3D Object Retrieval Approach for Better Perception Using View- Model Combination

M. A. Irshad (PG Scholar)¹ Dr. M. Narsing Yadav Ph.D Prof. HOĐ² R. Raja Kishore M.tech Asst. Prof.³
CERD, Department of Electronics and Communication Engineering
Mallareddy Institute of Engineering and Technology, Secunderabad, India,

mohammadabdulirshad@gmail.com¹ ecehod.mriet@gmail.com² rajakishore.r@gmail.com³

Abstract

Digital image processing and its associated technological development came long way to perceive the digital content in 3-dimensional (3D) from two-dimensional (2D). Although 3D representation of digital content has introduced digitalization in all digital image processing associated fields but finding the relevancy between 3D objects is still concerned area and extensive research must be done to achieve relevant relevancy between 3D objects. In literature, vast amount of algorithms are proposed based on model-based approach or view-based approach which results in inaccurate data for 3D representation of digital content. In this work, collaboration of both view-model relevance among 3D objects for retrieval and 3D objects perception is performed based on various graph structures. Object hypergraph structure (view information) is implemented at initial stage to perceive the 3D objects in multiple views and an object graph is constructed for model data for obtaining the information about the relationship between the different features of the obtained 3D objects. The proposed method registers good performance and reliability over traditional state of art methods and experiments are conducted on two data sets respectively.

Keywords: 3-Dimensional (3D), View-Model Relevance, Object Hypergraph Structure, View Star Expansion, Pair Wise Object Distances

1. INTRODUCTION

Digital image processing have witnessed revolution in terms of development and achieved success to introduce new technologies in medicine research field, computer graphics field, academia, robotics research field, genetics research field, etc. Although image processing achieved tremendous progress but still perception of the digital content in 3D view and achieving the relevancy between the 3D objects is still considered as area of concern in 3D object retrieval.

3D model retrieval methods can be divided into two categories: model-based methods and view-based methods. Early works are mainly model-based methods, in which low-level feature based methods (e.g. the geometric moment, surface distribution, volumetric descriptor, and surface geometry) or high-level structure-based methods are employed. Due to the requirement of 3D models, these methods are limited in the practical applications. Extensive research efforts have been dedicated to view-based 3D model retrieval methods because of the high discriminative property of multi-views for 3D object representation. Many view-based 3D object retrieval
methods (e.g. Light Field Descriptor (LFD), Elevation Descriptor (ED), Bag-of-Visual-Features (BOVF), and Compact Multi-View Descriptor (CMVD)) have been proposed these years. This is due to the fact that view based methods are with the highly discriminative property for object representation and visual analysis also plays an important role in multimedia applications. The advantages of the view-based method are twofold.

- It does not require the explicit virtual model information, which makes the method robust to real practical applications.
- Image processing has been investigated for many decades. The view-based 3D model analysis methods can be benefited from existing image processing technologies.

In this work, we propose to jointly employ both the model and the view information for 3D object relevance estimation. In the view part, representative views are firstly selected for each object, and then the view-level distances are calculated. An object hypergraph is constructed using the view star expansion. In the model part, the spatial structure circular descriptor is extracted and a simple graph is generated using the pair wise object distances. In this way, the view information and the model data can be formulated in two graph structures. Learning on the two graphs is conducted to estimate the relevance among 3D objects, in which the graph weights can be also optimized.

2. BACKGROUND

(A) 3D object retrieval

The rapid development of computer graphics hardware and 3D technologies has increasingly lead to the use of 3D objects in various applications, especially in the entertainment, medical, and architectural design industries. As a result, the need for effective and efficient 3D object retrieval methods has increased significantly as well. For instance, 3D object retrieval can help reduce the costs of model design by nearly 80 percent in the CAD field.

In general, 3D object retrieval methods can be divided into one of two categories based on either 3D models or multiple views. In 3D model-based methods, each 3D object is represented by a virtual 3D model, which can be created using statistics-, extension-, volume-, or surface-geometry-based methods, all of which use the 3D model data. Many practical applications cannot obtain a 3D model, however, so a virtual 3D model must be reconstructed. This approach is computationally expensive, and the poor performance of reconstruction methods often results in low-quality 3D models.

View-based 3D object retrieval methods, on the other hand, use a single view or multiple views for 3D object representation. These views can be obtained with either a group of cameras or a virtual camera array. Figure 1 shows several example views used to describe 3D objects. Such view-based methods do not require a 3D model, and the ubiquity of mobile devices with cameras makes it easy to obtain images of real objects. Online multi view data of 3D objects have become increasingly available as many e-business websites, such as Amazon and eBay, provide multiple views for most of their products.

(B) Retrieving 3D Objects with Multiple Views

We can define the view-based 3D object retrieval task as follows: Each object consists of one or more
views, and given one query object, the objective is to find all relevant and/or similar objects from the 3D object database under the view-based representation.

View-based 3D object retrieval has several main challenges.

- **View capture**: Views are the fundamental elements for view-based 3D object analysis. Most existing methods use a camera array that consists of a group of cameras capturing views from different directions.
- **Representative view selection**: Although a large number of views can provide rich information, they also introduce redundant and noisy data and result in high computational costs.
- **Feature extraction**: It is still difficult to extract features for multiple views because of the special characteristics of 3D data. The spatial correlation among different views should be taken into consideration, which still requires further investigation.
- **Object matching using multiple views**: Most of the existing image retrieval tasks are based on one-to-one image matching. View based 3D object retrieval, however, focuses on multiple view matching. Thus, it is challenging to determine how best to conduct many-to-many view matching and estimate the relevance among different 3D objects.

3. PROPOSED METHOD

(A) 3D-object retrieval and recognition with hypergraph analysis

View based 3-D object retrieval and recognition has become popular in practice, e.g., in computer aided design. It is difficult to precisely estimate the distance between two objects represented by multiple views. Thus, current view-based 3-D object retrieval and recognition methods may not perform well. In this paper, we propose a hypergraph analysis approach to address this problem by avoiding the estimation of the distance between objects: In particular, we construct multiple hyper graphs for a set 3-D objects based on their 2-D views. In these hyper graphs, each vertex is an object, and each edge is a cluster of views. Therefore, an edge connects multiple vertices. We define the weight of each edge based on the similarities between any two views within the cluster. Retrieval and recognition are performed based on the hyper graphs. Therefore, our method can explore the higher order relationship among objects and does not use the distance between objects.

(B) View-based hypergraph generation

Here the view-based hypergraph is generated following the method in [20] and briefly introduced as follows. Let \( O=\{O_1, \ldots , O_n\} \) denote then3D objects in the data set, and \( V_i=\{v_{i1}, \ldots , v_{in}\} \) denote the \( n_i \) views of the \( i_{th} \) 3D object \( O_i \). In this part, we aim to explore the relevance among 3D object with multiple view information.

Generally, although multiple views can represent rich information of 3D objects, they also bring in redundant data, which may cause much computational cost and even lead to false results. Here we first select representative views for each 3D object, and only these representative views are employed in the 3D object retrieval process.
Given then $n_i$ views $V_i = \{v_{i1}, \ldots, v_{im}\}$ of $O_i$, we conduct hierarchical agglomerative clustering (HAC) to group these views into view clusters. The HAC method is selected here due to that it can guarantee the intra cluster distance between each pair of views cannot exceed a given threshold. Here the widely employed Zernike moments are used as the view features, which are robust to image rotation, scaling and translation and have been used in many 3D object retrieval tasks. The 49-D Zernike moments are extracted from each view of 3D objects. With the view clustering results, one representative view is selected from each view cluster. Here we let $V_i = \{v_{i1}, \ldots, v_{im}\}$ denote the $m_i$-representative views for $O_i$. In our experiments, $m_i$ mostly ranges from 5 to 20.

Hypergraph has been used in many multimedia information retrieval tasks, such as image retrieval. Hypergraph has shown its superior on high-order information representation. In our work, we propose to employ star expansion to construct an object hypergraph with views to formulate the relationship among 3D objects. Here we denote the object hypergraph as $G_H = (V_H, E_H, W_H)$. For the $n$ objects in the dataset, there are $n$ vertices in $G_H$, where each vertex represents one 3D object.

![Figure 1: An illustration of hyper edge construction.](image)

In this figure, there are seven objects with representative views. Here one view from $O_4$ is selected as the centra view, and its four closest views are located in the figure, which are from $O_1$, $O_3$, $O_6$ and $O_7$. Then the corresponding hyper edge connects $O_1$, $O_3$, $O_4$, $O_6$ and $O_7$.

The hyper edges are generated as follows. We assume there are totally $nr$ representative views for all $n$ objects. We first calculate the Zernike moments-based distance between each two views, and the top $K$ closest views can be generated for each representative view. For each representative view, one hyper edge is constructed, which connects the objects with views in the top $K$ closest views. In our experiment, $K$ is set as 10. Figure 3 shows an example of hyper edge generation. Generally, $n_r$ hyper edges can be generated for $G_H$. The weight of one hyper edge $e_H$ can be calculated by

$$w_H(e) = \frac{1}{K} \sum \exp \left( \frac{d(V_c, V_x)^2}{\sigma_H^2} \right) \quad (1)$$

Where $V_c$ is the centra view of the hyper edge, $V_x$ is one of the top $K$ closest view to $V_c$, $d(V_c, V_x)$ is the distance between $V_c$ and $V_x$, and $\sigma_H$ is empirically set as the median of all view pair distances.

Given the object hyper graph $G_H = (V_H, E_H, W_H)$, the incidence matrix $H$ can be generated by

$$h(v_{e_H}) = \begin{cases} 1 & \text{if } V_c \in e_H \\ 0 & \text{if } V_x \notin e_H \end{cases} \quad (2)$$

The vertex degree of $V_c$ can be defined as

$$\rho(v_H) = \sum_{e_H \in E_H} o(e_H) h(v_H, e_H) \quad (3)$$

The edge degree of $e_H$ can be defined as

$$\rho(e_H) = \sum_{V \in V_H} h(V_H, e_H) \quad (4)$$
The vertex degree matrix and the edge degree matrix can be denoted by two diagonal matrices $D_v$ and $D_e$.

In the constructed hypergraph, when two 3D objects share more similar views, they can be connected by more hyper edges with high weights, which can indicate the high correlation among these 3D objects.

(C) Model-based graph generation

Given the model data of 3D objects, here we further explore the model-based object relationship. Here the spatial structure circular descriptor (SSCD) is employed as the model feature. SSCD aims to represent the depth information of the model surface on the projection minimal bounding box of the 3D model. The depth histogram is generated as the feature for the 3D model. Following [21], the bipartite graph matching is conducted to measure the distance between each two 3D models, i.e., $d_{SSCD}(O_i,O_j)$

Here, the relationship among 3D objects is formulated in a simple object graph structure $G=(V,E,W)$. Here each vertex in $G$ represents one 3D object, i.e., there are $n$ vertices in $G$. The weight of an edge $e(i,j)$ in $G$ is calculated by using the similarity between two corresponding 3D objects $O_i$ and $O_j$ as

$$W(V_i,V_j) = \exp\left(-\frac{d_{SSCD}(V_i,V_j)^2}{\sigma_s^2}\right) \quad (5)$$

Where $d_{SSCD}(V_i,V_j)$ is distance between $O_i$ and $O_j$, and $\sigma_s$ is set as the median of all modal pair distances.

(D) Learning on the joint graph

Now we have two types of formulation of relationship among 3D objects, i.e., view-based and model-based. Here these two formulations are jointly explored to estimate the relevance among 3D objects.

In this part, first we introduce the learning framework when the view-based and model-based information are regarded with equal weight, and then we propose a jointly learning framework to learn the optimal combination weights for each modality.

(1) The initial learning framework

Here we start from the learning framework which regards different modalities, i.e., model and view, as equal. The 3D object retrieval task can be formulated as the one-class classification work as shown in [51]. The main objective is to learn the optimal pairwise object relevance under both the graph and hypergraph structure. Given the initial labeled data (the query object in our case), an empirical loss term can be added as a constraint for the learning process. The transductive inference can be formulated as a regularization as

$$\arg\min_f \{\Omega_v(f) + \mu R(f)\} \quad (6)$$

In this formulation, $f$ is the to-be-learnt relevance vector, $\Omega_v(f)$ is the regularizer term on the view-based hypergraph structure, $\Omega_m(f)$ is the regularizer term on the model-based graph structure, $R(f)$ is the empirical loss. This objective function aims to minimize the empirical loss and the regularizers on the model-based graph and the view-based hypergraph simultaneously which can lead to the optimal relevance vector $f$ for retrieval. The two regularizers and the empirical loss term are defined as follows.

The view-based hypergraph regularizer $\Omega_v(f)$ is defined as
\[ \Omega_v(f) = \frac{1}{2} \sum_{e \in \mathcal{E}} \sum_{u,v \in \mathcal{V}} w_{eH} \left( \frac{f(u)}{p(eH)} \right)^2 \]

\[ = \frac{1}{2} \sum_{e \in \mathcal{E}} \sum_{u,v \in \mathcal{V}} w_{eH} \left( \frac{f(u)}{p(eH)} \right)^2 \]

\[ = \frac{1}{2} \sum_{e \in \mathcal{E}} \sum_{u,v \in \mathcal{V}} w_{eH} \left( \frac{f(u)}{p(eH)} \right)^2 \]

\[ = f^T \left( I - \theta_c \right) f \]

\[ = \sum_{u,v \in \mathcal{V}} w_{eH} \left( \frac{f(u)}{d(u)} - \frac{f(v)}{d(v)} \right)^2 \]

\[ = f^T \left( I - \Delta_s \right) f \]

Where \( \theta_c \) is defined as \( \theta_c = D^{-1/2}W D^{-1/2} \). Here we denote \( \Delta_s = I - \theta_c \). \( \Omega_m(f) \) can be written as

\[ \Omega_m(f) = f^T \Delta_s f \]

(10)

The empirical loss term \( R(f) \) is defined as

\[ R(f) = \| f - y \|^2 \]

(11)

Where \( y \) is the initial label vector. In the retrieval process, it is defined as an \( n \times 1 \) vector, in which only the query is set as 1 and all other components are set as 0.

Now the objective function can be rewritten as

\[ \arg \min_f \{ f^T \Delta_H f + f^T \Delta_s f + \mu \| f - y \|^2 \} \]

(12)

f can be solved by

\[ f = (I + \frac{1}{\mu} (\Delta_H + \Delta_s))^{-1} y \]

(13)

f is the relevance of all the objects in the dataset with respect to the query object. A large relevance value indicates high similarity between the object and the query. The higher the corresponding relevance value is, the more similar the two objects are. With the generated object relevance f, all the objects in the dataset can be sorted in a descending order according to f.

(2) Learning the combination weights

We noted that the view information and the model information may not share the same impact on 3D object representation. In some scenarios, the view information may be more important, and in some other cases, the model data may play an important role. Under such circumstances, we further learn the optimal weights for the view information and the model data. In this part, we introduce the learning framework embedding the combination weight learning. The objective for the learning process is composed of three parts, i.e., the graph/hypergraph structure regularizers, the empirical loss and the combination weight regularizer.
Here we let $\alpha$ and $\beta$ denote the combination weights for view-based and model-based information respectively, where $\alpha + \beta = 1$. After adding the l2-norm on the combination weights, the objective function can be further revised as

$$\min_{f, \alpha, \beta} \{ \alpha f^T \Delta_H f + \beta f^T \Delta_S f + \mu \| f - y \|^2 + \eta (\alpha^2 + \beta^2) \} \quad (14)$$

where $\alpha + \beta = 1$.

The solution for the above optimization task is provided as follows. To solve the above objective function, we alternatively optimize $f$ and $\alpha/\beta$. We first fix $\alpha$ and $\beta$, and optimize $f$.

Now the objective function changes to

$$\min_f \{ \alpha f^T \Delta_H f + \beta f^T \Delta_S f + \mu \| f - y \|^2 \} \quad (15)$$

According to Eq. (13), it can be solved by

$$f = \left( I + \frac{1}{\lambda} (\alpha \Delta_H + \beta \Delta_S) \right)^{-1} y \quad (16)$$

Then we optimize $\alpha/\beta$ with fixed $f$. Here we employ the Lagrangian method, and the objective function changes to

$$\min_{\alpha, \beta} \{ \alpha f^T \Delta_H f + \beta f^T \Delta_S f + \eta (\alpha^2 + \beta^2) + \xi (\alpha + \beta - 1) \} \quad (17)$$

Solving the above optimization problem, we can obtain

$$\xi = - \frac{f^T \Delta_H f + f^T \Delta_S f}{2} - \eta, \quad (18)$$

$$\alpha = \frac{1}{2} - \frac{f^T \Delta_H f - f^T \Delta_S f}{4 \eta} \quad (19)$$

The above alternative optimization can be processed under the optimal $f$ value is achieved, which can be used for the 3D object retrieval. With the learned combination weights, the model-based and view-based data can be optimally explored simultaneously and the relevance vector $f$ can be obtained. The main merit of the proposed method is that it jointly explore the view information and the model data of 3D objects in hypergraph/graph frameworks for 3D object retrieval.

4. RESULTS

Figure 2: Query image

Figure 3: Retrieved images
5. CONCLUSION

Joint View-based and model based 3D object retrieval is an essential topic with many emerging applications. The next stage of research in this field will need to not only focus on the key technologies for view-based object retrieval but also extend it to general domains, which can certainly benefit from the achievements of view-based object analysis. Combination of both view-model relevance among 3D objects for retrieval and 3D objects perception is performed based on various graph structures. Object hypergraph structure (view information) is implemented at initial stage to perceive the 3D objects in multiple views and an object graph is constructed for model data for obtaining the information about the relationship between the different features of the obtained 3D objects.

6. REFERENCES


