Journal of Southeast University (English Edition) Vol. 23, No. 2, pp. 221 – 225 June 2007 ISSN 1003—7985

Part-level 3-D object classification with improved interpretation tree

Xing Weiwei[1] Liu Weibin[2] Yuan Baozong[2]

([1] School of Software, Beijing Jiaotong University, Beijing 100044, China)

([2] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China)

Abstract: For classifying unknown 3-D objects into a set of predetermined object classes, a part-level object classification method based on the improved interpretation tree is presented. The part-level representation is implemented, which enables a more compact shape description of 3-D objects. The proposed classification method consists of two key processing stages: the improved constrained search on an interpretation tree and the following shape similarity measure computation. By the classification method, both whole match and partial match with shape similarity ranks are achieved; especially, focus match can be accomplished, where different key parts may be labeled and all the matched models containing corresponding key parts may be obtained. A series of experiments show the effectiveness of the presented 3-D object classification method.

Key words: 3-D object classification; shape match; similarity measure; interpretation tree

The goal of 3-D object classification is to assign an unknown object to a generic object class. In this paper, we address the object classification problem with a part-level method based on the improved interpretation tree. Most existing object classification methods are implemented at the surface patch level, which require complete surface models of query objects and complex computation[1-2].

3-D part representation has been used popularly in computer vision[3-6] since it is viewpoint independent, insensitive to local variations and supported by extensive psychological evidence. However, most previous research has been directed toward object recognition rather than generic classification.

In this paper, we propose a part-level 3-D object classification method, which consists of the improved constrained interpretation tree search and the shape similarity measure. In the tree search stage, a set of integrated features and the corresponding constraints are first presented, which reflect the individual parts' shape and the models' topological information among volumetric parts. Then the constraints are employed to direct a fast interpretation tree search by defining the efficient tree search rules. Finally, a shape similarity measure algorithm is developed to obtain the matched models and the assigned object class.

Received 2006-12-12.

Foundation items: The National Basic Research Program of China (973 Program) (No. 2006CB303105), the Research Foundation of Beijing Jiaotong University (No. K06J0170).

Biographies: Xing Weiwei (1980—), female, doctor, wwxing@bjtu.edu.cn; Yuan Baozong (1932—), male, doctor, professor, bzyuan@bjtu.edu.cn.

1 Problem Description

In this paper, the 3-D object classification can be regarded as the match between 3-D object data and models, which can be formulated as determining the correspondences subject to certain match constraints between object data parts and model parts. One of the most well-known algorithms for high-level matches in computer vision is the interpretation tree algorithm[7], which searches a tree for consistent model-to-data matching pairs. Let $\{m_1, m_2, \ldots, m_i, \ldots, m_p\}$ be a set of model parts and $\{d_1, d_2, \ldots, d_{j_i}, \ldots, d_q\}$ be the object data parts, where p and q are part numbers of the 3-D model and object data. Starting at a root node, we construct the tree in a depth-first fashion, assigning one model part to different object data parts at each level of the tree. To deal with the possible nonexistence of a feasible match between the current model part and the object data parts, the "wild card" is introduced to match a null object data part with a model part in the tree in order to improve robustness. A path through this tree represents a set of feasible correspondences, i. e., a consistent interpretation.

In this paper, superquadric-based geons (SBGs) are implemented to represent the volumetric parts, the constituents of 3-D objects. SBGs combine the superquadric quantitative parametric information and geon qualitative geometrical attributes, which have a powerful representative and discriminative capability.

2 Features, Constraints and Tree Search

The efficacy of the interpretation tree method as a matching algorithm is the use of some constraints to prune the branches of the tree which can lead to inconsistent interpretations. In this paper, we present a set of novel integrated part features with powerful discriminative capabilities, based on which the corresponding constraints and tree search rules are defined to fast find possible consistent interpretations.

2.1 Features and corresponding constraints

2.1.1 Unary features and constraints

All the following unary constraints forced by the corresponding features are applied to the newest pair (m_i, d_{j_i}) in an interpretation tree, where $i = 1, 2, \ldots, p$ and $j_i = 1, 2, \ldots, q$.

① Part connection number FU_{Pnum}: It is the number of parts connecting with the current part. Considering the $FU_{\text{Pnum}}^{m_i}$ of model part m_i and the $FU_{\text{Pnum}}^{d_{j_i}}$ of object data part d_{j_i}, we define the constraint for FU_{Pnum} as

$$FU_{\text{Pnum}}\text{-constraint}(i, j_i) = \text{True} \quad \text{iff } FU_{\text{Pnum}}^{m_i} = FU_{\text{Pnum}}^{d_{j_i}} \tag{1}$$

② Volume rate FU_{Vrate}: The ratio of current part volume to the whole model volume, reflects the part's spatial occupancy. The constraint forced by FU_{Vrate} is

$$FU_{\text{Vrate}}\text{-constraint}(i, j_i) = \text{True} \quad \text{iff } |FU_{\text{Vrate}}^{m_i} - FU_{\text{Vrate}}^{d_{j_i}}| \leqslant \varepsilon_{\text{v}} \tag{2}$$

This constraint says that the volume rate FU_{Vrate} of m_i must differ from that of d_{j_i} by no more than a bounded measurement tolerance.

③ 3-D spherical harmonic descriptor FU_{sph}: It is a 3-D rotation invariant describing volumetric part shape[8]. Since FU_{sph} is a vector, we let

$$FU_{\text{sph}}\text{-constraint}(i, j_i) = \text{True} \quad \text{iff} \| FU_{\text{sph}}^{m_i} - FU_{\text{sph}}^{d_{j_i}} \| \leqslant \varepsilon_{\text{s}} \tag{3}$$

④ Geon type FU_{geon}: Geon type is the qualitative attribute of volumetric parts. The related constraint is defined as

$$FU_{\text{geon}}\text{-constraint}(i, j_i) = \text{True} \quad \text{iff } FU_{\text{geon}}^{m_i} = FU_{\text{geon}}^{d_{j_i}} \tag{4}$$

⑤ Elongation FU_{elong}: This feature consists of two elements $FU_{\text{elong1}} = a_{\max}/a_{\text{med}}$ and $FU_{\text{elong2}} = a_{\text{med}}/a_{\min}$, in which $a_{\max}$, a_{med}, $a_{\min}$ are the maximal, medium and minimal superquadric size parameters of volumetric parts along x, y, z axes, respectively. Here, we let

$$FU_{\text{elong}}\text{-constraint}(i, j_i) = \text{True} \quad \text{iff} \| FU_{\text{elong}}^{m_i} - FU_{\text{elong}}^{d_{j_i}} \| \leqslant \varepsilon_{\text{e}} \tag{5}$$

2.1.2 Binary features and constraints

Suppose that current matching pairs are $\{(m_i, d_{j_i})\}, i = 1, 2, \ldots, k$ and given a new pair $(m_{k+1}, d_{j_{k+1}})$, the used binary features and corresponding constraints $(m_{k+1}, d_{j_{k+1}}; m_i, d_{j_i})$ are as follows:

① Connections FB_{connect}: This feature represents the connecting relationship of one part with other parts of the model and the related constraint can be written as

$$FB_{\text{connect}}\text{-constraint}(k+1, j_{k+1}; i, j_i) = \text{True} \quad \text{iff } FB_{\text{connect}}^{m_{k+1}, m_i} = FB_{\text{connect}}^{d_{j_{k+1}}, d_{j_i}} \tag{6}$$

② Connection type FB_{contype}: It reflects the number of intersections between two parts, which corresponds to FB_{connect}. The FB_{contype}-constraint is defined by

$$FB_{\text{contype}}\text{-constraint}(k+1, j_{k+1}; i, j_i) = \text{True} \quad \text{iff } FB_{\text{contype}}^{m_{k+1}, m_i} = FB_{\text{contype}}^{d_{j_{k+1}}, d_{j_i}} \tag{7}$$

③ Volume ratio FB_{Vratio}: It is defined as the ratio of one part volume to another part volume. The related constraint is the difference of $V_{m_{k+1}}/V_{m_i}$ and $V_{d_{j_{k+1}}}/V_{d_{j_i}}$. It must be restricted within a bounded measurement tolerance for $i = 1, 2, \ldots, k$, which is formulated as

$$FB_{\text{Vratio}}\text{-constraint}(k+1, j_{k+1}; i, j_i) = \text{True} \quad \text{iff } | FB_{\text{Vratio}}^{m_{k+1}, m_i} - FB_{\text{Vratio}}^{d_{j_{k+1}}, d_{j_i}} | \leqslant \varepsilon_{\text{vr}} \tag{8}$$

The above tolerances ε_{v}, ε_{s}, ε_{e} and ε_{vr} are predefined empirically.

The unary features and constraints represent the volumetric part shape, while the binary feature constraints mainly reflect the topological structure of the 3-D object, all of which are efficient for the tree search.

2.2 Constrained tree search

Given these feature constraints, the constrained tree search process consists of a depth first search. Suppose that the search process is currently at some node at level k in the interpretation tree and with a consistent partial interpretation given by

$$I_k = \{(m_1, d_{j_1}), (m_2, d_{j_2}), \ldots, (m_k, d_{j_k})\}$$

We now consider the next model part m_{k+1} and its possible assignment to object data part $d_{j_{k+1}}$, where j_{k+1} varies from 1 to $q+1$ and q is the number of object data parts. This leads to a potential new interpretation

$$I_{k+1} = \{(m_1, d_{j_1}), \ldots, (m_k, d_{j_k}), (m_{k+1}, d_{j_{k+1}})\}$$

Due to implementing volumetric parts instead of surface patches, the computational complexity is decreased greatly, which allows for defining the looser pruning in the tree search rules so that more possible consistent interpretations and matching results are obtained.

The following rules are defined and applied in the constrained tree search.

- If $(m_{k+1}, d_{j_{k+1}})$ is a wild card match, then the new interpretation I_{k+1} is consistent and we continue downward in our search.
- If m_{k+1} and $d_{j_{k+1}}$ are both real parts, we must verify that the unary constraints hold for the pair $(m_{k+1}, d_{j_{k+1}})$, and that the binary constraints hold for the pairs $[(m_{k+1}, d_{j_{k+1}}), (m_i, d_{j_i})]$, for $i = 1, 2, \ldots, k$.
- If N_1 unary constraints and N_2 binary constraints are true, where N_1 and N_2 can be predefined as the threshold according to the required matching precision, then the new interpretation I_{k+1} is a consistent interpretation, and we continue our depth first search. Other-

wise, I_{k+1} is an inconsistent interpretation; in this case, we increase the object data part index j_{k+1} by 1 and try again with a new I_{k+1}, until $j_{k+1} = q + 1$.

- If the number of wild card matches along one branch is beyond a predefined threshold, then this branch is considered to be an inconsistent interpretation.

However, if the search process is currently at some node at level k in the interpretation tree, and has an inconsistent partial interpretation given by

$$I_k = \{(m_1, d_{j_1}), (m_2, d_{j_2}), \ldots, (m_k, d_{j_k})\}$$

then it is in the process of backtracking.

Once the search process reaches a leaf of the interpretation tree, a consistent interpretation is achieved, which represents a feasible match between the model and the object data. Finally, one or more consistent interpretations are obtained as the feasible part matches between the model and the object data. We may get the best match according to the similarity measure computation, which is described in section 3.

3 Similarity Measure Computation

Based on the obtained consistent interpretations and the presented feature constraints, a similarity measure computation algorithm is developed to achieve matching results with similarity ranks. The algorithm consists of the following: ① Part similarity measure; ② Whole and partial match; ③ Focus match on the labeled key parts; ④ Classification.

Definition 1 Let one certain object class include r models M_l with $l = 1, 2, \ldots, r$, the unknown object data D and model M_l are represented as a set of parts

$$M_l = \{m_{l,1}, m_{l,2}, \ldots, m_{l,i}, \ldots, m_{l,p}\}$$
$$D = \{d_1, d_2, \ldots, d_{j_i}, \ldots, d_q\}$$

The nodes of the interpretation tree denote the matching pairs $(m_{l,i}, d_{j_i})$ of M_l and D for $i = 1, 2, \ldots, p$ and $j_i = 1, 2, \ldots, q$. Each part $m_{l,i}$ or d_{j_i} can be formulated as

$$\left.\begin{aligned} \boldsymbol{f}_m &= \{u_1, u_2, \ldots, u_k, \ldots, u_n\} \\ \boldsymbol{f}_d &= \{v_1, v_2, \ldots, v_k, \ldots, v_n\} \end{aligned}\right\} \tag{9}$$

where u_k or v_k may be a scalar or a vector in accordance with the features described in section 2.

3.1 Part similarity measure

Part similarity is computed on one node along a feasible path obtained from the constrained tree search stage, i. e. the matching pair $(m_{l,i}, d_{j_i})$ of one model part and the object data part. Due to the different effects of features on the shape similarity measure, it is natural to assign different weights $\{w_1, w_2, \ldots, w_k, \ldots, w_n\}$. The weight values are determined empirically, and reflect the experience by some tests on the importance of each feature.

First, for two corresponding features u_k and v_k, the feature similarity $\mathrm{Fsim}(u_k, v_k)$ can be calculated by Eq. (10) or (11). For $\mathrm{FU_{Pnum}}$, $\mathrm{FU_{geon}}$ and $\mathrm{FB_{connect}}$, $\mathrm{FB_{contype}}$, Eq. (10) is implemented,

$$\mathrm{Fsim}(u_k, v_k) = \begin{cases} 1 & \text{if } u_k = v_k \\ 0 & \text{if } u_k \neq v_k \end{cases} \tag{10}$$

And for the other features, Eq. (11) is employed,

$$\mathrm{Fsim}(u_k, v_k) = \left\| 1 - \frac{\| u_k - v_k \|}{\max(\| u_k \|, \| v_k \|)} \right\| \qquad 1 \leqslant k \leqslant n \tag{11}$$

Then the part similarity measure $\mathrm{Psim}(m_{l,i}, d_{j_i})$ of $m_{l,i}$ and d_{j_i} is calculated by

$$\mathrm{Psim}(m_{l,i}, d_{j_i}) = w_{\mathrm{un}} \sum_{k=1}^{s} w_k \mathrm{Fsim}(u_k, v_k) + w_{\mathrm{bi}} \sum_{k=s+1}^{s+t} w_k \mathrm{Fsim}(u_k, v_k) \tag{12}$$

where s and t are the numbers of unary and binary features, respectively; part similarity measure Psim is the sum of the former $\mathrm{Psim_U}$ and the latter $\mathrm{Psim_B}$ corresponding to the unary and binary feature constraints, respectively, in which w_{un} and w_{bi} are the related normalized weights. In particular, for the nodes at the first level, $\mathrm{Psim_B}$ is always assigned to 1; and the wild card matches have no contribution to the part similarity computation, and thus are assigned to 0.

3.2 Whole and partial similarity measure

Model shape similarity measure Msim_l of model M_l and object data D corresponds to an obtained feasible consistent interpretation. We define two kinds of matches for the model similarity measure: whole match and partial match. Whole match focuses on the whole shape similarity of M_l and D; while the part number q of object data D is less than the p of model M_l; i. e. the object is superimposed on the 3-D model. There also exists a partial match that emphasizes the local accurate correspondence.

Let the similarity of part pair $(m_{l,i}, d_{j_i})$ in a consistent interpretation be Psim_i. The whole similarity measure $\mathrm{Msim}_l^{\mathrm{W}}$ of M_l and D can be calculated by

$$\mathrm{Msim}_l^{\mathrm{W}} = \frac{\sum_{i=1}^{\max(p,q)} \mathrm{Psim}(m_{l,i}, d_{j_i})}{\max(p, q)} \tag{13}$$

where p and q are the part numbers of model and object data, respectively. The similarity measure $\mathrm{Msim}_l^{\mathrm{P}}$ of the partial match is evaluated by

$$\mathrm{Msim}_l^{\mathrm{P}} = \frac{\sum_{i=1}^{\max(p,q)} \mathrm{Psim}(m_{l,i}, d_{j_i})}{q} \tag{14}$$

where q is less than p.

3.3 Focus matching similarity

Generally, the parts concerned for matching of the same object data are different according to different tasks or aims, which is just the motivation of the focus match proposed here. In focus match, the different key parts of the object data may be labeled and all the models containing the corresponding key parts can be matched. Let the labeled key parts number of object data be N_{Dkey}. Then two supplementary constraints should be added to the tree search especially for focus match. First, only the model with part number p that is not less than N_{Dkey} may be passed onto the interpretation tree search process; secondly, the allowed number of wild card matching pairs along a path must not be more than $p - N_{\mathrm{Dkey}}$; otherwise, the branch will be pruned. These two constraints ensure that all the key parts of object data will have real matches. The focus matching similarity is achieved by

$$\mathrm{FOsim} = \frac{\sum_{i=1}^{N_{\mathrm{Dkey}}} \mathrm{Psim}(m_{l,i}, d_{j_i})}{N_{\mathrm{Dkey}}} \tag{15}$$

3.4 Classification

The unknown object is classified based on the calculated class similarity measures.

1) Selecting all the models with similarity measures more than a predefined threshold, the class similarity measure is computed by averaging the model similarity measures between all the selected models and the unknown object data.

2) The unknown object data are classified into the object class that has the highest class similarity measure obtained in 1).

4 Experimental Results and Evaluation

We construct a number of 3-D models for each object class according to the actual diversity of the class. The models within one object class are similar but still exhibit significant variation among exemplars. 24 object classes are built and used in the experiments. The domain of the objects explored is that of some man-made commodities. We match the query objects selected randomly from the 3-D library with all the models in the library and analyze how well the result ranks correspond to human similarity classification of the models.

In each figure below, the leftmost column shows the query object, the middle columns show the first five closest models in the library which are ranked from left to right according to shape similarity measures, and the rightmost column is the classification result with object class label and the class similarity measure.

4.1 Whole match and partial match

Fig. 1 shows three experiments on the whole match. In each result, the query object to be classified is the first of the similarity ranks and the corresponding similarity measure value is 1. It is obvious that the ranked models according to the obtained similarity measures are consistent with human visual perception on object shape similarity.

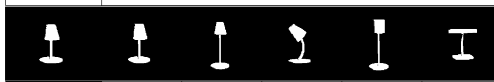
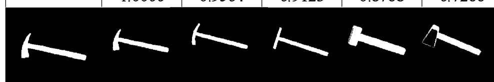
Query object	Model similarity rank					Object class
						Class 14 Lamp 0.8944
Query 1	lamp1	lamp4	lamp2	lamp5	lamp3	
	1.0000	0.9564	0.9123	0.8768	0.7266	
						Class 7 Hammer 0.8085
Query 2	hammer1	hammer4	hammer3	hammer2	ax1	
	1.0000	0.9258	0.9067	0.8900	0.8356	
						Class 22 Table 0.7625
Query 3	table1	table7	stool7	table9	table5	
	1.0000	0.9590	0.9319	0.8471	0.8384	

Fig. 1 Whole matching results of the objects belonging to the lamp class, hammer class and table class

Fig. 2 shows two experiments on cup class and stool class respectively. In the first experiment, the query object is a cup with only one part of the body. In the case of whole match, the matched models with higher similarity measures are the cups which consist of only one part since they have higher whole shape similarity to the query cup, such as cup 10 and cup 9 in the shaded cells. As for the partial match, it is obvious that cup 6 and cup 5 contain the more similar part, cup body, to the query and have better local shape similarity, and thus the higher similarity measures are achieved. In the second experiment on stool class, the obtained models by whole match have closer global topological structure and shape to the query, while the models from partial match contain the more similar "sub-model" to the query object in spite of different global structures. The experimental results clearly show the different focuses of whole match and partial match.

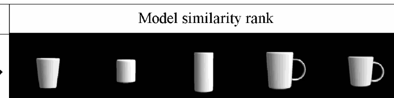
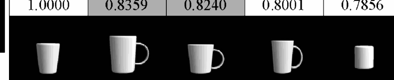
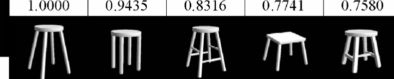
Query object	Model similarity rank					Object class
Whole match →						Class 1 Cup 0.6924
	cup11	cup10	cup9	cup6	cup5	
	1.0000	0.8359	0.8240	0.8001	0.7856	
						Class 1 Cup 0.7986
Partial match →	cup11	cup6	cup5	cup8	cup10	
	1.0000	0.8564	0.8481	0.8402	0.8359	
Whole match →						Class 21 Stool 0.6800
	stool16	stool14	stool6	table 1	table 7	
	1.0000	0.9435	0.8316	0.7741	0.7580	
						Class 21 Stool 0.7803
Partial match →	stool16	stool14	stool15	stool6	stool17	
	1.0000	0.9435	0.9209	0.8316	0.7808	

Fig. 2 Comparison experiments of whole and partial match

4.2 Focus match on key parts

Fig. 3 shows some experimental results of focus match. Take the chair class for instance. As for the same query chair, the labeled key parts are different, as illustrated in the left column of "key parts", and, hence, the different matching results are achieved. In the comparison experiments on the chair class, six key parts including one seat, four legs and one back bar, are labeled for the first test, and the six parts with a cross-bar between two legs instead of the back bar are labeled for the second test. In the results of the second experiment on chair class, the models with the corresponding key parts are matched successfully.

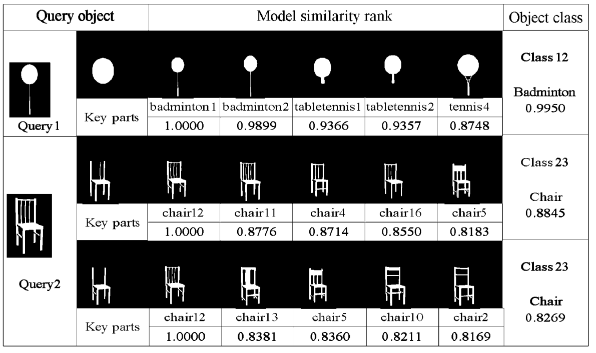


Fig. 3 Focus matching results of the objects belonging to the badminton racket class and the chair class

5 Conclusion

In this paper, a part-level method based on the improved interpretation tree is proposed to solve the 3-D object classification problem. The method mainly consists of the improved constrained interpretation tree search and the developed shape similarity measure computation. By this method, both whole match and partial match can be accomplished with shape similarity measures. In particular, the focus matching computation can be completed on different labeled key parts, which is suitable for different specific tasks.

References

[1] Tangelder J W H, Veltkamp R C. Polyhedral model retrieval using weighted point sets [J]. *International Journal on Image and Graphics*, 2003, **3**(1): 209 – 229.

[2] Csákány P, Wallace A M. Representation and classification of 3-D objects [J]. *IEEE Trans on Systems, Man, and Cybernetics—Part B: Cybernetics*, 2003, **33**(4): 638 – 647.

[3] Krivic J, Solina F. Part-level object recognition using superquadrics [J]. *Computer Vision and Image Understanding*, 2004, **95**(1): 105 – 126.

[4] Huber D, Kapuria A, Donamukkala R R, et al. Parts-based 3-D object classification [C]//*Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington DC, 2004, **2**: 82 – 89.

[5] Borges D L, Fisher R B. Class-based recognition of 3-D objects represented by volumetric primitives [J]. *Image and Vision Computing*, 1997, **15**(8): 655 – 664.

[6] Biederman I. Recognition-by-components: a theory of human image understanding [J]. *Psychological Review*, 1987, **94**: 115 – 147.

[7] Grimson W E L, Huttenlocher D P. On the verification of hypothesized matches in model-based recognition [J]. *IEEE Trans on Pat Anal Machin Intel*, 1991, **13**(12): 1201 – 1213.

[8] Xing Weiwei, Liu Weibin, Yuan Baozong. Superquadric similarity measure with spherical harmonics in 3-D object recognition [J]. *Chinese Journal of Electronics*, 2005, **14**(3): 529 – 534.

基于改进解释树的部件级三维物体分类

邢薇薇[1] 刘渭滨[2] 袁保宗[2]

([1]北京交通大学软件学院,北京 100044)
([2]北京交通大学信息科学研究所,北京 100044)

摘要:为了实现对未知物体的分类,提出了一种基于改进解释树的部件级三维物体分类方法. 采用部件级描述形式,使得对物体类的描述更加简洁. 所提的物体分类方法主要包括 2 个核心处理模块,即改进的约束解释树搜索和形状相似性度量计算. 利用该方法,不但能够进行未知物体与三维模型之间的全局匹配和部分匹配,得到具有形状相似度排序的分类结果;而且能够实现焦点匹配,即对同一个未知物体,为其标注不同的关键部件,通过焦点匹配便可以获得所有包含对应关键部件的三维模型. 大量的实验结果证明了所提出的部件级三维物体分类方法的有效性.

关键词:三维物体分类;形状匹配;相似性度量;解释树

中图分类号:TP391