IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 24, 2019, accepted January 12, 2020, date of publication January 21, 2020, date of current version January 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968460

View-Based 3D Model Retrieval by Joint Subgraph Learning and Matching

HAIPENG GUAN[1], QIUXIA ZHAO[1], YAN REN[1], AND WEIZHI NIE[2]
[1]Maths and Information Technology School, Yuncheng University, Yuncheng 044000, China
[2]School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

Corresponding authors: Yan Ren (renyan-2000@163.com) and Weizhi Nie (weizhinie@tju.edu.cn)

This work was supported by the National Natural Science Foundation of Shanxi under Grant 201601D021014, in part by the Higher Education Technology Innovation Projects Foundation of Shanxi under Grant 2019L0860, in part by the National Natural Science Foundation of China under Grant 61702124, and in part by the Subject Research Projects Foundation of Key Laboratory of Information Security Technology of Guangdong under Grant GDXXAQ2016-05.

ABSTRACT View-based 3D model retrieval is an important and challenging task in computer vision, which can be utilized in many applications such as landmark detection, image set classification, etc. Representation view selection and similarity measure are two key problem in view-based 3D model retrieval. Many classic algorithms were proposed to handle these two problems. However, they were often independent to consider these two problems while ignoring the contact with each other. In this paper, we proposed a joint subgraph learning & matching method (SGLM) via Markov Chain Monte Carlo (MCMC) to handle view-based 3D model retrieval problem, which effectively combine representation view extraction with similarity measure process to find the best matching result. The proposed (SGLM) can benefit: 1) considering the correlation between representation view selection and similarity measure, which can effectively improve the final performance of retrieval; 2) eliminating redundant visual information by subgraph learning; 3) learning representation views automaticly in similarity measure process. We validate the SGLM based on 3D model retrieval on ETH, PSB, NTU and MVRED datasets. Extensive comparison experiments demonstrate the superiority of the proposed method.

INDEX TERMS View-based 3D model retrieval, subgraph learning, Markov chain Monte Carlo, graph matching.

I. INTRODUCTION

Compared with 2D images, 3D models have the advantages to reappear the objects'spatial and structure information which is more suitable for the visual perception system of humans. Therefore, 3D objects are becoming ubiquitous in many application fields, such as computer-aided design, construction design, medical study and entertainment gaming [1]. With the development of 3D model construction technology, more and more 3D models are generated in kinds of applications. 3D model retrieval is also becoming a key technology in computer vision. However, it is hard to obtain the 3D model information and representation in real world according to the limitation of collecting equipments and collecting environment. Thus, 3D model retrieval is also one challenging problem in computer vision.

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

A. MOTIVATION AND OVERVIEW

In recent years, many effective algorithms have been proposed to handle 3D model retrieval problem. Generally, 3D object retrieval methods can be divided into one of two categories: model-based and view-based methods [2]. Model-based methods directly utilize 3D model data to extract effective low-level features [3]–[5] for retrieval, which can fully leverage the spatial and structure information of 3D model. However, model-based methods are limited in the practical applications because the limitation of 3D model reconstruction and generation technologies. Thus, researchers have been actively engaged in view-based methods [6].

In view-based methods, one 3D model can be converted to a set of views for representation, which can effectively avoid the challenge of 3D model reconstruc-

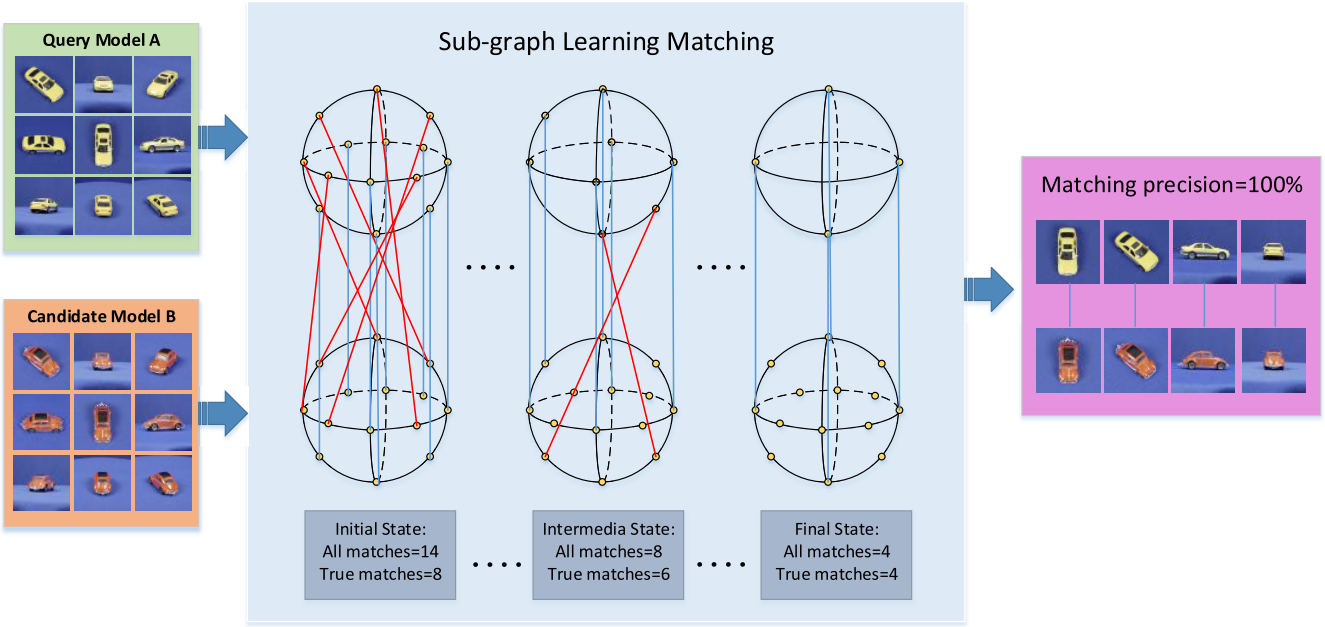


FIGURE 1. The framework of the sub-graph learning method on 3D model retrieval. Each yellow node represents each view of the view-based 3D model. We utilize MCMC model to select subgraphs from the query model to get accurate matching scores by multiple samples.

tion. Meanwhile, the model retrieval can be accomplished by measuring the similarity between different models using multi-view data [7]. Lu *et al.* [8] combined view-based with model-based methods in a graph-based framework for 3D model retrieval. The experimental results obviously demonstrate that view-based methods can achieve better performance than model-based methods. In general, view-based retrieval process mainly consists of four procedures: view capture, represent-view selection, feature extraction, and similarity measurement [9]. The last two steps are the key processes. Many methods focus on these two steps [10] and also demonstrate the effectiveness of view-based method. However, two challenges still remains:

- Representation views selection: many recent methods often apply some simple view clustering and center selection via visual low-level features. However, these methods are not robust and effective enough to filter redundant and noisy data. Both the missed inliers and the existence of outliers will have a negative influence on similarity measure step;
- Similarity measurement: statistical model, probability model are utilized to handle similarity measure between two different 3D models. However, these methods all focus on the distribution of visual feature and ignore the spatial and structure information of views. Graph matching-based methods often get better performance than statistical and probability model. However, the current graph matching-based methods often focus on the node-to-node mapping and ignore the global structure of graph model. It is not suitable for many real applications.

In order to handle these problems, we proposed a novel sub-graph learning matching method (SGLM) based on Markov Chain Monte Carlo framework, like in Fig.1, to handle 3D model retrieval problem, which can combine the representation views selection with similarity measure steps into one step and effectively improve the effectiveness of retrieval. The subgraph selection can effectively filter the existence of outliers or redundant information. The step of graph matching can consider the local and global structure information of graph model to help the subgraph selection. These two steps are complementary in our approach. Consequently, the proposed method can aid in redundant and noisy data elimination by strengthening inliers while suppressing outliers. The extensive comparison experiments demonstrate this method's superiority.

B. CONTRIBUTIONS

The contributions of this paper are followed as:

- The paper proposed a novel sub-graph learning matching methods that considers the spatial, structure, and visual information for 3D model retrieval. Theoretically, the proposed method can be considered as the improvement of a classic graph matching method, which can filter the negative influence of outlier and also preserve both global structure information of graph. In the learning process, the proposed method also fully considers correlation among nodes. The corresponding experiments also demonstrate our conclusion;
- For the sub-graph learning, we proposed a modified object function by adding two penalties. These two penalties can limit the size of sub-graph and also save

and strength the inner correlation among nodes, which can effectively filter outliers. In the learning process, Markov chain Monte Carlo (MCMC) is utilized to optimize the object function. The initial nodes of subgraph are generated by classic graph matching, which can guarantee the effectiveness of matching and also assure the accuracy of matching;
- The proposed method is extensively evaluated using three popular single-modal datasets and one novel multi-modal dataset. We discuss the influence of different parameters in object function and the physical meaning by some experimental results. We also compare the proposed method with some classic graph matching methods and some state-of-the-arts. Finally, we explore and compare their performances by varying the view numbers of each model to demonstrate the robustness of our approach.

The rest of this paper is organized as follows. Related work on 3D object retrieval is reviewed in Section II. The proposed method is provided in Section III. Experiments and discussion are given in Section IV and V. We conclude the paper in Section VI.

II. RELATED WORK

In this section, we discuss related works on 3D model retrieval and graph matching problem. We can find a number of different approaches that address 3D model retrieval task by using model-based or view-based techniques or even collaboration of all above process [11]–[13].

For model-based 3D model retrieval, the shape descriptors come first as representation. There are some kinds of shape descriptors including histogram-based, transform-based and graph-based method. Shape distribution [4], generalized shape distribution [14] and 3D Hough transform [15] can be classified into the first type. While 3D Fourier Transform [16] and Spherical Wavelet Transform [17] belong to the second class. One typical graph-based technique was exampled in [18], where the author developed a matching technique to compute the similarity between two objects based on the shape matching of their Reeb Graphs. In [19], the author proposed several lightweight local binary 3D shape features which were inspired by compact binary features for 2D images. The distance in Hamming space between these binary features could be computed efficiently. The proposed method in [20] was based on Bag of Features (BoF) paradigm and salient characteristics of objects. Fourier descriptor was adopted to represent each interest point calculated on the surface.

For view-based 3D model retrieval, extensive efforts have been focused on this area. Generally, there are two key issues, i.e. model representation and model similarity computing. As the first view-based 3D retrieval method, Lighting Field Descriptor(LFD) [21] utilized 10 views to represent each object and employed Zernike moments and Fourier descriptors as the view features. Subsequently, large numbers of specific descriptors came forth. For instance, Elevation Descriptor (ED) [22], Compact Multi-View Descriptor(CMVD) [23], Spatial Structure Circular Descriptor(SSCD) [24] and so on. Tree-based shape representations were presented in [25] as a novel framework of 3D model retrieval task. In order to make distance estimation between two groups of views, many distinguished methods have been proposed. In [26] [27], the comparison among two objects was formulated as a probabilistic model to compute the relationship. A Hausdorff distance learning approach was employed in [28], in which a Mahalanobis distance metric was learnt in view-level while Hausdorff distance was employed in object-level. Wang and Nie [29] utilized the reconstruction error to compute the relevance between the query and candidate.

Lu *et al.* [8] firstly proposed to jointly learn both the view and model relevance among 3D objects for retrieval. Hypergraph structure was employed to formulate the 3D object relationship via multiple views while SSCD feature was extracted to construct a graph to describe the relevance.

Graph matching has been widely used in vision tasks such as shape matching, feature matching and object retrieval. Most graph matching algorithms have been formulated as the Integer Quadratic Programming(IQP) [30] and most research has focused on developing efficient and robust algorithms. Spectral Matching(SM) [31] approximately found the optimal assignment through maximizing the dot-product with the principle eigenvector of the affinity matrix. Cho *et al.* [32] introduced a random walk algorithm to solve the optimization problem. A novel graph matching technique based on max-pooling [33] was proposed to improve the performance by evaluating each candidate match using the most promising nearby matches.

III. PROPOSED METHOD

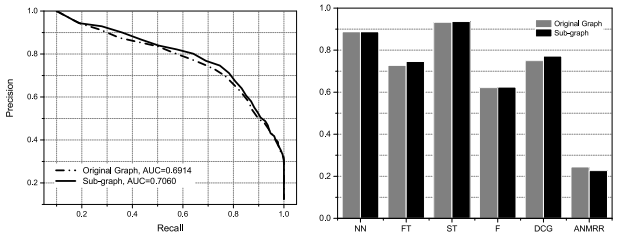
In this section, we will introduce the proposed method (SGLM) for view-based 3D model retrieval task, which combine the characteristic views extraction and similarity measure into a step and improve the effectiveness of retrieval. A Markov chain is utilized to sample subgraphs while a classic graph matching method is applied to realize graph matching based on the learning subgraphs. We detail these technologies in the next sections.

A. DEFINITION OF GRAPH MATCHING

Graph matching problem must consider the one-to-one matching results between different graphs. We utilize the $G_1 = \{V_1, E_1, A_1^v, A_1^e\}$ and $G_2 = \{V_2, E_2, A_2^v, A_2^e\}$ to represent graphs, where V, E, A^v and A^e denote a set of nodes, edges, node attributes and edge attributes, respectively. The goal of graph matching is to find the subset of node correspondences between G_1 and G_2 among all of the possible correspondences. $y \in \{0,1\}^{n_1 * n_2}$ is used to represent the matching result. Here, n_1 and n_2 represent the number of nodes in G_1 and G_2 respectively. y_i is 1 if *i-th* match is selected and 0, otherwise. Affinity matrix M consists of the relational similarity values between edges and nodes.

TABLE 1. The comparison between original graph and sub-graph.

	Original Graph	Sub-Graph
All Matching Views	41	6
True Matching Views	25	6
Matching Score	18.06	4.85
Matching Precision	61%	100%



(a) PR-Curve (b) Evaluation

FIGURE 2. The comparison between original graph and sub-graph in ETH.

Let us consider a simple example of graph matching in view-based 3D model matching. We selected two view-based 3D models from ETH dataset. Meanwhile, we also selected six views from the six faces of the cube as subgraph (representation views) to represent image set extracted from the original 3D model. The final matching results are shown in Table.1. We can find that the matching score of sub-graph is smaller than original graph. However, the precision of matching based on sub-graph is higher than that of original graph. Sub-graph obviously can get better matching results if the subgraph can effectively represent original view-based 3D model. We applied the same setting in all ETH dataset and utilized retrieval result to demonstrate our conclusion. The matching score is utilized as similarity between different objects. The final experimental results are shown in Fig.2.

By observing, the result of sub-graph outperforms that of original view-based 3D model, which means that IQP solution has higher matching score and does not guarantee all true matches. Thus, the matching score is also hard to guarantee its reliability. The sub-graph extracted from original data can filter redundant information and get more precision of matching.

B. MODIFICATION OF GRAPH MATCHING

In order to handle this problem of outliers, we design to select one subgraph without outliers from original graph by inspired by [34]. In the matching process, this design can effectively guarantee the precision and recall of matching. Thus, the object function of matching can be rewritten as:

$$\begin{aligned} y^* = \ & arg\max_y y(x)^T My(x) - \lambda_1||y||_0 + \lambda_2 x^T Wx, \\ & s.t.\ y \in \{0,1\}^{n_1 * n_2}, \quad x \in \{0,1\}^{n_1} \\ & \sum_j y_{ij} \le 1\ , \quad \sum_i y_{ij} \le 1. \end{aligned} \tag{1}$$

where W is the similarity matrix of nodes in graph G_1, $X = \{x | x \in \{0,1\}^{n_1}\}$, n_1 is the number of nodes in G_1. The function $y(\cdot)$ is defined as:

$$\begin{aligned} y(x) = \ & arg\max_y y^T My \\ & s.t.\ y \in \{0,1\}^{n_1 * n_2} \\ & \sum_j y_{ij} \le 1, \ \sum_i y_{ij} \le 1. \end{aligned} \tag{2}$$

Eq.2 can be handled by the tradition IQP optimization method [35] between two graph G_1' and G_2, where G_1' is the subgraph of G_1.

The goal of Eq.1 is to select the subgraph G_1' from original graph G_1, where the subgraph G_1' do not fully represent the information of graph G_1. However, it should represent the most of structure information. Meanwhile, based on the 3D model retrieval problem, we modified the original subgraph matching method [34]. We define the $x^T Wx$ is the positive term. The reason is that each node is a virtual view extracted from 3D model, which should have strong correlation. They have no outliers in 3D model retrieval problem. Thus, we make the above changes. In orther word, y represents the matching pairwise between G_1' and G_1. Thus, $||y||_0$ is leveraged to constrain the size of subgraph. λ_1 and λ_2 are used to control the weight of penalties respectively.

C. ALGORITHM

In this paper, we applied MCMC framework to solve the optimization problem. We define the target probability distribution on X as follows:

$$P(x) = e^{(y(x)^T My(x) - \lambda_1||y||_0 + \lambda_2 x^T Wx)}, \tag{3}$$

One state transition in the Markov chain is a selection of x in space X with its best graph matching with G_2 according to the Eq.2.

According to our work [34], the classic Metropolis-hastings algorithm is used to generated the Markov chain. The random proposal and state-transfer proposal are defined. In random proposal, we select one node from V_1 and flipped from active to inactive or vice versa:

$$q_{rand}(x \to x') = \begin{cases} 1/n_1 & if\ ||x' - x||_1 = 1 \\ 0 & otherwise \end{cases} \tag{4}$$

The state-transfer proposal is designed based on the similarity between subgraph and original graph model. If the subgraph has been selected, the structure should be stable and and similarity with original graph is hard bo be changed. Thus, the state-transfer proposal can be defined as:

$$q_{data}(x \to x') = \begin{cases} \frac{1}{Z} e^{-\frac{|x'^T Wx' - x^T Wx|}{\tau}} & if\ ||x' - x||_1 = 1 \\ 0 & otherwise \end{cases} \tag{5}$$

where Z and τ are the normalizing constants. We sample a node according to the distribution Eq.5, which will make the sampling constrained in a small sets of nodes and find the optimal subgraph structure quickly.

Finally, the acceptance ratio is obtained as follows:

$$a(x \rightarrow x') = min(\frac{p(x')q(x' \rightarrow x)}{p(x)q(x \rightarrow x')}, 1), \quad (6)$$

where $q(\cdot)$ is defined as Eq.4 or Eq.5, according to the proposal type.

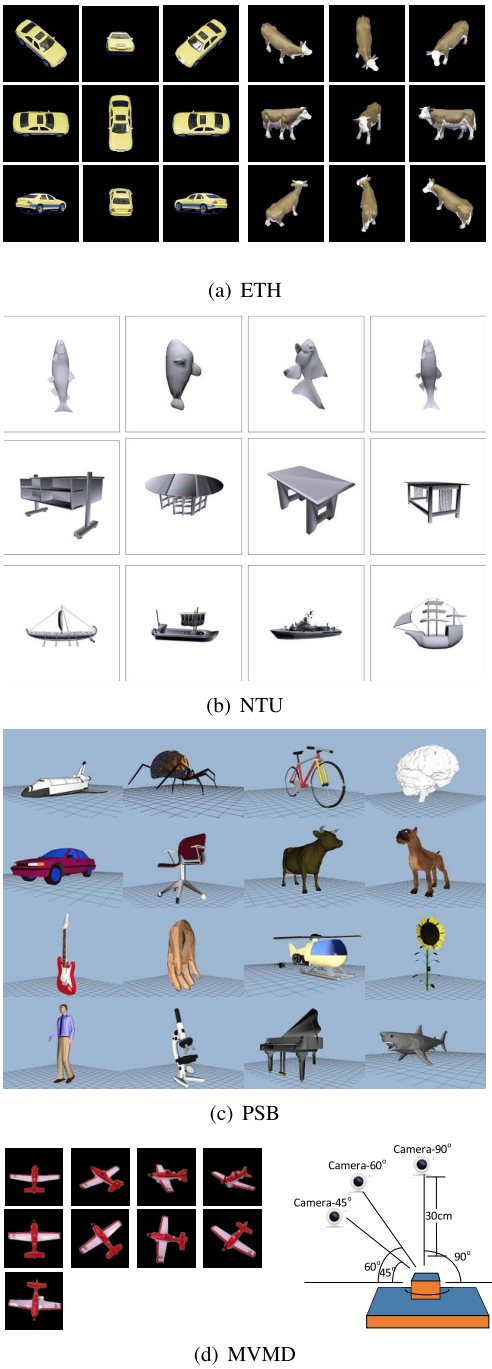
IV. EXPERIMENTAL METHODS

In order to test the performance of the proposed method on 3D model retrieval, we conduct experiments on four datasets and compare our approach with several state-of-the-art methods. In this section, we first introduce the testing datasets, followed by evaluation criteria and settings.

A. TESTING DATASETS

To evaluate our proposed approach, we conduct experiments on four datasets: ETH 3D object collection (ETH) [36], National Taiwan University 3D Model database (NTU) [21], Princeton Shape Benchmark (PSB) [37] and Multi-view RGB-D Object Dataset (MV-RED) [38].

- ETH: This is a real-world 3D object multi-view database which contains 80 objects belonging to 8 categories. Each object from ETH dataset has 41 views spaced evenly over the upper viewing hemisphere, and all the positions for cameras are determined by subdividing the faces of an octahedron to the third recursion level.
- NTU: This is a virtual 3D model dataset which is composed of 549 models from 47 categories. We select 210 models divided into 22 categories, such as boat, car, gun, truck, plane and so on. For each category, there are 8 to 10 models. To generate the multiple views of each 3D model, a virtual camera array including 60 cameras is employed. These 60 virtual cameras are set on the vertices of a polyhedron with the same structure of Buckminsterfullerene (C60).
- PSB: This dataset contains 1,814 models from 161 classes. We choose 30 classes with 10 models in each class, thus constructing a 300-object dataset. The 60 cameras used in the NTU dataset are employed here to capture the 60 views for each 3D model in the PSB.
- MVRED: This dataset is recorded by Kinect cameras and contains 505 models of the real world objects categorized into 61 classes. For each object, the RGB and depth information were recorded simultaneously. Thus, each object has multi-view and multi-modal information. In our experiment, 311 objects are selected as queries to retrieve all 505 objects in the dataset.



(a) ETH

(b) NTU

(c) PSB

(d) MVMD

FIGURE 3. 3D model samples from different datasets.

B. EVALUATION CRITERIA

In our experiments, the following popular criteria are employed as the measures of the retrieval performance.

- Precision-Recall Curve (PR-Curve) is able to comprehensively demonstrate the retrieval performance, which illustrates the precision and recall measures by varying the threshold for distinguishing relevance and irrelevance in model retrieval.
- Nearest Neighbor (NN) is the correct rate of the first returned result.
- First Tier (FT) is defined as the recall of the top τ results, where τ is the number of relevant samples for the query.
- Second Tier (ST) is defined as the recall of the top 2τ results, where τ is the number of relevant samples for the query.
- F-measure (F) jointly evaluates precision and recall for a fixed number of top returned results. In our experiments,

F considers the first 20 retrieved models for every query objects and calculates the precision and recall over those results.

- Discounted Cumulative Gain (DCG) is a statistic that assigns relevant results near the front of the list higher weights under the assumption that a user is less likely to consider results near the end of the list.
- Average Normalized Modified Retrieval Rank (ANMRR) is another objective measure to evaluate the retrieval performances, where a lower ANMRR value indicates a better performance.

Here, PR-Curve evaluate the precision and recall of the retrieval result. It can show the global performance. However, it also focused on the results at the back. Thus, it's result is not effective. On the contrary, NN focus on the first result, it can directly affect the performance of the retrieval methods. FT and ST show the precision of retrieval in top τ results and top 2τ results respectively, which is similar to NN. F-measure, DCG and ANMRR are the comprehensive evaluation metrics, which can effectively represent the performance of retrieval.

C. EXPERIMENT SETTINGS

Zernike moment was extracted as the view feature in our experiment because of its robustness to scaling and rotation. We set four different experiments on four popular datasets to verify the effectiveness of our method.

- We vary the weights λ_1 in Eq.1 from 0.1 to 2.0 to explore the optimal number of subgraphs and their structural meaning;
- We vary the weights λ_2 in Eq.1 from 0.1 to 1.0 to find the optimal parameter for 3D retrieval performance;
- We compare our method with several classic graph matching methods in the ETH, NTU, PSB and MVRED dataset;
- We compare our method with some state-of-the-art view-based 3D model retrieval methods in the ETH, NTU, PSB and MVRED dataset;
- We vary the view numbers from 10 to 60 with the step 10 in MVRED dataset to evaluate its effect on the performance.

D. THE PERFORMANCE ON DIFFERENT PARAMETER λ_1

λ_1 is a key parameter in the object function, which can effectively control the construct and number of nodes in subgraph. In order to test the performance on different parameter λ_1, we conduct experiments on the ETH dataset. The value of λ_1 is tuned from 0.1 to 2.0. We fixed $\lambda_2 = 0.5$. The average number of nodes in subgraph is shown in Fig.4(a). From this figure, we can find that the number node of subgraph declines when λ_1 increases. The corresponding retrieval results are shown in Fig.4(b). In Fig.4(b), we can find that the best retrieval result is appeared when $\lambda_1 = 0.9$. Meanwhile, the average number of nodes in subgraph is 30 when $\lambda_1 = 0.9$. (The number of views is 41 on ETH dataset for each

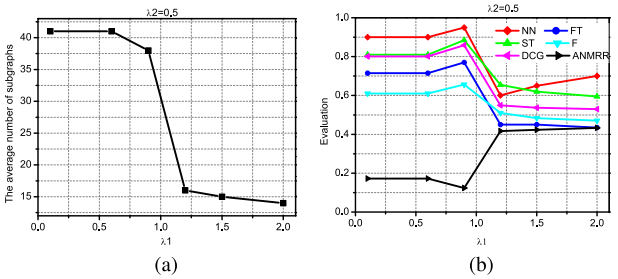


FIGURE 4. (a)The average number of subgraphs in ETH. (b)The evaluation with different λ_1 in ETH.

object). We also can find that the retrieval result has a rapid decline when the value of λ_1 declines. The reason is that λ_1 limits the number of views in subgraph which leads that more useful information is removed and accuracy of matching was affected.

At the same time, all of views are used to measure the similarity between different models which is also hard to guarantee the accuracy of matching. The reason is that more views must include more visual redundant information. More interference is hard to get more accuracy matching result. Our approach can effectively remove visual redundant information while utilizing subgraph to represent the original graph model. Fig.5 shows the characteristic of our approach. Figure.5(a) shows the final subgraph learned by our approach. Fig.5(b) shows the subgraph or represent views learned by traditional 3D retrieval method [39]. K-means was utilized to cluster all of views and center view of each cluster was utilized as represent view [39]. This method reduced the number of matching view and improved the effective of matching process. However, it only considered the correlation of views in feature space and ignored the spatial and structure information. Thus, we can find that the uniformity of the classification results in Fig.5. Meanwhile, our approach considers the spatial information of different views to learn subgraph. The final results also show the structure information of object. Each node of subgraph represents each view, which has obvious difference with other nodes. Meanwhile, we also can find that these nodes of subgraph have obvious difference in spatial information. All of this condition also demonstrate the superiority of our approach.

E. THE PERFORMANCE ON DIFFERENT PARAMETER λ_2

λ_2 is utilized to control the weight of internal correlation of subgraph in order to guarantee that subgraph can effectively represent the structure of original graph. Here, we fixed $\lambda_1 = 0.9$ and the value of λ_2 is tuned from 0.1 to 1.0 to evaluate its effect on the sub-graph matching method. The evaluation metrics of retrieval 1.0are utilized to find the best parameter λ_2. The experiment is implemented on the ETH dataset. The final experimental results are shown in Fig.6. From these expermental results, The upper bound retrieval result are got when $\lambda_2 = 1.0$. The performance will degrade when λ_2 either increases or decreases after the optimal λ_2 is achieved, which

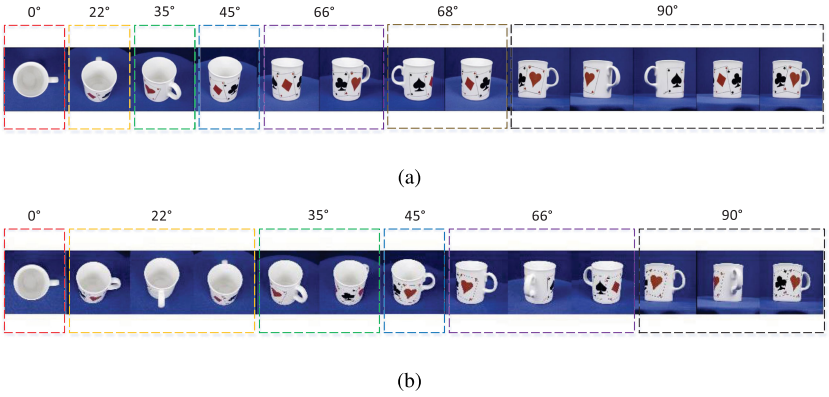


FIGURE 5. (a)The subgraphs derived from our method. (b)The representative views generated by K-means. The views selected by our approach have more differences than traditional methods. Our approach can remove more redundant information.

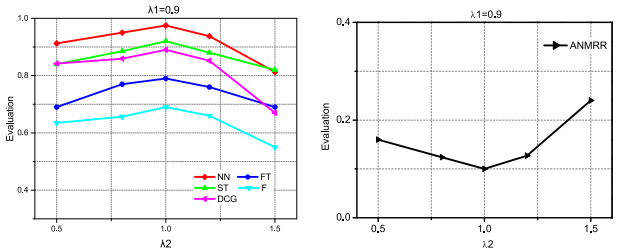


FIGURE 6. The evaluation with different λ_2 in ETH.

demonstrate that the internal correlation should be considered. However, overemphasized weight of internal correlaiton will lead that subgraph learned focuses on otherness of nodes and ignores much useful information. Meanwhile, neglectful weight of internal correlation also reduce the effect of inner strucutre of subgraph and lead more redundant information in subgraph learned. Thus, one appropriate parameter λ_2 is very improtant for the proposed method.

F. COMPARISON WITH PREVIOUS WORK

The proposed work is inspired by our previous work [34]. The goal of $x^T Wx$ represents the correlation among nodes of subgraph, which is used to guarantee that the subgraph includes more nodes without outliers. However, in this paper, we define the $x^T Wx$ is the positive term. The reason is that each node is a virtual view extracted from 3D model, which should have strong correlation and cannot be seen as the outliers. We hope the positive term can strength the correlation among nodes of subgraph and effectively reduce the redundant information. The removal of redundant information can directly improve the accuracy the final graph matching. We compared with previous methods on NTU and PSB dataset. The related experimental results are shown in Table.2 and Table.3.

Here, No-$x^T Wx$ meets that we remove $x^T Wx$ from the Equation.3. SGLM is the proposed method in this paper.

TABLE 2. Comparison with previous work on NTU dataset.

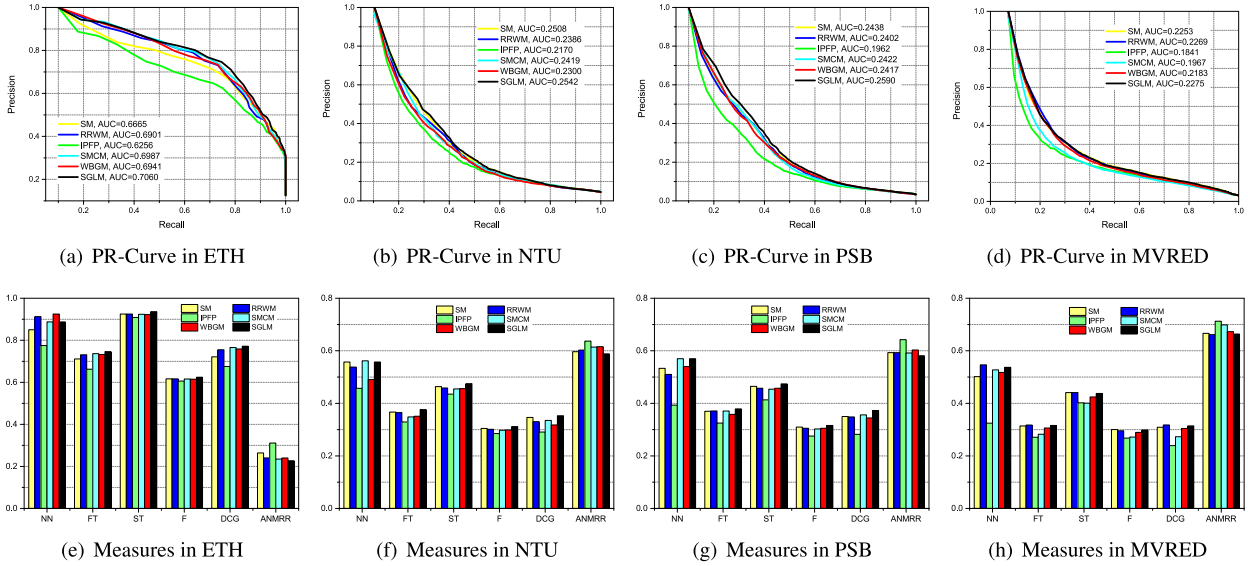
	NN	FT	ST	F_measure	DCG	ANMRR
No-$x^T Wx$	0.4921	0.4257	0.5109	0.4113	0.4674	0.5604
[34]	0.4432	0.3287	0.4367	0.3854	0.3669	0.6428
SGLM	**0.5625**	**0.3875**	**0.4738**	**0.3278**	**0.3595**	**0.5822**

TABLE 3. Comparison with previous work on PSB dataset.

	NN	FT	ST	F_measure	DCG	ANMRR
No-$x^T Wx$	0.5054	0.3527	0.5239	0.4337	0.4596	0.5924
[34]	0.4565	0.3351	0.4405	0.3498	0.3653	0.6405
SGLM	**0.5672**	**0.3834**	**0.4697**	**0.3239**	**0.3681**	**0.5901**

We can find that [34] get the worse result. SGLM outperforms the other methods. We have the following observations.

- No-$x^T Wx$ does not consider the correlation among nodes. The final performance is decided by the sampling and the optimization method. Meanwhile, $||y||$ controls the size of subgraph. It is hard to guarantee the effectiveness of nodes in subgraph. The processing of subgraph generation may remove usefulness information and influences the final experimental results;
- [34] applied the $x^T Wx$ to reduce the correlation among nodes. The goal of this design is to make the subgraph does not focus on the small size, which hopes the nodes of subgraph should have difference and represent the structure of original graph. However, in our problem, the goal of subgraph is to focus on the represent views and remove the redundant information. Thus, the previous design is not suitable for the this work;
- Our approach achieves the best performance. The reason is that we define the $x^T Wx$ is the positive term, which strengths the correlation among nodes from subgraph. It can effectively make the information focus and reduce the redundant information. Thus, we should design corresponding solutions according to different problems.



(a) PR-Curve in ETH (b) PR-Curve in NTU (c) PR-Curve in PSB (d) PR-Curve in MVRED

(e) Measures in ETH (f) Measures in NTU (g) Measures in PSB (h) Measures in MVRED

FIGURE 7. The PR-Curve and measures compared with graph matching methods in ETH, NTU, PSB and MVRED.

G. COMPARISON WITH CLASSIC GRAPH MATCHING METHODS

In order to evaluate the performance of the proposed method, some classic graph matching methods are selected for comparison:

- Spectral Matching (SM) [31]: This method drops the constraints entirely and assumes that the leading eigenvector of affinity matrix is close to the optimal discrete solution. It then finds the discrete solution by maximizing the dot-product with the leading eigenvector of affinity matrix.
- Reweighted Random Walk Graph Matching (RRWM) [32]: Matching between two graphs is formulated as node selection on an association graph whose nodes represent candidate correspondences between the two graphs. The solution is obtained by simulating random walks with reweighting jumps enforcing the matching constraints on the association graph.
- Integer Projected Fixed Point Method (IPFP) [40]: This algorithm optimizes the quadratic score in the discrete domain. It gives excellent results either by itself or by starting from the solution returned by any graph matching algorithm.
- Weighted Bipartite Graph Matching (WBGM) [41]: Representative views are first selected from both the query model and the candidate model. The initial weights of representative views are initialized and further updated based on the correlations among them. Then the weighted bipartite graph is built and the proportional max-weighted bipartite matching algorithm is employed to calculate the best matching score which is utilized to measure the similarity between two 3D models.
- Sequential Monte Carlo Graph Matching (SMCM) [42]: This approach is developed based on the sequential Monte Carlo framework. By constructing a sequence of intermediate target distributions, it sequentially performs a sampling and importance resampling to maximize the graph matching objective.

The PR-Curve and the quantitative evaluation comparison between these methods on ETH, NTU, PSB and MVRED are shown in Fig.7. With these experimental results, we have the following observations.

- In the ETH, the proposed method can achieve a gain of 4% - 15%, 1% - 13%, 1% - 3%, 1% - 3%, 1% - 14% in terms of NN, FT, ST, F-measure, DCG, and also achieve a decline of 3%-27% in terms of ANMRR. In the NTU, the proposed method can achieve a gain of 4%-22%, 3%-14%, 2%-9%, 2% - 9%, 2% - 21% in terms of NN, FT, ST, F, DCG, and achieve a decline of 1%-8% in terms of ANMRR. In the PSB, the proposed method can achieve a gain of 6% - 45%, 2% - 17%, 2% - 15%, 2% - 15%, 5% - 32% in terms of NN, FT, ST, F, DCG, and achieve a decline of 2% - 10% in terms of ANMRR. In the MVRED, the proposed method can achieve a gain of 2% - 65%, 1% - 17%, 3% - 9%, 1% - 12%, 2% - 31% in terms of NN, FT, ST, F, DCG, and achieve a decline of 0.3% - 7% in terms of ANMRR.
- WBGM only considers the relationship between the nodes from two graphs while ignores the relevance among nodes in the same graph. Our method takes both node-to-node affinity and edge-to-edge affinity into consideration. Therefore, our method can theoretically and experimentally outperforms WBGM.
- IPFP takes any continuous or discrete solution as inputs and iteratively improve the solution generated by SM method. However, the iterative operation needs long cost time. It is not suitable in many real applications.

TABLE 4. Cost time of different graph matching methods on NTU dataset.

	SM	RRWM	IPFP	SMCM	WBGM	SGLM
Cost Time(s)	2±0.2	2.1±0.1	5.3±0.3	1.3±0.06	2.1±0.1	1.2±0.11

IPFP and SM only consider the correlation between nodes while ignore the correlation between edges. Thus, they do not get the best retrieval results.

- RRWM applies random walk algorithm to handle optimization problem of object function. It needs to find the best matching pairwise by jumping between nodes. However, this design only guarantee the local optimum and ignore the global structure of graph. Thus, the final retrieval results of RRWM is worse than that of our approach. However, RRWM consider the correlation between edges. Thus, the final results is better than that of IPFP and SM.
- SMCM is similar with our approach. We both utilize MCMC to sample random nodes in order to find the best matching results. However, our approach utilize penalty to consider the inter correlation of graph. This design can make our approach get the best and robust subgraph to represent original graph and guarantee the accuracy of matching result. SMCM ignores this problem. Thus, our approach outperforms SMCM.
- All of graph matching methods except our approach need to extract represented views before the step of matching. Thus, they ignore the effect of matching object and easily be influenced by outliers or redundant information. Meanwhile, our approach considers the represented views extraction in the process of matching. It can fully consider the effect of matching object. Therefore, our approach can theoretically and experimentally outperform other graph matching methods.

In order to demonstrate the performance of our approach, we also discuss the cost time of each graph matching method. We randomly select 20 pairs of samples to calculate the matching time and calculate the average and standard deviation to measure the performance of each algorithm. The related experimental results are shown in Table.4. From the experiments, we can find that IPFP improves the final accuracy by Iterative operation. Thus, it needs the long cost time for more accuracy. WBGM and RRWM are the traditional graph matching methods. They have the same computational complexity. The cost time is determined by the number of nodes. More nodes mean more calculation time. Our approach can effectively reduce the number of nodes by subgraph selection. Thus, we have less calculation time. Finally, SMCM and our approach applied the same optimization method. We have the similar cost time in experiments. However, $x^T Wx$ helps us to remove the redundant information and improve the final matching accuracy. The experimental results also demonstrate the conclusion.

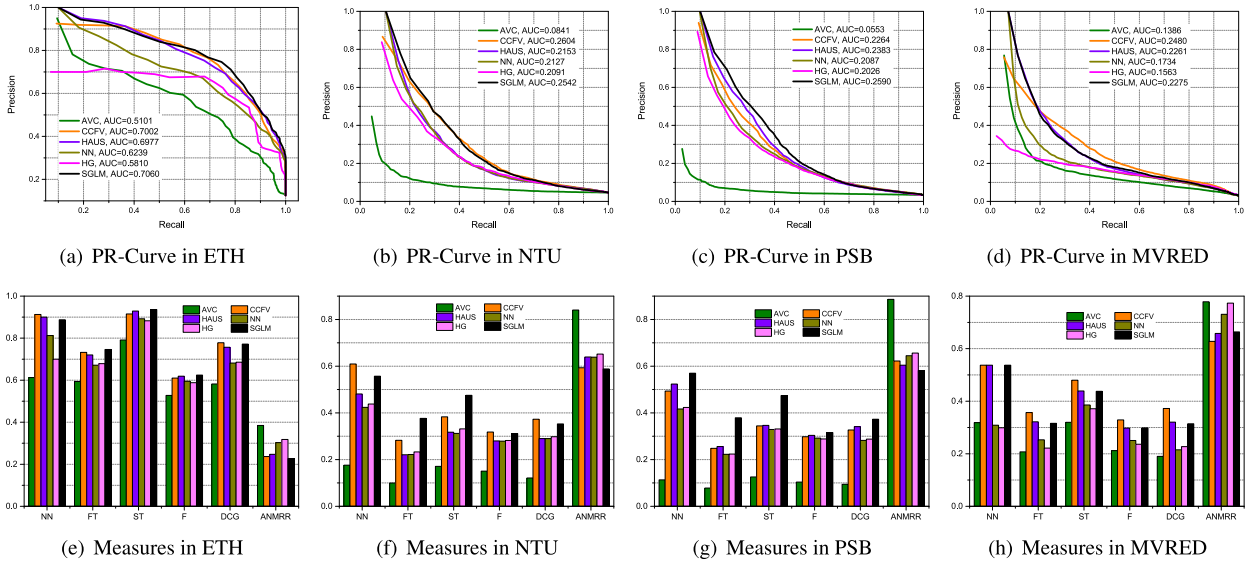
H. COMPARISON WITH OTHER STATE-OF-THE-ARTS

Several existing state-of-the-art methods are selected for comparison. The details about these methods are given as follows:

- Adaptive view clustering (AVC) [26]: AVC selects the optimal 2D characteristic views of a 3D model based on the X-means clustering algorithm and then utilizes a probabilistic Bayesian method for 3D model retrieval.
- Camera constraint free view (CCFV) [27]: Firstly, all views of a query are grouped to generate representative views. Then, a positive and negative matching mode are individually trained with positive and negative matched samples. Finally, the CCFV model is generated on the basis of the query Gaussian models by combining the positive and negative matching model.
- Model with nearest neighbor (NN) [43]: It consists of two level comparison for view-based 3D model retrieval. For view level, the matching between two view sets is analyzed. In the model level, all comparisons between two view sets are taken into account to generate the final matching result. For similarity measure, Euclidean distance is used to measure the minimum distance between a set and its nearest point in the other set.
- Model with Hausdorff distance (HAUS) [43]: A two-level comparison scheme for view-based 3D model retrieval is adopted as HAUS. Different from NN, HAUS leverages the Hausdorff distance between the characteristic views of two model for similarity measure.

The PR-Curve and the quantitative evaluation comparison between these methods on ETH, NTU, PSB and MVRED are shown in Fig8. With these results, we have the following observations.

- In the ETH, the proposed method can achieve a gain of 9% - 45%, 2% - 26%, 1% - 18%, 1% - 18%, 2% - 33% in terms of NN, FT, ST, F-measure, DCG, and also achieve a decline of 4%-41% in terms of ANMRR. In the NTU, the proposed method can achieve a gain of 16%-216%, 33%-277%, 24%-178%, 11% - 107%, 22% - 190% in terms of NN, FT, ST, F, DCG, and achieve a decline of 1%-30% in terms of ANMRR. In the PSB, the proposed method can achieve a gain of 9% - 403%, 48% - 384%, 37% - 277%, 4% - 204%, 9% - 296% in terms of NN, FT, ST, F, DCG, and achieve a decline of 4% - 34% in terms of ANMRR.In the MVRED, the proposed method can achieve a gain of 69% - 80%, 25% - 52%, 13% - 37%, 0.3% - 41%, 46% - 65% in terms of NN, FT, ST, F, DCG, and achieve a decline of 9% - 15% in terms of ANMRR.
- CCFV and AVC can be seen as probability-based method. AVC utilized Bayesian model and CCFV applied Gaussian model to simulate the direction of feature extracted from views. However, Bayesian model relies on the performance of feature from single view and also ignores the structure and spatial information of views. The Gaussian model of CCFV relies on the



(a) PR-Curve in ETH (b) PR-Curve in NTU (c) PR-Curve in PSB (d) PR-Curve in MVRED

(e) Measures in ETH (f) Measures in NTU (g) Measures in PSB (h) Measures in MVRED

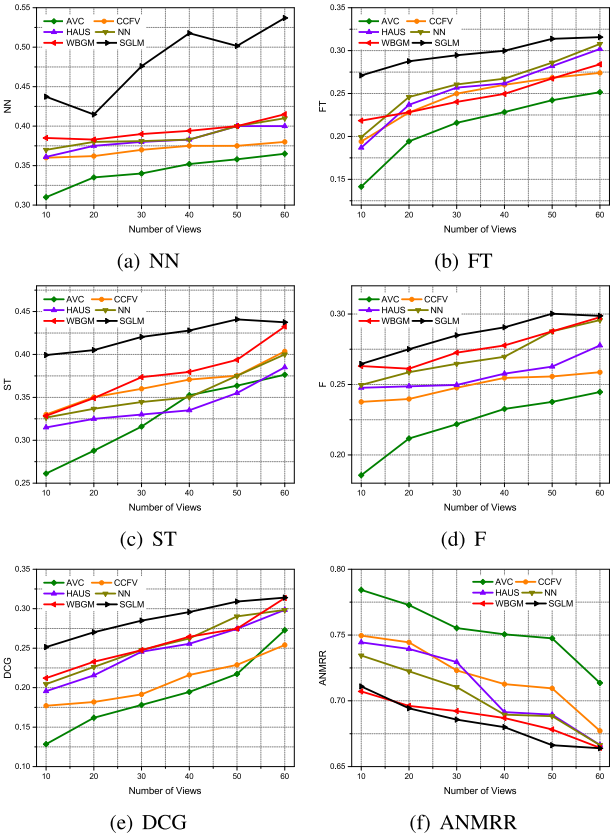
FIGURE 8. The PR-Curve and measures compared with other 3D model retrieval methods in ETH, NTU, PSB and MVRED.

TABLE 5. Cost time of different comparison methods on NTU dataset.

	AVC	CCFV	NN	HAUS	HG	SGLM
Cost Time(s)	2±0.2	2.1±0.1	5.3±0.3	1.3±0.06	2.1±0.1	1.2±0.11

training data, which is hard to be adapted to all of view-based models. The final experimental results also demonstrate the conclusion. CCFV get unstable results on different datasets. Meanwhile, the long cost time of training also limit the application of CCFV.

- NN and HAUS are distance-based methods. NN measures the Euclidean distance between pairwise views which relies on the performance of feature. Comparatively, HAUS measure the set-to-set distance and is more robust to NN. Therefore, HAUS outperforms NN on different dataset. However, HAUS also relies on the property of feature and also ignores the correlation among views. Thus, our approach also outperforms HAUS.
- We also make the comparison experiment on cost time. The corresponding experiments are shown in Table.5. Form the experiment, we can find that NN and HAUS have the least cost time. AVC and CCFV have the similar cost time. However, they have the worst retrieval results. HG costs the longest time for similarity measure. Our approach has the similar cost time like AVC and CCFV. However, we achieve the best retrieval results;
- Generally speaking, graph-based methods outperform the probability-based methods. The graph-based methods usually select the representative views or meaningful subgraphs to keep the important information from different views and then leverage graph matching for similarity measure. Comparatively, the probability-based methods usually learn the statistical model to represent a cluster of view set (HAUS/NN) or one model (CCFV/AVC). They easily ignore the spatial information among nodes. Thus, graph-based methods outperforms the probability-based methods.



(a) NN (b) FT (c) ST (d) F (e) DCG (f) ANMRR

FIGURE 9. Comparison by varying view numbers in MVRED.

I. COMPARISON BY VARYING VIEW NUMBERS

In real applications, fewer views are expected to achieve better performance. To further show the robustness of the

our method, we compare with other representative methods. We vary view number from 10 to 60 with step 10 to evaluate its effect on the retrieval result. λ_1 is set as 0.9 and λ_2 is set as 0.4 according to the experiment above. The quantitative results are shown in Fig9. From Fig9, we have the following observations.

- The performances of all methods are improved when the view number is increased. It is reasonable that more views may convey more structural characteristics of 3D models and consequently achieve better performance.
- Our method can consistently outperform the competing methods in terms of NN, FT, ST, F-measure, DCG and ANMRR when varying N from 10 to 60. Our method can achieve a gain of 23%, 17%, 10%, 13%, 25%, 7% in terms of NN, FT, ST, F-measure, DCG and ANMRR when N is tuned from 10 to 70. It can outperform the second best with a average gain of 22%, 20%, 16%, 3%, 12%, 0.6% in terms of NN, FT, ST, F-measure, DCG and ANMRR.

V. CONCLUSION

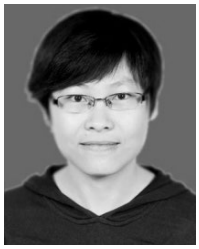
In this paper, we propose a novel view-based 3D model retrieval method based on subgraph learning matching method. In the process of subgraph learning, MCMC sampling method is utilized to effectively combine subgraph learning process with matching process to find the best matching result. Our method can preserve global attributes of one graph while removing redundant and noisy information. The experimental results indicate that the proposed method (SGLM) effectively improve the efficiency of matching and retrieval. SGLM successfully outperforms the other classic graph matching and 3D model retrieval methods.

REFERENCES

[1] S. Yoon, T. Schreck, and G. Yoon, "Sparse coding based feature optimisation for robust 3D object retrieval," *Electron. Lett.*, vol. 48, no. 9, p. 493, 2012.
[2] W. Nie, Q. Cao, A. Liu, and Y. Su, "Convolutional deep learning for 3D object retrieval," *Multimedia Syst.*, vol. 23, no. 3, pp. 325–332, Jun. 2017.
[3] J. W. H. Tangelder and R. C. Veltkamp, "Polyhedral model retrieval using weighted point sets," *Int. J. Image Graph.*, vol. 3, no. 1, pp. 209–229, Jan. 2003.
[4] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Trans. Graph.*, vol. 21, no. 4, pp. 807–832, 2002.
[5] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.
[6] Y. Lian and X. Shen, "Image-based hierarchical method of 3D model retrieval," in *Proc. Int. Conf. Virtual Reality Vis.*, Aug. 2014, pp. 160–165.
[7] B. Leng, Y. Liu, K. Yu, X. Zhang, and Z. Xiong, "3D object understanding with 3D convolutional neural networks," *Inf. Sci.*, vol. 366, pp. 188–201, Oct. 2016.
[8] K. Lu, N. He, J. Xue, J. Dong, and L. Shao, "Learning view-model joint relevance for 3D object retrieval," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1449–1459, May 2015.
[9] Y. Gao and Q. Dai, "View-based 3D object retrieval: Challenges and approaches," *IEEE Multimedia Mag.*, vol. 21, no. 3, pp. 52–57, Jul. 2014.
[10] Z. Zhu, X. Wang, S. Bai, C. Yao, and X. Bai, "Deep learning representation using autoencoder for 3D shape retrieval," in *Proc. Int. Conf. Secur., Pattern Anal. (SPAC)*, 2014, pp. 279–284.
[11] F. Shaikh, V. Kharpate, D. Mahale, and A. Wadhwa, "Approach to enhance 3D object retrieval system by considering various characteristics," in *Proc. IEEE Global Conf. Wireless Comput. Netw. (GCWCN)*, Dec. 2014, pp. 84–87.
[12] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[13] J. Han, X. Ji, X. Hu, D. Zhu, K. Li, X. Jiang, G. Cui, L. Guo, and T. Liu, "Representing and retrieving video shots in human-centric brain imaging space," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2723–2736, Jul. 2013.
[14] Y. Liu, H. Zha, and H. Qin, "The generalized shape distributions for shape matching and analysis," in *Proc. IEEE Int. Conf. Shape Modeling Appl. (SMI)*, Jul. 2006, pp. 1–16.
[15] T. Zaharia and F. Preteux, "Shape-based retrieval of 3D mesh models," in *Proc. IEEE Int. Conf. Multimedia Expo*, vol. 1, Jun. 2003, pp. 437–440.
[16] H. Dutagaci, B. Sankur, and Y. Yemez, "Transform-based methods for indexing and retrieval of 3D objects," in *Proc. 5th Int. Conf. 3-D Digit. Imag. Modeling (3DIM)*, 2005, pp. 188–195.
[17] H. Laga, H. Takahashi, and M. Nakajima, "Spherical wavelet descriptors for content-based 3D model retrieval," in *Proc. IEEE Int. Conf. Shape Modeling Appl. (SMI)*, Jul. 2006, p. 15.
[18] T. Tung and F. Schmitt, "The augmented multiresolution reeb graph approach for content-based retrieval of 3D shapes," *Int. J. Shape Model.*, vol. 11, no. 1, pp. 91–120, Jun. 2005.
[19] T. Matsuda, T. Furuya, and R. Ohbuchi, "Lightweight binary voxel shape features for 3D data matching and retrieval," in *Proc. IEEE Int. Conf. Multimedia Big Data*, Apr. 2015, pp. 100–107.
[20] H. Haj Mohamed and S. Belaid, "3D-model retrieval using bag-of-salient local spectrums (BOS)," in *Proc. 6th Int. Conf. Soft Comput. Pattern Recognit. (SoCPaR)*, Aug. 2014, pp. 174–179.
[21] 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.
[22] 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, Jan. 2007.
[23] P. Daras and A. Axenopoulos, "A 3D shape retrieval framework supporting multimodal queries," *Int. J. Comput. Vis.*, vol. 89, nos. 2–3, pp. 229–247, Sep. 2010.
[24] Y. Gao, Q. Dai, and N.-Y. Zhang, "3D model comparison using spatial structure circular descriptor," *Pattern Recognit.*, vol. 43, no. 3, pp. 1142–1151, Mar. 2010.
[25] V. Garro and A. Giachetti, "Scale space graph representation and kernel matching for non rigid and textured 3D shape retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1258–1271, Jun. 2016.
[26] T. F. Ansary, M. Daoudi, and J.-P. Vandeborre, "A Bayesian 3-D search engine using adaptive views clustering," *IEEE Trans. Multimedia*, vol. 9, no. 1, pp. 78–88, Jan. 2007.
[27] Y. Gao, J. Tang, R. Hong, S. Yan, Q. Dai, N. Zhang, and T.-S. Chua, "Camera constraint-free view-based 3-D object retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2269–2281, Apr. 2012.
[28] Y. Gao, M. Wang, R. Ji, X. Wu, and Q. Dai, "3-D object retrieval with Hausdorff distance learning," *IEEE Trans. Ind. Electron.*, vol. 61, no. 4, pp. 2088–2098, Apr. 2014.
[29] X. Wang and W. Nie, "3D model retrieval with weighted locality-constrained group sparse coding," *Neurocomputing*, vol. 151, pp. 620–625, Mar. 2015.
[30] A. Berg, T. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jul. 2005, pp. 26–32.
[31] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Oct. 2005, pp. 1482–1489.
[32] M. Cho, J. Lee, and K. M. Lee, "Reweighted random walks for graph matching," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, Sep. 2010, pp. 492–505.
[33] M. Leordeanu, R. Sukthankar, and M. Hebert, "Unsupervised learning for graph matching," *Int. J. Comput. Vis.*, vol. 96, no. 1, pp. 28–45, Jan. 2012.
[34] W. Nie, H. Ding, A. Liu, Z. Deng, and Y. Su, "Subgraph learning for graph matching," *Pattern Recognit. Lett.*, to be published.

[35] W.-Z. Nie, A.-A. Liu, Z. Gao, and Y.-T. Su, "Clique-graph matching by preserving global & local structure," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4503–4510.
[36] A. Ess, B. Leibe, K. Schindler, and L. Van Gool, "A mobile vision system for robust multi-person tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
[37] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The princeton shape benchmark," in *Proc. Shape Modeling Appl.*, Nov. 2004, pp. 167–178.
[38] Y. Gao, A. Liu, W. Nie, Y. Su, Q. Dai, F. Chen, Y. Chen, Y. Cheng, S. Dong, and X. Duan, "3D object retrieval with multimodal views," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2015, pp. 99–106.
[39] J. Hartigan and M. Wong, "A k-means clustering algorithm: Algorithm as 136," *Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979.
[40] M. Leordeanu, M. Hebert, and R. Sukthankar, "An integer projected fixed point method for graph matching and map inference," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1114–1122.
[41] Y. Gao, Q. Dai, M. Wang, and N. Zhang, "3D model retrieval using weighted bipartite graph matching," *Signal Process., Image Commun.*, vol. 26, no. 1, pp. 39–47, Jan. 2011.
[42] Y. Suh, M. Cho, and K. M. Lee, "Graph matching via sequential Monte Carlo," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 624–637.
[43] Y. Gao, J. Tang, H. Li, Q. Dai, and N. Zhang, "View-based 3D model retrieval with probabilistic graph model," *Neurocomputing*, vol. 73, nos. 10–12, pp. 1900–1905, Jun. 2010.

HAIPENG GUAN received the bachelor's degree from Shanxi Normal University, Linfen, China, in 2004, and the M.Eng. degree from Hunan Normal University, Changsha, and South China Normal University, in 2006 and 2009, respectively. She is currently a Lecturer with Yuncheng University. Her research interests include embedded application, data science, and information security.



QIUXIA ZHAO received the B.S. and M.S. degrees from Yunnan University, Kunming, China, in 2009 and 2012, respectively. She is currently a Teaching Assistant with Yuncheng University. Her research interests include data science and information security.

YAN REN received the bachelor's degree from Shanxi Normal University, Linfen, China, in 2004, the master's degree from Shaanxi Normal University, Xi'an, China, in 2010, and the Ph.D. degree from Guangzhou University, Guangzhou, China, in 2014. She is currently with Yuncheng University. Her primary research interest is applied cryptography, in particular, cryptographic protocols, encryption and signature schemes, and anonymous systems. She has published some interesting results of signature schemes.

WEIZHI NIE received the Ph.D. degree from Tianjin University, Tianjin, China. He was a Visiting Scholar with the NExT Center, National University of Singapore, where he was involved along with Prof. T.-S. Chua. He is currently an Associate Professor with the School of Electronic Information Engineering, Tianjin University. His research interests include computer vision, machine learning, and social networks.