2014

Efficient View-Based 3-D Object Retrieval via Hypergraph Learning

Yue Gao
School of Computing, National University of Singapore, Singapore.

Qionghai Dai
Department of Automation, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China.

Follow this and additional works at: https://tsinghuauniversitypress.researchcommons.org/tsinghua-science-and-technology

Part of the Computer Sciences Commons, and the Electrical and Computer Engineering Commons

Recommended Citation

This Research Article is brought to you for free and open access by Tsinghua University Press: Journals Publishing. It has been accepted for inclusion in Tsinghua Science and Technology by an authorized editor of Tsinghua University Press: Journals Publishing.
Efficient View-Based 3-D Object Retrieval via Hypergraph Learning

Yue Gao and Qionghai Dai*

Abstract: View-based 3-D object retrieval has become an emerging topic in recent years, especially with the fast development of visual content acquisition devices, such as mobile phones with cameras. Extensive research efforts have been dedicated to this task, while it is still difficult to measure the relevance between two objects with multiple views. In recent years, learning-based methods have been investigated in view-based 3-D object retrieval, such as graph-based learning. It is noted that the graph-based methods suffer from the high computational cost from the graph construction and the corresponding learning process. In this paper, we introduce a general framework to accelerate the learning-based view-based 3-D object matching in large scale data. Given a query object $Q$ and one object $O$ from a 3-D dataset $D$, the first step is to extract a small set of candidate relevant 3-D objects for object $O$. Then multiple hypergraphs can be constructed based on this small set of 3-D objects and the learning on the fused hypergraph is conducted to generate the relevance between $Q$ and $O$, which can be further used in the retrieval procedure. Experiments demonstrate the effectiveness of the proposed framework.

Key words: view-based; 3-D object retrieval; hypergraph learning

1 Introduction

The recent advances in graphic hardware, rapid networks, and computer techniques have led to various applications of 3-D objects\cite{1-3} in different fields, such as 3-D graphics, Computer-Aided Design (CAD), and entertainment. Under these circumstances, the generated large scale 3-D object data require effective and efficient management and indexing and retrieval technique.

3-D object retrieval has attracted extensive research efforts in recent years\cite{4,5}. Existing works can be divided into two categories: model-based and view-based methods. Early methods\cite{6-8} mainly are model-based, in which the model information is explicitly required for feature extraction. However, unlike the CAD applications, the model information in many real applications is not always available, which limits the use of model-based methods. The other type of methods is view-based, in which a set of views are employed to represent the 3-D model. As introduced in Ref. [9], “view-based methods have the advantages of being highly discriminative, can work for articulated objects, can be effective for partial matching, and can also be beneficial for 2-D sketch-based and 2-D image-based queries.”

In view-based 3-D object retrieval, a main challenge is to measure the similarity/distance between two 3-D objects. It is different from the traditional image matching, which is based on single image pair comparison. Some existing methods employ the view matching to measure the distance between two objects\cite{9-12}, such as the Hausdorff distance\cite{13,14} and the Earth Mover’s Distance\cite{15}. These methods can work well if the two objects share similar views. It is noted that it is difficult to guarantee that the multiple
views can be exactly matched. Also, the underneath structure of 3-D objects requires deep investigation.

Recently, learning-based methods have been introduced in 3-D object retrieval. In our previous work, a hypergraph learning method\cite{16} is employed to exploit the higher order relationship among 3-D objects. In this method, the correlation among 3-D objects is formulated in a hypergraph structure in which the hyperedges are generated using the similarity among views. A learning procedure is conducted to estimate the relevance among 3-D objects which can be used in the retrieval process. However, this hypergraph-based method is limited in the high computational cost in the learning process, which is difficult to handle large scale data.

In this work, we introduce a general framework to learn the pairwise object relevance in large scale 3-D dataset. As shown in Fig. 1, given a query object $Q$ and one object $O$ from a large set $D$ of 3-D objects, a small set of 3-D objects are selected as the candidate relevant samples for object $O$. Then multiple hypergraphs can be constructed based on this small set of 3-D objects and the learning on the fused hypergraph is conducted to estimate the relevance between $Q$ and $O$. Experiments on the NTU dataset are provided to demonstrate the effectiveness of the proposed framework.

2 Related Work

In view-based 3-D object retrieval, a set of multiple views are employed for 3-D object representation. Some existing methods employed predefined camera arrays to capture views, such as Lighting Field Descriptor (LFD)\cite{17} and Elevation Descriptor (ED)\cite{12}. In LFD, the views were obtained using the camera array on 20 vertices of a dodecahedron over a hemisphere surrounding the 3-D model, in which 10 binary images were employed as the LFDs. In this method, Zernike moments and Fourier descriptors were extracted as the view features. ED\cite{12} employed six range views to represent 3-D models, in which each elevation view described the altitude information of the model from one direction. In ED, 2-D gray images were used to represent each view and a feature histogram was extracted as the feature. Seven characteristic views were extracted in Ref. [18] for 3-D model representation. A panoramic view was introduced in Ref. [19], in which the panoramic view was obtained by projecting the model to the lateral surface of a cylinder located at the centroid of the model and aligned with one of the model’s three principal axes. A compact multi-view descriptor was introduced in Ref. [9] which captured 18 views from the vertices of a 32-hedron. In this method, both the binary images and the depth images were employed and Polar-Fourier Coefficients, Zernike Moments, and Krawtchouk Moments were extracted as image features. Different from these methods, some other methods selected representative views from a large pool of views, such as Adaptive Views Clustering (AVC)\cite{10}. In AVC, 320 views were firstly captured and the representative views were selected using view clustering. A probabilistic method, Camera Constraint Free View-based method (CCFV), was proposed in Ref. [20], in which all query views were clustered to generate view clusters. Both a positive matching model and a negative matching model were trained respectively. The CCFV model was generated on the basis of the query Gaussian models by combining both the positive matching model and the negative matching model. Ohbuchi et al.\cite{21} proposed to employ the bag-of-visual-features for 3-D model representation. In this method, SIFT features were extracted from all the views and then quantized into visual words using a pre-trained visual vocabulary. The Kullback-Leibler divergence was employed to measure the distance between two bag-of-words features. A Query View Suggestion (QVS) method was introduced in Ref. [11], in which the query views were incrementally selected based on the user relevance feedback information. For the newly selected query view, a distance metric was learned to enhance the discriminative property of query views.

Learning-based methods have been introduced in view-based 3-D object retrieval recently. A hypergraph learning method was introduced in Ref. [16], in which the relationship among 3-D objects was formulated in a hypergraph structure. The learning on the hypergraph was conducted to estimate the relevance among 3-D
objects, which were used for 3-D object retrieval. To learn the optimized matching metric, a Hausdorff distance learning method was introduced in Ref. [22]. In this method, all 3-D objects were modeled in a graph structure, and the Hausdorff distance was employed to measure the distance between each two 3-D objects. A joint learning on the relevance among 3-D objects and the optimized Hausdorff distance metric was conducted to achieve the optimal results. In this process, a Mahalanobis distance metric was learned for view distance calculation procedure.

3 The Proposed Method

In our previous work [18], a hypergraph structure is introduced to formulate the relevance among 3-D objects. All the views of 3-D objects are first grouped into clusters and each cluster generates one hyperedge to connect the vertices. By varying the number of generated clusters, multiple hypergraphs can be constructed. The learning on these hypergraphs is conducted to estimate the relevance among these 3-D objects. However, this method suffers from the high computational cost of the learning on the large hypergraph. Here, we aim to reduce the cost in the learning process for pairwise object relevance estimation.

We notice that only a small group of samples play an important role for the relevance estimation of one specific sample in the learning part. Therefore, we propose to first extract a small set of highly relevant samples, \( S' \), from the large dataset for one 3-D object \( O \) and further conduct the learning on this small set, \( \{ Q, O, S' \} \), for pairwise object relevant estimation, i.e., the relevance between \( O \) and query \( Q \).

3.1 Candidate relevant 3-D object selection

In this step, we select a small set of candidate relevant 3-D object \( S' \) for \( O \). Here, a rough 3-D object set \( S' \) is first extracted which can cover most of related objects of \( O \). The Zernike Moments [23] are employed as the visual feature in our experiments, which has been widely applied to 3-D object retrieval and recognition tasks [10, 17, 24] due to the robustness to image translation, scaling, and rotation. The distance between another object \( O' \) with \( O \) is measured according to

\[
d(O, O') = \sum_{v_i \in O} \min_{v_j \in O'} d(v_i, v_j) \tag{1}
\]

The top \( n \) 3-D objects with smallest distances to \( O \) can be selected for further processing.

It is noted that the rough candidate set \( S' \) usually contains very few related objects due to the limitation of the employed distance measure. Hence we need to further refine the candidates to extract a compact set \( S'' \) consisting of a few representative objects highly relevant to \( O \). Here, we model the whole process as the subgraph extraction problem on the graph \( G = (V, E, w) \) constructed by \( S' \).

Here, \( V = \{ v_1, \cdots, v_{n1} \} \) is a set of \( n1 \) vertices where each vertex \( v_i \) corresponds to each individual object \( O_i \), \( E \) is edge, and \( w \) is the edge weight. Here, we define \( w \) as the similarity between two 3-D objects of the corresponding edge and whether two vertices are connected is determined by their similarity. The similarity between two objects is calculated according to Eq. (1).

With the constructed graph \( G \), the fast algorithm in Ref. [25] can be used to extract the compact and highly related object set \( S'' \) for \( O \). As introduced in Ref. [25], a dense subgraph should have a large average affinity. Let \( H \) denote an embedding from graph \( G \) to a simplex space \( \mathcal{G} \rightarrow \Delta: H(G_{sub}) = x \), where \( x_i = 0 \) if \( v_i \notin V_{sub} \) and \( x_i = \frac{1}{I} \) otherwise, where \( I \) is the number of vertices in \( G_{sub} \). Here, \( G_{sub} \) is a subgraph, \( V_{sub} \) is the set of vertices in \( G_{sub} \), \( x \) is used to represent \( G_{sub} \), which lies in a simplex space \( \Delta = \{ x \in \mathbb{R}^n : x \geq 0 \text{ and } \|x\|_1 = 1 \} \) where \( \|x\|_1 = \sum_{i=1}^{n} |x_i| \).

For a subgraph \( G_{sub} \) represented by \( x \), the density \( g(x) \) is calculated according to

\[
g(x) = \sum_{i,j} x_i a_{ij} x_j = x^T A x \tag{2}
\]

where \( a_{ij} \) is consistent with the weight of edge between vertices \( i \) and \( j \). The objective of dense subgraph extraction can be formulated as the following quadratic optimization problem,

\[
x = \arg \max_{x \in \Delta} g(x) \tag{3}
\]

As \( x \) is in a discrete space, it can be relaxed into the continuous space \( \Delta \).

To optimize Eq. (3), Lagrangian multipliers \( \lambda \) and \( \mu_1, \cdots, \mu_n, \mu_j \geq 0 \) can be introduced as

\[
L(x, \lambda, \mu) = g(x) - \lambda \left( \sum_{i=1}^{n} x_i - 1 \right) + \sum_{i=1}^{n} \mu_i x_i \tag{4}
\]

The subgraph \( G_{sub} \) represented by \( x \) should satisfy the Karush-Kuhn-Tucker (KKT) conditions. Let \( \Psi \)
denote a set of all KKT points \( x^* \). It is impossible to achieve the whole set as the size of \( \Psi \) is usually very large. Following Ref. [25], a “Shrinking and Expansion” algorithm can be used to extract the KKT points. In the shrinking phase, a KKT point \( x^* \) in a smaller subgraph \( G_s \) is achieved and the subgraph shrinks to the subgraph represented by \( x^* \). Then, in the expansion phase, the vertices with strong relations with \( G_s \) are added and generate a new subgraph \( G_s \). The shrinking and expansion phases can be repeated until \( x^* \) is a KKT point of graph \( G \). In our experiments, we initially set the component corresponding to the input object of \( x \), which is \( O \), as 1 and the rest are 0. In the dense subgraph extraction procedure, we can set the required number of samples as \( n_2 \). Then, we have a small set \( S^i \) with \( n_2 \) samples for \( O \), which is denoted by \( S^i = \{ O_{s1}, O_{s2}, \ldots, O_{sn_2} \} \).

3.2 Learning on the hypergraph

With \( S^i \), the hypergraph learning method\[16\] can be employed to generate the pairwise object relevance for 3-D object retrieval. For a hypergraph \( G = (V, E, u) \), each vertex denotes one 3-D object in \( S^i \) and the hyperedges are constructed by view clustering using K-means method\[26\]. There are \( n_2 \) vertices in total. The diagonal matrices of vertex degrees and edge degrees, \( D_v \) and \( D_e \), and the incidence matrix \( H \) can be generated, respectively.

By varying the parameter \( K \) in the K-means clustering procedure, \( n_g \) hypergraphs can be constructed, i.e., \( G_1 = (V_1, E_1, u_1) \), \( G_2 = (V_2, E_2, u_2) \), \( \ldots \), and \( G_{n_g} = (V_{n_g}, E_{n_g}, u_{n_g}) \). The learning on the fused hypergraph is conducted to estimate the relevance between \( O \) and query \( Q \), where \( \alpha_i \) denotes the weight of the \( i \)-th hypergraph and set as \( \alpha_i = \frac{1}{n_g} \).

As introduced in Ref. [16], the relevance learning is formulated as a one-class classification problem\[27\], which can be written by

\[
\arg\min_f \left\{ \lambda R_{\text{emp}}(f) + \Omega(f) \right\}
\]

where the regularizer term \( \Omega(f) \) is defined by

\[
\Omega(f) = \frac{1}{2} \sum_{i=1}^{n_g} \sum_{e \in E_i, u, v \in V_i} \frac{w_i(e) h_i(u, e) h_i(v, e)}{d_i(e)} - \frac{(f(u) - f(v))^2}{d_i(u) d_i(v)} + f^T \sum_{i=1}^{n_g} \alpha_i \left(I - \frac{1}{2} H_i W_i D_e^{-1} H_i D_v^{-1}\right) f.
\]

where \( f \) is the to-be-learned relevance score vector.

Here, we let \( \Theta_i = D_v^{-\frac{1}{2}} H_i W_i D_e^{-1} H_i D_v^{-\frac{1}{2}} \) and \( \Delta = \sum_{i=1}^{n_g} \alpha_i (I - \Theta_i) = I - \Theta \), where \( \Theta = \sum_{i=1}^{n_g} \alpha_i \Theta_i \).

The loss function term \( \lambda R_{\text{emp}}(f) \) is defined according to

\[
\lambda R_{\text{emp}}(f) = \|f - y\|^2 = \sum_{u \in V} (f(u) - y(u))^2
\]

where \( y \) is the \( n_2 \times 1 \) label vector and all \( y \) values are zeros except query object \( Q \), which is 1.

According to Ref. [16], the optimized \( y \) can be achieved by

\[
f = \left(I + \frac{1}{\lambda \Delta}\right)^{-1} y
\]

With the obtained \( y \), the relevance between \( Q \) and \( O \) can be generated. This process can be repeated to calculate all the relevance scores between \( Q \) and other 3-D objects. The generated relevance scores can be used to re-rank the 3-D objects in a descending order to get the final retrieval results.

In this work, the Zernike moments are employed as the view feature. It is noted that any other features can be used here. To select the candidate relevant 3-D objects, the dense subgraph extraction method\[25\] is employed here due to its efficiency. It can be replaced by other approaches due to different scenarios.

4 Experimental Results

In this section, the experimental setups, including the testing datasets, evaluation criteria and compared methods, and experimental results are provided.

4.1 Experimental settings

We conducted experiments on a subset of the “National Taiwan University” (NTU) 3-D model dataset\[17\]. The selected testing dataset is composed of two parts. The first part includes 500 3-D models from 50 categories, including aqua, ball, bed, bike, Bird, boat, bomb, book, bottle, car, chair, chip, cube, cup, door, driver, drum, facemask, finger, flower, glasses, guitar, gun, hat, head, helicopter, house, knife, lamp, man, map, motorcycle, ocd, pen, phone, plane, plant, pot, starship, stick, submarine, sword, table, table-oneleg, tank, train, tree, truck, weed, and wheel. All 3-D models from this part are employed as the query in the 3-D object retrieval experiments. The second part includes 5000 3-D models from other categories, which are employed as the noise data for the queries.
Following the experimental settings in Ref. [16], 20 views are captured for each 3-D model from the vertices of a regular dodecahedron. Figure 2 shows 3-D model examples in the NTU dataset.

### 4.2 Evaluation criteria

In our experiments, the following evaluation measures are employed which have been widely used in existing 3-D object retrieval works\cite{9,10,12,17}:

1. **Precision-Recall curve**\cite{10}. The Precision-Recall curve comprehensively demonstrates retrieval performance which is assessed in terms of average recall and average precision.

2. **F-Measure** ($F$). Here, the $F$-measure is defined as \( F = \frac{2 \times P_{20} \times R_{20}}{P_{20} + R_{20}} \), where $P_{20}$ and $R_{20}$ are the precision and the recall of the top 20 retrieval results.

3. **Discounted Cumulative Gain (DCG)**\cite{28}. DCG is a statistic that assigns relevant results at the top ranking positions with higher weights under the assumption that a user is less likely to consider lower results.

4. **Average Normalized Modified Retrieval Rank (ANMRR)**\cite{29}. ANMRR is a rank-based measure, and it considers the ranking information of relevant objects among the retrieved objects. A lower ANMRR value indicates a better performance, i.e., relevant objects rank at top positions.

### 4.3 Compared methods

To evaluate the performance of the proposed method, the following state-of-the-art methods are employed for comparison:

1. **Adaptive Views Clustering (AVC)**\cite{10}. AVC is a view-based 3-D object retrieval method which captured 320 initial views and selected representative views by adaptive view clustering. In AVC, the similarity between two 3-D objects are measured in a probabilistic way.

2. **ED**\cite{12}. ED is another view-based method which captured six range views for 3-D object representation. The altitude information of the 3-D object were included in the range views, named elevation descriptor.

3. **Extension Ray-based Descriptor (ERD)**\cite{30}. ERD is a model-based method, which employed concentric spheres to extract the surface information of the 3-D model. In this method, a nearest sphere surface with a corresponding value was provided for each sampling surface point as the descriptor of the surfaces.

4. **Multiple Hypergraph Learning (MHL)**\cite{16}. In this method, all the 3-D objects are formulated in a set of hypergraphs. The learning on the fused hypergraph is conducted to estimate the relevance among 3-D objects.

5. **Multiple Hypergraph Learning with Candidate Relevant Object Selection (MHL-CROS)**, i.e., the proposed framework in this paper. In the experiments, the number of selected candidate relevant objects is set as $n_2 = 200$.

### 4.4 Results and comparison

The Precision-Recall curves are shown in Fig. 3, and the evaluation on $F$, DCG, and ANMRR is provided in Fig. 4. As shown in these results, the proposed MHL-CROS method outperforms AVC, ED, and ERD, while it is slightly worse than MHL.

From these results, we have the following...
observation.

1. The proposed MHL-CROS can lead to an improvement of 18.02%, 82.97%, and 55.69% in terms of $F$, an improvement of 7.61%, 54.97%, and 30.67% in terms of DCG, and an improvement of 7.91%, 15.93%, and 11.75% in terms of ANMRR from AVC, ED, and ERD, respectively. This indicates that the proposed method is effective on 3-D object retrieval.

2. In comparison with MHL, the proposed method can achieve a close result with very small degradation. This result shows that the proposed method is comparable with MHL while the object pairwise relevance estimation can be more efficient due to the reduction of the hypergraph dimension.

We further evaluate the influence of the number of selected candidate relevant objects, i.e., $n_2$, which is important in the hypergraph learning procedure. Figure 5 illustrates the performance comparison with the variance of the number of selected candidate relevant objects $n_2$ from 50 to 300. As shown in the results, the increasing of $n_2$ can lead to the improvement of 3-D object retrieval performance, which can be close to the performance of MHL. When $n_2$ is large enough, such as $n_2 > 200$, the performance becomes steady. This can indicate that (1) more employed samples can improve the generated structure and (2) only a small set of samples are essential for any selected sample.

5 Conclusions

In this paper, we introduce an accelerated framework for learning-based view-based 3-D object retrieval. In the proposed framework, a set of candidate relevant objects are first selected for each object which are further employed in the hypergraph-based learning method to generate the pairwise object relevance. Experiments were conducted to evaluate the performance of the proposed framework. As shown in the results, the proposed method can achieve comparable results to the multiple hypergraph learning method and better results compared with other state-of-the-art methods. This work is just an attempt towards efficient learning-based 3-D object retrieval methods. The employed dense subgraph extraction method can be changed to other methods based on different scenarios.

References


Qionghai Dai received the BS degree in mathematics from Shaanxi Normal University, China, in 1987, and the MEng and PhD degrees in computer science and automation from Northeastern University, China, in 1994 and 1996, respectively. Since 1997, he has been with the faculty of Tsinghua University, Beijing, China, and is currently professor and the director of the Broadband Networks and Digital Media Laboratory. His research areas include signal processing, broadband networks, video processing, and communication.

Yue Gao received the BS degree from Harbin Institute of Technology in 2005, and the MEng degree and PhD degree from Tsinghua University in 2008 and 2012, respectively. He is currently a research fellow in School of Computing at National University of Singapore. His research interests include multimedia information retrieval, pattern recognition, and social media.