IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 27, 2019, accepted September 17, 2019, date of publication September 27, 2019, date of current version November 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944307

Skeleton-Based 3D Object Retrieval Using Retina-Like Feature Descriptor

XUEQING ZHAO[1,2,3], XIN SHI[1,2], BO YANG[1,2], QUANLI GAO[1,2], ZHAOFEI YU[3,4], JIAN K. LIU[4,5], YONGHONG TIAN[3,4], AND TIEJUN HUANG[3,4]

[1]Shaanxi Key Laboratory of Clothing Intelligence, School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China
[2]National and Local Joint Engineering Research Center for Advanced Networking and Intelligent Information Service, Xi'an Polytechnic University, Xi'an 710048, China
[3]National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China
[4]PengCheng Laboratory, Shenzhen 518040, China
[5]Centre for Systems Neuroscience, Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester LE1 7HA, U.K.

Corresponding authors: Zhaofei Yu (yuzf12@pku.edu.cn) and Yonghong Tian (yhtian@pku.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61806160, Grant 61806011, Grant U1611461, and Grant 61806159, in part by the Shaanxi Association for Science and Technology of Colleges and Universities Youth Talent Development Program Grant 20190112, in part by the International Talent Exchange Program of Beijing Municipal Commission of Science and Technology under Grant Z181100001018026, in part by the China Postdoctoral Science Foundation under Grant and Grant 2018M630036, in part by the National Postdoctoral Program for Innovative Talents under Grant BX20180005, and in part by the Youth Innovation Team of Shaanxi Universities.

ABSTRACT Skeleton-based 3D object retrieval is a very efficient method to query the sketch databases in numerous applications. However, few skeleton images are found so far in existing sketch benchmarks. In this paper, we provide an initial benchmark dataset consisting of skeleton sketches, including hand-drawn skeletons and skeletons extracted from 3D objects, and both of them are used to form a generic object class. Then we present a method for skeleton-based 3D object retrieval using a retina-like feature descriptor (S3DOR-RFD) based on the structural property of the human retina for processing complex visual information in a very efficient way. As part of the S3DOR-RFD algorithm, we combine artificial bee colony (ABC) in support vector machine (SVM) so as to improve the performance with automatic parameter selection, where one can make full use of the advantages of ABC and SVM to further improve the accuracy rate of 3D object retrieval. Experimental results indicate that skeleton sketches can be automatically distinguished from perspective sketches, and that the proposed S3DOR-RFD method works efficiently for selected object classes.

INDEX TERMS 3D object retrieval, feature descriptor, skeleton, retina, feature extraction.

I. INTRODUCTION

In recent years, with the rapid development of 3D object capturing techniques and computer graphics hardware, the amount of 3D objects databases are dramatic increasing in a wide range of application domains, such as medical industry [1], computer-aided design [2], virtual reality [3] and bio-informatics [4], which lead to an urgent need to propose more effective and efficient 3D object retrieval algorithms.

A main purpose of retrieval is to search for objects automatically and to assess the similarity between any pair of objects in the content [5]. For each query, such similarity is very important to implement effective retrieval algorithms, which is required to return a list of complete 3D objects retrieved from a database and ranked according to their similarity with the query. According to the types of querying objects, content-based 3D object retrieval algorithms can be divided into two major categories: model-based algorithms and view-based algorithms. Most of the early 3D object retrieval algorithms are largely belong to model-based 3D object retrieval [6], where low-level features can be directly extracted and digital represented of objects in the database [7]. Thus, there is little about human perception. In addition, high-level features can also be employed to retrieval the model-based 3D object retrieval. Many high-level features have been designed for general or specific content-based 3D object retrieval algorithms [8], and indeed some of them show good retrieval performance. However, the gap between low-level features and high-level features of the objects is the major obstacle for better retrieval

The associate editor coordinating the review of this manuscript and approving it for publication was Xiping Hu.

performance, as the 3D object information is not so easy to be obtained from real objects directly [9]. Therefore, these limitations severely affect the practical applications of model-based 3D object retrieval methods.

In recent years, view-based 3D object retrieval methods have been studied intensively, because of the high flexibility and easy implementation of 3D object representation from multiple views, which achieve a better object retrieval performance [10]. In the scheme of view-based 3D object retrieval, a 3D object is described by a single view or multiple views with more features, which can be obtained from different feature spaces, such as moment-based descriptors [11], shape-based descriptors [12], light field descriptors [13] and elevation descriptors [14]. Generally, view-based 3D object retrieval process consists of the following four parts: view capturing, view representation, feature extraction and object matching [15].

View capturing. The foundation phase of view-based 3D object retrieval is to capture a single view or multiple views, which can be directly obtained by a group of cameras or a virtual camera array from different perspectives.

View representation. More detailed information from multiple-views can be represented in a certain way to describe a 3D object.

Feature extraction. To better describe the data with sufficient accuracy and easy for human to understand interpretation, generally, only relevant features can be extracted, due to these features are highly redundant as the input data, therefore, the following two stages can be used for feature extraction: reduced representation and selected features. The complete initial data are transformed into reduced feature vectors in the first stage, and then the reduced feature vectors are selected so that they can contain all necessary information about the input data.

Object matching. It is challenging to conduct many views matching, and estimate the relevance among different 3D objects in order to obtain the ''best matching''. Distance and similarity estimation between 3D objects are usually used, like sum distance [16], Hausdorff distance [17], and bipartite graph matching [18].

In summary, to improve the accuracy performance of content-based 3D object retrieval algorithms, the essential part is to improve the feature representation. In this paper, we present a novel approach to this point. Our main contributions are as follows:

(1) Providing an initial benchmark dataset comprises skeleton,contour and perspective sketches where skeleton sketches contains not only hand-drawn skeletons but also skeletons extracted from 3D objects.

(2) Designing a retina-like feature descriptor based on the property of the human retina to easily analyze and efficiently process complex visual information. Such descriptor can be used to extract the query feature in the retrieval phase.

(3) Improving the traditional support vector machine (SVM) by using the strong global search capability of artificial bee colony algorithm to optimize the parameters for SVM.

We tested our proposed method for skeleton-based 3D object retrieval using a retina-like feature descriptor (S3DOR-RFD) on skeleton dataset, contour-sketch dataset, and perspective-sketch dataset. Experimental results indicate that skeleton sketches can be automatically distinguished from perspective sketches, and that the proposed S3DOR-RFD method works efficiently for selected object classes.

The rest of the paper is organized as follows. The proposed S3DOR-RFD is introduced in Section II; experiments and analysis are presented for retrieval capability of the proposed S3DOR-RFD in Section III. Finally, the conclusions of this paper are given in Section IV.

II. METHODS

Our work is closely related to two main topics in the literature: human retina-like feature descriptor and skeleton-based 3D object retrieval algorithms.

A. RETINA-LIKE FEATURE DESCRIPTORS (RFD)

With the feature defined as an ''interesting'' part of a 3D object, descriptors are the important tool to find out the connection between features contained in a 3D object and what humans recall after observing the 3D object or a group of 3D objects. In order to have a better human interpretation, feature descriptors have been proposed in recent years. Most of the descriptors are divided into two main groups: the first group is general information-based descriptors, in which most of the low-level features about a 3D object can be described, such as color, shape, textures, motions and some of regions [19]; the second group is specific domain information-based descriptors. They can describe some events of the specific scene. Such descriptors are most applied to face recognition and personal re-identification [20].

The retina is part of the central visual system in the brain as illustrated in Fig. 1. It is very important to assimilate and interpret information from the light in the visible spectrum, and build a representation of the surrounding environment. Many complex tasks are carried out by our visual system, for instance, light reception, monocular representation, nuclear binocular perception, objects identification, assessing distances between objects and movement guiding [21]. The retina has been suggested to play a more complex functional role for computations of the visual information [22]. Following the light information in the retina, the act of seeing object starts when the cornea and lens refract light into a small image and shine it on the retina that transduces light into electrical signals by using rods and cones cells. These electrical pulses are passed through the optic canal by optic nerves, then reached the optic chiasm. The perceived information by the retina is further processed via different parts of the cortex [23].

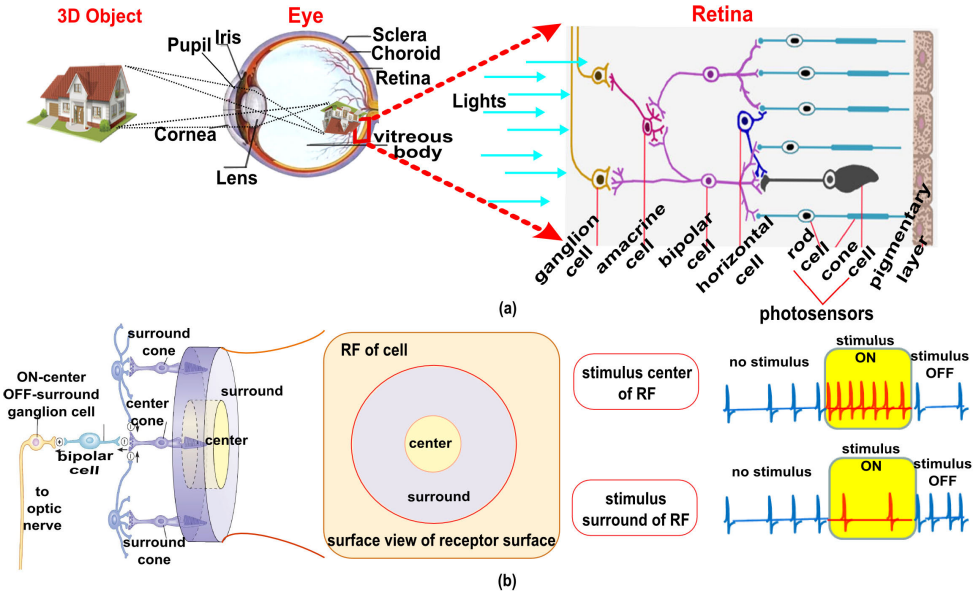


FIGURE 1. **Retina senses 3D objects. (a) eyes translate 3D objects from the outside light into keypoint features. (b) Example ganglion cell with an ON-center receptive filed.**

Traditional feature descriptors have usually been designed to describe general information and specific domain information, in which they work independently each other. However, human retina-inspired characterization can provide an innovative method to overcome the drawbacks of traditional feature descriptors. The human retina is quite different, due to its space-variant distribution of photo-sensitive cells, which can encode a large field of perceptible view with variable spatial resolution [24]. Moreover, it has been shown that the retina and visual cortex are approximated by some logarithmic-polar law [25], which results in a better property for redundancy compression and in-variance of scaling and rotation for the object. Meanwhile, those properties are beneficial for improving the efficiency of feature description for the visual object. Therefore, the human retina-like feature descriptor is more significant for designing a 3D object retrieval algorithm.

The human retina can easily analyze and efficiently process complex visual information. According to the structural property of human retina, the sampling pattern of neurons perceives visual features by increasing radii circular in terms of density of neurons. Two similar feature descriptors, FREAK (fast retina keypoint) [26] and DERF (distinctive efficient robust feature) [27] have been proposed to mimic the human retina. For the FREAK, sampling points come from a higher density of points in the center, and features are kept with a greedy selection process; while for the DERF, sampling points of the descriptor are obtained with an exponential manner. Both of them can present a set of features with adaptability and robustness.

In the retina, photoreceptors (cones and rods cells) in the first layer receive a variety of light stimuli, and at the output of the retina, light information is represented by the action potential of retinal ganglion cells (RGCs) that serve as an ultimate code for the cortex [28].

RGCs vary significantly in terms of their sizes, connections and responses to visual stimulation. There are two major regions in the retina: the central fovea retina and the remaining peripheral retina. For the central fovea, the highest RGC densities are found in a horizontally oriented and elliptical ring whose half-height extends by 0.4-2.0 mm from the foveal center along different meridians of different eyes. In contrast, the remaining peripheral retina consists of less dense ganglion cells [29]. One main feature of the RGC is its visual receptive field (RF) as a crucial role in feature description. One can use the classical linear-nonlinear model to estimate the retinal response from the perceived information in the RF [30]:

$$O^{\omega} = F(R^{\omega}) \tag{1}$$

where O is the output of retinal neuron at the position (x, y). For each cell, the input is a filtered result as

$$R^{\omega} = \omega \sum_{x^2+y^2 \le r^2} I(x, y) * G(x, y) \tag{2}$$

where $I(x, y)$ is the stimulus of the retina, ω is the strength parameter, and $G(x, y)$ is the RF function, and here the weighted Difference of Gaussians (DOG) is used as a model for RGC-receptive field at the location of (x, y) in the RF circular domain of radius r.

The unique feature of retinal cells is that there is a center-surround structure such that the center and surround respond to light with an opposite polarity. For instance, the OFF RGC refers to the cell with an OFF-center and ON-surround RF [30], which responses to bright light in the center while as dark light in the surround. In contrast, the ON RGC works

TABLE 1. Retina-like feature descriptor algorithm.

Algorithm 1 RFD algorithm

Input:
Query skeletons.

Output:
A feature vector to character the query skeletons.

Step 1: Initialization: The number of sampling circles is 6, and the details of feature extraction phase are presented in Fig. 2.

Step 2: Create a matrix with the sampling points: The sampling points are the key features of the retinal response from the perceived information in the RF, which come from the increasing radii circles. The schematic diagrams are shown in Fig. 1(b)

Step 3: DoG convolution: After a matrix with sampling points is obtained by the Step 2, ON-center and OFF-surround RF responding model is used to finish DoG convolution by using Eqs. (4)-(6), and the scale of the DoG is decided by radii of the concentric circles.

Step 4: Assemble the feature vector: In order to get a better discriminative feature vector, a hard threshold is computed by a mean value of each column in the sampling points matrix, and then decide the feature vector. The schematic diagram is shown in Fig. 2.

with an ON-center and OFF-surround RF. Both types of cells can be modeled as:

$$F(R^{\omega}) = \begin{cases} \dfrac{1}{1-\delta R^{\omega}} & if\ R^{\omega} < 0 \\ 1+\delta R^{\omega} & if\ otherwise \end{cases} \tag{3}$$

where the parameter ω defines the OFF and ON types of RGCs as ω is $+1$ and -1, respectively. δ is the control parameter.

The center-surround response of RGC can be defined as:

$$O^{\omega} = \sum_{x^2+y^2 \le r_c^2} I(x,y) * G_c(x,y) - \sum_{x^2+y^2 \le r_s^2} I(x,y) * G_s(x,y) \tag{4}$$

where r_c and r_s are radius of the center and surround, $G_c(x,y)$ and $G_s(x,y)$ are the DOG function [27]:

$$G_c(x,y) = \frac{1}{2\pi r_c^2} exp(-\frac{x^2+y^2}{2r_c^2}) \tag{5}$$

and

$$G_s(x,y) = \frac{1}{2\pi r_s^2} exp(-\frac{x^2+y^2}{2r_s^2}) \tag{6}$$

Retina-like features descriptor (RFD) used in this paper can be obtained as a series of combined RFs similar to a modified version of FREAK with a central RF consisting of 8 to 12 surrounding neighbors [30]. The details of the proposed RFD algorithm is listed in the Table 1, and illustrated in Fig.2. In this study, we have tested both FREAK and DERF descriptors, and have found that there is no real difference in terms of performance of 3D object retrieval.

B. SUPPORT VECTOR MACHINE BASED ON ARTIFICIAL BEE COLONY

To optimize the process, we use the support vector machine based on the artificial bee colony method.

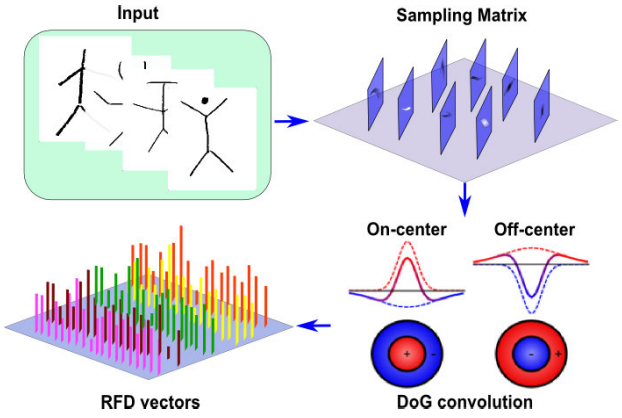


FIGURE 2. Retina-like feature descriptor.

1) SUPPORT VECTOR MACHINE

Support Vector Machines (SVM) is a set of related supervised learning methods to address classification problems [31]. SVM is widely used in classification and regression analyses, such as pattern recognition, machine learning, and data mining, whose execution is very effective, the general problems of classification and regression are greatly simplified, its decision function is determined by only a few support vectors, and the complexity depends on the number of support vectors not the dimension of the sample space, In a sense, SVM can avoid ''dimensional disaster'' very well. Moreover, such simple method has a good ''robustness'' as well. Recently, Deep Neural Network (DNN) exhibits state-of-the-art performance in many recognition fields [32], where big dataset is used in training. However, DNN commonly shows worse performance than SVM method with small dataset, and such limitation prevents the application because collecting big dataset in some field is still a challenge. Besides, DNN merely rely on the the activation functions to deal with nonlinear problems, which results in the high dependency of the selection of activation functions. Therefore, we use the SVM method with small benchmark dataset in this paper. The basic idea of SVM used in classification is described below.

An input space X has an n-dimensional object, $X = (x_1, x_2, \ldots, x_n)$, where $X \in R_n$, an output space $Y = -1, 1$ determines the learning type, such that each x_i belongs to a class Y. In a training set of objects and their classification results, $T = (x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$, any hyper plane in the space S can be described by:

$$w \cdot x + b = 0 \tag{7}$$

where $w \in S$ and $b \in R$. The SVM training is correspondingly transformed into a dual presentation problem with bound constraints:

$$\begin{cases} \max\limits_{\alpha} \Phi(\alpha) = \sum_{i=1}^{N} \alpha_i - \dfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j k(x_i, x_j) \alpha_i \alpha_j \\ s.t. \quad \sum_{i=1}^{N} y_i \alpha_j = 0, \quad 0 \le \alpha_i \le C, \ i = 1, 2, \ldots, n \end{cases} \tag{8}$$

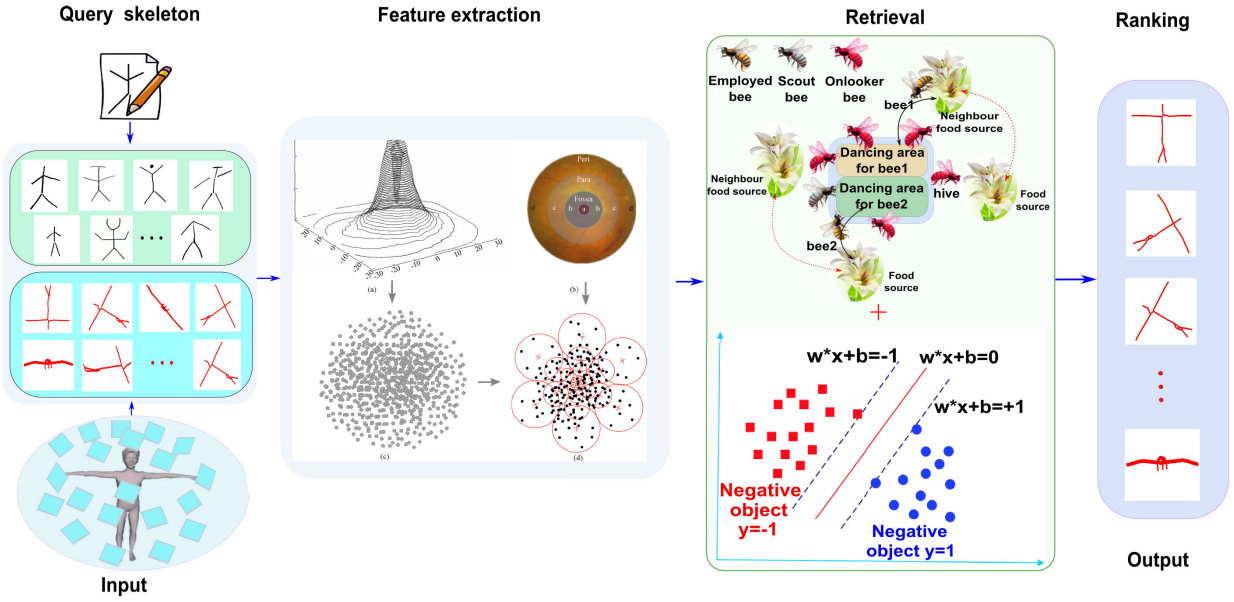


FIGURE 3. Overview of skeleton-based 3D object retrieval method.

where α_i is the Lagrange function, and C is penalty factor. $K(x_i, y_i)$ is a kernel function, here we use:

$$K(x_i, x_j) = exp(\frac{-\|x_i - x_j\|^2}{\sigma^2}) \tag{9}$$

the value of $K(x_i, x_j)$ is affected by kernel function parameters σ. This shows that the key problem is to find the optimal penalty factor C and kernel function parameters σ, in order to maximize the correct rate of classification.

In this algorithm, two different categories are separated by a linear plane. Essentially, the classification ability of SVM is greatly influenced by the kernel function and the parameters of the linear plane (see the retrieval part of in Fig.3). In most applications, the optimal parameters are difficult to be determined. In order to address this problem, the artificial bee colony-based (ABC) parameter optimization method is used in this paper, and we make the most of ABC method, that is not only the strong global search capability but also is a very simple implementation. The framework of the ABC-SVM is shown in Fig. 4.

2) ARTIFICIAL BEE COLONY

The artificial bee colony algorithm is one of the swarm intelligence methods inspired by the foraging behavior of honey bees [33]. It has been successfully applied in many fields due to its simple implementation with only a few control parameters, such as neural network training, combinatorial optimization, computer system optimization, system and engineering design [33]. The flowchart of ABC is shown in Fig. 4.

The ABC algorithm simulates three different foraging artificial bees: employed bees, onlooker bees and scout bees. Employed bees search for food and share the information around the food sources. Onlooker bees waiting on dance are to get information and make a decision to choose the food sources found by employed bees. Scout bees are those employed bees performing a random search. In ABC, the number of bees employed in the colony is equal to the number of onlooker bees. Meanwhile, the number of food sources is equal to the sum of employed bees and onlooker bees. The main steps of ABC are presented as the following parts:

Initialization. In ABC, a food source represents a candidate solution, and the nectar amount of the associated food sources is the fitness value of the solution, which is an optimization problem. The initial number of food sources (SN) is generated randomly, and each candidate solution consists of a D-dimensional parameter vector, i.e. $x_i(x_i^1, x_i^2, \ldots, x_i^D), i = 1, 2, \ldots, SN$. In order to cover the search space as much as possible, the initial food sources are uniformly placed within the search space constrained by the predefined minimum and maximum parameter bounds, i.e. $x_{min}(x_{min}^1, x_{min}^2, \ldots, x_{min}^D)$ and $x_{max}(x_{max}^1, x_{max}^2, \ldots, x_{max}^D)$. Therefore, the randomly generated SN can be defined as

$$x_i^j = x_{min}^j + rand(0, 1) \times (x_{max}^j - x_{min}^j), \tag{10}$$

where $i = 1, 2, \ldots, SN$ is the number of food sources, and $j = 1, 2, \ldots, D$, D is the number of parameters. x_{min}^j is the minimum and x_{max}^j is the maximum values of parameter j. $rand(0, 1)$ is a random value in the range of [0,1].

Employed bees. Every employed bee is associated with a specific food source. Meanwhile, For each food source i, its employed bees produce a neighboring search to generate a new food source V_i, where V_i is a new vector by updating its vector X_i. This phase can be described by the following equation:

$$v_i^j = x_i^j + \phi_i^j \times (x_i^j - x_k^j), \tag{11}$$

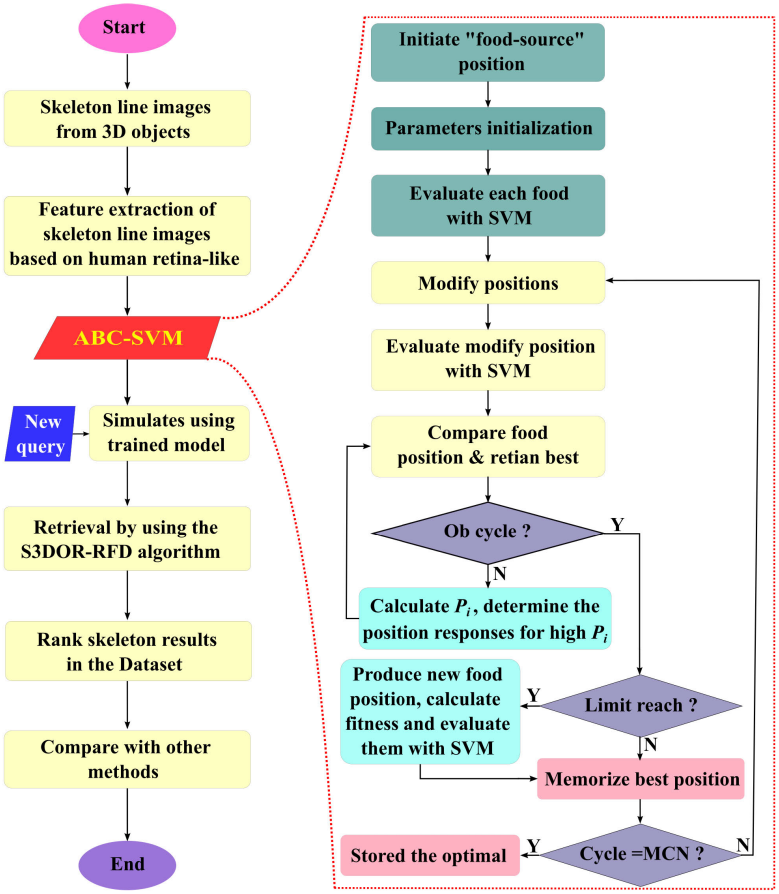


FIGURE 4. Flow chat of the S3DOR-RFD method. In the ABC-SVM flow chat, the steps of green boxes are initial bees stage, steps of yellow boxes are employed bees stage, steps of light yellow boxes are onlooker bees stage, and steps of light red are scout bees stage.

where $j \in [1, 2, \ldots, D]$ and $k \in [1, 2, \ldots, SN]$ are randomly selected parameter and neighborhoods, respectively. ϕ_i^j is a random value within $[-1, 1]$. After producing V_i, it is compared with X_i, the nectar amount (fitness value) of the food source v_i computed by the following equation:

$$fitness_i = \begin{cases} \dfrac{1}{1+fit_i} & fit_i \geq 0 \\ 1 + abs(fit_i) & fit_i < 0 \end{cases} \tag{12}$$

where fit_i is the fitness value of the food source v_i, which measures the quality of the candidate solution. That is, a greedy selection is applied to between V_i and X_i food source. If the fitness value of v_i is greater than x_i, the employed bee memorizes v_i as a new food source and leaves x_i. Otherwise, the employed bee continues to keep the current food source. When x_i is not improved, its counter representing the number of trials is incremented by 1; otherwise, it is set to 0.

Onlooker bees. Every onlooker bee selects a food source via employed bees waggle dance on the dance area, and further searches its neighboring area. The phase is completed based on evaluating the probability of the nectar amount from the shared information by employed bees. In ABC, the probability is computed by:

$$p_i = \frac{fitness_i}{\sum_{i=1}^{SN} fitness_i}. \tag{13}$$

The probability p_i is used to increase the chance of selecting the food sources with the high nectar amount, which can be produced by a positive feedback feature in ABC. Therefore, the chance of finding the most valuable food sources is also increased.

Scout bees. Scout bee can decide a food source as an abandoned food source or not. Abandoned food source can be replaced with a new randomly produced food sources, after the employed bees and the onlooker bees have finished their searches. In this stage, the number of continuously failed trials, $trial_i$ is checked. Once it is greater than the predefined threshold limit, this food source i is considered as an abandoned food source, and subsequently, a scout bee randomly discovers a new food source, which is replaced by a new abandoned one, and its $trial_i$ will be reset to 0. The property of this process is a negative feedback in the ABC algorithm to generate a food source randomly as specified.

The proposed ABC-based SVM parameters optimization algorithm is listed in the Table 2.

TABLE 2. ABC-based SVM parameters optimization algorithm.

Algorithm 2 ABC-based SVM parameters optimization algorithm

Input:
Original parameters from SVM.
Output:
Optimal parameters for SVM.
Step 1: Initialization:
The initial population food sources (SN) is 20, the maximum number of cycles for food sources limit is 50, and the number of cycles terminated MCN is 100.
Step 2: Employed Bee:
Produce a new food resource V_i for the employed bee of the food resource X_i using Eq. (7) and calculate its amount.
Greedy selection is applied to choose the better one between V_i and X_i food source.
Step 3: Onlooker Bee:
Evaluate the probability values p_i using Eq. (8) to increase the chance of selecting the SN with the high nectar amount and to get a better solution by fitness criteria.
Generate a new V_i food place using Eq. (9) for onlooker bees.
Exploit a greedy choice procedure V_i and X_i and choose the superior one by using the modify $trial_i = trial_i + 1$.
Step 4: Scout Bee:
Replace X_i by means of a new randomly created solution by using $max(trial_i) > limit$.
Step 5: Memorize the best solution achieved so far:
Remember the finest solution obtained so far.
Step 6: Until a termination criterion is satisfied:
Until (cycle = max cycle number)

C. SKELETON-BASED 3D OBJECT RETRIEVAL ALGORITHMS

Object skeleton is a succinct and effective geometric tool for shape analysis in that it can capture the topological structure of the primitive shape. It has been widely used in shape deformation [34], object registration [35], human dynamics recognition [36], object retrieval [37], symmetry detection [38] and sketch-based modeling [39], due to the advantages of invariance to different views, high-level information abstraction, robustness to illumination and clustered background [40]. Thus, it is shown that skeleton-based methods can outperform view-based methods for the same classification tasks [41].

In terms of 3D object retrieval, skeleton-based algorithms also attract considerable attention, and are widely used in human action recognition [42], 3D shape matching, and retrieval [43]. For human action recognition, the skeleton data usually can be summarized as a set of human dynamics in the video, where the skeleton structures reflect the individual appearance of the human body. For 3D shape matching and retrieval, the skeleton is a useful geometric tool for 3D shape representation, due to the following major advantages:

Succinct representation: Skeleton can represent the 3D shape with high-level information without any irrelevant signals, and redundant information is minimized.

Invariant viewpoint: Skeleton can feasibly select the potential viewing angles from the human user to draw a 3D object including as many views as possible.

Effective matching: Skeleton-based matching can be adapted to as a part of 3D object matching, e.g. whether the object to be queried can be matched as a part of a large object.

TABLE 3. Proposed S3DOR-RFD retrieval algorithm.

Algorithm 3 S3DOR-RFD algorithm

Input:
Query skeletons.
Output:
Retrieval results.
Step 1: Extract skeleton features by using the Algorithm 1. See the RFD algorithm listed in Table 1.
Step 2: Use K-means to cluster the features and set K to be 100 in this paper.
Step 3: Classify skeletons by using SVM classifier, which is optimized by ABC.
Step 4: Ranking skeletons according to the similarity between the query skeleton and other skeletons come from dataset.

The proposed S3DOR-RFD retrieval algorithm is listed in table 3. The full chart flow of Skeleton-based 3D object retrieval process is illustrated in Fig. 4.

III. EXPERIMENTS

In this section, we present extensive experimental results for our proposed S3DOR-RFD on a series of skeletons for 14 object classes from the Konstanz database [44] and EITZ database [45]. In order to compare the performance of our proposed S3DOR-RFD from three aspects: the first one is the effectiveness of the extracted features, we select two traditional, classic and effective methods, HOG-SVM [46] and SIFT-SVM [47], which are widely used to extract feature and performance comparison, and retina-based feature extraction method REF-SVM without ABC parameter optimization is also used to compare in our experimental results; the second one is the effectiveness of SVM-based methods with our small benchmark dataset; the third one is the robustness to noises. All selected experiments were simulated with Matlab on PC with the following specifications. CPU: Intel(R) Core(TM) i7-6500U 2.50 GHz; RAM: 8GB DDR3L; OS: Windows10 SP1 of 64 bits.

This method not only is simple in algorithm, but also has good ''robustness''.

A. DATABASES

Drawn skeletons benchmark: a test dataset was compiled by inviting 50 users to draw skeleton line images for a list of 14 selected classes and for which skeleton line images can be meaningfully created. Our users were given the list of class names with no further information except to ask them to draw a perspective, contour and a skeleton line sketch for 14 object classes (see Fig. 5), see our previous work [48] and [49]. Each one of the collected hand-drawn images was scanned, cropped, filtered for noise, deblurred and finally converted to a binary image.

Contour and perspective sketch benchmark: we obtained the contour sketch dataset and perspective sketch dataset by using manual classification from a large number of object sketches in Eitz database [45], the number of classified contour sketches and perspective sketches is the same as the number of drawn skeletons, and the number of species is 14.

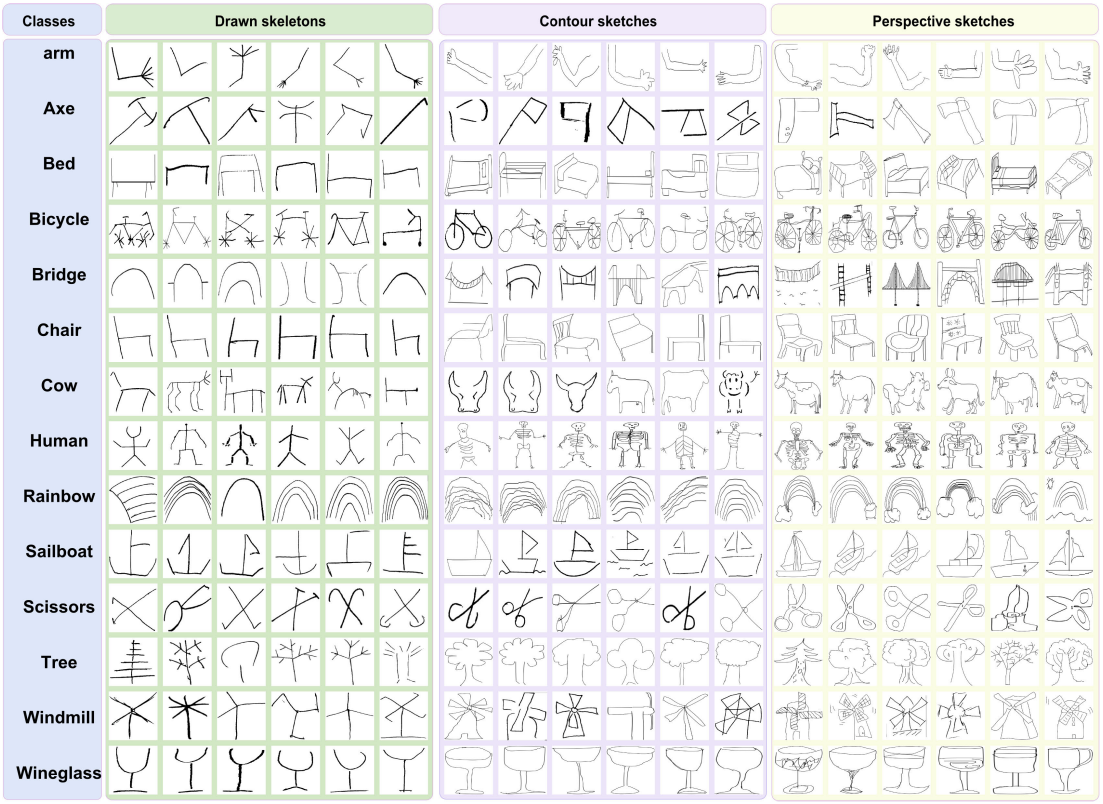


FIGURE 5. Drawn skeletons (left), contour sketches (middle), and perspective sketches (right) for 14 selected classes (top to bottom).

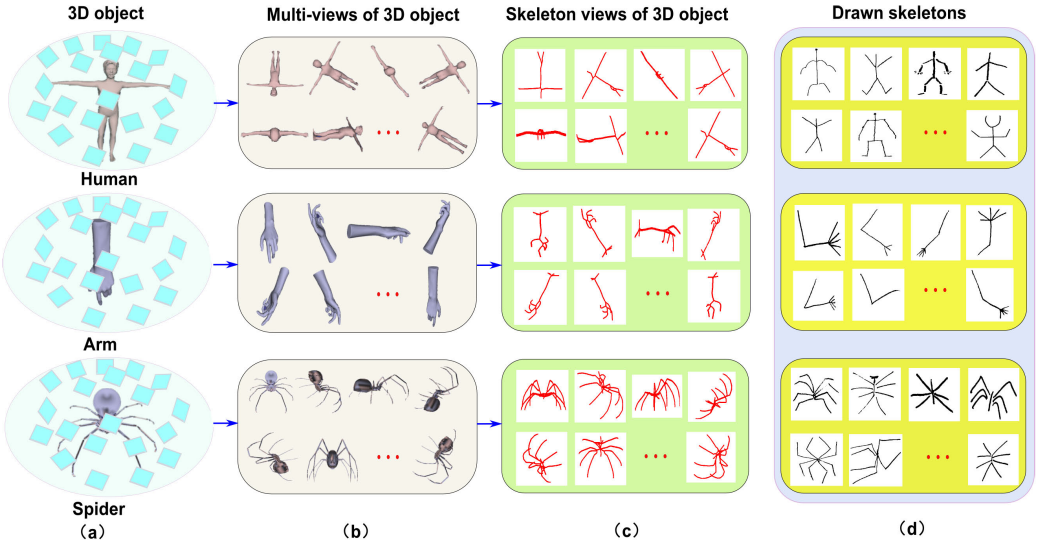


FIGURE 6. View examples of 3D objects and skeletons.

We described the contour sketch and the perspective sketch as the following:

Contour sketch: That is contour drawing, the contour of a subject was drawn in lines, which are essentially an outline, it can express a 3D perspective, length and width, in order to emphasize the shape and structure of the subject rather than the details.

Perspective sketch: Perspective is an approximate representation, generally on a flat surface, like an image as it is seen by the eye, that objects appear smaller as their distance

from the observer increase. In the perspective sketches, more details features are included, which can be better express more information of the 3D object.

Skeletons extract from 3D objects benchmark: We chose a subset of classes from the Konstanz database [44] that match the classes of our skeleton images. From these 3D models, we generated skeleton images from 27 different views of each 3D object to compare each of them against the retina-like features which were extracted from the query, see Fig. 6(a-c).

B. VIEWS EXAMPLES OF 3D OBJECTS

We extracted a number of skeletons from 27 different views of each 3D object, which come from the Konstanz database, in order to express the 3D structures of objects. Three randomly selected 3D objects (human, arm and spider) are shown in Fig. 6 (a) their corresponding 27 views and different skeleton views are shown in Fig. 6 (b) and Fig. 6 (c). Each of skeleton view of size 64×64, which can better describe the details of the 3D object.

C. COMPARISON OF THE DRAWN SKELETON VIEWS AND SKELETONIZATION BASED ON 3D CLOUD POINT VIEWS

Three selected example 3D objects (human, arm and spider) and their corresponding drawn skeletons are shown in Fig. 6(d), where the hand drawn skeleton views better describe the structural details of the 3D objects. For simple shapes such as human and arm, the hand-drawn skeleton view is highly efficient at describing the structural from top to bottom. For complex shapes, like spider, the hand-drawn skeleton views provide a better representation of some different views of the shape and detailed features of the connection part.

D. 3D OBJECT RETRIEVAL BASED ON SKELETON LINE IMAGE

In order to test our proposed S3DOR-RFD, we implement three comparative experiments: 1) retrieval in skeleton dataset obtained from 3D objects in Konstanz database and drawn skeletons; 2) retrieval in contour-sketch dataset; 3) retrieval in perspective-sketch data set. At present, we have collected 691 skeleton-based sketches, 1822 contour and perspective sketches. In order to compare the performance of our proposed S3DOR-RFD, HOG-SVM [46], SIFT-SVM [47], REF-SVM are chosen as comparative methods.

1) RETRIEVAL IN SKELETON DATA SET

The first stage that we compare the four methods, including HOG-SVM, SIFT-SVM, REF-SVM and our proposed S3DOR-RFD by retrieving the skeletons in skeleton-based sketches itself. As shown in Fig. 7(a), S3DOR-RFD outperforms the other three methods with an average accuracy rate of 86.78%. From Fig. 7(a), we can easily see that the proposed S3DOR-RFD has better accuracy rate.

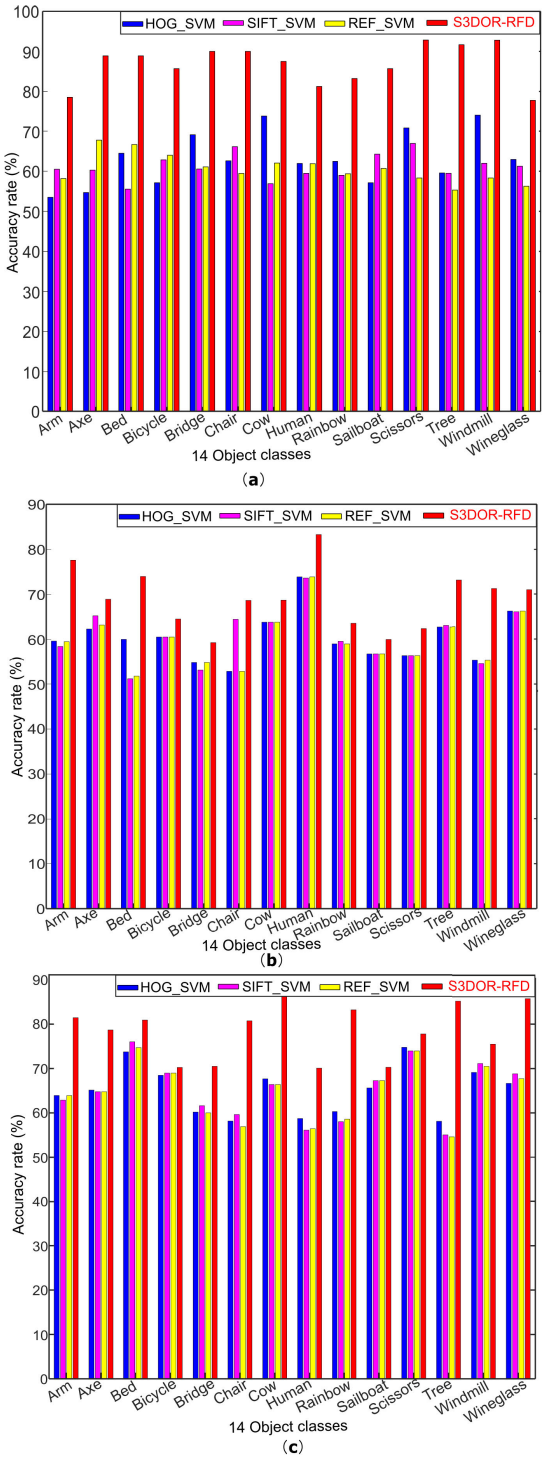


FIGURE 7. Results of retrieval in (a) skeleton data set, (b) contour-sketch data set, (c) perspective-sketch data set.

2) RETRIEVAL IN CONTOUR-SKETCH DATA SET

The second stage is done in the contour-sketch data set, which the skeleton can be used in retrieval by HOG-SVM, SIFT-SVM, REF-SVM and our proposed S3DOR-RFD.

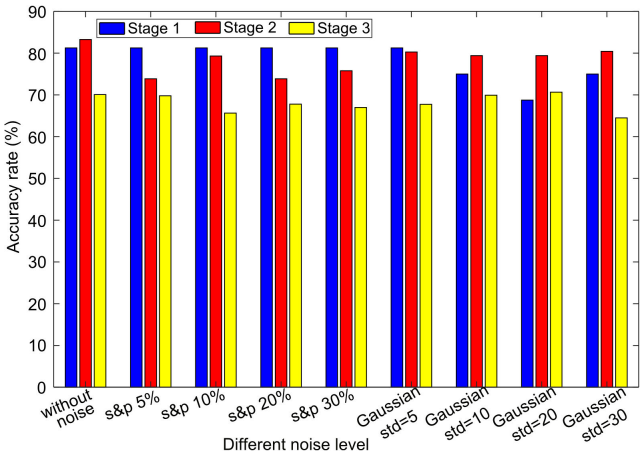


FIGURE 8. Robustness of S3DOR-RFD at the different level of noise on above three comparative experiments of skeleton (Stage 1), contour-sketch (Stage 2), perspective-sketch (Stage 3) data sets.

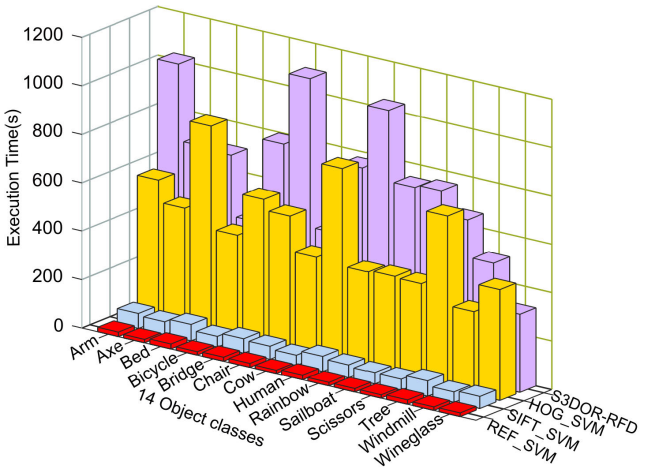


FIGURE 9. Execution Time for 14 object classes.

As shown in Fig. 7(b), our proposed S3DOR-RFD outperforms the other three methods, and the accuracy rate is higher than others from Fig. 7(b).

3) RETRIEVAL IN PERSPECTIVE-SKETCH DATA SET

The third stage is completed in the perspective-sketch data set. We query the data set by using the skeleton, and the retrieval-ling results are shown in Fig. 7(c). Meanwhile, the comparative methods are used to retrieval in the same data set, like HOG-SVM, SIFT-SVM, REF-SVM and our proposed S3DOR-RFD. As shown in Fig. 7(c), the performance of our proposed S3DOR-RFD is significantly better than others.

E. ROBUSTNESS ANALYSIS

In order to investigate whether the proposed S3DOR-RFD method is robust to noise, we randomly add different level Salt & Pepper and white Gaussian noise to the retrieval stages. The mean accuracy values of the retrieval are shown in Fig. 8. For stage 1, retrieval in skeleton data set, it can be easily seen that it is very stable for the Salt & Pepper intensity 5% increased to 30%, and for low level white Gaussian noise like the standard deviation is 5. However, when the Gaussian noise density is increased to 10, 20 and 30, it lacks stability to some extent; for stage 2, the changes to Salt & Pepper and white Gaussian noise are obvious; for stage 3, the changes to the different level of Salt & Pepper and white Gaussian noise remain basically stable, and the amount of drop is smal which shows the noise-resistance of our algorithm.

F. EXECUTION TIME

A comparison of execution time for the proposed S3DOR-RFD is simulated on 14 objects, and the comparative methods are also executed in our simulation, Fig. 9 shows execution time relative to the different retrieval stages. As Fig. 9 shows, the execution time of REF-SVM is fast, and the second one is SIFT-SVM. For HOG-SVM and our proposed S3DOR-RFD are longer compared to the above two methods. For our proposed S3DOR-RFD, the time efficiency is lower because the artificial bee colony algorithm is used to optimal parameters during the retrieval stage.

IV. CONCLUSION

In this paper, we presented a method for skeleton-based 3D object retrieval using human retina-like feature descriptor, shortened for S3DOR-RFD. We collected and provided an initial benchmark for skeleton-based 3D object retrieval, which is based on more than 2500 user-drawn sketches. Moreover, a human retina-like feature descriptor is used to extract features in retrieval for skeleton sketch queries. Meanwhile, the artificial bee colony-based parameter optimization method of support vector machine is used to classify the skeletons in datasets. Experimental results show that our proposed S3DOR-RFD has the efficiency for some 14 classes in the benchmark.

Future work is needed to extend the classification of skeleton images against perspective images. In addition, human retina-like feature descriptor can be useful for retrieval in combination with an appropriate skeletonization. One also needs to optimize the retrieval for skeleton sketch queries. Many sketch-based methods have been proposed, which need to be evaluated for our data. Also, our provided data set is a starting point but more classes are needed. We expect that, for skeleton sketches, a reasonably optimized skeleton retrieval will be able to outperform standard sketch retrieval based e.g., on perspective rendering like Suggestive Contours, simply because the rendering step is closer to the abstraction made by the user when submitting sketch-based queries. Eventually, a full retrieval system should include a classification stage that detects the type of sketch including perspective, orthogonal or skeleton, and applies the best-suited view generation and feature extraction to carry out the search.

REFERENCES

[1] S. Tulsiani, A. Kar, J. Carreira, and J. Malik, "Learning category-specific deformable 3D models for object reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 719–731, Apr. 2017.

[2] N. Bilalis and E. Maravelakis, "Computer-aided design," in *Advances in Manufacturing and Processing of Materials and Structures*. Boca Raton, FL, USA: CRC Press, 2018, pp. 15–50.

[3] T. Teo, M. Norman, M. Adcock, and B. H. Thomas, "Data fragment: Virtual reality for viewing and querying large image sets," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2017, pp. 327–328.

[4] C. T. Yang, D. Ghosh, and F. Beaudry, "Detection of gelatin adulteration using bio-informatics, proteomics and high-resolution mass spectrometry," *Food Additives Contaminants A*, vol. 35, no. 4, pp. 599–608, 2018.

[5] B. Bustos, D. Keim, D. Saupe, and T. Schreck, "Content-based 3D object retrieval," *IEEE Comput. Graph. Appl.*, vol. 27, no. 4, pp. 22–27, Jul./Aug. 2007.

[6] M. Günther, T. Wiemann, S. Albrecht, and J. Hertzberg, "Model-based furniture recognition for building semantic object maps," *Artif. Intell.*, vol. 247, pp. 336–351, Jun. 2017.

[7] D. V. Vranic, D. Saupe, and J. Richter, "Tools for 3D-object retrieval: Karhunen–Loeve transform and spherical harmonics," in *Proc. IEEE 4th Workshop Multimedia Signal Process.*, Oct. 2001, pp. 293–298.

[8] C. Leng, H. Zhang, B. Li, G. Cai, Z. Pei, and L. He, "Local feature descriptor for image matching: A Survey," *IEEE Access*, vol. 7, pp. 6424–6434, 2019.

[9] H. Zeng, Q. Wang, and J. Liu, "Multi-feature fusion based on multi-view feature and 3D shape feature for non-rigid 3D model retrieval," *IEEE Access*, vol. 7, pp. 41584–41595, 2019.

[10] D. Roobaert and M. M. Van Hulle, "View-based 3D object recognition with support vector machines," in *Proc. 9th IEEE Signal Process. Soc. Workshop Neural Netw. Signal Process.*, Aug. 1999, pp. 77–84.

[11] S. Li, M.-C. Lee, and C.-M. Pun, "Complex Zernike moments features for shape-based image retrieval," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 39, no. 1, pp. 227–237, Jan. 2009.

[12] D. Zhang and G. Lu, "Shape-based image retrieval using generic Fourier descriptor," *Signal Process., Image Commun.*, vol. 17, no. 10, pp. 825–848, Nov. 2002.

[13] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, "On visual similarity based 3D model retrieval," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 223–232, Sep. 2003.

[14] J.-L. Shih, C.-H. Lee, and J. T. Wang, "A new 3D model retrieval approach based on the elevation descriptor," *Pattern Recognit.*, vol. 40, no. 1, pp. 283–295, 2007.

[15] S. Zhao, H. Yao, Y. Zhang, Y. Wang, and S. Liu, "View-based 3D object retrieval via multi-modal graph learning," *Signal Process.*, vol. 112, pp. 110–118, Jul. 2015.

[16] M. Tom and M. S. Sunitha, "Sum distance in fuzzy graphs," *Ann. Pure Appl. Math.*, vol. 7, no. 2, pp. 73–89, 2014.

[17] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[18] K. Riesen and H. Bunke, "Approximate graph edit distance computation by means of bipartite graph matching," *Image Vis. Comput.*, vol. 27, no. 7, pp. 950–959, 2009.

[19] S. Gauglitz, T. Höllerer, and M. Turk, "Evaluation of interest point detectors and feature descriptors for visual tracking," *Int. J. Comput. Vis.*, vol. 94, no. 3, p. 335, 2011.

[20] Q. Leng, M. Ye, and Q. Tian, "A survey of open-world person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

[21] C. A. Curcio, K. R. Sloan, Jr., O. Packer, A. E. Hendrickson, and R. E. Kalina, "Distribution of cones in human and monkey retina: Individual variability and radial asymmetry," *Science*, vol. 236, no. 4801, pp. 579–582, 1987.

[22] T. Gollisch and M. Meister, "Eye smarter than scientists believed: Neural computations in circuits of the retina," *Neuron*, vol. 65, no. 2, pp. 150–164, Jan. 2010.

[23] M. R. Hee, J. A. Izatt, E. A. Swanson, D. Huang, J. S. Schuman, C. P. Lin, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography of the human retina," *Arch. Ophthalmol.*, vol. 113, no. 3, pp. 325–332, Mar. 1995.

[24] J. Najemnik and W. S. Geisler, "Optimal eye movement strategies in visual search," *Nature*, vol. 434, no. 7031, pp. 387–391, 2005.

[25] J. Cao, Q. Hao, W. Xia, Y. Peng, Y. Cheng, J. Mu, and P. Wang, "Design and realization of retina-like three-dimensional imaging based on a MOEMS mirror," *Opt. Lasers Eng.*, vol. 82, pp. 1–13, Jul. 2016.

[26] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 510–517.

[27] D. Weng, Y. Wang, M. Gong, D. Tao, H. Wei, and D. Huang, "DERF: Distinctive efficient robust features from the biological modeling of the P ganglion cells," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2287–2302, Aug. 2015.

[28] Z. Jiang, W. W. S. Yue, L. Chen, Y. Sheng, and K.-W. Yau, "Cyclic-nucleotide- and HCN-channel-mediated phototransduction in intrinsically photosensitive retinal ganglion cells," *Cell*, vol. 175, no. 3, pp. 652–664, Oct. 2018.

[29] E. M. Martersteck, K. E. Hirokawa, M. Evarts, A. Bernard, X. Duan, Y. Li, L. Ng, S. W. Oh, B. Ouellette, J. J. Royall, M. Stoecklin, Q. Wang, H. Zeng, J. R. Sanes, and J. A. Harris, "Diverse central projection patterns of retinal ganglion cells," *Cell Rep.*, vol. 18, no. 8, pp. 2058–2072, Feb. 2017.

[30] B. A. Rheaume, A. Jereen, M. Bolisetty, M. S. Sajid, Y. Yang, K. Renna, L. Sun, P. Robson, and E. F. Trakhtenberg, "Single cell transcriptome profiling of retinal ganglion cells identifies cellular subtypes," *Nature Commun.*, vol. 9, no. 1, 2018, Art. no. 2759.

[31] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 2013.

[32] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[33] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Appl. Soft Comput.*, vol. 8, pp. 687–697, Jan. 2008.

[34] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Robust video surveillance for fall detection based on human shape deformation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 611–622, May 2011.

[35] M. Uenohara and T. Kanade, "Vision-based object registration for real-time image overlay," in *Computer Vision, Virtual Reality and Robotics in Medicine*. Springer, 1995, pp. 13–22.

[36] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1473–1488, Nov. 2008.

[37] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1470–1477.

[38] N. J. Mitra, L. J. Guibas, and M. Pauly, "Partial and approximate symmetry detection for 3D geometry," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 560–568, 2006.

[39] L. Olsen, F. F. Samavati, M. C. Sousa, and J. A. Jorge, "Sketch-based modeling: A survey," *Comput. Graph.*, vol. 33, pp. 85–103, Feb. 2009.

[40] F. O. Bochud, C. K. Abbey, and M. P. Eckstein, "Statistical texture synthesis of mammographic images with clustered lumpy backgrounds," *Opt. Express*, vol. 4, no. 1, pp. 33–43, 1999.

[41] A. Yao, J. Gall, G. Fanelli, and L. Van Gool, "Does human action recognition benefit from pose estimation?" in *Proc. 22nd Brit. Mach. Vis. Conf. (BMVC)*, 2011, pp. 67.1–67.11.

[42] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 1, no. 2, pp. 4263–4270.

[43] J. Xie, G. Dai, F. Zhu, E. K. Wong, and Y. Fang, "Deepshape: Deep-learned shape descriptor for 3D shape retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1335–1345, Jul. 2017.

[44] B. Bustos, D. A. Keim, D. Saupe, T. Schreck, and D. V. Vranić, "Feature-based similarity search in 3D object databases," *ACM Comput. Surv.*, vol. 37, no. 4, pp. 345–387, Dec. 2005.

[45] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects?" *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–44, Jul. 2012.

[46] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 1, no. 1, pp. 886–893.

[47] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Sep. 1999, pp. 1150–1157.

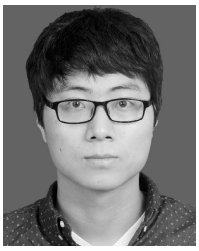
[48] Z. Xueqing, R. Gregor, P. Mavridis, and T. Schreck, "Sketch-based 3D object retrieval with skeleton line views—Initial results and research problems," in *Proc. EG Workshop 3D Object Retr.*, 2017, pp. 55–58.

[49] Z. Xueqing, M. Pavlos, S. Tobias, and R. Matthias, "A high-performance approach for enhancing the clarity of hand-drawn sketch images," *Basic Sci. J. Textile Univ.*, vol. 31, no. 2, pp. 112–120 and 135, 2018.

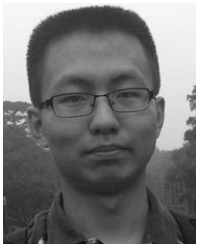
XUEQING ZHAO received the B.Sc. degree in electronic information science and technology, the M.Sc. degree in biomedical/medical engineering, and the Ph.D. degree in computer software and theory from the Shaanxi Normal University, Xi'an, China, in 2007, 2010, and 2013. She was a Visiting Scholar with the National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, from 2018 to 2019. She was a Visiting Scholar with the Institute for Computer Graphics and Knowledge Visualization, Graz University of Technology, Austria, from 2016 to 2017. She is currently a Lecturer of computer science with Xi'an Polytechnic University, Xi'an, China. Her current research interests include visual computation, machine learning, 3D objects recognition, and digital image processing.

XIN SHI received the master's degree in digital media technology from Nanyang Technological University, Singapore, in 2013. She joined the School of Computer Science, Xi'an Polytechnic University, Xi'an, China, as a Laboratory Tutor, in 2013. Her current research interests include but are not limited to pattern recognition, image processing, and brain-like calculation.

BO YANG received the Ph.D. degree in computer science and technology from Xi'an Jiaotong University, Xi'an, China, in 2017. Since June 2017, he has been a Postdoctoral Fellow with the School of Computer Science, Northwestern Polytechnical University, Xi'an. He joined the School of Computer Science, Xi'an Polytechnic University, as a Lecturer, in 2017. His current research interests include artificial intelligence, machine learning, data mining, pattern recognition, and bioinformatics.



QUANLI GAO received the B.S. degree in information and computer science, in 2010, and the Ph.D. degree from the Department of Information Science and Technology, Northwest University, Xi'an, China. He is currently a Lecturer of computer science with Xi'an Polytechnic University, Xi'an. His research interests include recommender systems, machine learning, and he has participated in several national research projects.



ZHAOFEI YU received the B.S. degree from the Hong Shen Honors School, College of Optoelectronic Engineering, Chongqing University, Chongqing, China, in 2012, and the Ph.D. degree from the Automation Department, Tsinghua University, Beijing, China, in 2017. He is currently a Postdoctoral Fellow with the National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing. His current research interests include artificial intelligence, brain-inspired computing, and computational neuroscience.

JIAN K. LIU received the Ph.D. degree in mathematics from UCLA, in 2009. He is currently a Lecturer with the Centre for Systems Neuroscience, University of Leicester, U.K. His areas of research include computational neuroscience and brain-like computation.

YONGHONG TIAN received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005. He is currently a Full Professor with the National Engineering Laboratory for Video Technology and the Cooperative Medianet Innovation Center, School of Electronics Engineering and Computer Science, Peking University, Beijing. He has authored or coauthored over 160 technical articles in refereed journals and conferences. He has owned over 55 Chinese and U.S. patents. His research interests include machine learning, computer vision, and multimedia big data.

TIEJUN HUANG received the master's and bachelor's degrees in computer science from the Wuhan University of Technology, Wuhan, China, in 1995 and 1992, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from Huazhong (Central China) University of Science and Technology, Wuhan, in 1998. He is currently a Professor with the School of Electronic Engineering and Computer Science, Peking University, Beijing, China, where he is also the Director of the Institute for Digital Media Technology. His research area includes video coding, image understanding, digital right management, and digital library. He has authored or coauthored over 100 peer-reviewed articles and three books. He is a member of the Board of Director for Digital Media Project, the Advisory Board of the IEEE Computing Society, and the Board of the Chinese Institute of Electronics.