SVM: Sketch-based 3D Retrieval Application Using Classification Method

Wen ZHOU* and Jin-yuan JIA

Software School of Engineering, Tongji University, Shanghai, China

*Corresponding author

Keywords: Sketch-based 3D model retrieval, SVM classification, K-means cluster, Adaptive thinning.

Abstract. With the rapid development of 3D technology, the demand to use and retrieve the 3D model is becoming more and more urgent. In this case, it is more and more import and necessary that sketch-based 3D model retrieval. The user provides a hand-drawn sketch, then the system can be provided the possible list of models, then the user can select the need model from this list. In this paper, it’s proposed a framework of using SVM classification methods and K-means cluster algorithm. In preprocess part; the adaptive thinning algorithm is used to process the input sketch. In the offline part, the model is projected to multi-view images, and then SVM classifier has been used to classify these images. Moreover, we extract the images features, and use K-means algorithm to cluster these features. Finally, the feature dictionary will be acquired. In online process part, it only needs to extract the input sketch feature, and computes the distance between dictionary and their features. Besides, the experiment is realized to verify the feasibility of the approach. Finally, it was compared with other’s approaches; the result shows that the approach is viable and robustness.

Introduction

3D model retrieval has been a hot research topic in computer graphics, information retrieval and pattern recognition in recent years. How to identify the 3D models data, retrieval, reuse and re-model has become a common concern subject to designers, engineers and researchers. Because it is subjective and biased; the defect of text-based retrieval has been more and more apparent. Two different people will describe a same object totally different, due to their cultural backgrounds, their world view, their live environment and even their emotional states etc. the traditional text-based retrieval, which need to make manual annotation and is a very tedious and difficult task with the explosive growth of storage capacities, and hence this situation makes more focus been put on alternatives for retrieval using text keyword, which describe the content of the model only rely on text annotation.

Therefore, sketch-based retrieval has also been a major research field. One salient characteristic of a sketch is the stroke orientation. Orientation is a characteristic that had been exploited widely and richly showing outperforming results in tasks like object recognition and object categorization etc. Furthermore, due to lack of features in a sketch, and then, it is necessary that more robust descriptors should been used to exploit the relationship between a sketch and 3D model.

The outline of this paper was as followed: in Sec.2, we introduced the related work. In Sec.3, we introduced some descriptors and their derivatives. In Sec.4, it explained our framework. Experimental results and comparative evaluation were discussed in Sec.5. Sec.6 concluded the work.
**The Related Research**

More efforts in this research fields, it was called Content Based Image Retrieval (CBIR). [1] proposed the method, called Query by Visual Example (QVE). And NIBLACK et al.[2] also developed the system, called Query by Image and Video Content System (QBIC). According to Hu et.al [3], “a key challenge in SBIR is overcoming the ambiguity inherent in sketch”. In fact, a sketch included fewer features than image or example, only had the contour and some pixels, and there existed very large differences due to the level of the user’s hand drawing sketch, such as professional and amateur level.

Currently, a relatively complete sketch based 3D model retrieval system mainly was in following systems [4]. In the system of sketch based 3D model retrieval, there was two key points: the 2D transformation and the extraction of the sketch feature of 3D model. The quality of these two steps directly determined the accuracy of the search results.

Funkhouser et al. [7] proposed a 3D model retrieval engine, which supported the switch between 3D and 2D. The method of 3D spherical harmonic was used. Eitz et al. [8] realized 2D/3D based retrieval algorithm by using Bag-of-words and HOG. But these methods didn’t the pre-process before retrieval. It maybe affected the result due to the ambiguity stroke of sketch or amateur drawing level caused the sketch error express the user purpose. Therefore, Li et al. [11] proposed the step of doing pre-process operate before retrieval starting; it would check the user hand-drawing sketch and display the possible sketch which tally with the user demand.

So far, sketch retrieval had formed 8 benchmark, they were as following: Snoggrass and Vanderwart[12] proposed the standard line drawings(1980), Cole et al[13]’s line drawing benchmark (2008), Saavedra and Bustos[14]’s sketch dataset (2010), Yoon et al[15]’s sketch-based 3D model retrieval benchmark (2010), Eitz et al.[9]’s sketch-based shape retrieval benchmark, Eitz et al.’s [10] sketch recognition benchmark(2012), Small-scale benchmark[11]: Shrec’12 Sketch Track Benchmark, large-scale: Shrec’13 Sketch Track Benchmark [16].These benchmarks played an important role in the research and application of sketch retrieval.

**The Descriptors**

**Classification Methodology**

We often divided the methods by three different views. They were that online and offline methods, global and local methods, statistical and Graph-based methods.

**Online and Offline Methods.** High real time or low latency was an important criterion; therefore, some steps should be operated in offline environment, such as, large scale images generated features or the best view images were acquired. If these processes were doing in online environments, it would seriously influence and low the system effect, what was more, it would consume large times to tackle with these operate, which lead to enlarge latency and the user cannot be patient due to too long waiting time; therefore, the system would become useless and valueless.

In case where the query was introduced online, the time information, namely the order in which sketch points had been introduced, was available and can be used as additional information, within this paradigm, a sketch was considered as a set of strokes, each defined as the set of points sampled between pens down and pen up events. When the sketch was introduced offline, the time information was not preserved and the sketch was dealt with as a bitmap. Online methods that rely on the time information suppose that the sketch database had been introduced online as well; hence they were not adequate retrieval on a database introduced offline.
Global and Local Methods. Methods for sketch-based image retrieval can be classified as global and local methods. Global methods generate features based on the coarse information of the sketch image, and hence did not carry much information about the local details surrounding a sketch point. As for local methods, they generated features using the point local neighborhood; therefore, they were more capable of capturing fine details of the sketch.

However, global features were robust against noise, yet not very distinctive. On the other hand, local features can provide more distinctiveness, since they captured fine details. But they were less robust against noise. This had motivated approaches combining global and local information in order to achieve good distinctiveness and keep fair robustness against noise.

Statistical and Graph Based Methods. Another classification was the methods that used a statistical feature representation and methods that used a graph-based feature representation. In statistical approaches, sketch images were represented by feature vectors using methods from the rich statistical repository such as shape pixel distribution and shape signatures, and matching was done using a distance or similarity function. Graph-based methods represent a sketch image using a graph-like structure, and matching was done by using graph matching methods.

Statistical approaches were less sensitive to noise than graph-based methods, while graph-based methods pave the way for allowing partial matching. Graph-based approaches offered only a limited repertoire of algorithmic tools needed to solve classification and clustering problems. By contrast, the statistical methods were mathematically well founded and offer many tools. Hence, both were complementary to each other[17].

State-of-the-art Descriptor

Dalal [23] et al. proposed the descriptor of histograms of gradient (HOG), it can capture edge of gradient structure that was very characteristic of local shape. Besides, translations or rotations made very little difference if they were smaller that the local spatial or orientation bin size. However, due to HOG followed a pixel-wise strategy, the represent of sketch image always produced many zeroes in the final histogram, because the sketch was sparse by nature. Saavedra [22] proposed the improved descriptor of histograms of edge local orientations (HELO), HELO was a cell-wise strategy; therefore, it seems to be very appropriate for representing sketch-like images. Saavedra proposed the soft computer of HELO (S-HELO), it computer cell orientations in a soft manner using bilinear and tri-linear interpolation, and took into account spatial information, then, it computed an orientation histogram using weighted votes from the estimated cell orientations. Fu et al. [18] also improved the HOG descriptor, namely, binary HOG descriptor (BHOG). It can be faster than Hog descriptor to compute the feature vectors and take up less memory.

In order to enhance the robust against noise in sketch images, Chatbri et al. [20] introduced an adaptation framework based on scale space filtering. Firstly, the sketch images were filtered by the Gaussian filter to smooth the sketch image, then, it extracted skeleton of sketch image. Weiss et al.[19] proposed the spectral Hashing algorithms (SHA). SHA sought compact binary codes of feature data so that the Hamming distance between code words correlated with semantic similarity. Wang[21] et al. did a review paper which introduced the hashing methods in detail.
Proposed Framework

![Proposed Framework Diagram](image)

In preprocess step, the adaptive thinning algorithm is used to process the input sketch in order to eliminate the ambiguity and noise of sketch image, and be more accurate when the feature vectors were extracted. Besides, SVM classifier has been used to classify many view images, K-means algorithm also has been used to cluster the feature. Finally, the result was acquired (see Fig. 1).

Adaptive Thinning Algorithm

Adaptive thinning method can produce thinned version of the input sketch image, a one-pixel-wide skeleton is visually similar to the original image, and preserves connectivity between foreground pixels. The input sketch should be a binarized or gray-scale image with only one channel, and the skeleton should be represented by black pixels.

![Neighborhood Pixel Layout](image)

The threshold, \( th \), for binarization:

\[
\begin{align*}
th &= \frac{1}{N_b} \sum_{i=1}^{N} \sum_{j=0}^{M} f(i,j) \\
f(i,j) &= \begin{cases} 
0, & \text{if } I_G(i,j) = 255 \\
I_G(i,j), & \text{otherwise}
\end{cases}
\end{align*}
\]

where, \( N_b \) is the number of none-white pixels, \( M \) and \( N \) are the dimensions of image \( I \), \( f \) is the function of \( I \). \( I_G \) was a gray-scale image. Besides, the sensitivity measure is the performance measurement to select the best thinning image in the scale space.

\[
S_m(I_{th}) = \frac{1}{n} \sum_{i=1}^{N} \sum_{j=0}^{M} S_1(i,j)
\]

where

\[
S_1(i,j) = \begin{cases} 
1, & T_{BW}(i,j) > 2 \\
0, & \text{otherwise}
\end{cases}
\]

58
\[ T_{BW}(i, j) = \sum_{k=1}^{8} \text{transition}(P_k) \]

\[ \text{Transition}(P_k) = \begin{cases} 
1 & \text{if } (P_k = 1) \text{ and } (P_{(k+1) \mod 8} = 0) \\
0 & \text{otherwise}
\end{cases} \]

here, if it was black foreground pixel, \( P_k = 1 \), otherwise, \( P_k = 0 \). The output was the image having the minimal value of \( S_m(I_{th}) \).

**SVM Classifier**

SVM classifier has been used to classify the multi-view images, which is projected from the model in many different view-points. This method was adopted in Eitz[9]. We uniformly put 102 cameras on the bounding sphere of the model; so that a model can be projected into multiple view images (see Fig.3). However, of course, many of these images are what we don't need, that is bad view images. Therefore, we need to train an intelligent SVM classifier to classify these view images. This can eliminate the negative interference of the bad viewpoint image for our retrieval results. This method is adopted by Zhao et al.[25] to acquire the best-view images of a model. Here, we only want to get the good view images, not best-view image. The difference is that the best view is generally only one, and the number of good points of view, the different models, the number is completely different, but certainly the count is more than one (see fig.4).

The steps are as following. First, we set up a mapping relationship between a class of sketches and a model. In fact, some model datasets had finished this work, such as Li, etc. [16]. Second, according to some perspectives, the model is projected into multiple view images, we extract HOG features (it is said that it is the best sketch feature), and then train the specific classifier. Next, we will use this classifier to classify the view images of other model. Last but not least, we will get all good view images; therefore, we will do our best to eliminate the negative effect on retrieval result.

![Figure 3. The process of acquired the multi-view images.](image)

![Figure 4. The overview of classification process.](image)
K-means Algorithm Cluster

Because the count of every model's good viewpoint is not totally same, it's necessary to cluster the extract features. On the one hand, it can reduce the compute time; on the other hand, it can make the size of each model’s features is the same. In this paper, we adopt the K-means algorithm, the main reason is that this algorithm is robustness and low time-consume.

![Figure 5. The overview of K-means cluster features.](image)

\[ K = \min\{\text{FEATURES. HEIGHT, VISUAL\_WORD\_SIZE}\} \]  \hspace{1cm} (4)

In this paper, we set the \text{VISUAL\_WORD\_SIZE} = 50. According to equation 4, the value of K is decided by the height of features space and \text{VISUAL\_WORD\_SIZE}.

The steps are as following. First, we extract the features of the model sets, like 4.2. Second, we randomly select center point and the count of iteration. Here, we set the number of iterations to five. The value of K is the minimum value of \text{VISUAL\_WORD\_SIZE} and the height of features space. As a result, we can get a new feature space. This is the feature dictionary we need. The process of retrieval is to obtain the results after comparing the features of the input sketch and the feature dictionary.

The Experiments & Result Analysis

We compare our result using the dataset proposed by National Taiwan University[24]. The dataset was composed of 10119 3D models, besides; the dataset of sketches came from the sketch dataset of EITZ which contained 20000 query sketches. The method presented in this paper has been implemented using C++ and Java program languages, and is executed on PC under Windows 7 OS, Intel core i5-M580 processor, 4G memory size. In this section, all results are got in this machine. Moreover, it has been applied to several models successfully.

In order to improve the retrieval speed of the system, the low time consumption of the preprocessing stage is very important. Not only adaptive thinning algorithm is higher than the thinning in the performance, but also is lowers time consumption. The test of time consumption is executed on PC under Windows 7 OS (see Fig.6.).

In order to evaluate the performance of retrieval system, we compare the PR figure with the other methods.PR figure has been widely used in retrieval domain, including text-based retrieval.

Assume there are n models in the dataset, precision P is to measure the accuracy of the relevant models among the top K (1\leq K\leq n) ranking results, while recall R is the percentage of the relevant class that has been retrieved in the top K results. This is called the PR curve plot (see fig.7).

From the figure 7, we find that our method has a certain advantage. In particular, the statistical result has been finished in Taiwan National University model data sets. We realize others method in this datasets, and then compare these methods.
Conclusions & Future Work

In this paper, we propose a framework to realize the sketch-based 3D model retrieval. In the preprocess stage, adaptive thinning algorithm is adopted in order to eliminate the sketch ambiguity. In the offline stage, we use the SVM classifier and K-means algorithm to finish the features dictionary. Our design requirements are fully achieved. In the future, many problems still need to solve, such as, extracting better sketch features.

Acknowledgment

This research work was supported by the National Science Foundation of China (No.61272276, No. 61305091).

References


