Multimed Tools Appl (2017) 76:4091–4104
DOI 10.1007/s11042-015-2840-x

3D object retrieval based on Spatial+LDA model

Weizhi Nie[1] · Xixi Li[1] · Anan Liu[1] · Yuting Su[1]

Received: 3 May 2015 / Revised: 29 June 2015 / Accepted: 21 July 2015 /
Published online: 8 August 2015

Abstract Latent Dirichlet Allocation (LDA) is one popular topic extraction method, which has been applied in many applications such as textual retrieval, user recommendation system and video cluster. In this paper, we apply LDA model for visual topics extraction and utilized the topic distribution visual feature of image to handle 3D object retrieval problem. Different from the traditional LDA model, we add the spatial information of visual feature for document generation. First, we extract SIFT features from each 2D image extracted from 3D object. Then, we structure the visual documents according to the spatial information of 3D model. Finally, LDA model is used to extract the topic model for handling the retrieval problem. We further propose a multi-topic model to improve retrieval performance. Extensive comparison experiments were on the popular ETH, NTU and MV-RED 3D model datasets. The results demonstrate the superiority of the proposed method.

Keywords 3D model retrieval · LDA · Topic model · Similarity measure · Topic extraction

1 Introduction

Recently, a rapidly growing technologies of 3D models have been widely used in diverse fields, such as computer-aided design [25] , bioinformatics [32], medicine [13], and the entertainment industry [31]. Owing to the large number of 3D objects, 3D model retrieval [2, 30] is becoming a burgeoning emerging technology. An effective and efficient 3D model retrieval method can expand the utilized occasion of virtual model and improve the efficiency of resource use.

✉ Anan Liu
anan0422@gmail.com

1 School of Electronic Information Engineering, Tianjin University, Tianjin, China

According to the querying type, 3D object retrieval methods can be divided into two paradigms: model-based retrieval [11, 19] and view-based retrieval [1, 8]. In model-based style, the features of 3D model are directly extracted from the virtual 3D model. Early works on 3D object retrieval are mainly based on 3D models, by employing low-level feature-based methods or high-level structure-based methods [18]. The other category is based on views, in which each 3D object is represented by a set of views and the model feature are extracted from these views. Several view-based 3D object retrieval methods, e.g., light-field descriptors (LFDs) [5], elevation descriptors (EDs) [26], bag of visual features (BoVF) [21, 22], and compact multiview descriptors (CMVDs) [12], have been proposed in recent years. Experimental results showed in Bustos et al. [4] and Shilane et al. [27] demonstrated that view-based methods can effectively leverage image processing algorithm of traditional 2D image and improve the efficiency of algorithm. Consequently, view-based 3D object retrieval has become a crucial and promising research field in recent year.

View-based 3D object retrieval has been positively studied on account of its high flexibility and easy implement of using multiple views for object representation [20]. In this scheme, a 3D object is described by a set of views with various features, focusing on primordial characteristics, such as surface distributions [23], Zernike moment [16], and HOG descriptors [7]. Then comparing the features in different feature spaces, seeking for the object which is the most similar one to the query. However, under this circumstance, we ignore semantic associations among the rough features.

In this paper, we utilize topic model to formulate such hidden relationships that incorporate the distribution of topics within a 3D object as a more discriminative feature, which obtains semantic meaning in a degree. In order to make LDA model is suitable for visual word, we add the spatial information of visual word into the LDA model and proposed to use spatial-LDA model for topics extraction. Then, the topic distribution of document is used to represent the 3D model and compute the similarity between different 3D models. Finally, the similarity is utilized to address the retrieval problem.

The advantages of the proposed method are twofold:

- The spatial information of 3D model is utilized to generate visual document, which can effectively strengthen the correlation between similar visual words and improve the performance of LDA model in visual feature space;
- LDA model is utilized in 3D model retrieval problem. Topic distribution is leveraged to represent 3D object;
- A new dataset named Multi-view RGB-D Object Dataset (MV-RED) is used in this study, which was recorded by Kinect cameras from different angles. The details will be introduced in next section.

The rest of this paper is organized as follows. In Section 2, related work and state-of-the-art works will be introduced. The proposed 3D model retrieval algorithm using LDA model is detailed in Section 3. Experimental results and discussions are shown in Section 4. Conclusions and acknowledgements are stated in Sections 5 and 6 respectively.

2 Related work

Thanks to the fast development of RGB-D camera and machine vision technology, the ever-growing 3D models necessitate the speedy progress of 3D model retrieval. There are generally two paradigms for existing 3D object retrieval methods: model-based retrieval and view-based retrieval.

Most early methods belong to model-based retrieval, which takes advantage of existing virtual 3D model information. Researchers extracted low-level features such as the geometric moment [24], the surface distribution, the volumetric descriptor [9, 29], and the surface geometry [15] or more complicated high-level features [28]. But the huge computation complexity and severe request of accuracy limit practical applications.

View-based methods extract feature descriptors from a set of 2D views captured by various camera arrays. Next significant step is computing the similarity of each 3D model from database to query. Some algorithms focus on the way to extract features, which is an integral part in any multimedia retrieval task. Jau-Ling Shih et al.[26] proposed a novel feature, called elevation descriptor(ED) . The elevation descriptor is invariant to translation and scaling of 3D models and it is robust for rotation. First, six elevations are obtained to describe the altitude information of a 3D model from six different views. Each elevation is represented by a gray-level image which is decomposed into several concentric circles. The elevation descriptor is obtained by taking the difference between the altitude sums of two successive concentric circles. An efficient similarity matching method is used to find the best match for an input model. In [22], the bag of visual features(BoVF) method was applied in view-based 3D object retrieval. For each range image, a set of 2D multi-scale local visual features is computed by using SIFT. To reduce cost of distance computation and feature storage, a set of local features describing a 3D model is integrated into a histogram using the Bag-Of Features approach. And Kullback Leibler (KL) divergence is employed to measure the distance between two 3D objects. Daras et al. [8] put forward compact multi-view descriptors (CMVDs).The method supports multimodal queries (2D images, sketches, 3D objects) by introducing a novel view-based approach able to handle the different types of multimedia data. More specifically, a set of 2D images (multi-views) are automatically generated from a 3D object, by taking views from uniformly distributed viewpoints. For each image, a set of 2D rotation-invariant shape descriptors is produced. The global shape similarity between two 3D models is achieved by applying a novel matching scheme, which effectively combines the information extracted from the multi-view representation. View-based methods have been shown to be superior in 3D object representation, in which only need a few views but achieves excellent performance.

However, most extracted features merely represent visual, textured or geometric characters of 3D model, while the relationships between features are neglected. The relationships can be described as some latent information between the original features and the semantic label of the 3D object [33]. We come up with a novel idea that regarding the original features as visual words, generate the topics distribution of each 3D model as semantic features and incorporate topic model to deal with the data.

Latent Dirichlet allocation (LDA), a generative probabilistic topic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities [3]. Hu [14] shows that LDA is not only useful in the text domain, but also in the image and music domain. In particular, they discuss algorithms that extend LDA to accomplish tasks like document classification for text, object localization for images, and automatic harmonic analysis for music.

3 Framework

The frame work includes three steps: 1) Feature extraction. We extract SIFT feature from each 2D image. 2) Generate Topic model. The structure of 3D model is utilized to generate documents. LDA model is applied to generate the final visual topic model and topic distribution. 3) Similarity measurement. Object distribution is used to compute the similarity between different 3D models. In the next subsections, we will detail the spatial-LDA model and similarity measurement.

4 Spatial+LDA model

We consider the spatial information of 3D model to construct the document and apply LDA model to generate the distribution of documents as feature representation of 3D model. Thus, the proposed method includes two steps: 1) document generation and 2) Topic model generation, which will be detailed in the next sections.

4.1 Document generation

In view-based 3D model retrieval, each 3D model consists of a set of 2D images, which are recorded from different angles. These 2D images much have redundancy between each other, when the the amount of data reaches a certain value. So we try to group these 2D images which are close in space into the same documents. Each group can be treated as a document instead of one 3D model. However, in practice, it is hard to define the group information for some special images or some robust images. In order to handle this problem, we adopt the sparse-subspace clustering method to discover the neighbor nodes of individual center point and is neighbors.

we first applied K-means to detect some representation views or characteristic images as the center point of each group. The individual center point can be represented by the feature vector $f_i(f_i \in \mathbb{R}^d)$ and all images can be represented by the feature set, $\{f_i\}_{i=1}^{I}$. motivated by sparse representation, f_i can be reconsructed by the linear combination of the other images $F = \{f_j\}_{j=1,2,....,I}\, j \neq i$.

$$f_i = Fr_i \quad s.t. \quad r_{ii} = 0, \tag{1}$$

where $r_i = [r_{i1}, ..., r_{iI}]$; r_{ik}, the k^{th} dimension of coefficient r_i, denotes the weight of the k^{th} node and implicitly means the relatedness between the i^{th} and k^{th} images; the constraint $r_{ii} = 0$ avoids the trivial solution of reconstructing f_i by itself.

Motivated by the theory of sparse representation, sparse constraint can be imposed to achieve a sparse solution, whose non-zero entities correspond to the related nodes. Ideally, the L_0-normal of r_i can be intuitively imposed to induce sparsity. However, this will lead to the general NP-hard problem of finding the sparsest representation of the given node. Therefore, we consider minimizing the tightest convex relaxation of the L_0-norm of r_i with L_1-norm. The objective functions can be formulated as:

$$min\|R\|_1 \quad s.t. \quad F = FR,\ diag(R) = 0, \tag{2}$$

where $R \triangleq [r_1 \ r_2 \ldots r_I]$ is the matrix, whose i^{th} column corresponds to the sparse representation of f_i. Equation (2) can be efficiently solved with convex programming tools.[1] According to the matrix R, we selected the top k images which can well reconstruct the image i in all of 2D images from 3D model. These k images and the image i as one group can be seen as one document for the next process.

4.2 Topic model

Through the previous process, 3D model can be represented by a set of documents. Thus, LDA model can be used to generate the final topic model. We assume that there are M documents in each 3D model and N words in each document. $w_{j,i}$ is the observed value of word i in document j. All the words and documents will be clustered into K topics. The generative procedure of LDA is described as follows.

- For a topic k, a multinomial parameter ϕ_k is sampled from Dirichlet prior $\phi_k \sim Dir(\alpha)$, which shows the word distribution in this topic;
- For a document j, a multinomial parameter φ_j over the K topics is sampled from Dirichlet prior $\varphi_j \sim Dir(\beta)$, which shows the topic distribution in document j;
- For a word i in document j, a topic label z_{ij} is sampled from discrete distribution $z_{ij} \sim Mult(\varphi_j)$;
- The value w_{ij} of word i in document j is sampled from the discrete distribution of topic z_{ij}, $w_{ij} \sim Mult(\phi_{z_{ij}})$;
- α and β are the known variables, which also are defined by statistical experiments. Gibbs sampling procedure is utilized to compute the φ_j and $\phi_{z_{ij}}$.

In this process, w is one observed variables, α and β are the known variables or prior variables. z, ϕ and φ is the unknown variables, which need to be learned by observed data. Here, we can get the joint distribution of all parameters.

$$p(w_i, z_i, \phi_i, \varphi_j|\alpha, \beta) = \Pi_{j=1}^{N} p(w_{ij}|\phi_{z_{ij}}) p(z_{ij}|\varphi_j) p(\varphi_j|\beta) p(\phi_i|\varphi), \tag{3}$$

where N represents the number of words in document j, $p(w_{ij}|\phi_{z_{ij}})$ represents the distribution $Mult(\phi_{z_{ij}})$, $p(z_{ij}|\varphi_j)$ represents the distribution $Mult(\varphi_j)$. $p(\varphi_j|\beta)$ and $p(\phi_i|\varphi)$ represents the Dirichlet prior according to the parameter α and β respectively. Equation (3) also can be rewritten as:

$$p(w, z|\alpha, \beta) = p(w|z, \beta) p(z|\alpha), \tag{4}$$

where $p(z|\alpha) = \Pi_{m=1}^{M} \frac{\Delta(n_m+\alpha)}{\Delta(\alpha)}$, $n_m = \{n_m^k\}_{k=1}^{K}$, where n_m^k represents frequency of occurrence of topic k in document m. $p(w|z, \beta) = \Pi_{z=1}^{k} \frac{\Delta(n_z+\beta)}{\Delta(\beta)}$, $n_z = \{n_z^t\}_{t=1}^{V}$, where n_z^t represents frequency of occurrence of word t in topic k. Finally, (4) can be rewritten as:

$$p(w_i, z_i, \phi_i, \varphi_j|\alpha, \beta) = \Pi_{m=1}^{M} \frac{\Delta(n_m + \alpha)}{\Delta(\alpha)} \Pi_{z=1}^{k} \frac{\Delta(n_z + \beta)}{\Delta(\beta)}, \tag{5}$$

According to (5), we can compute the conditional distribution of topic z based on the observed variable w.

$$p(z_i = k|z_{-i}, w) = \frac{p(w, z)}{p(w, z_{-i})}, \tag{6}$$

[1] http://cvxr.com/cvx/

where z_{-i} represents all of topics excluding the topic k. Finally, (6) can be rewritten as:

$$p(z_i = k|\alpha, \beta) \propto \frac{n_{m,-i}^{k} + \alpha_k}{\sum_{k=1}^{K}\left(n_{m,-i}^{k} + \alpha_k\right)} \cdot \frac{n_{k,-i}^{t} + \beta_t}{\sum_{t=1}^{V}\left(n_{k,-i}^{t} + \beta_t\right)}, \tag{7}$$

where $n_{m,-i}^{k}$ is the number of tokens assigned to topic k excluding w_i. $n_{k,-i}^{t}$ is the number of word w assigned to topic k excluding word x_i. We sampling according to (7) to compute the unknown variable ϕ and φ. The values of ϕ and φ will converge after several iterations.

In the retrieval process, when given a query object Q, the topic distribution ϕ_Q can be also obtained via the Gibbs sampling. The sampling process according to:

$$p(z_i = k|\alpha, \beta) \propto \frac{n_{Q,-i}^{k} + \alpha_k}{\sum_{k=1}^{K}\left(n_{Q,-i}^{k} + \alpha_k\right)} \cdot \frac{n_k + n_{k,-i}^{t} + \beta_t}{\sum_{t=1}^{V}\left(n_k + n_{k,-i}^{t} + \beta_t\right)}, \tag{8}$$

where n_k is the number of word w assigned to topic k. $n_{Q,-i}^{k}$ represents the number of tokens assigned to topic k excluding x_i within the query object Q.

4.3 Similarity measurement

Based on the feature learned by LDA, we applied (9) to compute the similarity between different 3D model.

$$S(Q, M) = \frac{1}{\sqrt{(\phi_Q - \phi_M)^2}}, \tag{9}$$

where ϕ_Q and ϕ_M represent the topic distribution of 3D model Q and M respectively.

$$\phi_M = \left[\phi_M^1, \phi_M^2, \ldots\ldots, \phi_M^N\right], \tag{10}$$

where M represents the idx of 3D model, N represents the number of documents in 3D model M, $S(Q, M)$ represents the similarity between 3D model Q and M. The retrieved model with the highest similarity score can be achieved by

$$M^* = arg \max_{M_i \in \widetilde{M}} S(M_i, Q). \tag{11}$$

where $\widetilde{M}$ represents the candidate model set, M_i is the candidate model, Q is the query model, $S(M_i, Q)$ is the similarity between the query model and candidate model.

5 Experiment

5.1 Dataset

The proposed method was evaluated on following datasets:

- ETH [17]: The ETH dataset is a real-world 3D object multi-view database. It contains 80 objects categorized into 8 classes, and each class has 10 objects. There are 41 multiple views for each object spaced evenly over the upper viewing hemisphere, and all positions for cameras are determined by subdividing the faces of an octahedron to the third recursion level.
- NTU [5]: The NTU dataset contains 500 objects in total. Virtual cameras are employed to capture initial views for 3D objects. The camera array contains 60

cameras, which are set on the vertices of a polyhedron with the same structure with Buckminsterfullerene (C60). Therefore, there are 60 views for each object.

- The Multi-view RGB-D Object Dataset (MV-RED):[2] The MV-RED is a real world 3D object dataset with multimodal views, which contains 505 objects divided into 60 categories, such as apple, cap, scarf, cup, mushroom and toy. Each object contains RGB image to represent color information and depth image to represent shape information. We select the categories containing no less than 10 objects as queries, giving rise to 311 queries altogether. Therefore, there are 311 queries and 505 tests in the experiment setting.Both RGB and depth image were collected for each object by 3 Microsoft Kinect sensors (the 1st generation) located in 3 different directions. The difference between these two settings lies in the directions for view acquisition, which increases the difficulties in view matching. Camera 1 and Camera 2 sampled 360 both RGB and depth images respectively while Camera 3 captured just one RGB and depth images from the above view. Thus each object covers 721 RGB images and 721 depth images in total. The image resolution of RGB and depth image is 640×480. Then we uniformly captured the images from Camera 1 and 2 with the step of 10 degrees to produce a simple dataset with 73 RGB and 73 depth images for each object.

5.2 Evaluation

In order to evaluate the proposed method, the following criteria are employed as the evaluation measures of the retrieval performance.

- Nearest neighbor (NN).It is the percentage of the closest match models belonging to the querys category.
- First tier (FT).It is the recall for the first K relevant match samples, where K is the cardinality of the querys category.
- Second tier(ST).It is the recall for the first 2K relevant match samples, where K is the cardinality of the querys category.
- F-measure (F).It is a synthetical measurement of precision and recall for a fixed number of retrieved results.
- Discounted cumulative gain (DCG). It is a statistical measure that assigns higher weights to relevant results occupying the top-ranking positions.
- Average Normalized Modified Retrieval Rank (ANMRR) .It measures the rank performance given a ranking list, which considers the ranking information of relevant objects among the top retrieved objects.
- Precision-recall curve(PR). It is a crucial indicator that shows the relationship between the precision and the recall.

5.3 Comparison to LDA

In this study, we compare our approach with traditional LDA mode. We select different topic number (50, 75, 100) to test the performance of the proposed method. ETH dataset is selected as the evaluation dataset. The experimental results are shown in Fig. 1.

[2] http://media.tju.edu.cn/mvred/dataset1.html

Figure 1a shows the precision-recall curves on ETH by different methods. Figure 1b shows the performances by different methods on ETH dataset. From these experimental results, we can find that the overall retrieval is becoming better with the decrease of the topic number. The reason is that more topics dispersed more visual information, which is not better for 3D object representation. Meanwhile, the proposed method outperforms the traditional LDA model when topic number is 50 and when topic number changed to 75 and 100, the performance compared to LDA is almost the same level, which means that our design can improve the performance of the final retrieval results in the case of suitable topic number.

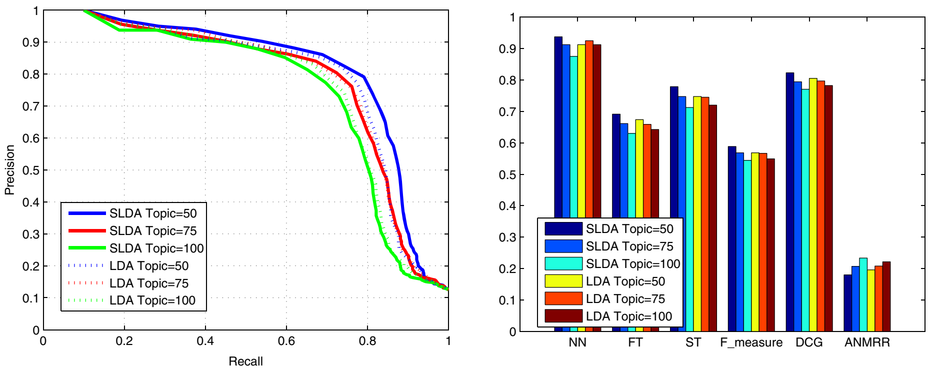
5.4 Comparison to Multiple-SLDA

We also compare our approach with multiple topic model generated by Spatial+LDA. ETH dataset also is selected as the evaluation dataset. The experimental results are shown in Fig. 2. Figure 2a shows the precision-recall curves on ETH by different methods. Figure 2b shows the performances by different methods on ETH dataset. From these experimental results, the multi-LDA topic models outperforms the single-LDA topic model, which means the performance of multiple topics model provides the more accuracy feature representation than single-LDA model when topic number increases. Meanwhile, multiple topic models can utilize advantages of single topic models to provide more effective feature representation. Thus, topic number is an essential factor that guarantee the similarity measurement is accuracy.

5.5 Comparison to other methods

In order to evaluate the performance of the proposed method, some state-of-the-art methods are selected for comparison:

- Latent Dirichlet Allocation (LDA) [3]: Latent Dirichlet allocation(LDA) is a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is



(a) (b)

Fig. 1 Comparison to LDA on ETH

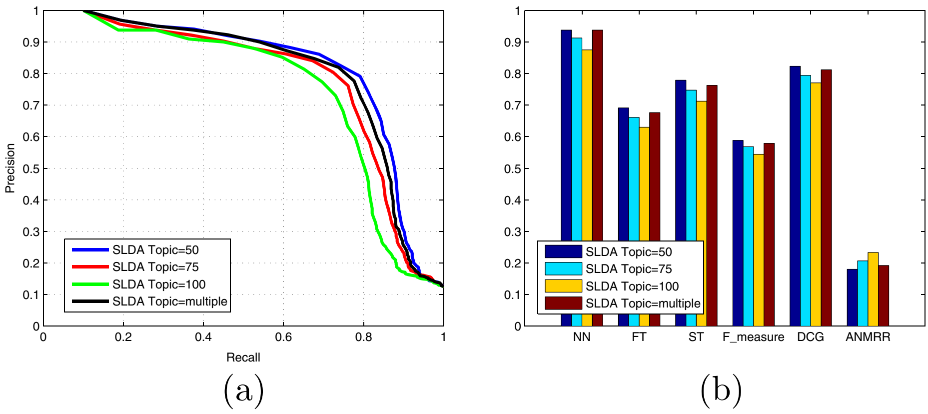


Fig. 2 Comparison to Multiple-SLDA on ETH

modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities. In the context of text modeling, the topic probabilities provide an explicit representation of a document. The method presents efficient approximate inference techniques based on variational methods and an EM algorithm for empirical Bayes parameter estimation.

- Nearest Neighbor(NN) [6]:The nearest neighbor decision rule assigns to an unclassified sample point the classification of the nearest of a set of previously classified points. This rule is independent of the underlying joint distribution on the sample points and their classifications, and hence the probability of error R of such a rule must be at least as great as the Bayes probability of error R–the minimum probability of error over all decision rules taking underlying probability structure into account. For any number of categories, the probability of error of the nearest neighbor rule is bounded above by twice the Bayes probability of error.
- Hausdorff Distance(HAUS) [6] :This efficient method estimate the distance between discrete 3D surfaces represented by triangular 3D meshes. The metric used is based on an approximation of the Hausdorff distance, which has been appropriately implemented in order to reduce unnecessary computations and memory usage. Results show that when compared to similar tools, a significant gain in both memory and speed can be achieved.
- Bipartite graph matching (WBGM) [10]: In this method, weighted bipartite graph matching is employed for comparison between two 3D models. Each 3D model is represented by a set of 2D views. Representative views are selected from the query model and the corresponding initial weights are provided. These initial weights are further updated based on the relationship among these representative views. The weighted bipartite graph is built with these selected 2D views and the matching result is used to measure the similarity between two 3D models.
- Camera constraint-free view-based (CCFV) method [12]: In the camera constraint-free view method, each object is represented by a free set of views, which implies that these views can be captured from any direction without camera constraints. For each query object, all query views are clustered to generate the view cluster, which is then

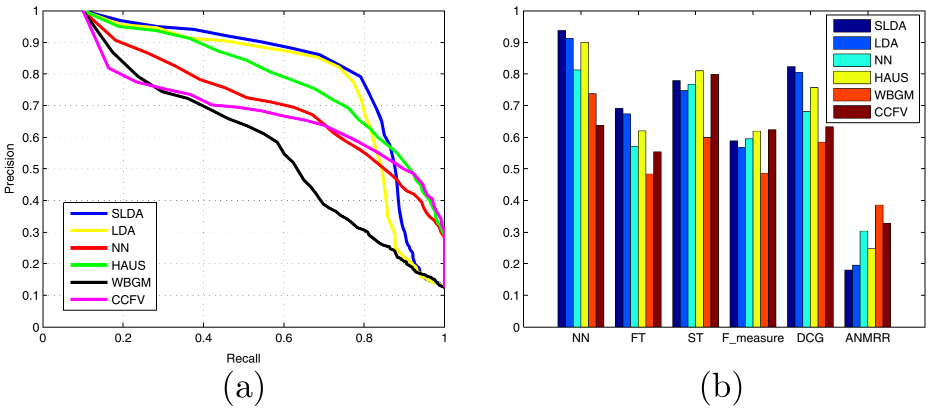


Fig. 3 The experimental results on ETH

used to build the query models. For a more accurate 3-D object comparison, a positive matching model and a negative matching mode are individually trained using positive and negative matched samples, respectively. The CCFV model is generated on the basis of the query Gaussian models by combining the positive matching model and the negative matching model. The CCFV method removes the constraint of static camera array setting for view capturing and can be applied to any view-based 3D object database.

The experimental results are shown in Figs. 3, 4 and 5. Figures 3a, 4a and 5a respectively show the precision-recall curves on ETH, NTU and MV-RED by different methods. Figures 3b, 4b and 5b show the performances by different methods on both datasets respectively. From these experimental results, the proposed method outperforms the other comparative methods, which also demonstrate the superiority of our approach.

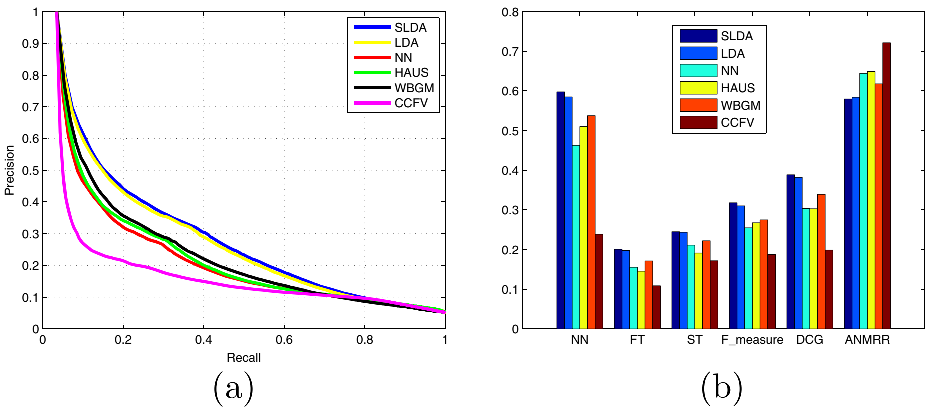


Fig. 4 The experimental results on NTU

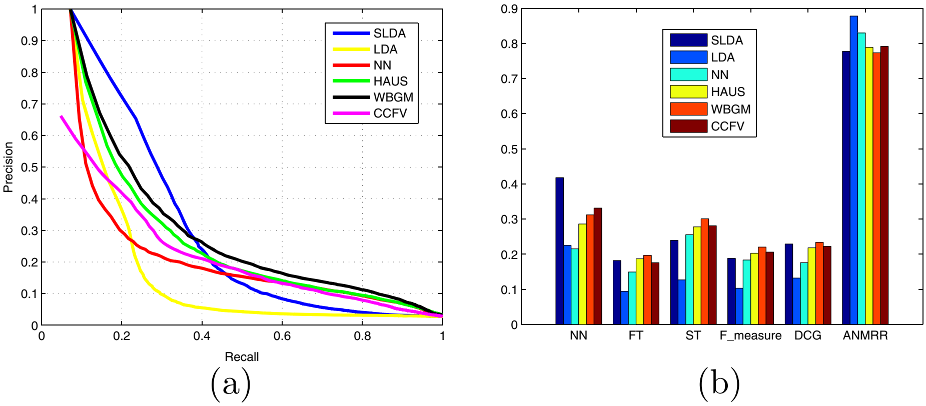


Fig. 5 The experimental results on MV-RED

6 Conclusion

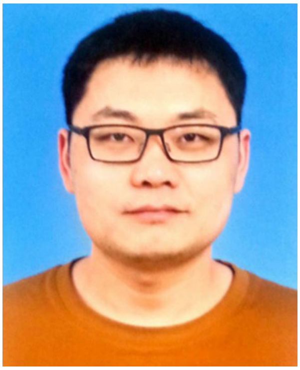
In this paper, we proposed an interesting 3D retrieval method based on spatial+LDA model. Each 3D object can be represented by a set of visual documents. Meanwhile, we consider the spatial information of visual word in order to get the accurate topic model. In this paper, we applied document distribution as feature for 3D object representation. The extensive experiments on ETH,NTU and MV-RED also demonstrate the superiority of this method.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China (61472275, 61170239, 61303208), the Tianjin Research Program of Application Foundation and Advanced Technology (15JCYBJC16200), and the grant of Elite Scholar Program of Tianjin University (2014XRG-0046).

References

1. Ansary TF, Daoudi M, Vandeborre J-P (2007) A bayesian 3-d search engine using adaptive views clustering. IEEE Trans Multimed 9(1):78–88
2. Bimbo AD, Pala P (2006) Content-based retrieval of 3d models. ACM Trans Multimed Comput Commun Appl (TOMM) 2(1):20–43
3. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022
4. Bustos B, Keim DA, Saupe D, Schreck T, Vranić DV (2005) Feature-based similarity search in 3d object databases. ACM Comput Surv (CSUR) 37(4):345–387
5. Chen D-Y, Tian X-P, Shen Y-T, Ouhyoung M (2003) On visual similarity based 3d model retrieval. In: Computer graphics forum, vol 22. Wiley Online Library, pp 223–232
6. Cover TM, Hart PE (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27
7. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition, 2005. CVPR 2005, vol 1. IEEE, pp 886–893
8. Daras P, Axenopoulos A (2010) A 3d shape retrieval framework supporting multimodal queries. Int J Comput Vis 89(2–3):229–247
9. Funkhouser T, Min P, Kazhdan M, Chen J, Halderman A, Dobkin Dd, Jacobs D (2003) A search engine for 3d models. ACM Trans Graph (TOG) 22(1):83–105

10. Gao Y, Dai Q, Wang M, Zhang N (2011) 3d model retrieval using weighted bipartite graph matching. Sig Proc Image Comm 26(1):39–47
11. Gao Y, Dai Q, Zhang N-Y (2010) 3d model comparison using spatial structure circular descriptor. Pattern Recogn 43(3):1142–1151
12. Gao Y, Tang J, Hong R, Yan S, Dai Q, Zhang N, Chua T-S (2012) Camera constraint-free view-based 3-d object retrieval. IEEE Trans Image Process 21(4):2269–2281
13. Guétat G, Maitre M, Joly L, Lai S-L, Lee T, Shinagawa Y (2006) Automatic 3-d grayscale volume matching and shape analysis. IEEE Trans Inf Technol Biomed 10 (2):362–376
14. Hu DJ Latent dirichlet allocation for text, images, and music. University of California, San Diego. Retrieved April 26, 2013
15. Ip CY, Lapadat D, Sieger L, Regli WC (2002) Using shape distributions to compare solid models. In: Proceedings of the seventh ACM symposium on solid modeling and applications. ACM , pp 273–280
16. Kim W-Y, Kim Y-S (2000) A region-based shape descriptor using zernike moments. Signal Process Image Commun 16(1):95–102
17. Leibe B, Schiele B (2003) Analyzing appearance and contour based methods for object categorization. In: CVPR, vol 2, pp 409–415
18. Leng B, Qin Zg, Cao X, Wei T, Zhang Z (2009) Mate: a visual based 3d shape descriptor. Chin J Electron 18(2):291–296
19. Leng B, Xiong Z (2011) Modelseek: an effective 3d model retrieval system. Multimed Tools Appl 51(3):935–962
20. Li W, Bebis G, Bourbakis NG (2008) 3-d object recognition using 2-d views. IEEE Trans Image Process 17(11):2236–2255
21. Ohbuchi R, Furuya T (2009) Scale-weighted dense bag of visual features for 3d model retrieval from a partial view 3d model. In: 2009 IEEE 12th international conference on computer vision workshops (ICCV Workshops). IEEE, pp 63–70
22. Ohbuchi R, Osada K, Furuya T, Banno T (2008) Salient local visual features for shape-based 3d model retrieval. In: IEEE international conference on shape modeling and applications, 2008. SMI 2008. IEEE, pp 93–102
23. Osada R, Funkhouser T, Chazelle B, Dobkin D (2002) Shape distributions. ACM Trans Graph (TOG) 21(4):807–832
24. Paquet E, Rioux M, Murching A, Naveen T, Tabatabai A (2000) Description of shape information for 2-d and 3-d objects. Signal Process Image Commun 16(1):103–122
25. Regli WC, Cicirello VA (2000) Managing digital libraries for computer-aided design. Comput-Aided Des 32(2):119–132
26. Shih J-L, Lee C-H, Wang JT (2007) A new 3d model retrieval approach based on the elevation descriptor. Pattern Recogn 40(1):283–295
27. Shilane P, Min P, Kazhdan M, Funkhouser T (2004) The princeton shape benchmark. In: Shape modeling applications, 2004. Proceedings. IEEE, pp 167–178
28. Sundar H, Silver D, Gagvani N, Dickinson S (2003) Skeleton based shape matching and retrieval. In: Shape modeling international, 2003. IEEE, pp 130–139
29. Tangelder JWH, Veltkamp RC (2003) Polyhedral model retrieval using weighted point sets. Int J Image Graphics 3(01):209–229
30. Wang F, Li F, Dai Q, Er G (2008) View-based 3d object retrieval and recognition using tangent subspace analysis. In: Electronic Imaging 2008. International Society for Optics and Photonics, pp 68220I–68220I
31. Wong H-S, Ma B, Yu Z, Yeung PF, Horace HIp (2007) 3-d head model retrieval using a single face view query. IEEE Trans Multimed 9(5):1026–1036
32. Yeh J-S, Chen D-Y, Chen B-Y, Ouhyoung M (2005) A web-based three-dimensional protein retrieval system by matching visual similarity. Bioinformatics 21(13):3056–3057
33. Zeng J, Leng B, Xiong Z (2014) 3-d object retrieval using topic model. Multimed Tools Appl:1–23



Weizhi Nie received the B.S. and M.S. degrees in electronic engineering from Tianjin University of China. He is currently pursuing the Ph.D. degree from Tianjin University. His research interests include multiple object tracking, computer vision and location-based social network.

Xixi Li received the M.S. degrees in electronic engineering from Tianjin University of China. His research interests include multiple object tracking, computer vision and 3D object retrieval.

Anan Liu received the B.S., M.S. and Ph.D. degrees in electronic engineering from Tianjin University of China. He is currently a Associate Professor at the School of Electronic Information Engineering in Tianjin University. His research interests include computer vision, gesture recognition and medical image processing.

Yuting Su received the B.S., M.S. and Ph.D. degrees in electronic engineering from Tianjin University of China, in 1995, 1998 and 2001, respectively. He is currently a Professor at the School of Electronic Information Engineering in Tianjin University. His research interests include digital video coding, digital watermarking and data hiding multimedia forensics, and multimedia retrieval.