Semantic-Oriented 3D Model Classification and Retrieval Using Gaussian Processes

Boyong GAO 1,2, Sanyuan ZHANG 1,†, Xiang PAN 3

1College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China
2College of Information Engineering, China Jiliang University, Hangzhou 310018, China
3College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310014, China

Abstract

The need of retrieving 3D models is constantly emerging. To improve the performance of a shape-based 3D model retrieval system, an approach is introduced to classify and retrieve 3D model by integrating shape features and semantic information. First, a new type of shape feature based on 2D views (called ZA) is proposed. Then we use Gaussian processes as supervised learning to model the mapping from low-level features to query concepts. At last the method ranks models by the overall distance determined by a weighted sum of the semantic distance and the shape feature distance. Experimental results show that the performance of the 3D model’s multiclass classifier using proposed method is significantly higher than those of other supervised learning methods, and the retrieval can capture the query model’s high-level semantics, the retrieval performance is improved.

Keywords: 3D Model Retrieval; Gaussian Processes; Semantics; Supervised Learning

1. Introduction

With the improvement of the 3D modeling tools and scanning devices, as well as the development of computer software and hardware technology, 3D objects become a type of important multimedia data with many applications, whose amount increases at geometric series. Roughly, beginning by 2000, 3D objects increasingly came into focus of multimedia retrieval research [1]. Reusing the models by retrieving 3D models in a large database becomes an important issue in the entertainment, CAD/CAM, game design and medical imaging. This has led to the research and development of 3D shape retrieval methods [2]. Obviously, the 3D models can’t be easily and precisely described only by text, content-based 3D model retrieval (below CB3DR) system was developed. These systems use the color [3], texture and shape information. The shape information is represented by rotation-, scaling-, translation-invariable feature descriptors based on topology [4], views [5] and shape [6, 7, 8]. Part matching uses local features [9].

Despite several years of research in this field, the performance of CB3DR is very limited. Firstly, feature descriptors capture the different characteristics of 3D model; different models are best represented by different descriptors. So it restricts the retrieval performance of large-scale general-purpose database.
including a variety of different models. Secondly, the statistical features and the distance are often not enough, they rely on low-level computable features such as color, shape and texture. People tend to use keywords, text descriptors and other high-level feature (concept) to explain and determine the degree of similarity between the objects. In general, there are no direct link between the high-level concepts and the low-level features. The semantically similar objects may have completely different low-level eigenvalue; search in the low-level feature space often leads to erroneous results. The discrepancy between the limited ability of low-level feature of 3D objects and the richness of user semantics forms the semantic gap. In order to exploit and to reuse the knowledge of 3D model database, it is an urgent need to retrieve precisely by semantic similarity with user queries, so high-level semantic processing in 3D model retrieval is now an active and challenging area of research [10].

Now there are many ways to reduce the semantic gap, we divide them into three categories [11]: (1) introducing relevance feedback into the retrieval loop for continuous on-line learning of users' intention [6, 7, 12]; (2) using supervised or semi-supervised learning methods to create the link between low-level features and the concept of queries [13]; (3) using the high-level ontology to define the concepts of 3D objects [14]. Some approaches use two or more of the above-mentioned methods [8]. Compared to relevance feedback, machine learning methods compute effectively, and are easy to classify in low-level feature space. For learning the concept, Xu et al. [15] used neural network, and Lü et al. [13] used Support Vector Machine (SVM) [16]. Compared to neural networks and SVM, Gaussian processes can establish the probability model and derive parameters.

Based on the analysis of the current 3D model retrieval systems’ low-level feature extraction and semantic learning method, the new method and new application of 3D model semantic retrieval are proposed. First, a new shape feature based on 2D views, named ZA, is proposed by combining Zernike moment of each view and the ratio of views area. Second, applying Gaussian process to 3D model semantic classification is proposed. Then the dissimilarity calculation formula is proposed. Finally, the method is compared with the 3D model of classification methods of using BP neural network.

This paper is organized as follows. Next section reviews the use of machine learning in the context of shape-based retrieval of 3D models. Section 3 proposes a method based on Gaussian process. Experiments and their results are described in Section 4. The conclusions and future work are discussed in the last section.

2. Previous Work

2.1. Low-level Feature Extraction of 3D Model

Feature extraction is to extract different characteristics of 3D models, which can distinguish between 3D models. This is the basis for retrieval. Now there are many complex 3D model’s feature extraction algorithms.

Shape features are the most important features of 3D model, so they are most widely used in the existing retrieval system. The advantage of shape-based retrieval technology is overall comparing the differences between shapes and ignoring details. This is similar to human visual recognition. The shape is the relatively well defined concept. The present studies mainly use the following features: the geometric relations between the points of model (the relationship of distance, angle, and the normal direction), the curvature
distribution of the model’s vertices, the statistical moments and the transform coefficients. For example, reference [7, 8] extracted a number of geometric parameters (such as area, volume and some special vector) as a feature vector.

2.2. Machine Learning for Three-dimensional Model Semantic Retrieval

In most cases, high-level semantic features is derived from low-level features by using supervised or semi-supervised machine learning techniques [15], such as SVM, neural network.

For the majority classification models in a high-dimensional space spanned by complex low-level features, the higher the dimension of feature vector is, the higher the class separability will be, but the more the estimated parameters are. When feature dimensions are greater than a certain number, the model's generalization performance will gradually decrease, the Hughes phenomenon appears [17].

For this reason, machine learning methods use kernel methods (KMs). KMs map the data from the original input feature space into a higher-dimensional kernel feature space through kernel functions. KMs can solve the nonlinearly separable problem, and the curse of dimensionality is bypassed. SVM is the kernel function approach, but there are some open issues, such as: no efficient practical method for selecting the best kernel function in specific problems; kernel function parameter selection, which usually only uses the experience method or the cross-validation method; how to select the appropriate penalty terms to prevent over-fitting; the output that is not a probability.

2.3. Gaussian Process

Gaussian process (GP), known as normal random process, is a collection of random variables, any finite number of which have a joint Gaussian distribution [18]. GP is an important non-parametric classification tool, and it does not require any assumptions on the form of data’ class-conditional density.

At the same time GP is also a kernel-based approach, but with full Bayesian formulation that can be defined for probability modeling. And compared to neural networks and SVM, GP provides probabilistic prediction estimates, so that the results are easier to explain.

GP is flexible, its computation is simple and effective for small sample problem. Such properties are desirable to 3D model semantic classification.

3. Semantic Processing Approach

The proposed 3D model retrieval method based on semantic classification follows three steps: first is low-level feature extraction; second is using supervised learning to derive high-level semantic classification; the last is the semantic information retrieval. The following describes the various steps, focusing on the semantic classification using Gaussian process.

3.1. Shape Descriptor of Three-dimensional Model

The proposed low-level feature (ZA) belongs to based-view feature, whose main idea is: if two 3D models look similar from the perspective, the two models are similar. In the paper, whether the two models similar to the corresponding viewing angles is determined by the orthogonal moments and the area ratio of the projection. The shape descriptor is calculated as follows:
1) Pretreatment.

It consists of two major steps and they are translation and scaling. Prior to the extraction of the ZA descriptor, we compute the model’s centroid using the area-weighted approach to improve the robustness [19]. Every triangle is weighed proportionally to its surface area. The centroid of model is translated to the coordinate origin so that translation invariance is achieved as the centroids of all 3D models coincide. Then the farthest distance from the vertex and centroid scales into the unit length, so the 3D model is normalized to the unit ball to ensure translation and scale invariance;

2) Obtaining the three principal directions by PCA, and getting their view by projecting on the these directions;

3) Calculating Zernike moment for each view;

4) Calculating the view’s respective percentage of total area by using the area obtained in PCA;

5) Combining the results of 3) and 4) to form the new shape features, named ZA, as the low-level features of the model.

This low-level feature extraction algorithm is view based similarity method, calculation is simple, fast and robust.

3.2. 3D Model Semantic Classification Using Gaussian Process

Supervised learning using Gaussian process has two purposes. One is semantic classification. For a new query model, Gaussian process can calculate the semantic class label the object most reasonably belongs to. Another purpose is to improve the retrieval accuracy. But the semantic categories need to have samples in both the training sets and the test set.

It is assumed here that the data in each category share only one semantic concept. Given a set of 3D model training samples $D=\{(x_i, y_i)\}_{i=1}^{n}$, where $n$ indicates the number of samples; $X=\{x_i\}_{i=1}^{n}$ is extracted from the unknown distribution of independent and identically distributed samples, and every sample point $x_i \in \mathbb{R}^d$ is the low-level feature vectors of training samples, $\mathbb{R}^d$ represents the d dimensional low-level feature space of 3D model, shape features will be treated as random variables; $y_i \in \{1, 2, \ldots, C\}$ corresponds to the semantic label of each model. Because the values of semantic category labels are not continuous, GP can not be directly used for the classification task. Thus, a latent function $f(x)$ is employed to infer the semantic category labels. The latent function values of $n$ training samples for $C$ classes form vector

$$f = (f_1^1, \ldots, f_a^1, \ldots, f_1^c, \ldots, f_a^C)^T,$$

Where $x_i$ has $C$ latent functions $f_i = (f_1^i, \ldots, f_a^i)$.

In order to reason function $f(\cdot)$ under Bayesian framework, a priori must be given about possible forms of the probability distribution of function $f(\cdot)$ in function space. The GP prior is therefore placed over $f(x)$:

$$f \sim p(f|X)=GP(m(x), k(x, x'))].$$

A Gaussian process is completely specified by its mean function $m(x)$ and covariance function $k(x, x')$. The mean function $m(x) = 0$ is generally assumed, so the selected covariance function can determine a GP, and it
encodes our assumptions about the function which we wish to learn. A variety of covariance function can be chosen for the GP classification [18]. The squared exponential function commonly used for classification is selected:

\[
k(x, x') = \exp(-\frac{1}{2\sigma^2}(x - x')^2),
\]

where \(\sigma\) and \(l\) are the free parameters.

According to Bayes’ rule, the prior distribution of the function \(f\) can be updated into the posteriori \(p(f|D)\) by using the training data \(D\):

\[
p(f \mid D) = \frac{p(y \mid f)p(f \mid X)}{p(D)},
\]

where likelihood function \(p(y|f)\) converts the value of \(f\) to the corresponding semantic categories of the 3D model by logistic function, or probit function.

The above Bayesian framework depends on the kernel parameter \(\sigma\) and \(l\) of covariance function (3) which control the kernel shape. These parameters can be integrated into the hyper-parameter vector \(\theta\). According to the training sample set \(D\), the optimal hyper-parameters are got when maximizing the evidence, i.e. \(\theta^* = \arg \max_{\theta} p(D|\theta)\). If the posteriori distribution estimation \(p(f|D)\) is approximated as a Gaussian, \(p(D|\theta)\) can be analytically calculated. Then the gradient-based optimization method is applied to maximize \(p(D|\theta)\) [18].

After calculating the posteriori, the predictive distribution of the latent value can be computed by integrating out \(f\):

\[
p(f_* \mid x_*, D) = \int p(f_* \mid f, D)p(f \mid D)df,
\]

Then the predictive distribution of the class probability \(y_*\) can be computed by integrating out \(f\):

\[
p(y_* \mid x_*, D) = \int p(y_* \mid f_*)p(f_* \mid x_*, D)df_*.
\]

The semantic category of 3D model is \(\arg \max_{c} p(y(x_*, y^*) \mid x_*, D)\), \(c = 1, ..., C\).

### 3.3. Retrieval

Through the supervised learning, the new model can also be labeled with a semantic category, which now is reflected to search results. Our object is putting the model which is not similar on low-level feature but is semantically similar before the model which is similar on low-level features but not semantically similar. The probability of the new model belonging to the semantic categories has been calculated, now can be used by the combination of probability and low-level feature.

Dissimilarity distance function of query model \(i\) and database model \(j\) is defined as

\[
d_{ij} = w_s \cdot d_{sy} + w_l \cdot d_{ly},
\]

where \(w_s\) and \(w_l\) are the semantic weight and the low-level feature weight respectively, and \(w_s + w_l = 1\); \(d_{ly}\) is the two models’ low-level feature distance, \(L_1\) distance is adopted; \(d_{sy}\) is the semantic distance for the two models, defined as
ds_{ij} = (1-p_{ij})(1-p_{ji}), \quad (8)

Where \( p_{ij} \) is the probability of the model \( i \) being the model \( j \)'s semantic category calculated by Gaussian process.

4. Experimental Results and Analysis

3D model library used in our experiments is the Princeton Shape Benchmark (PSB) [20]. PSB includes 1814 models collected from the Internet, and the models are manually pre-classified into 161 classes.

4.1. Semantic Classification

Two experiments about semantic classification are done. First experiment uses the 5 categories of biplane, sword, shelves, rectangular table and the handgun selected in Ref. [15]. The 78 models are used as the training set, and the remaining 78 models as the test set. The second experiment uses 5 to 15 largest categories of the PSB, and randomly selects 66.67% of the models as a training set, and the rest as a test set.

To investigate the impact of the dimensionality of shape descriptor \( Z_A \) we successively added the descriptor’s dimensionality. Figure 1 shows the detail of the first experiment we did by varying the number of low-level feature dimensions. When the dimension is equal to 15, the classification error rapidly descends to 15.38%; the dimension is 39, the error is 7.69%. This graph shows that the best performance is attained at different dimensionality: 333, 366, 399, for instance. Because of a kernel-based approach, the method also attains the best performance when the dimension is 975. However, the plot in Fig.1 indicates that after reaching a certain dimensionality, the change of classification error rate is not significant, when adding dimensionality. Higher dimensionality requires more memory to store the vectors and more times required for descriptor extraction and matching process, so we set the dimensionality to 93 for the following experiments. The classification accuracy rate of the first experiment is 94.87% when the dimensionality is 93, which is lower than the best value (96.15%), but is higher roughly 34.87% than 60% in Ref. [15].

The results of the second experiment are shown in figure 2, the error rate is higher with the number of the training samples’ semantic classes (\( n_c \)) to increase. Lû et al. [13] used the multiclass SVM classification method on the collection of 1000 models, the average error rate of 4-class is 15%, the error rate of 6-class is 19%; Liu et al. [21] randomly selected 217 models for training from 326 models which are collected from the Internet and manually classified into four categories, 109 models for testing, and adopted the
adaptive-weighted asymmetric AdaBoost HMM classification method, the error rate of 4-class was 9.7%; the average classification error rate of 6-class was 4.76%, the error of 4-class rate was 2.97% in our second experiment.

![Fig.2 The Classification Error with Variable Number of Categories](image)

4.2. Retrieval Performance Evaluation

The retrieval experiment on 78 models selected from 5 classes of PSB in Ref. [15] compares two kinds of methods: using the proposed geometric features, the joint use of geometric features and the semantic classification information based on Gaussian process. And $w$, in (7) is a constant which is set to be 0.9, then $w=0.1$, so the second method relies mainly on semantic classification information.

The joint PR curve of precision and recall is used to describe 3D model retrieval system performance, as shown in Fig. 3. To further quantify the performance of the compared methods, other performance indexes are compared on Table 1. The results clearly show that a significant increase in the retrieval performance when using semantic classification information obtained from Gaussian process.

![Fig.3 The PR Curve of Two Methods](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>NN</th>
<th>FT</th>
<th>ST</th>
<th>EF</th>
<th>DCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZA</td>
<td>0.846</td>
<td>0.519</td>
<td>0.715</td>
<td>0.426</td>
<td>0.831</td>
</tr>
<tr>
<td>combine semantics</td>
<td>0.949</td>
<td>0.873</td>
<td>0.939</td>
<td>0.599</td>
<td>0.962</td>
</tr>
</tbody>
</table>

An example of 3D model retrieval is shown in Fig. 4, where (a) using the geometry feature ZA, the m705 and m715 do not belong to "handgun" category which the query model m656 belongs to; (b) using the method combining semantic information, the top 9 models are the same semantic concept with the query model.
5. Conclusion and Future Work

This paper proposed a new kind of low-level features ZA; Gaussian processes, a type of nonparametric Bayesian method, are introduction to the 3D models semantic classification and retrieval. Compared with the neural network methods in other literature, semantic classification accuracy is significantly improved, and retrieval performance is superior to using only low-level features.

There are some avenues of future exploration. For PSB has a small overall size and small individual category size, we would explore an active learning method with Gaussian process to reduce the number of samples and computation. In addition, we would generate a new covariance function for learning.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (No. 60703001), Key Project of Chinese National Programs for Fundamental Research and Development (973 program) (No. 2009CB320804), Program for production-university-academe of Guangdong Province joint with Ministry of Education(No.2010B090400193) and the Science and Technology Program of Department of Education of Zhejiang Province of China (No. Y200702635).

References


