International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.10, No. 2 (2017), pp.111-120
http://dx.doi.org/10.14257/ijsip.2017.10.2.10

Handwritten Templated Sketch Recognition Based on a Hierarchical Method

Xin Yuan[1], Cao Weiqun[2]
[1.] *School of Information Science and Technology, Beijing Forestry University, Beijing, China*
[2.] *School of Information Science and Technology, Beijing Forestry University, Beijing, China*
E-mail: [1]da.xinyuan@163.com, [2]weiqun.cao@126.com

Abstract

A hierarchical method for handwritten templated sketch recognition which combined bag of features (BOF) and perceptual hashing is proposed in this paper. The method takes both the overall properties and local characteristics of the sketch into account, in order to overcome the defection aroused by the strong randomness and much freedom of handwritten input. Firstly, we set up the rectangular bounding box for every sketch to get the corresponding sketch image and then adjust it to a square, the regularized sketch image. Secondly, divide every regularized sketch image uniformly to little patches, and take the bag of features as the local characteristics and use support vector machine (SVM) classifier to do the first level classification. Thirdly, resort Top 10 of the initial classification results using the perceptual hashing algorithm which reflects the differences of the objects on overall properties. For our experimental objects are sketches bearing certain structures (9 types), we take these structures as the additional overall features improving the recognition rate.We realize the recognition of 150 kinds of templated sketches, and the average recognition rate is 92.9% (Top1), and 100% (Top5) respectively. The experimental results show that the method is robust and has higher recognition rate.

Keywords: *Handwritten sketch recognition; Bag of features method; K-means cluster; Perceptual hashing technique; Support vector machine*

1. Introduction

In recent years, with the widely use of touch screen mobile phone and tablet PC, handwriting input mode becomes more and more popular, and research on handwritten sketch recognition becomes extremely active and turns to be an important research direction of artificial intelligence.

Most of the existing work on sketch recognition focuses on the stroke and structural features [1-5]. Besides, there are some works focusing on the visual appearance of shapes and symbols [6]. Zhang L S *et al.*, [7] presented a sketch-based graphics input system for conceptual design, which was based on online graphic recognition and dynamic user modeling. Mathias Eitz *et al.*, [8] analyzed the distribution of non-expert sketches of everyday objects, and developed a bag-of-features sketch representation, which recognized objects from 250 categories and achieved 56% accuracy of recognition. Delaye and Anquetil [9] extracted a set of 49 features, called HBF49, for the representation of hand-drawn symbols, based on which, the sketch recognition was implemented. Guo Y P *et al.*, [10] proposed a templated sketch recognition system that helps children to practice sketching by using perceptual hashing [11] to analyze the visual features of the sketch. Because perceptual hash functions are analogous if image features are similar, it's widely used for multimedia content identification, retrieval,

ISSN: 2005-4254 IJSIP

authentication, *etc.*, [12]. Methods based on the visual appearance are independent from the stroke sequence, and then are more robust for templated sketch recognition.

This paper presented a hierarchical method for handwritten templated sketch recognition, which combined bag of features (BOF) and perceptual hashing to take both the overall properties and local characteristics of the templated sketch into account. Some of the templated sketches are shown in Figure 1.

Figure 1. Templated Sketch Samples

2. The Basic Idea

The templated sketch is a special kind of sketch, whose structure and stroke sequence are regularized instead of generalized. For each type of object, the corresponding templated sketches have fixed overall properties, such as the structure, contour etc. Therefore the overall properties can be used to recognize the input sketch. However, the free drawing style of sketching results in the imprecision and variability of local characteristics of the same type of sketch [13]. Then, the local characteristics of the sketch should be taken into account as well for the recognition.

Pan H *et al.*, [12] used perceptual hash technology as the core part of image search engine, because perceptual hash functions are analogous if image features are similar. In our presented approach, perceptual hashing techniques was applied to the classification of templated sketches with the similar structure. The experimental results showed that it worked well for the templated sketches of small group of objects, but turned worse as the number of the candidate objects was more than 20. That is because the sketches composed of simple strokes and symbols and the information contained in its image is little and sparse, thus the perceptual hash function has less effective information and cannot work properly when the number of the candidate objects is big.

What’s more, we adopted BOF methods (bag of feature) from computer vision and represented sketches using local feature vectors that encode distributions of sketch local properties [8,9]. By using the clustering method based on statistical model, we got the feature vocabulary which can represent all the local features of the sketches. According to the feature vocabulary, all the local features of a sketch were classified, which can eliminate the difference of local features effectively. At this point, the sketch is repented as a frequency histogram of feature words, which is not sensitive to the individual local features in the sketch, but rather to the overall local characteristics. The experiment shows that the BOF method is not sensitive to the scale of the candidate objects. It should be pointed out that with BOF, the “similar sketches” is recognized in statistics, and then the consistency of frequency histogram of local features does not mean that the type of sketches are the same. But for the same type of sketches, the frequency histogram of the local features must be consistent.

Thus, a hierarchical method for templated sketch recognition which combined bag of features (BOF) and perceptual hashing was proposed in this paper. Firstly, used the BOF method to form the sequence of the whole candidate objects according to its similarity to the input sketch. Secondly, resorted smaller group of candidate objects on the top of the sequence (according to the experimental results, we took top10 of the sequence.) with the perceptual hashing algorithm.

For we worked on templated sketches with regularized structures (9 types) [10], each sketch has certain structure. By analyzing the experimental results, we found that certain structure information could further improve the recognition accuracy, and then we took the certain structure as additional overall features for recognition.

The flowchart of the proposed approach is shown in Figure 2.

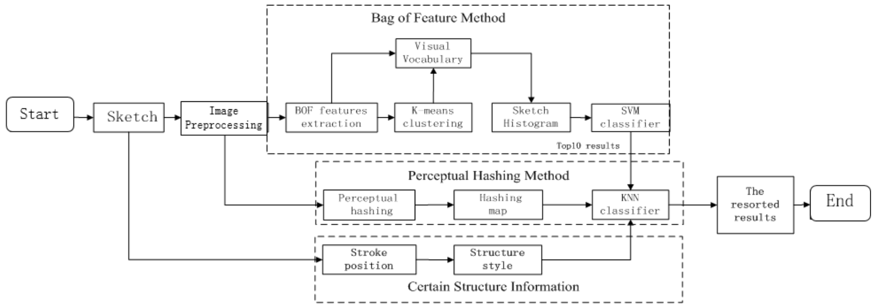


Figure 2. Sketch Recognition Flowchart

3. Key Steps

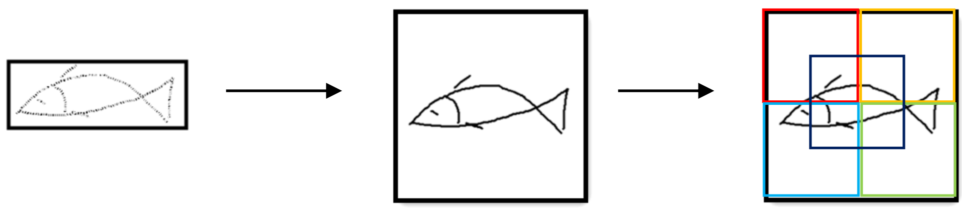
As shown in Figure 2, the proposed approach has following key steps: sketch image preprocessing; first level classification with BOF method [8]; second level classification with perceptual hashing method; additional classification with the certain structure information.

3.1. Preprocessing

The input image of the BOF method is required to be of the same size, while the shape and size of the input sketch are various. During preprocessing, normalize all the sketch images to a square firstly, and then isotropically scale them to the same size, 256×256 [8].

Moreover, the perceptual hashing method do the matching between images on the basis of sparse sampling, and the recognition result is relevant to the location of the sketch object in the sketch image. So we calibrate the sketch object at the center of the square image as well during the preprocessing.

Here is the process to form normalized sketch image for a given sketch object: (1) Create the rectangular bounding box for the sketch object and convert to the corresponding sketch image by Melkman algorithm;(2) Keeping the rectangular sketch image at the center, resize the canvas to a square, as shown in Figure 3.



(a) Bounding box (b) Regularized sketch image (c) Patches of sketch image

Figure 3. Process of Sketch Image Regularization

3.2. The First Level Classification

The BOF method is commonly used in the field of object recognition. It mainly includes three steps: (1) local feature extraction; (2) visual vocabulary cluster and histogram construct; (3) SVM classifier training.

3.2.1. Feature extraction: Traditional BOF method usually extracts SIFT features [14] or DIFT features of an image. These two methods define a feature space for object recognition using its bitmap representation, and work not so well for sketch images whose information is quite sparse. So, we extract gradient feature of the sketch image for the BOF method, which encodes stroke orientation in the sketch.

Given a sketch image S, here is the process to extract its local feature [9]: (1) calculate the gradient value of every pixel, and the results are save as ▽S. (2) Quantize the orientations of the stroke Θ to 4 bins O1(when Θ∈[0, 45)∪(315, 360]), O2(when Θ∈[45, 90)∪(270, 315]), O3(when Θ∈[90, 135)∪(225, 270]), O4(when Θ∈[135, 225]), and the quantized ▽S is denoted as ▽Sq . Each pixel of S has a corresponding quaternion (O1, O2, O3, O4) in ▽Sq. And because a pixel has a certain gradient orientation, $\sum_{i=1}^{4} O_i = 1$. (3) Sample small patches uniformly from the sketch image: sample 32×32 pixels at the upper left corner of the sketch image as the first patch, and sample continuously every 8 pixels in both horizontal and vertical direction respectively to get 28×28 patches; (4) In each patch, assemble the neighbored 8×8 pixels to get the patch of 4×4, where each sub-patch accumulates the quaternion of all the 8×8 pixels. Thus we can get 64=4×4×4 dimensional local feature for each patch.

3.2.2. Visual vocabulary cluster and histogram construction: Using a training set of n preprocessed sketch images, we construct a visual vocabulary using k-means clustering for generated n×784 local features, which partitions them into K disjunct clusters Ci (i = 1,2,…,K). And then, for each sketch image in training set, map its 784 patches to the visual vocabulary according to the Gaussian distance to construct its histogram.

Step 1: Obtain K cluster centers from N×784 local features d j= (0, N×784) through the K-means clustering algorithm. We define visual dictionary ν as the set of cluster center vectors {μ}. The center of each cluster is treated as a visual word and all visual words are combined to the visual vocabulary with capacity K.

$$\nu = \operatorname{argmin} \sum_{i=1}^{k} \sum_{d_{j \in C_i}} ||d_j - \mu_i||^2 \tag{1}$$

Step 2: For each sketch image in training set, map its 784 patches to the visual words in the visual vocabulary according to the Gaussian distance between them to construct the histogram H(ui), (i=1,2,…K), i.e. each patch is assigned to the visual word of smallest Gaussian distance with it.

$$K(d, \mu) = \exp(-||d - \mu||^2 / 2\sigma^2) \tag{2}$$

Where, d is the given local features, u is the visual word, and σ is set to be 0.1 [9].

Step 3: Normalize the histogram of each sketch image, and result in the normalized histogram (as shown in Figure 4).

$$HN(\mu i) = H(\mu i)/784 \tag{3}$$

Where, I = 1, 2,…, K.

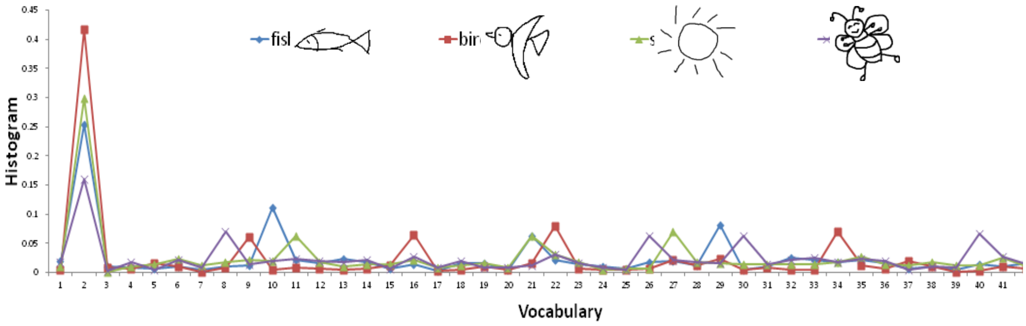


Figure 4. BOF Histograms of Sketches

3.2.3. SVM classifier: We use SVM to do the classification, and RBF is used as the kernel function. The first level classification results in a sequence of the whole candidate objects according to its similarity to the input sketch, and the most similar candidate is on the top of the sequence.

3.3. The Second Level Classification

According to the experimental results, for an input sketch object, the correct result is contained in the top10 of the sequence definitely (as descripted in 4.3). So we use perceptual hashing technology to resort the top10 candidates in the sequence to do the optimization according to the overall properties.

Step 1 Resize the sketch image to the 32×32 and get a picture of 1024 pixels (P1).Then convert P1 into a gray image (P2), which can be saved as a 32×32 matrix, the representation of the given sketch image.

Step 2 Calculate the DCT transform for P2.

$$\boldsymbol{F} = \boldsymbol{Af}\left(i, j\right)\boldsymbol{A}^{\mathrm{T}} \tag{4}$$

$$\boldsymbol{A}\left(i, j\right) = c\left(i\right)\cos\left[\frac{\left(j+0.5\right)\pi}{N} i\right] \tag{5}$$

F is the 32×32 matrix after DCT transformation, f (i , j) (i, j =1,2,…32) represents the matrix of gray image P2; A is a 32×32 matrix, and A(i, j) (i, j =1,2,…32) is the coefficient in the formula of the two-dimensional discrete cosine transform(DCT).

Step 3 Calculate the hash encoding sequence.

After DCT transformation, the information of lower frequency is transformed to the upper left corner, so we take only the upper left corner matrix F' (8×8) of F(32×32) into account. If F' (i, j) is greater than the average, it is recorded as 1, and 0, otherwise. Thus, each sketch is converted to a 64bits hash encoding sequence, H [64].

Step 4 Picture matching. Compare the hash encoding sequences between the input sketch and the top10 results, and calculate the Hamming distance. The smaller is the Hamming distance, means the two objects are more similar. If the distance is 0, we think, the two sketches are exactly the same.

3.4. Certain Structure Information

For our experimental objects are sketches with regularized structure (9 types) [10], we take the certain structure information as additional overall feature, which improve further the recognition rate. According to the relatively position of the first stroke and second stroke, sketches were divided into 9 types (upper, lower, left, right, upper left, upper right, lower left , lower right, inside and outside). For each candidate in the Top10 sequence resulted from 3.3, analyze its structure, and moved it to the end of the sequence if its structure is different from the input sketch.

(a) Structure types (b) Structure type: outside (c) Structure type: inside

Figure 5. Certain Structure Information of Sketches: (a) Nine Kinds of Structure Types, (b) Structure Type: Outside, (c) Structure Type: Inside

4.1. Experimental Environments

The experiment is carried out with Visual Studio 2013 in Windows 7 on a PC with CPU Intel （R） Core (TM) i5-2300, 4G RAM and NVIDIA GeForce GTX 460 graphics.

In this paper, we proposed an interactive human templated sketch recognition system that helps children to practice sketching as shown in Figure 6.

Figure 6. Templated Sketch Recognition System

4.2. Experimental Data

So far, we defined 150 kinds of templated sketches, and collected 100 samples of 5 different drawing levels (20 samples for each drawing level) for each kind. Among all the 15000 objects, 7500 were used as training samples, and the remaining 7500 were used as test samples.

4.3. Recognition Results

In the first level classification, the experiment results are: (1) the accuracy of top1 is 80.6%; (2) the accuracy of top 10 is 100%.

In the second level classification, we use perceptual hashing algorithm and structure recognition to resort Top 10 of the initial classification to reflect the differences of the objects on overall properties. The experiment results show that the average recognition accuracy is 92.9% (Top1) and 100% (Top5) respectively (as shown in Figure 7).

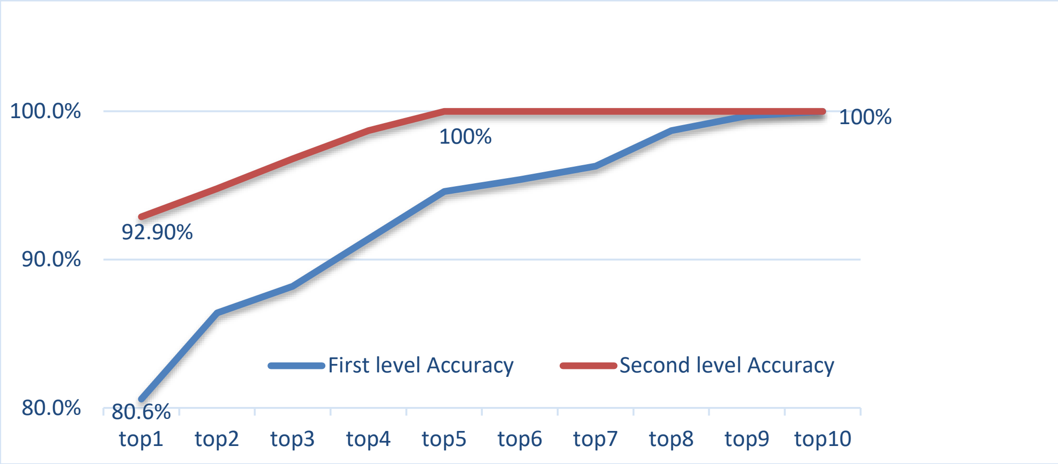


Figure 7. Average Recognition Accuracy

4.4. Results Analysis

Actually, the parameters of the BOF method such as the size of image patch, the visual vocabulary capacity and the orientation bins could be adjusted, which have rather a large influence on the average recognition accuracy. We choose the size of patch image 32×32, cluster centers K=40 and orientation bins O=4, with considering the recognition accuracy and time costing [9].

For most of the templated sketches, we have higher recognition rate. The results are shown in Table 1.

Table 1. Representative Sketches from 150 Categories with High Recognition Accuracy

Sketch categories recognition	sun	snowman	fish	hill	bird	octopus	house
Recognition accuracy	96%	98%	97%	95%	96%	96%	97%

However, in terms of the sketches with similar overall properties, the method mentioned in this paper could not distinguish them completely. As it is shown in table 2, we have three sketch groups whose objects share similar overall properties (such as shape and structure), which are prone to confuse with each other. This is the main problem we try to solve in the future work.

Table 2. Representative Sketches from 150 Categories with Low Recognition Accuracy

Groups	The first sketch group			The second sketch group		The third sketch group	
Sketch categories	Firefly	Bee	Locust	Chick	Duck	Bridge	Hat
Recognition accuracy	93%	91%	93%	92%	93%	92%	91%
Mistakes Recognition	Bee 4%	Firefly 5%	Ant 3%	Bird 4%	Chick 4%	Hat 3%	Bridge 4%

Groups	The first sketch group			The second sketch group		The third sketch group	
	Ant 1%	Butterfly3%	Firefly 1%	Duck 3%	Bird 3%	Bathtub 2%	Hat 2%

5. Conclusion and Future Work

In this paper, a hierarchical method is utilized to recognize sketches, which takes both their overall properties and local characteristics into account. In the first classification step, we make use of statistical vectors rather than structural representations because the former method enjoys more comprehensive characteristics to distinguish different drawing styles. In the second classification step, the perceptual hashing algorithm is used to reflect the differences of the objects in overall property. For our experimental objects are sketches bearing certain structures (9 types), we take these structures as the additional overall features improving the recognition rate. We realize the recognition of 150 categories of templated sketches, and the average recognition rate is 92.9% (Top1), and 100% (Top5) respectively. The experimental results show that the method is more reasonable, sound and enjoys higher recognition rate.

As it is reflected in the templated sketch, the temporal stroke order is quite certain and corresponding. We will make full use of it as an additional property to improve the recognition accuracy in further work. What is more, some similar sketches have their own certain features, such as chick, duck and bird with different tails in Table 2. As for these categories with low recognition performance, we can utilize them to further improve their recognition accuracy in the future.

Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Universities (NO. 2015ZCQ-XX）.

References

[1] X. L. Chang, “Research on hand-drawing geometrics recognition”, Wuhan University of Technology, **(2009)**.
[2] K. Yang, Q. Y. Hua and X. S. Chen, “A Simple Approach to Recognise Geometric Shapes Interactively”, Computer Technology & Development, vol. 18, no. 3, **(1999)**, pp. 21-27.
[3] S. Wang, G. Wang and H. Yu, “Interpretation of online multi-stroke sketching with straight lines based on time-space relationship”, Computer Engineering & Applications, vol. 48, no. 14, **(2012)**, pp. 198-202.
[4] T. M. Sezgin and R. Davis, “HMM-based efficient sketch recognition”, Proceedings of the 10th international conference on intelligent user interfaces, San Diego, California, USA, **(2005)**, January 9-12.
[5] J. Sivic and A. Zisserman, “Video Google: A Text Retrieval Approach to Object Matching in Videos”, IEEE Computer Society, vol. 2, **(2003)**, pp. 1470–1477.
[6] T. Y. Ouyang and R. Davis, “A visual approach to sketch symbol recognition”, Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-09), **(2009)**, pp. 1463-1468.
[7] L. S. Zhang, Z. X. Sun, R. H. Zhou and X. G. Xu, “A Method of Graph-Based Composite Sketchy Graphics Recognition”, Computer Science, vol. 31, no. 4, **(2004)**, pp. 147-150.
[8] E. Mathias, H. James and A. Marc, “How do humans sketch objects”, ACM TOG (Proceedings SIGGRAPH), vol. 31, no. 4, **(2012)**, pp. 44-54.
[9] A. Delaye and E. Anquetil, “HBF49 feature set: A first unified baseline for online symbol recognition”, Pattern Recognition, vol. 46, no. 1, **(2013)**, pp. 117-130.
[10] Y. P. Guo and W. Q. Cao, “Handwritten sketch recognition based on sketch entity and perceptual hashing”, Journal of Image and Graphics, vol. 20, no. 9, **(2015)**, pp. 1222-1229.
[11] S. Voloshynovskiy, O. Koval, F. Beekhof and T. Pun, “Robust perceptual hashing as classification problem: decision-theoretic and practical considerations”, IEEE Workshop on Multimedia Signal Processing, **(2007)**, pp. 345-348.
[12] H. Pan, G. Zheng, X. H. Hu and H. T. Ma, “Performance analysis of image content identification on perceptual hashing”, Journal of Computer-Aided Design and Computer Graphics, vol. 24, no. 7, **(2012)**, pp. 925-931.
[13] R. Davis, “Sketch Understanding in Design: Overview of Work at the MIT AI Lab”, Sketch Understanding, **(2002)**, pp. 24-31.

[14] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, vol. 60, no. 2. **(2004)**, pp. 91-110.

Authors

Xin Yuan, he is a graduate student in the School of Information Science and Technology, Beijing Forestry University, Beijing, China. His research interest is Computer Graphics and Sketch Recognition.

Cao Weiqun, she is currently a professor in the School of Information Science and Technology, Beijing Forestry University, Beijing, China. Her research interest is mainly in the area of Computer Graphics, Pattern Recognition and Digital Forestry. She has published several research papers in scholarly journals in the above research areas.

