A novel hand-drawn sketch descriptor based on the fusion of multiple features

Peng Zhao^{1,2,*}, GuoQin Wu^{1,2}, Yijuan Lu³, XianWen Wu^{1,2}, Sheng Yao^{1,2}

¹Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei 230039, China

²Co-Innovation Center for Information Supply & Assurance Technology, Anhui University, Hefei 230601, China

³Dept. of Computer Science, Texas State University San Marcos, TX, 78666, USA

Abstract: Constructing a distinctive and robust sketch descriptor is one of the most challenging problems in sketch based applications. In this paper, a new hand-drawn sketch descriptor is proposed. The proposed descriptor utilizes the statistic information of multiple features and bag-of-features representation to achieve translation and scale invariance and rotation robustness. The proposed descriptor also encodes the information entropy distribution of information point located on the contour, which can describe the intrinsic property of sketch better and is more robust to noises. Experimental results demonstrate the validity and efficiency of the proposed sketch descriptor.

Keywords: sketch descriptor; bag-of-features; information entropy; clustering

1. Introduction

With the popularity of touch phone, touch pad, hand-drawn sketch based image retrieval shows promising application potential and attracts a lot of attention in recent years. Users can sketch a rough shape of an object or a scene with a mouse or fingers on touch-enabled device and retrieve images with similar objects or scenes back. Different from full color and black-white images, hand-drawn sketch is a more abstract binary image with simple contour/shape information only. Moreover, hand-drawn sketch contains more noises since most of users are not professional drawers. When they draw sketches on the touch devices, their shaking hands always generate a lot of noises. How to construct a distinctive and robust sketch descriptor is still a challenging problem in the sketch-based image retrieval and other sketch based applications.

Mathias et al. started the first large scale exploration of human sketches, which analyzed the distribution of non-expert sketches of everyday objects. Many shape description methods have been presented in the past decades [2]. In general, sketch descriptors can be classified into four categories: contour-based descriptor [3-6], region-based descriptor [7-9], skeleton-based descriptor [10], and hybrid-based descriptor [11-13]. Yang Cao et al. [4] described a visual word using a triple (x, y, Θ) of the position (x, y) of an edge pixel (edgel) and the edgel orientation Θ at that position. By converting $\langle \text{image}, \text{edgels} \rangle$ representation to $\langle \text{document}, \text{words} \rangle$ representation, it leverages an inverted index-like structure to speed up the sketch-based image search. However, this representation inevitably results in the loss of position-invariant feature. Houssem Chatbri et al. [9] introduced a descriptor which combines feature information in near and far support regions defined for each sketch point. Mathias Eitz et al. [14-15] developed a bag-of-features sketch representation and represented sketches using local feature vectors that

* corresponding author. Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei 230601, China. Tel.: +8613033082994.
E-mail addresses: zhaopeng_ad@163.com

encode distributions of line orientation. Xin Shu et al. [16-17] represented shape with contour points distribution histogram (CPDH), which is evolved from shape context and is based on the distribution of points on object contour under polar coordinate. The circular shift and mirror matching scheme are adopted in measuring the distance between CPDHs to resolve the rotation invariance problem partially, which lead to high computational complexity. Qiang Jing et al. [11] described the sketch with the weighted combination of contour, region and skeleton features.

The hybrid-based sketch descriptor shows very good descriptive and discriminative power. But most existing hybrid-based descriptors just concatenate the feature vectors obtained with different description techniques and suffer from high computational cost and correspondence missing problems when matching two sketches by finding correspondence between points.

In this paper, we propose a novel hand-drawn sketch descriptor based on the fusion of multiple features (FMFD), which combines the contour and skeleton information. The proposed descriptor utilizes the statistic information of multiple features and bag-of-features representation to achieve translation and scale invariance and rotation robustness. The proposed descriptor and three other state-of-the-art sketch descriptors are implemented and tested on several sketch datasets. The experimental results show that the proposed sketch descriptor achieves higher classification accuracy and better retrieval precision.

The major contribution of the paper is that we combine the intrinsic properties implied in the sketch contour and skeleton based on the statistic information to construct the distinctive and robust descriptor efficiently.

The rest of the paper is organized as follows. Section 2 presents the details of the proposed approach and describes how to construct the sketch descriptor. In section 3, the invariance property and the computational complexity of the proposed approach are analysed. Section 4 shows the experimental results. Finally, the conclusion and future work are given in section 5.

2. The proposed sketch descriptor

In this section, we describe how to generate a sketch descriptor which is robust to sketch affine transformation (translation, scaling and rotation) and noise corruption. There are five steps: feature point extraction, information point extraction, construction of feature point descriptor, codebook generation, and construction of sketch descriptor. Firstly, the end points and branch points of the skeleton and vertices of minimum enclosing rectangle of sketch are obtained as feature points. Secondly, a set of points sampled from sketch contour are defined as information points. Thirdly, the descriptor of a feature point encodes the information entropy distribution of relative locations of information points to the feature point. Then, a codebook is learned from all the feature point descriptors of the training sketches. Finally, each sketch is mapped into the codebook space and a sketch descriptor is constructed. We discuss each step in more detail in this section.

2.1 feature point extraction

The skeleton and the minimum enclosing rectangle of a sketch are two typical features of the sketch. The skeleton represents the significant geometrical and topological features of a sketch. The minimum enclosing rectangle of a sketch indicates the orientation of a sketch. In the proposed approach, we define the vertices of the minimum enclosing rectangle of a sketch and the end points and branch points of the skeleton as feature points. In our experiments, skeletonization

algorithm in [11] is used to extract the skeleton of a sketch. For the original sketch S , we obtain its skeleton S' .

Fig. 1. Procedures and results of skeleton feature point extraction

The results of the skeleton feature point extraction are summarized in Fig.1. Figure 1 (a) is the original sketch with an airplane shape. Figure 1 (b) is the skeleton of the airplane and (c) is the feature points of the skeleton marked with red-cross symbol.

Then the resultant feature points of the skeleton can be represented as $F_1 = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, (x_i, y_i) \in R^2$, where n is the number of feature points of the skeleton.

Fig. 2. Feature points of the minimum enclosing rectangle

We extract the four vertices of the minimum enclosing rectangle as the feature points. Fig.2 shows a demonstration, in which four vertices marked with red cross symbol are extracted as the feature points. Then the resultant feature points of the minimum enclosing rectangle can be represented as $F_2 = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}, (x_i, y_i) \in R^2$.

Therefore the set of feature points F is the union of F_1 and F_2 : $F = F_1 \cup F_2$.

2.2 information point extraction

The contour of a sketch represents the shape information of a sketch. To avoid too many points of contour, the contour points are sampled with maximum distance method, in which the farthest N pairs of points on the contour are extracted. The resultant points on the contour are defined as information points. Then the set of information points can be represented as $I = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, where $(x_i, y_i) \in R^2$, $m = 2*N$, which is the number of information points.

The results of the information point extraction are summarized in Fig.3. Figure 3(a) is the original sketch with an airplane shape and (b) is the sampled result.

Fig. 3. Results of information point extraction

2.3 feature point descriptor

2.3.1 information entropy: Information entropy is widely used to measure the quantity of information in information theory, which represents the uncertainty of the information. In the proposed approach, the sketch information entropy is utilized to describe the information point distribution based on each feature point.

The information entropy of a sketch region is calculated as follows:

step1: calculating the positive index ρ_1 and negative index ρ_0 using Eq.(1) and Eq.(2) respectively.

$$\rho_1 = n/A \quad (1)$$

$$\rho_0 = 1 - \rho_1 \quad (2)$$

Where n is the number of information points in the region and A is the area of the region. ρ_1 depicts the distribution density of the information points, while ρ_0 depicts the distribution density of the blank space.

step2: calculating the information entropy p of the region using Eq.(3).

$$p = -\rho_0 * \log_2 \rho_0 - \rho_1 * \log_2 \rho_1 \quad (3)$$

2.3.2 construction of feature point descriptor: We denote the feature point as the origin and maximum value of distance between the origin and information point as radius to build a minimum circumscribed circle. An example is shown in Fig. 4. The point A is a feature point, while the point B is the farthest information point from A. We take A as the origin and the distance between A and B as the radius.

Fig. 4. Illustration of constructing feature point descriptor under polar coordinate

In this paper, the region of the minimum circumscribed circle is divided into sixty sectors using five concentric circles and twelve equal interval angles. The sectors are labelled anticlockwise from inside to outside. Each sector can be described as a triple $B_i = (r_i, \theta_i, p_i)$, where r_i denotes the radius of the concentric circle, θ_i denotes the angle space and p_i denotes the information entropy of the sector.

The descriptor of feature point encodes the information entropy distribution of relative locations of information points to the feature point. The distribution is encoded as a log-polar histogram with 5 bins for logarithmic distance and 12 bins for angles.

To achieve robustness against rotation, 1-D discrete Fourier transform of that histogram is applied with Eq.(4) and the set of magnitude $\|P(u)\|$ forms the final descriptor of the feature point as $(\|P(0)\|, \|P(1)\|, \dots, \|P(59)\|)$, where

$$P(u) = \frac{1}{60} \sum_{i=0}^{59} p(i) e^{-j2\pi ui/60} \quad (4)$$

The algorithm of constructing the feature point descriptor can be summarized as follows:

Algorithm 1: construction of the feature point descriptor

Input: the feature point (x_c, y_c) and the set of information points(I)

Output: the feature point descriptor

(1) Set the feature point as origin and convert $I = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ to polar coordinates $I_p = \{(r_1, \theta_1), (r_2, \theta_2), \dots, (r_m, \theta_m)\}, (r_i, \theta_i) \in R^2$.

Where $r_i = ((x_i - x_c)^2 + (y_i - y_c)^2)^{1/2}$ is the distance between the information point (x_i, y_i) and the feature point (x_c, y_c) , $\theta_i = \arctan((y_i - y_c)/(x_i - x_c))$ is the angle between (x_i, y_i) and x-axis.

(2) Set $r_{\max}$ as the radius of the minimum circumscribed circle, where $r_{\max} = \max(r_i)$, $1 \leq i \leq m$, partition the area into 5*12 sectors with 5 bins for $r_{\max}$ and 12 bins for θ .

(3) For $i=0$ to 59

Calculate the information entropy p_i of sector i using Eq.(3).

End

$$H' = (p_0, p_1, \dots, p_{59})$$

(4) Do Discrete Fourier transform with Eq.(4) and obtain the final descriptor of the feature point as:

$$H = (\|P(0)\|, \|P(1)\|, \dots, \|P(59)\|)$$

2.3 codebook generation

A large number of different feature points could be extracted from the sketches. A sketch usually has dozens of feature points and each feature point is represented as a sixty-dimension vector. If all the descriptors of the feature points in the same sketch are concatenated to form the sketch descriptor, the sketch descriptor will be a high and unfixed dimensional vector which will cause difficulty in the similarity calculation. To reduce the dimension of the sketch descriptor and construct a fixed-length sketch descriptor, we adopt a bag-of-features approach. So a fixed sized visual codebook of visual words is generated, in which perceptually similar features are quantized to the same visual word.

The codebook is learned from the training sketches. We extract feature points for each sketch and perform k -means clustering on the resulting set of all the feature point descriptors.

Codebook is denoted as $\mu = (\mu_1, \mu_2, \dots, \mu_C)$, where $\mu_i (i=1, 2, \dots, C)$ is the cluster center and is called a sketch visual word. C is the size of the codebook. In the clustering, the number of feature points is much smaller than the number of contour points or skeleton points, which largely reduces the computational time of k -means clustering. The setting of C (codebook size) is an important problem in k -means clustering. We will test various codebook sizes in experiments to evaluate the influence of the codebook size on the descriptor performance.

2.4 sketch descriptor construction

A sketch, namely S , is described as a frequency histogram of sketch visual words as:

$$h(S) = \frac{1}{|F|} \sum_{f_i \in F} \frac{q(f_i)}{\|q(f_i)\|_l} \quad (5)$$

Where $|F|$ is the number of feature points on the S , $q(f_i)$ is vector-valued quantization function which quantizes a feature point f_i in the codebook space.

In this paper, there are two different quantization strategies employed: hard quantization [18] and soft quantization [19]. In hard quantization, the feature point f_i is mapped to the nearest visual word as:

$$q(f_i) = [0, 0, \dots, \max\{K(f_i, u_j)\}, \dots, 0]^T \quad (6)$$

In soft quantization, the feature point f_i is mapped to a weighted combination of visual words as:

$$q(f_i) = [K(f_i, u_1), K(f_i, u_2), \dots, K(f_i, u_C)]^T \quad (7)$$

We use a Gaussian kernel to measure the similarity between feature point f_i and visual word μ_j with Eq.(8)

$$K(f_i, \mu_j) = \exp(-\|f_i - \mu_j\|^2 / 2\sigma^2) \quad (8)$$

As mentioned above, the whole algorithm of constructing the sketch descriptor can be summarized as follows:

Algorithm 2: construction of the sketch descriptor

Input: the set of sketches

Output: the sketch descriptors

(1) constructing the feature point descriptors using Algorithm 1.

(2) performing k -means clustering on the resulting set of feature point descriptors and generating the codebook $\mu = (\mu_1, \mu_2, \dots, \mu_C)$.

(3) for the feature point set F_s of each sketch S

(a) for each $f_i \in F_s$

for each $\mu_j \in \mu$

Calculate $K(f_i, \mu_j)$ using Eq.(8)

(b) adopting two different quantization strategies to quantize each f_i into the codebook μ :

① hard quantization using Eq.(6)

② soft quantization using Eq.(7)

(c) constructing the sketch descriptor using Eq.(5)

3. invariance analysis and computational complexity

The invariance of the sketch descriptor and the efficiency of sketch descriptor construction are very important in the large-scale sketch relative application. In this section, the invariance property of the proposed descriptor and the computational complexity of the descriptor construction are discussed. As described above, the feature point is set as the origin of the polar coordinate and information entropy distribution is employed in the representation of the feature point. So the invariance to scale and translation are the intrinsic properties of the FMFD. Additionally, the Discrete Fourier transform makes the FMFD more robust to the rotation of the sketch [20].

The construction of the proposed sketch descriptor mainly involves extraction of feature points and information points, codebook generation and descriptor construction. The quick sort algorithm is used in information point extraction, whose time complexity is $O(n \log n)$. In feature point extraction, the complexity of skeletonization is $O(n^2)$ and computation complexity of obtaining minimum enclosing rectangle is $O(n \log n)$. The n is the number of the contour points. In codebook generation, the complexity of K -means clustering algorithm is $O(f * k * t)$, where f is the number of feature points, k is the size of codebook and t is the number of iterations. Both k

and t are much smaller than f . The codebook generation is an off-line process. Therefore, the computational complexity of the descriptor construction is $O(n^2)$.

4. Experimental results and analysis

The most similar work to ours can be found in paper [16] and [9], which have evolved from shape context [21]. But our work is different from these two approaches in two aspects: (1) the proposed approach uses the end points, branch points of the skeleton, and vertices of minimum enclosing rectangle of sketch as feature points. Other approaches all rely on points on the sketch contour; (2) the sampled contour point distribution is calculated relative to the feature points, not the center of the sketch or other contour points. We fuse multiple significant features with the spatial relationship among different features, not just directly concatenating the vectors of different features obtained with different description techniques.

The performance of the proposed descriptor (FMFD) and three state-of-the-art descriptors (shape context [21], CPDH [16-17] and Support Regions Descriptor (SRD) [9]) are evaluated and compared on several public sketch datasets. These three sketch descriptors are selected due to their popularity and similarity to the proposed approach. Two variants of the proposed approaches are also evaluated: the proposed FMFD with hard quantization (denoted by FMFD-hard) and the proposed FMFD with soft quantization (denoted by FMFD-soft).

In our experiments, both the distribution of shape context and CDPH are encoded as log-polar histogram with 5 bins for logarithmic distance and 12 bins for angles. In SRD, near and far support regions are defined for each sketch point. Two types of experiments are designed: classification experiment and sketch retrieval experiment.

4.1 classification experiment on sketch dataset

(a) example classes in sketch dataset

(b) airplane

(c) cactus

Fig. 5. Examples in sketch dataset

The testing sketch dataset is the largest public human sketch dataset, which contains 20,000 sketches over 250 object categories that we commonly encounter in daily life. Examples of these

sketches are shown in Fig.5(a). Figure 5 (b) and (c) demonstrate the intra-class variation and inter-class ambiguity. In our experiments, the sketches are resized to 256*256 pixels.

The classification experiments are designed to evaluate the distinctiveness and robustness of the different descriptors. LIBSVM software package [22] is used to train a sketch classifier and the query sketch is recognized with the learned classifier. The classification accuracy is used to evaluate the performance, which is the number of classified correct sketch over the total number of sketches in that class.

18 classes from the sketch dataset are randomly chosen and each class contains 80 sketches. In our experiments, 80% of sketches (64 sketches in each class) are used for training and 20% of sketches (16 sketches in each class) are used for testing. So there are 1152 sketches in training set and 288 sketches in testing set in total. The experiments are repeated 10 times with a random selection at each run. The average accuracy is taken as the result.

Classification performance of SVM classifier can be quite sensitive to user-chosen model parameters. In this experiment, RBF kernel function is applied to train SVM classifier. There are several parameters need to be tuned. Through the cross-validation, the best penalty factor Cost and kernel parameter Gamma are obtained. Table 1 shows the optimal parameters-Cost and Gamma.

Table 1 Optimal parameters for SVM model using different descriptors

	Cost	Gamma
FMFD-soft	8	8
FMFD-hard	2	0.007813
CDPH	2048	0.000488
Shape Context	512.0	8.0
SRD	32768	0.125

To identify how the codebook size influences the classification accuracy, the following range of codebook sizes are tested: 100, 250, 500, 750, 1000. It turns out that the classification accuracy depends on the codebook size. From Table 2, we can see that the performance becomes more stable as the codebook size becomes greater than 500. In addition, the computational complexity of k -means clustering is $O(f*k*t)$, where k is the codebook size. Therefore the computational complexity is linearly increased as the codebook size increases. So the codebook size is set to 500 as a compromise in the following experiments.

In these experiments, the effectiveness of the four vertices of the minimum enclosing rectangle as feature points in FMFD is also validated. We take the four vertices of the minimum enclosing rectangle out of the set of feature points in FMFD with hard quantization, which is denoted by FMFD'-hard. From Table 2, compared with the FMFD'-hard, the performance of FMFD-hard improves slightly. It shows that the minimum enclosing rectangle of the sketch indicates the orientation of the sketch. The vertices of the minimum enclosing rectangle of sketch can capture more geometrical information of the sketch.

Table 2 Codebook size vs. computational classification accuracy

CodeBook Size	FMFD-soft	FMFD-hard	FMFD' -hard
100	0.80	0.73	0.71
250	0.80	0.75	0.73
500	0.86	0.79	0.78
750	0.86	0.79	0.78
1000	0.86	0.80	0.79

To evaluate the influence of the sample point number to the classification performance, different numbers of information points in FMFD and contour points in three other descriptors are extracted respectively. The following range of sample point number is tested: 50, 100, 150, 200, 250, 300, 350, 400.

Table 3 displays the classification accuracies of different descriptors with different sample point numbers. When the number of sample points is small, the classification accuracy is low. When the number of sample points increases, the classification accuracy increases accordingly. But after the number hits a certain point, the classification accuracy decreases with fluctuation. It reveals that fewer sample points cannot fully represent the shape information of the sketch. But more sample points may introduce more noise and more computation time. So we take a compromise which can not only capture the shape information but also reduce the influence of noise and the computing time.

Table 3 Sample point size vs. computational classification accuracy

Number of Sample Points	CDPH	SRD	shape context	FMFD-hard	FMFD-soft
50	0.40	0.56	0.69	0.67	0.75
100	0.47	0.58	0.71	0.76	0.82
150	0.52	0.63	0.74	0.81	0.83
200	0.47	0.67	0.76	0.79	0.86
250	0.49	0.67	0.74	0.77	0.84
300	0.49	0.66	0.75	0.79	0.83
350	0.46	0.66	0.74	0.80	0.84
400	0.47	0.67	0.75	0.78	0.84

From Table 3, we find that no matter with soft quantization or hard quantization, the performance of the proposed FMFD is better than the other descriptors. Shape context encodes the distribution of sample point locations on the contour relative to each of the other sample points. CDPH only considers the distribution of contour points based on sketch centre point, which is sensitive to the rotation. SRD combines feature information in near and far support regions defined for each sketch point. There are only two feature values extracted for each point, corresponding to near and far support regions from the point’s perspective. In addition, SRD utilizes the feature values of all the contour points and encodes the features anti-clockwise. That is why SRD is sensitive to noises and rotation. The proposed descriptor employs the information entropy distribution based on each of the feature points which are significant points of the sketch and can be used to describe the intrinsic property of sketch better. Moreover, there are sixty feature values extracted for each feature point in the proposed descriptor. So the proposed descriptor can capture more fine shape information of the sketch. The FMFD with soft quantization outperforms the FMFD with hard quantization. We attribute this performance boost to the ability of soft quantization to overcome the information lose in hard quantization.

4.2 classification experiments on kimia-216

Kimia-216 consists of 216 binary images from 18 categories. There are 12 different shapes in each category. Examples of these binary images are shown in Fig. 6. We extract the shape contour as hand-drawn sketch.

Fig. 6. Examples in kimia-216

In these experiments, the image set is divided into five parts and cross validation method is adopted. Table 4 reveals that the performance of FMFD-soft and shape context is similar. The images in kimia-216 are more normative and have less noise than the sketches in the previous sketch dataset. So the classification accuracies of the five descriptors are high and fluctuate within a small scope. Compared with Table 3, it is shown that the proposed descriptor is less sensitive to noises than the other descriptors. The reason is that FMFD takes the vertices of the minimum enclosing rectangle of sketch, the end points and branch points of the skeleton as feature points, which are less sensitive to noise than contour points. Since the shaking hand motions always generate a lot of noises, FMFD is more suitable to describe the hand-drawn sketch than the other three descriptors.

Table 4 Sample point size vs. computational classification accuracy

Sample Points	CDPH	SRD	shape context	FMFD-hard	FMFD-soft
50	0.88	0.88	0.93	0.89	0.94
100	0.90	0.88	0.94	0.89	0.94
150	0.92	0.91	0.96	0.90	0.95
200	0.92	0.92	0.96	0.91	0.96
250	0.89	0.93	0.95	0.91	0.96
300	0.93	0.92	0.97	0.92	0.97
350	0.93	0.92	0.96	0.94	0.96
400	0.94	0.95	0.96	0.92	0.95

4.3 classification experiments on Mpeg-7 Part B

Mpeg-7 Part B includes 1400 images which cover 70 categories. Each category comprises 20 images with different shapes. This image set is divided into five parts and cross validation method is adopted. In each image, the shape contour is extracted as a hand-drawn sketch. This dataset is challenging due to high inter-class ambiguity (images in different categories are very similar) and intra-class variation (images in the same category are dissimilar).

Table 5 shows the experimental results. The classification accuracy of FMFD-soft outperforms the other descriptors. Compared with the other descriptors, the proposed descriptor has more distinctive power and can deal with the intra-class variation and inter-class ambiguity better. This is due to that in FMFD the end points and branch points of the skeleton can more

columns are the top 5 matches using the proposed descriptor, which are ranked according to their rating from 1(left) to 5(right). From Fig.7, we can see most of the retrieved sketches are in the same semantic category as the query sketch. Although some results (e.g. the result sketches with the blue frame) are not correct, they look similar with the query sketch in the shape. The results of the retrieval also indicate our descriptor is robust to the rotation (shown in the result sketches with the green frame).

Fig. 8. An example of a original sketch (upper-left) and its transformed versions

To observe the sensitivity of the proposed FMFD to different transformations more clearly, five sketches from each class are randomly chosen and some transformations (scale and rotation) are performed on them. Fig.8 shows an example of transformation on a sketch, which includes an original sketch, two scaled versions created by two scale factors of 0.75 and 0.5, and seven rotated versions created by rotation in steps of 45° . These original sketches and their transformed versions compose a new dataset and we perform the sketch retrieval with the proposed FMFD on this dataset again.

Fig. 9. Illustration of top 10 sketches retrieved. The sketch with the red frame is the query sketch. The sketches with the black frame are the transformed versions of the query sketch. The sketches with the blue frame are not the transformed versions of the query sketch.

Fig.9 demonstrates a typical query result of the sketch retrieval. In Fig.9, the upper-left sketch with the red frame is the query sketch and the rest sketches are the top 10 matches using the proposed descriptor, which are ranked according to their rating by row from left to right. From Fig.9, we can see the top 10 matches include all scaled versions and 5 rotated versions. It depicts that the proposed FMFD is scale invariance and robust to rotation.

5. Conclusion and future work

In this paper, we develop a novel hand-drawn sketch descriptor based on the fusion of

multiple features. We encode the information entropy distribution of information point location on the contour relative to the feature point to construct the descriptor of the feature point and take the bag-of-features representation to map the sketch into codebook space. The main innovation is the efficient fusion of multiple features to capture the intrinsic and distinctive property of the sketch. The proposed descriptor is not only concise, but also insensitive to affine transformation and robust to noise. The experiments show that the performance of the proposed descriptor is superior to many state-of-the-art descriptors.

In the future work, we will apply the FMFD to the large-scale sketch-based image retrieval. Most of existing sketch-based image retrieval systems rate the geometric similarity of the query sketch and real images to find the best matches. But people often sketch an object in a simplified or exaggerated way, which may cause both the shape and proportion of the sketch very different from the corresponding real object. Therefore we will explore to apply the proposed FMFD in the sketch annotation and use the tag to search for the semantic similar images on Internet or in an image collection. The results will be more meaningful to users. There are a limitless variety of sketch categories in real word. How to index and organize the sketches efficiently is one of the key challenges in the large-scale sketch-based image retrieval. Recently, learning based hashing techniques have attracted broad research interests due to their ability to greatly increase the scale at which existing algorithms may operate[23-25]. We will also explore to combine the FMFD with the learning based hashing methods for making the pairwise comparisons more efficiently in the large-scale sketch dataset, which will be a good future research direction.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant Nos. 61472001, 61402005, 61202227), Natural Science Foundation of Anhui Province (Grant No. 1408085MF122, 1508085MF127), Natural Science Foundation of the Education Department of Anhui Province (Grant No. KJ2016A041) and Foundation of Center of Information Support & Assurance Technology (Grant No. ADXXBZ2014-5, ADXXBZ2014-6). This work is also supported by USA Army Research Office grant W911NF-12-1-0057 and NSF CNS 1305302 to Dr. Yijuan Lu.

References:

- [1] Mathias Eitz, James Hays, Marc Alexa, How Do humans Sketch Objects?, ACM Transactions on Graphics 31(4)(2012) 44:1-44:10.
- [2] Dengsheng Zhang, Guojun Lu, Review of shape representation and description techniques, Pattern Recognition 37 (1) (2004) 1–19.
- [3] Farzin Mokhtarian, Sadeh Abbasi, Shape similarity retrieval under affine transforms, Pattern Recognition 35(1) (2002) 31-41.
- [4] Yang Cao, Changhu Wang, Liqing Zhang, Edgel Index for Large-Scale Sketch-based Image Search, Proceedings of 24th IEEE International Conference on Computer Vision and Pattern Recognition (2011)761-768.
- [5] Yang Cao, Hai Wang, Changhu Wang, MindFinder: Interactive Sketch-based Image Search on Millions of Images, Proceedings of the international conference on Multimedia (2010)1605-1608.

- [6] Glauco Vitor Pedrosa, Celia A. Zorzo Barcelos, Marcos Aurelio Batista, A shape contour descriptor based on salience points, 18th IEEE International Conference on Image Processing (2011)1049 – 1052.
- [7] Mirosław Bober, MPEG-7 visual shape descriptors, IEEE Transaction on Circuits and Systems for Video Technology 11(6)(2001)716-719.
- [8] Whoi-Yul Kim, Yong-Sung Kim, A region-based shape descriptor using Zernike moments, Signal Processing: Image Communication 16(1-2)(2000)95-102.
- [9] Housseem Chatbri, Keisuke Kameyama, Sketch-based image retrieval by shape point description in support regions, 20th International Conference on systems, Signals and Image Processing (2013) 19-22.
- [10] Wai-Pak Choi, Kin-Man Lam, Wan-Chi Siu, Maximal disk based histogram for shape retrieval, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (2003)9-12.
- [11] QIAN Jing, WANG Zengfu, A method for freehand sketch retrieval based on affine adaptive skeleton, Journal of University of Science and Technology of China 40(10)(2010)1043— 1053.
- [12] QIAN Jing, WANG Zeng- fu, CAO Yang, Query-by-sketch Semantic Object Image Retrieval, Journal of Chinese Computer Systems, 31(10)(2010)2079-2087.
- [13] Rakesh S., Kailash Atal, Ashish Arora, et al. Face Image Retrieval Based on Probe Sketch Using SIFT Feature Descriptors, Proceedings of the First Indo-Japan Conference Perception and Machine Intelligence (2012)50-57.
- [14] Mathias Etiz, Kristian Hildebrand, Tamy Boubekeur, et al. Sketch-Based Image Retrieval: Benchmark and Bag-of-Features Descriptors, IEEE Transactions on Visualization and Computer Graphics 17(11)(2011)1624-1636.
- [15] Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur, et al. A descriptor for large scale image retrieval based on sketched feature lines, Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling (2009)29-38.
- [16] Shu Xin, Xiao-Jun Wu, A novel contour descriptor for 2D shape matching and its application to image retrieval, Image and Vision Computing 29(4)(2011)286-294.
- [17] Shu Xin, Xiao-Jun Wu, Shape Retrieval based on CPDH and Dynamic Programming Algorithm, Journal of Image and Graphics 16(5)(2011)844-849.
- [18] Josef Sivic, Andrew Zisserman, Video Google: a text retrieval approach to object matching in videos, Proceedings of 9th IEEE International Conference on Computer Vision (2003) 1470–1477.
- [19] James Philbin, Ondrej Chum, Michael Isard, et al. Lost in quantization: improving particular object retrieval in large scale image databases, IEEE Conference on Computer Vision and Pattern Recognition (2008)1-8.
- [20] A. Chalechale, G. Naghdy, Alfred Merlins, Sketch-based Image Matching Using Angular Partitioning, IEEE Transaction on Systems, Man and Cybernetics, Part A35(1)(2005) 28-41.
- [21] Serge Belongie, Jitendra Malik, Jan Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence 24(24) (2002) 509-521.
- [22] Chih-Chung Chang, Chih-Jen Lin, LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2(3)(2011)1-27.

- [23] Fumin Shen, Chunhua Shen, Qinfeng Shi, Anton van den Hengel, Zhenmin Tang, Inductive hashing on manifolds, IEEE Conference on Computer Vision and Pattern Recognition (2013) 1562-1569.
- [24] Fumin Shen, Chunhua Shen, Qinfeng Shi, Anton van den Hengel, Zhenmin Tang, Heng Tao Shen, Hashing on nonlinear manifolds, IEEE Transaction on Image Processing 24(6)(2015)1839-1851
- [25] Fumin Shen, Chunhua Shen, Wei Liu, Heng Tao Shen, Supervised Discrete Hashing, arXiv:1503.01557v3 [cs.CV]