrieval by viewpoint entropy-based adaptive view clustering. In: Proceedings eurographics workshop on 3D object retrieval, pp 49–56
32. 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, Scherer M, Yanagimachi T, Yoon G, Yoon S (2012) SHREC12 track: sketch-based 3D shape retrieval. In: Eurographics workshop on 3d object retrieval, pp 109–118
33. Macrini D, Shokoufandeh A, Dickenson S, Siddiqi K, Zucker S (2002) View based 3D object recognition using shock graphs. In: Proceedings international conference on computer vision
34. Natarajan BK (1995) Sparse approximate solutions to linear systems. SIAM J Comput 24:227–234
35. Olshausen BA, Field D (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381(6583):607–609

36. Saavedra J, Bustos B (2010) An improved histogram of edge local orientations for sketch-based image retrieval. In: Proceedings 32nd annual symposium of the German association for pattern recognition, pp 424–441
37. Saavedra J, Bustos B, Scherer M, Schreck T (2011) STELA: sketch-based 3d model retrieval using a structure-based local approach. In: ACM International conference on multimedia retrieval, pp 26:1–26:8. Peer-reviewed poster paper
38. Shilane P, Min P, Kazhdan MM, Funkhauser TA (2004) The princeton shape benchmark. In: SMI, pp 167–178
39. Sipiran I, Bustos B (2011) Harris 3D: a robust extension of the harris operator for interest point detection on 3d meshes. The Visual Computer. Online first
40. Tangelder JWH, Veltkamp RC (2008) A survey of content based 3D shape retrieval methods. Multimedia Tools Appl 39(3):441–471
41. Wright J, Yang A, Ganesh A, Sastry S, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell
42. Yoon SM, Kuijper A (2011) Sketch-based 3D model retrieval using compressive sensing classification. IET Electronic Lett 47(21):1181–1183
43. Yoon S, Scherer M, Schreck T, Kuijper A (2010) Sketch-based 3D model retrieval using diffusion tensor fields of suggestive contours. In: Proceedings ACM multimedia. ACM, pp 193–200
44. Yoon SM, Schreck T, Yoon GJ (2012) Sparse coding based feature optimization for robust 3D object retrieval. IET Electron Lett 48(9):493–495
45. Zhou R, Chen L, Zhang L (2012) Sketch-based image retrieval on a large scale database. In: ACM multimedia, pp 973–976
46. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Stat Soc 65(2):301–320

Sang Min Yoon received B.E. and M.E. degrees in electronics engineering from the Korea University, Seoul, Korea, in 2000 and 2002, respectively, and a Dr.-Ing. degree in Informatik, TU Darmstadt, Darmstadt, Germany in 2010. Then, he joined the Digital Human Research Center, Advanced Industrial Science and Technology(AIST), Tokyo, Japan, as post-doctoral researcher. He was with Samsung Advanced Institute of Technology (SAIT) as a researcher in the area of face recognition and video surveillance from 2002 to 2005. Currently, he has been working at the school of computer science at Kookmin University, South Korea, as an assistant professor since 2012. His main interests are human motion analysis and understanding, 3D model retrieval, and the human computer interface (HCI).

Gang-Joon Yoon is a researcher at National Institute for Mathematical Sciences in Korea. He joins the National Agenda Project of Korea on Surveillance System. His research interests are in Compressive Sensing and Computer Vision based on Mathematics, and in Functional Analysis also. He obtained a Master's degree and a PhD in Mathematics from Korea Advanced Institute of Science and Technology in 1995 and in 1999.

Tobias Schreck is an Assistant Professor for Visual Analytics with the Department for Computer and Information Science at University of Konstanz, Germany. Between 2007 and 2011, he was a Postdoc researcher at Technische Universitaet Darmstadt, Germany, where he headed a research group on visual search and analysis. He obtained a Ph.D. in Computer Science in 2006, and a Master of Science degree in Information Engineering in 2002, both from University of Konstanz. Tobias Schreck works in the areas of 3D Object Retrieval, Digital Libraries, and Visual Analytics. His research interests include visual search and analysis in time-oriented, high-dimensional, and 3D object data, with applications in data analysis and multimedia retrieval.