Neurocomputing 195 (2016) 40–49

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

3D object retrieval with multi-feature collaboration and bipartite graph matching

Yan Zhang [a], Feng Jiang [a,*], Seungmin Rho [c], Shaohui Liu [a], Debin Zhao [a], Rongrong Ji [b]

[a] School of Computer Science, Harbin Institute of Technology, Harbin 150001, China
[b] School of Information Science and Technology, Xiamen University, 361005, China
[c] Department of Multimedia, Sungkyul University, Anyang, Korea

ARTICLE INFO

Article history:
Received 6 April 2015
Received in revised form
23 August 2015
Accepted 19 September 2015
Available online 3 March 2016

Keywords:
View-based retrieval
Multi-feature
Bipartite graph

ABSTRACT

In this paper, we propose a novel 3D object retrieval with features collaboration and bipartite graph matching strategies. We explored the essential characters of 3D object in a view-based retrieval framework, which extracts complement descriptors from both the contour and the interior region of 3D object effectively. Specifically, a greedy bipartite graph matching algorithm is employed. With the bipartite graph matching and feature concatenation, significant performance improvement is achieved in the 3D object retrieval task. The proposed method is evaluated by the third party on the data set comprising more than 500 3D objects and achieves the best performance for SHREC'15 challenge.

1. Introduction

With the rapid development of 3D technologies, computer graphics hardware and networks, 3D objects have been widely explored in plenty of applications [23,24], especially in architecture design [1], movie production, 3-D graphics, and the medical industry, which leads to the eager requirement of effective and efficient 3D object retrieval. Based on different data type adopted, 3D object retrieval methods can be roughly categorized into two groups [8,13]: 3D models based [2,3] and multiple views based [4,5].

In 3D model-based methods, each 3D object is represented by a virtual 3D model with geometry-based methods. To describe the information of 3D models, 3D objects are described with model-based features, such as low-level feature (e.g. the volumetric descriptor [25], the surface distribution [26] and surface geometry [27–29]). With the 3D model data, 3D model-based methods can preserve the global spatial information of 3D objects; while in some cases when we want to search the objects in the world, 3D model information is not available. For example, when the tourist finds some interesting things and wants to find similar ones in the dataset, it is hard to obtain the model information but just take several pictures. Some method [17] employs a set of 2D images to constructs 3D model, but it is both time consuming and fine sampling. Due to the expensive computational costs and the limitation of obtaining explicit 3D information, the poor performance of reconstructions methods often results in low-quality 3D models, which restricts the development of model-based 3D parsing methods in some practical applications.

Different from the 3D model-based methods, the view-based 3D object retrieval methods use a group of images from different directions for 3D object representation [30,31] and the matching between two 3D objects is accomplished via multiple-view matching. These views may be captured with a static camera array or without such camera array constraint. Such view-based methods release the restriction of 3D model, and the ubiquity of mobile devices with cameras makes it convenient to capture real objects images. Besides, online multiview data of 3D objects have become increasingly available on websites, which facilitates the practical application for view-based method. Due to convenient obtainment and bargain price of equipment, plenty of researchers pitch into multi-view based 3D object retrieval methods [6–8], recently. On the one hand, View-based retrieval may learn nutrition from large quantity studies of visual parsing techniques, like search [9,10], segmentation [11] and tracking [12] etc. On the other hand, it's greatly flexible to represent a 3D object by a set of 2D views. For location-based mobile applications, view based methods also provide new search opportunities with the help of cameras. Compared with model-based methods, view-based methods is more discriminative for 3D objects, which can lead to better object retrieval performance[32,33].

Here, we focus on the recent progress in view-based 3D object retrieval, which has been widely used in CAD applications. For

* Corresponding author.
E-mail addresses: Y.zhang@hit.edu.cn (Y. Zhang), fjiang@hit.edu.cn (F. Jiang), korea.smrho@gmail.com (S. Rho), shliu@hit.edu.cn (S. Liu), dbzhao@hit.edu.cn (D. Zhao), rrji@xmu.edu.cn (R. Ji).

http://dx.doi.org/10.1016/j.neucom.2015.09.118
0925-2312/

example, We first survey the key technologies and challenges in view-based 3D object retrieval and then discuss the state-of-the-art methods and future research directions in the field. Different from general classifier [38], an ideal view-based 3D object retrieval should has the abilities of exploring effective strategy to conduct multi-view matching and estimate the relevance among different 3D objects. Aiming at utilizing the collaborated features for multiple views and conducting multi-view matching and estimating the relevance among different 3D objects, this paper proposes a novel 3D object retrieval with features collaboration and bipartite graph matching strategies, and the main contributions of this paper are summarized as follows:

- Inspired by the essential characters of 3D object, a view-based retrieval framework with multi-feature collaboration and bipartite graph matching is proposed, which extracts complement descriptors from both the contour and the interior region of 3D object effectively.
- A Greedy Search (GS) algorithm is proposed to calculate the similarity of query object and object, and three bipartite graphs are employed to obtain the optimal match of each bipartite graph pair.

Our proposed method participated the SHREC'15 challenge and achieved the state-of-the-art performance in 0.6174 (top 10 precision). The SHREC'15 challenge is designed for 3D object retrieval which is modeled by multi-view to simulate the true environment.

The remainder of this paper is organized as follows. Related work is reviewed in Section 2. The detail descriptions of the proposed method are presented in Section 3. Extensive experimental results are reported in Section 4. Section 5 concludes the paper.

2. Related work

Due to convenient obtainment and bargain price of equipment, plenty of researchers pitch into multi-view based 3D object retrieval methods [6,7], recently. On the one hand, View-based retrieval may learn nutrition from large quantity studies of visual parsing techniques, like search [9,10], segmentation [11] and tracking [12] etc. On the other hand, it is greatly flexible to represent a 3D object by a set of 2D views. And abundant information can be fetched by multi-view object representation. Meanwhile it is a big challenge to compare two groups of views [36,37]. In [34], Firstly, the views of the query object are grouped into clusters where query models are trained by the hierarchical agglomerative clustering method [14]. And a positive matching model and a negative matching model are trained according to positive matched samples and negative matched samples. Then, the generated query model is learned for 3-D object retrieval. To calculate the distance between Query object Q and Dataset object M, similarity is defined as following:

$$S(Q,M) = p(M|Q,\Delta=1) - p(M|Q,\Delta=0) \tag{2.1}$$

where $p(M|Q,\Delta=1)$ denotes that the probability of M given Q, when M is relevant to Q and $p(M|Q,\Delta=0)$denotes the probability of M given Q, when M is not relevant to Q . For Positive Matching Model, the similarity of two object will be calculated by the similarity of feature vector, which conditional density is defined as $p(\phi|\varphi,\Delta=1)=N(\phi|\varphi,\sigma_{pos}^2)$. Where σ_{pos}^2 is the positive variance when ϕ and φ are relevant. The conditional density of which object Q and M are relevant becomes

$$p(M|Q,\Delta=1) = p\left(\left\{\phi_1^{\overline{M}},\phi_2^{\overline{M}},\ldots,\phi_\tau^{\overline{M}}\right\}\middle|\left\{\varphi_1^Q,\varphi_2^Q,\ldots,\varphi_m^Q\right\},\Delta=1\right) \tag{2.2}$$

And for negative matching model, the equation above becomes

$$p(\phi|\varphi,\Delta=0) = N(\phi|\varphi,\sigma_{neg}^2)$$
$$p(M|Q,\Delta=0) = p\left(\left\{\phi_1^{\overline{M}},\phi_2^{\overline{M}},\ldots,\phi_\tau^{\overline{M}}\right\}\middle|\left\{\varphi_1^Q,\varphi_2^Q,\ldots,\varphi_m^Q\right\},\Delta=0\right) \tag{2.3}$$

And the optimal retrieval result should satisfy

$$\text{argmax}S(Q,M) = \text{argmax}p(M|Q,\Delta=1) - p(M|Q,\Delta=0) \tag{2.4}$$

Because precise distance between two objects is hard to estimate, hypergraph analysis approach is employed in [35], in which distance calculating is avoidance. In hypergraph $G=(V,E,\omega)$ the vertices V represent the objects from database , edges E denote clusters which gathering the view of all objects into, And the weight ω of an edge describes the similarity of two views. The degree of vertex and edge is defined as following respectively

$$d(v) = \sum_{e\in E}\omega(e)h(v,e)$$
$$d(e) = \sum_{v\in V}h(v,e)$$
$$h(v,e) = \begin{cases} 1, v\in e \\ 0, v\notin e \end{cases} \tag{2.5}$$

Suppose there are n hypergraphs $G_1=(V_1,E_1,\omega_1)$,$G_2=(V_2,E_2,\omega_2)$,...,$G_n=(V_n,E_n,\omega_n)$ and $\{H_1,H_2,\ldots,H_n\}$,$\{D_{v1},D_{v2},\ldots,D_{vn}\}$, $\{D_{e1},D_{e2},\ldots,D_{en}\}$ denote the incidence matrices, vertex degree matrices, and hyperedge degree matrices, correspondingly. And α_i (≥ 0) represents the weight of ith hypergraph where $\sum_{i=1}^n\alpha_i=1$. Then The retrieval object function becomes

$$\underset{f}{\text{argmin}} f^T\Delta f + \lambda\|f-y\|^2 \tag{2.6}$$

where y is the label vector, λ is the weighting parameter and Δ is the (positive semidefinite) hypergraph Laplacian.

To our best knowledge, there is no work in the literature modeling the complementation of region shape description and contour context of 3D object, especially for view based 3D object retrieval. It should be noted that these two characters demonstrate the essences of 3D object form distinctive and complementary aspects.

Most recently, due to the appealing performance, bipartite graph matching draw lots of attentions. For instance, bipartite graph optimal matching (OM) [15] and WBGM [16]. In WBGM, there are two sub-sets which are composed of weighted bipartite graph, and comparing two 3D objects is to calculate the max-weighted bipartite matching. Other relative method of BoVF methods like bag-of-region-words (BoRW) consider region information in BoVF.

Inspired by the methods above, a Greedy Search (GS) algorithm is proposed. To calculate the similarity of query object and object from database, three bipartite graphs are employed. To give a mark of similarity of two objects, we calculating the optimal match of each bipartite graph pair greedily, which show a great performance.

3. Multi-feature collaboration and bipartite graph matching based 3D retrieval

In this section, the proposed view-based 3D object retrieval method is introduced. Firstly three descriptors are extracted. Secondly three bipartite graph are constructed on each descriptor between two objects. At last, combine the three bipartite graphs to one descriptor which is used to calculate the similarity of the two objects.

3.1. 3D objects descriptor extraction

To improve the retrieval performance a combination of different shape descriptors has been proposed lately. In this paper, we employ two region shape descriptors (Zernike moments [18] descriptor and BoVW descriptor), one contour shape descriptor (Fourier descriptor). For each 3D object, firstly we extract each descriptor, secondly the similarity matrix is constructed, then we sort the objects by calculate the score with the similarity matrix. The detailed procedure of computing the weight vector are summarized in Algorithm 1.

Algorithm 1. Proposed multi-feature collaboration and bipartite graph matching.

Input: 3D objects Data set $[D_1, D_2, \ldots, D_n]$, query object Q
Output: Data set objects in new order $\left[\tilde{D}_1, \tilde{D}_2, \ldots, \tilde{D}_n\right]$
For $i=1{:}n$ //processing on each object in data set
  Extract Zernike Moment feature matrix $D_i^{Zernike}$ from D_i
  Extract Bag-of-Visual-Words feature matrix D_i^{BoW} from D_i
  Extract Fourier Descriptor features matrix D_i^{FD} from D_i
end
Extract features $Q^{Zernike}$,Q^{BoW},Q^{FD} of Q with the same way mentioned above
For $i=1{:}n$ //construct similarity matrix
  For $j=$[Zernike Moment, Fourier, Bag-of-Visual-Words]
  Construct similarity matrix $M_j^i = \left(Q^j\right)^T \bullet D_i^j$
  end
end
For $i=1{:}n$ //Score each object
  $M_{mix}^i = \lambda_1 M_{BoW}^i + \lambda_2 M_{FD}^i + \lambda_3 M_{Zernike}^i$
  $Score_i = \sum M_{mix}^i(i,j)$
End
Order the object $[D_1, D_2, \ldots, D_n]$ by $Score_i(i=1\cdots n)$

Note that each (d=Zernike, BoVW, FD) is normalized, which means , (p is one of the view to represent an object).

3.1.1. Zernike moments (ZM)

ZM is composed of a set of complex polynomials which describe the interior of unit circle (i.e., $x2+y2=1$) by a complete orthogonal set:

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho)\exp(jm\theta) \tag{3.1}$$

n and m must be integers . Besides $n-|m|$ must be even and $|m| \le n$. To make m meaningful n must be positive or zero. ρ is the polar radius and θ is the polar angle. The radial polynomial $R_{nm}(\rho)$ is defined below:

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}\rho^{n-2s} \tag{3.2}$$

It is easy to see that $R_{n,m}(\rho) = R_{n,-m}(\rho)$. These orthogonal polynomials satisfy

$$\iint_{x^2+y^2\le 1} [V_{nm}(x,y)]^* V_{pq}(x,y)dxdy = \frac{\pi}{n+1}\delta_{np}\delta_{mq} \tag{3.3}$$

where $\delta_{ab}=1$, if $a=b$, otherwise $\delta_{ab}=0$. Zernike moments decompose the image function by basic orthogonal functions. For a continuous image function $f(x,y)$, Zernike moment with the order n with repetition m should vanish outside the unit circle:

$$A_{nm} = \frac{n+1}{n}\iint_{x^2+y^2\le 1} f(x,y)V_{nm}^*(\rho,\theta)dxdy \tag{3.4}$$

For a discrete image this function should be re-write as

$$A_{nm} = \frac{n+1}{n}\sum_x\sum_y f(x,y)V_{nm}^*(\rho,\theta), x^2+y^2 \le 1 \tag{3.5}$$

For a given image, the center is the origin and pixel coordinate projects to the range of unit circle ($x^2+y^2=1$), and pixel which outside the unit circle would be ignored.

3.1.2. BoVW/SIFT descriptor

Inspired by the BoVW and SIFT descriptor [19], a BoVW/SIFT representation is proposed in this section retrieval of 3D objects, which comprise of SIFT descriptor extraction, Dictionary building, Bag-of-Visual-Words descriptor construction.

- SIFT Descriptor Extraction in Each Object
  On account of no label for objects, we consider each object as a category. And each view in the object is parsed into SIFT descriptors.
- Dictionary Building
  Due to some SIFT descriptors are too common to distinguish one object from another, excluding those should be time and memory saving. Using large vocabularies causes the number of parameters to grow too large. Hence, we propose to identify groups of visual words and share weights among visual words within each group. Similar to previous works that relate visual word frequency with visual word importance (e.g., using idf), we also decide to associate similar frequencies with similar weights. In practice, the k-means clustering is adopted on the remaining descriptors.
- Bag-of-Visual-Words Descriptor Construction.

We use BoVW representation for retrieval of word images. This is motivated by multiple factors (i) Bag of visual Words (BoVW) representation has been the most popular representation and perform excellently for recognition and retrieval tasks in images and videos. (ii) Being a loose representation, BoVW representation can retrieve sub words, which is difficult with the popular vector space models. After quantifying the set of SIFT descriptors of the object by the dictionary, we can get the Bag-of-Visual-Words descriptor of each object.

Accordingly, the multiview of a 3D object can be represented by an unordered set of no distinctive discrete visual words. In retrieval phase, a 3D object is retrieved by computing the histogram of visual word frequencies, and returning the word image, with the closest histogram. This can also be used to rank the returned3D objects. A benefit of this approach is that, matches can be effectively computed without delay. Besides, the discriminative ability of the descriptor facilitates the accuracy and robustness of 3D object retrieval.

3.1.3. Fourier descriptor (FD)

As a frequently-used descriptor, Fourier descriptor [20,21] represents the object shape by boundary chain code and can be calculate effectively and efficiently. Firstly, compute the time consumption on traversing a particular link a_i, on the assumption of constant speed.

$$\Delta t_i = 1 + \left(\frac{\sqrt{2}-1}{2}\right)(1-(-1)^{a_i}) \tag{3.6}$$

And it is to t_p traverse the first p links of the chain

$$t_p = \sum_{i=1}^{p} \Delta t_i \tag{3.7}$$

The variations projecting to the coordinate axis can be calculated as

$$\Delta x_i = \mathrm{sgn}(6-a_i)\mathrm{sgn}(2-a_i)$$

$$\Delta y_i = \mathrm{sgn}(4-a_i)\mathrm{sgn}(a_i) \tag{3.8}$$

where sgn(•) is an indicative function,

$$\mathrm{sgn}(Z) = \begin{cases} 1 & Z>0 \\ 0 & Z=0 \\ -1 & Z<0 \end{cases} \tag{3.9}$$

Selecting continuous p links on the chain at well, the projections on coordinate axis are:

$$x_p = \sum_{i=1}^{p} \Delta x_i$$

$$y_p = \sum_{i=1}^{p} \Delta y_i \tag{3.10}$$

Then the Fourier series expansion of the whole chain code on the x projection can be obtained:

$$x(t) = A_0 + \sum_{n=1}^{\infty} a_n \cos\frac{2n\pi t}{T} + b_n \sin\frac{2n\pi t}{T}$$

where

$$A_0 = \frac{1}{T}\int_0^T x(t)dt$$

$$a_n = \frac{2}{T}\int_0^T x(t)\cos\frac{2n\pi t}{T}dt$$

$$b_n = \frac{2}{T}\int_0^T x(t)\sin\frac{2n\pi t}{T}dt \tag{3.11}$$

The time derivative $\hat{x}(t)$ can be expressed as follow:

$$\hat{x}(t) = \sum_{n=1}^{\infty} -\frac{2n\pi}{T}a_n \sin\frac{2n\pi t}{T} + \frac{2n\pi}{T}b_n \cos\frac{2n\pi t}{T}$$

where

$$a_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta x_p}{\Delta t_p}\left[\cos\frac{2n\pi t_p}{T} - \cos\frac{2n\pi t_{p-1}}{T}\right]$$

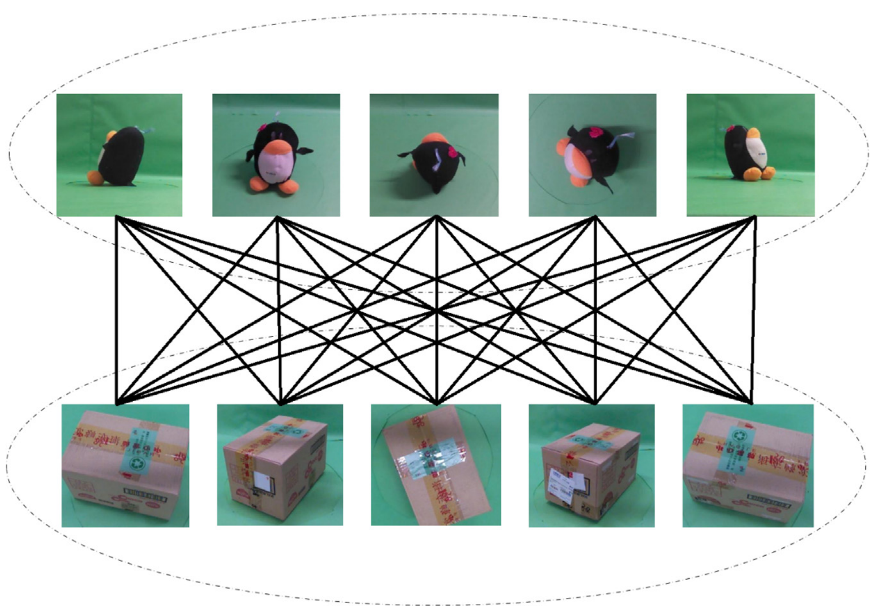


Fig. 1. Bipartite graph a graph which divides the whole space into 2 parts and each end of edges belongs to different part.

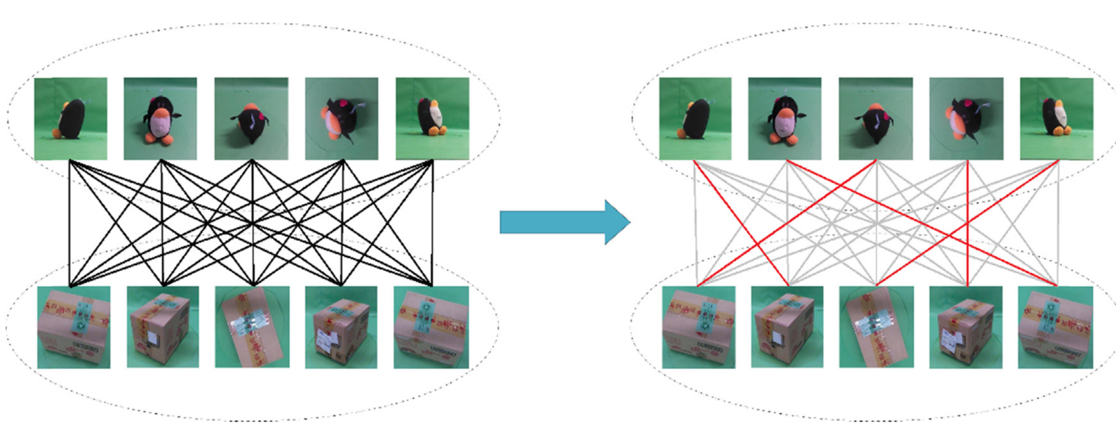


Fig. 2. Bipartite graph matching.

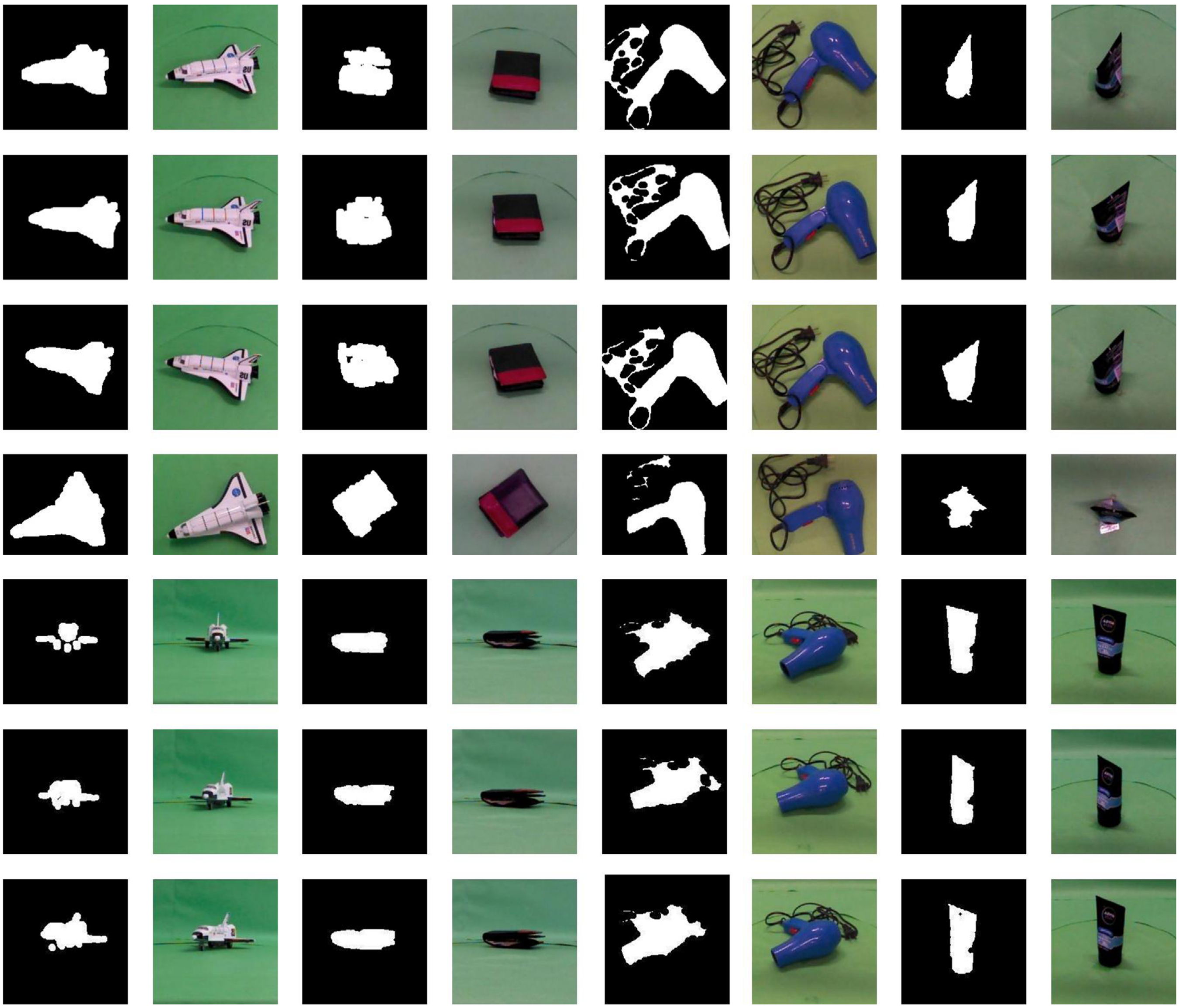


Fig. 3. SHREC'15 dataset.

$$b_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta x_p}{\Delta t_p}\left[\sin\frac{2n\pi t_p}{T} - \sin\frac{2n\pi t_{p-1}}{T}\right] \tag{3.12}$$

Equivalently the following formulation can be obtained

$$y(t) = C_0 + \sum_{n=1}^{\infty} c_n \cos\frac{2n\pi t}{T} + d_n \sin\frac{2n\pi t}{T}$$

$$\hat{y}(t) = \sum_{n=1}^{\infty} -\frac{2n\pi}{T}c_n \sin\frac{2n\pi t}{T} + \frac{2n\pi}{T}d_n \cos\frac{2n\pi t}{T}$$

where

$$c_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta y_p}{\Delta t_p}\left[\cos\frac{2n\pi t_p}{T} - \cos\frac{2n\pi t_{p-1}}{T}\right]$$

$$d_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta y_p}{\Delta t_p}\left[\sin\frac{2n\pi t_p}{T} - \sin\frac{2n\pi t_{p-1}}{T}\right] \tag{3.13}$$

And

$$A_0 = \frac{1}{T}\sum_{p=1}^{k}\left[\frac{\Delta x_p}{2\Delta t_p}\left(t_p^2 - t_{p-1}^2\right) + \xi_p\left(t_p - t_{p-1}\right)\right]$$

$$C_0 = \frac{1}{T}\sum_{p=1}^{k}\left[\frac{\Delta y_p}{2\Delta t_p}\left(t_p^2 - t_{p-1}^2\right) + \delta_p\left(t_p - t_{p-1}\right)\right]$$

where

$$\xi_p = \sum_{j=1}^{p-1}\Delta x_j - \frac{\Delta x_p}{\Delta t_p}\sum_{j=1}^{p-1}\Delta t_j$$

$$\delta_p = \sum_{j=1}^{p-1}\Delta y_j - \frac{\Delta y_p}{\Delta t_p}\sum_{j=1}^{p-1}\Delta t_j \tag{3.14}$$

3.2. Bipartite graphs construction

Bipartite graphs for each descriptor of two object are constructed to measure the similarity in the matching procedure. For

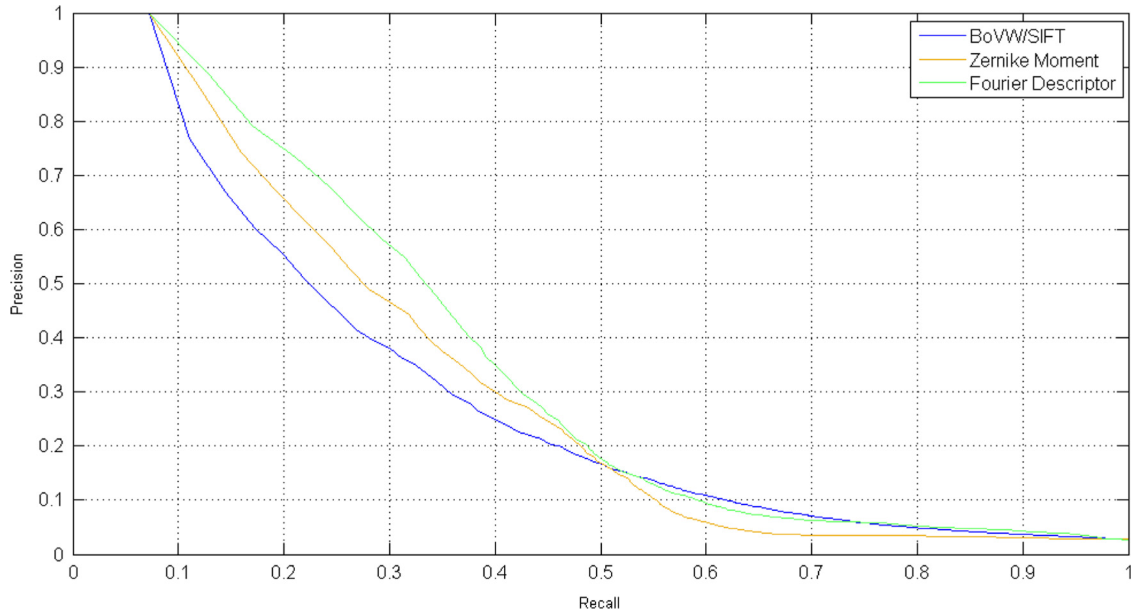


Fig. 4. PR-curve of BoVW/SIFT, Zernike moment, and Fourier descriptor.

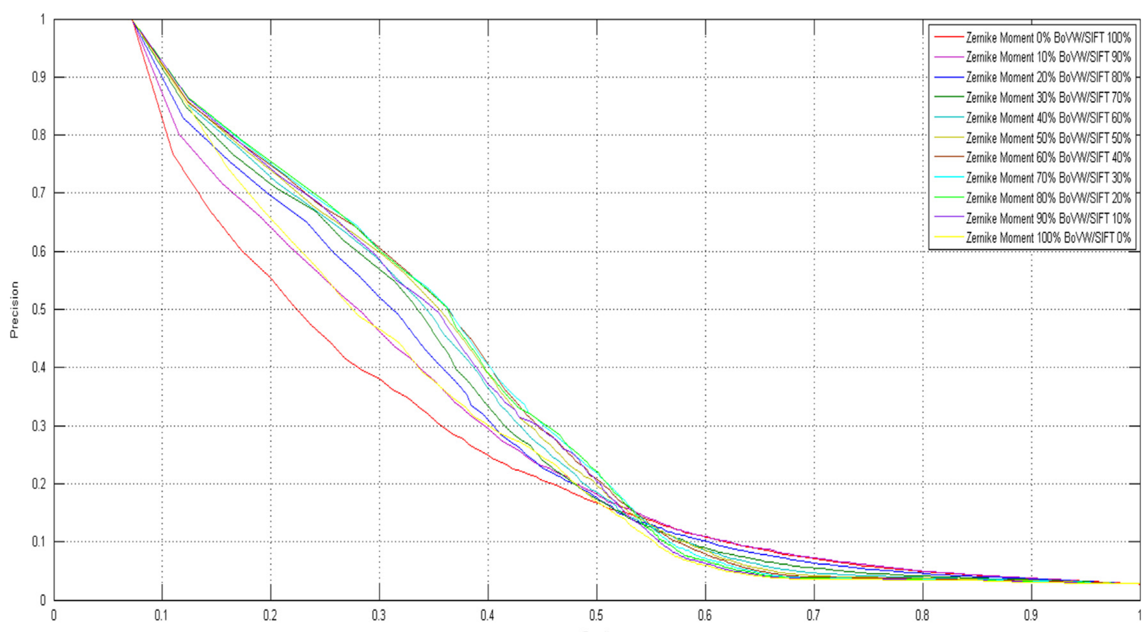


Fig. 5. PR-curve of collaboration of Zernike moment and BoVW/SIFT.

Fourier descriptor, assuming two objects O_1 and O_2, whose feature matrixes are $\mathrm{F}_1 = \left\{d_1^1, d_2^1, \cdots, d_{n_1}^1\right\}$ and $\mathrm{F}_2 = \left\{d_1^2, d_2^2, \cdots, d_{n_2}^2\right\}$ (d_i^j means the ith view and jth object), the Euclidean distance M_{pq}^T is employed to compute the similarity of d_q^2 and d_p^1, stored in matrix $\mathrm{M}^T \in \Re^{n1 \times n2}$. The optimal match of bipartite graphs is calculated with following optimization process

$$X^* = \underset{X}{\operatorname{argmax}} X \odot M^T$$

$$s.t.\ X = \{0, 1\}^{n1 \times n2} \tag{3.15}$$

We greedily acquire the optimal match of the bipartite graphs, and ensure that each view will not be selected more than twice, which facilitates making each bipartite graph sparse.

3.3. Bipartite graphs combination

For each bipartite graph of the same objects, Eq. (3.15) has been done to get the incidence matrix ($X_{BoVW}^*, X_{FD}^*, X_{Zernike}^*$) and the similarity matching score ($M_{BoVW}^*, M_{FD}^*, M_{Zernike}^*$), respectively.

$$M_{BoVW}^* = X_{BoVW}^* \odot M_{BoVW}^T$$
$$M_{FD}^* = X_{FD}^* \odot M_{FD}^T$$
$$M_{Zernike}^* = X_{Zernike}^* \odot M_{Zernike}^T \tag{3.16}$$

Having obtained the similarity matching score for each descriptor, these three bipartite graphs are combined into a new one by following procedure, then the final matching score of two objects can be calculated as

$$Score = \sum M_{mix}$$

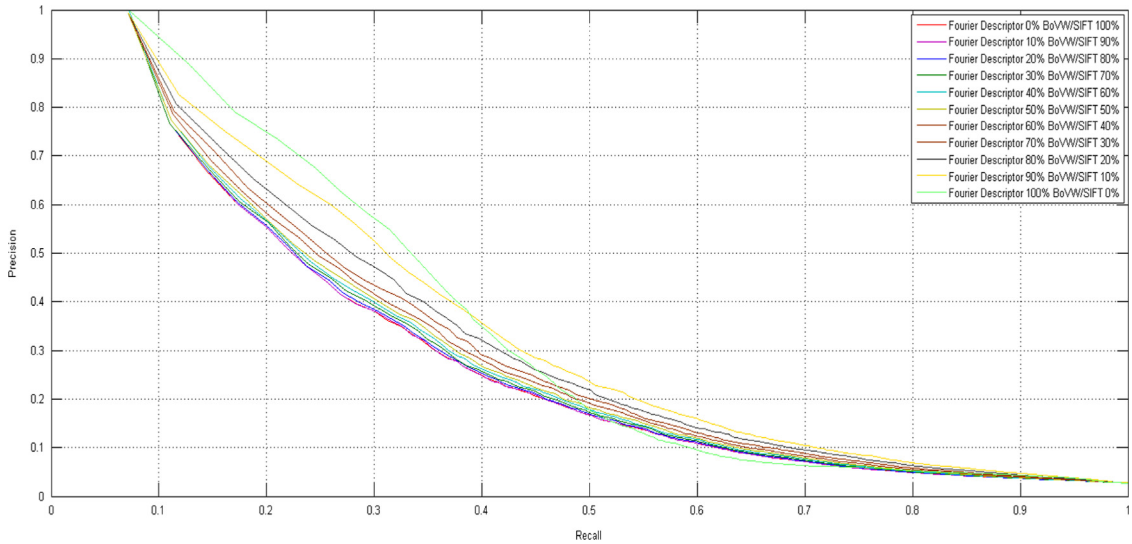


Fig. 6. PR-curve of collaboration of Fourier descriptor and BoVW/SIFT.

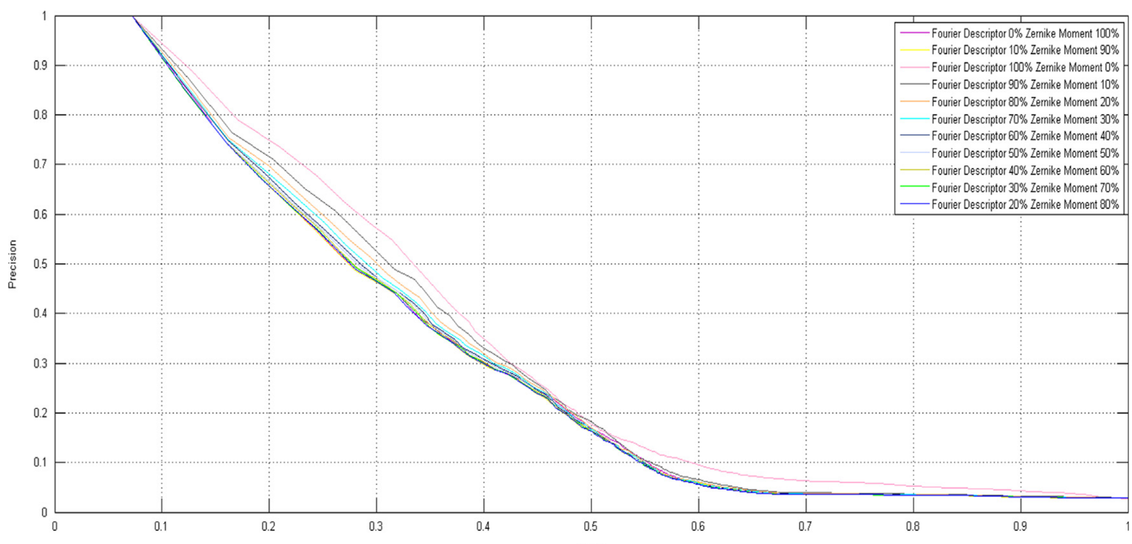


Fig. 7. PR-curve of collaboration of Zernike moment and Fourier descriptor.

$$M_{mix} = \lambda_1 M^*_{BoVW} + \lambda_2 M^*_{FD} + \lambda_3 M^*_{Zernike} \tag{3.17}$$

For a certain query, higher the score is, more similar they are. Sorting the score in descending order would be the final ranking.

4. Experimental results

In this section, extensive experiment results are presented to evaluate the proposed multi-feature collaboration and bipartite graph matching based 3D retrieval.

4.1. SHREC data set

SHREC'15 challenge data set is aimed at exploring the optimal retrieval algorithm of view-based 3D model. There are two tasks in this competition. Both of them contain 505 3D objects of the data set in which each object is explained by RGB images and depth images correspondingly with frame size 640×480. There are 311 objects being chosen as query to inspect the performance of methods submitted. Six groups were attracted to this competition, and 26 methods were committed.

Three Kinect sensors were used to collect the information in two different ways. To simulate the real world environment, there are 202 object collected in one way, and 303 in the other. The different between these two settings are the quantity of information, and one is few, the other is plenty. To evaluate the performance seven criteria are employed, such as PR curve, FT, ST etc. These evaluation results may guide the researchers improving their methods in some way.

4.2. Performance of different 3D objection features

We evaluate the performance of the proposed method on SHREC'15 dataset with different 3D objection features and the results are reported in Fig. 1. If only the Zernike Moments is used for retrieval, the average 0.4434 (top 10 precision). If only the BoVW/SIFT feature is adopted, the average 0.3958 (top 10 precision). If only FD is used, the 0.4785 (top 10 precision) per sample/object (Figs. 2 and 4).

To obtain the optimal performance, a variety of configurations of parameters was evaluated. We can see that Fourier descriptor

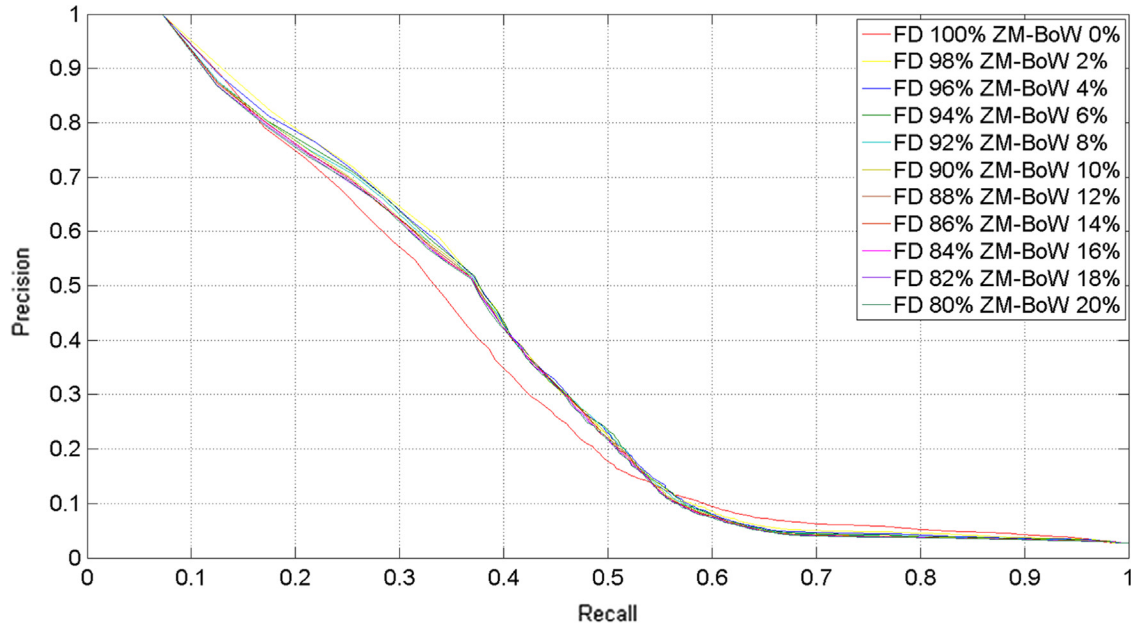


Fig. 8. PR-curve of collaboration of 3 descriptors.

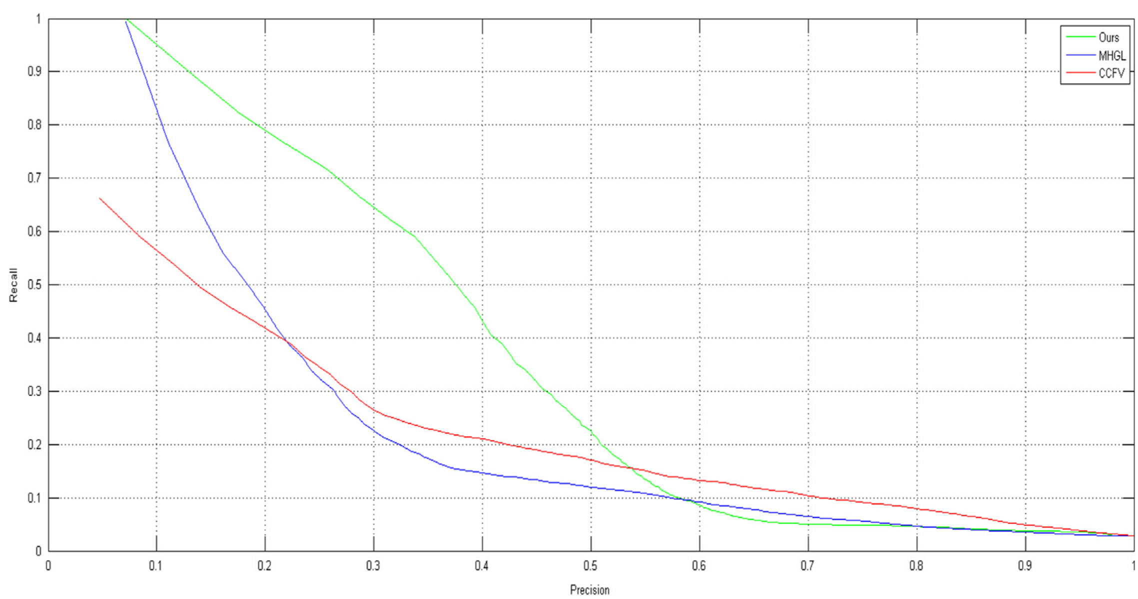


Fig. 9. PR-curve of CCFV, hypergraph method (MHGL), and ours.

Table 1
The parameters used in the proposed method.

Parameters and description	$M_{mix} = \lambda_1 M^{*}_{BoW} + \lambda_2 M^{*}_{FD} + \lambda_3 M^{*}_{Zernike}$
λ_1	0.006
λ_2	0.980
λ_3	0.014

(FD) is the best and Zernike moment (ZM) is better. We decide to combine each two of them to find the optimal combination.

From the Figs. 5–7, we can see that when combination each two of the three descriptors, the FD should get the largest percentage and ZM should get the larger one. Also we set the combination of ZM and BoVW by 7:3, then we test them with FD as follow

Then we get the optimal combination which is FD 98%, ZM 1.4%, BoVW 0.6%. From these comparison results, we can see that the proposed method achieves high recognition accuracy while having low computational complexity. The reason is that with the proposed collaboration strategy, these complement features can fix shortcoming with each other which can express the object in all directions (Fig. 8).

4.3. Comparison with recent representative methods

We compare the proposed method with other recent representative methods on SHREC'15 dataset (Fig. 3) and the parameters used in the proposed method are listed in Table 1.

From Fig. 9 we can see that our method get a better performance than CCFV [34] and Hypergraph [35]. Also our method won one of the optimal results in the compact of SHREC'15 Track [22].

5. Conclusion

In this paper, we present a view-based 3D model retrieval algorithm on multi-feature by bipartite graph matching. The proposed method extracts three descriptors and combine them through bipartite graph. The user feedback information is effectively explored to achieve better performance. We compare our method on the SHREC'15 data set with other methods. Experimental results and comparison show that the proposed method outperforms the other methods for 3D model retrieval.

Acknowledgments

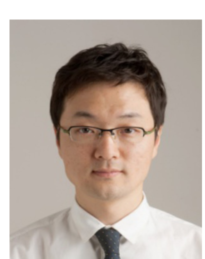
We would like to acknowledge the editors and reviewers, whose valuable comments greatly improved the manuscript. This work was supported in part by the Major State Basic Research Development Program of China (973 Program 2015CB351804) and the National Natural Science Foundation of China under Grant nos. 61572155 and 61272386.

References

[1] W. Hu, G. Liu, H. Zhou, Web-based 3D control laboratory for remote real time experimentation, IEEE Trans. Ind. Electron. 60 (10) (2013) 4673–4682.
[2] Y. Liu, X.-L. Wang, H.-Y. Wang, H. Zha, H. Qin, Learning robust similarity measures for 3D partial shape retrieval, Int. J. Comput. Vis. 89 (2/3) (2010) 408–431.
[3] Y. Gao, Q. Dai, N. Zhang, 3D model comparison using spatial structure circular descriptor, Pattern Recognit. 43 (3) (2010) 1142–1151.
[4] T.F. Ansary, M. Daoudi, J.P. Vandeborre, A Bayesian 3D search engine using adaptive views clustering, IEEE Trans. Multimedia 9 (1) (2007) 78–88.
[5] J. Shih, C. Lee, J. Wang, A new 3D model retrieval approach based on the elevation descriptor, Pattern Recognit. 40 (1) (2007) 283–295.
[6] D.Y. Chen, X.P. Tian, Y.T. Shen, M. Ouhyoung, On visual similarity based 3D model retrieval, Comput. Graph. Forum 22 (3) (2003) 223–232.
[7] P. Daras, A. Axenopoulos, A 3D shape retrieval framework supporting multi-modal queries, Int. J. Comput. Vis. 89 (2/3) (2010) 229–247.
[8] Yue Gao, Qionghai Dai, View-based 3-D object retrieval: challenges and approaches, IEEE Multimedia Mag. 21 (3) (2014) 52–57.
[9] Z. Stejic, Y. Takama, K. Hirota, Relevance feedback-based image retrieval interface incorporating region and feature saliency patterns as visualizable image similarity criteria, IEEE Trans. Ind. Electron. 50 (5) (2003) 839–852.
[10] T. Takeguchi, S. Kaneko, Using depth aspect images for robust and efficient search of multiple objects in optomechatronic sensing, IEEE Trans. Ind. Electron. 52 (4) (2005) 1041–1049.
[11] H. Zhuang, K. Low, W. Yau, Multichannel pulse-coupled-neural network-based color image segmentation for object detection, IEEE Trans. Ind. Electron. 59 (8) (2012) 3299–3308.
[12] C.-Y. Chang, W. Lie, Real-time visual tracking and measurement to control fast dynamics of overhead cranes, IEEE Trans. Ind. Electron. 59 (3) (2012) 1640–1649.
[13] Yue Gao, Qionghai Dai, View-Based 3-D Object Retrieval, Elsevier, ISBN 978-0-12-802419-5, 2014.
[14] M. Steinbach, G. Karypis, V. Kumar, A comparison of document clustering techniques, In: Proceedings of the KDD Workshop Text Mining, 2000, pp. 1–2.
[15] W.S. Xiao. Graph Theory and Its Algorithms, 1993.
[16] Y. Gao, Q. Dai, M. Wang, N. Zhang, 3d model retrieval using weighted bipartite graph matching, Signal Processing: Image Commun. 26 (1) (2011) 39–47.
[17] S. Avidan, A. Shashua, Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence, IEEE Trans. Pattern Anal. Mach. Intell. 22 (4) (2000) 348–357.
[18] ALIREZA KHOTANZAD, Y.A.W. HUA HONG, Invariant image recognition by Zernike moments, IEEE Trans. Pattern Anal. Mach. Intell. 12 (5) (1990).
[19] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[20] R. Bracewell, The fourier transform and iis applications, New York, 1965.
[21] Pkuhl Frank, Elliptic Fourier features of a closed contour, Comput. Graph. Image Proc. 18 (1982) 236–258.
[22] SHREC'15 Track: 3D Object Retrieval with Multimodal Views, EUROGRAPHICS 2015.
[23] A. Bimbo, P. Pala., Content-based retrieval of 3D models, ACM Trans. Multimedia Comput. Comm., Appl. 2 (1) (2006) 20–43.
[24] B. Bustos, et al., Feature-based similarity search in 3D object databases, ACM Comput. Surv. 37 (4) (2005) 345–387.
[25] J. Tangelder, R. Veltkamp, Polyhedral model retrieval using weighted point sets, Int. J. Image Graph. 3 (1) (2003) 209–229.
[26] R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin, Shape distributions, ACM Trans. Graph 21 (4) (2002) 807–832.
[27] A.E. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3D scenes, IEEE Trans. Pattern Anal. Mach. Intell. 21 (5) (1999) 433–449.
[28] C.Y. Ip, D. Lapadat, L. Sieger, W.C. Regli, Using shape distributions to compare solid models, In: Proceedings of the 7th ACM Symposium on Solid Modeling Applications, 2002, pp. 273–280.
[29] A. Makadia, K. Daniilidis, Spherical correlation of visual representations for 3D model retrieval, Int. J. Comput. Vis. 89 (2–3) (2010) 193–210.
[30] R. Ohbuchi, K. Osada, T. Furuya, T. Banno, Salient local visual features for shape-based 3D model retrieval, In: Proceedings of the IEEE International Conference on Shape Modeling and Applications, 2008, pp. 93–102.
[31] W. Li, G. Bebis, N.G. Bourbakis, 3D object recognition using 2D views, IEEE Trans. Image Process. 17 (11) (2008) 2236–2255.
[32] Y. Yubin, L. Hui, Z. Yao, Content-based 3D model retrieval: A survey, IEEE Trans. Syst., Man, Cybern. C, Appl. Rev. 37 (6) (2007) 1081–1098.
[33] K. Lu, Q. Wang, J. Xue, W. Pan, 3D model retrieval and classification by semi-supervised learning with content-based similarity, Inf. Sci. 281 (2014) 703–713.
[34] Yue Gao, Jinhui Tang, Richang Hong, Shuicheng Yan, Qionghai Dai, Naiyao Zhang, Tat-Seng Chua, Camera constraint-free view-based 3D object retrieval, IEEE Trans. Image Proc. 21 (4) (2012) 2269–2281.
[35] Yue Gao, Meng Wang, Dacheng Tao, Rongrong Ji, Qionghai Dai, 3D object retrieval and recognition with hypergraph analysis, IEEE Trans. Image Proc. 21 (9) (2012) 4290–4303.
[36] Yue Gao, Meng Wang, Rongrong Ji, Xindong Wu, Qionghai Dai, 3D object retrieval with Hausdorff distance learning, IEEE Trans. Ind. Electron. 61 (4) (2014) 2088–2098.
[37] Yue Gao, Meng Wang, Zhengjun Zha, Qi Tian, Qionghai Dai, Naiyao Zhang, Less is more: efficient 3D object retrieval with query view selection, IEEE Trans. Multimedia 11 (5) (2011) 1007–1018.
[38] Chen Lin, Wenqiang Chen, Cheng Qiu, Yunfeng Wu, Sridhar Krishnan, Quan Zou, LibD3C: ensemble classifiers with a clustering and dynamic selection strategy, Neurocomputing 123 (2014) 424–435.

Yan Zhang received the B.S. and M.S. degrees from Department of Mathematics and School of Computer Science and Technology, Harbin Institute of Technology (HIT), Harbin, China, in 2007 and 2009, respectively. He is now working towards the Ph.D. degree at School of Computer Science and Technology, HIT.

Feng Jiang received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology (HIT), Harbin, China, in 2001, 2003, and 2008, respectively. He is now an Associated Professor in the Department of Computer Science, HIT and a visiting scholar in the School of Electrical Engineering, Princeton University. His research interests include computer vision, pattern recognition and image and video processing.



Seungmin Rho received his M.S. and Ph.D. degrees in Computer Science from Ajou University, Korea, in Computer Science from Ajou University, Korea, in 2003 and 2008, respectively. In 2008–2009, he was a Postdoctoral Research Fellow at the Computer Music Lab of the School of Computer Science in Carnegie Mellon University. In 2009–2011, he had been working as a Research Professor at School of Electrical Engineering in Korea University. In 2012, he was an assistant professor at Division of Information and Communication in Baekseok University. Dr. Rho is currently a faculty of Department of Multimedia at Sungkyul University. His current research interests include database, big data analysis, music retrieval, multimedia systems, machine learning, knowledge management as well as computational intelligence.

Shaohui Liu received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology (HIT), Harbin, China, in 2000, 2002, and 2007, respectively. He is now an Associated Professor in the Department of Computer Science, HIT and his research interests include data compression, pattern recognition and image and video processing.

Debin Zhao received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology (HIT), Harbin, China, in 1985, 1988, and 1998, respectively.

He is now a Professor in the Department of Computer Science, HIT. He has published over 200 technical articles in refereed journals and conference proceedings in the areas of image and video coding, video processing, video streaming and transmission, and pattern recognition.

Rongrong Ji is currectly wokring at Xiamen University since 2013, where he directs the Intelligent Multimedia Technology Laboratory and serves as a Dean Assistant in the School of Information Science and Technology. Before that, he used to be a Postdoc research fellow in the Department of Electrical Engineering, Columbia University from 2010 to 2013, worked with Professor Shih-Fu Chang. He obtained his Ph.D. degree in computer science from Harbin Institute of Technology, under supervision of Professor Hongxun Yao. He had been a visiting student at University of Texas of San Antonio worked with Professor Qi Tian, and a research assistant at Peking University worked with Professor Wen Gao in 2010, a research intern at Microsoft Research Asia, worked with Dr. Xing Xie from 2007 to 2008.