Sketch-based Image Retrieval Using Contour Segments

Yuting Zhang [#1], Xueming Qian [*2], Xianglong Tan [#3]

[#] *SMLESLAB of Xi'an Jiaotong University, Xi'an CN710049, China*

[1] zhangyuting@stu.xjtu.edu.cn

[2] qianxm@mail.xjtu.edu.cn

[3] xjtuicemaple@sina.com

***Abstract*—The paper presents a sketch-based image retrieval algorithm. One of the main challenges in sketch-based image retrieval (SBIR) is to measure the similarity between a sketch and an image in contour with high precision. To tackle this problem, we divided the contour of image into two types: the first is global contour, suggesting that we can use it to reduce the similarity between the images with complex background. The second, called salient contour, is helpful to retrieve images with objects similar to the query. Besides, we propose a new descriptor, namely angular radial orientation partitioning (AROP) feature, which makes full use of the gradient orientation information to decrease the gap between sketch and image. Using the two contours as candidate contours for feature extraction could increase the retrieval rate dramatically. Finally an application of retrieval system based on this algorithm is established. The experiment on 0.42 million image dataset shows excellent retrieval performance of the proposed method and comparisons with other algorithms are also given.**

I. INTRODUCTION

Developments in Internet and mobile devices have increased the demand for powerful and efficient image retrieval tools. Content-based image retrieval (CBIR) mainly uses the text or an image as a query. Text features are less accurate and might take mismatch between the user's expression and the user's expectation. Although the image-based search technology develops rapidly and works well, there are some trouble in obtaining relevant images when the user does not have the query images and text. To avoid this problem, the user could draw a sketch and then use the sketch as the input for an image retrieval system, this becomes more and more convenient for users. Sketch-based image retrieval (SBIR) technology becomes an active research area.

SBIR methods use a hand-drawn sketch composed of rough and simple black and white to retrieve the corresponding images. Although SBIR had been studied since 1990s, it still remains challenge to measure the similarity between a sketch and an image with high precision. Image retrieval must deal with the ambiguousness in the query sketch caused by a lack of semantic, besides a large majority of potential users fail to precisely express fine details in their drawings [14, 15, 16]. To improve the precision, many descriptors are proposed. Thus many studies have been focussed on how to choose a good descriptor. Some works focus on global descriptors, but the other works focus on local descriptors. Some researchers design a robust global descriptor to represent the sketch and image individually. Global features can be better used in image analysis, matching, and classification, such as HOG (histogram of gradients) [1], EHD (edge histogram descriptor) [2], and ARP (angular radial partitioning) [3]. However, global features are unsatisfactory as they are unreliable under affine variations. To overcome such drawbacks, Eitz et al. [4], [5] use local descriptors to achieve state-of-art performance. And QVE (query by visual example) [6] is a typical method using blocks and local features. Cao et al. also propose a local feature method, edgel index method [7], for sketch-based image search by converting a shape image to a document-like representation.

In order to develop an image retrieval system which is able to find out more images with objects similar to the query, we develop a global feature based on the global and salient contours. The global contour is a global feature, and is defined to find the relevant image with simple background. The salient contour is local feature, and is defined to tackle the problem that one object is similar to the query. Besides, the AROP feature is refined the ARP feature, and makes full use of contour orientation to constrain the shape information.

The main contributions of this paper are summarized as follows. 1) We propose the global contour, which introduces the salient region to make the contour more discriminative. 2) We propose the salient contour to make the retrieve images with objects similar to the query. 3) The orientation partitioning scheme is introduced based on the original ARP feature. Thus AROP feature contains more information, which makes the retrieval result more accurate and reliable.

The remainder of this paper is organized as follows. Work related to sketch-based retrieval is reviewed in Section II. We describe the proposed approach in Section III, our experiments in Section IV, and the discussion in Section V. Finally, we present our conclusions in Section VI.

II. RELATED WORK

There have been a lot of studies in sketch-based image retrieval system recently and sketch based image retrieval techniques have been well discussed in [18]. In the following, we briefly describe some approaches which are widely used in SBIR system.

2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), Oct 19-21, 2015, Xiamen, China.
978-1-4673-7478-1/15/$31.00 ©2015 IEEE

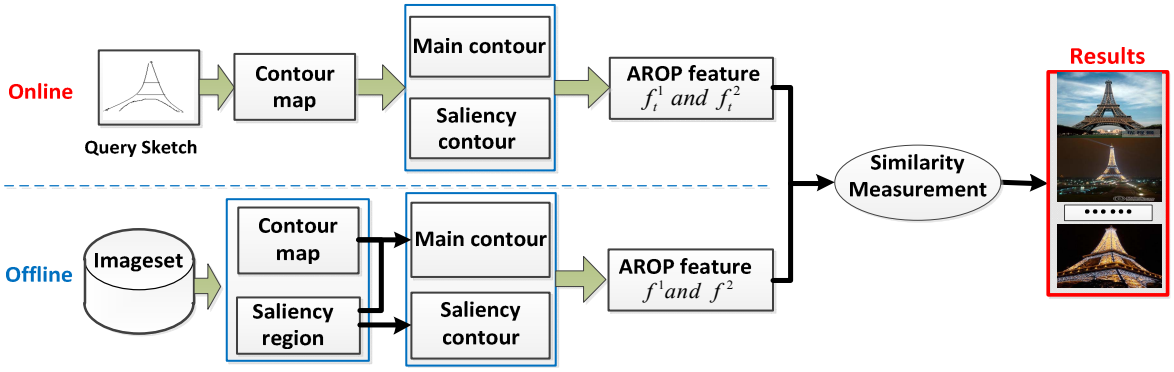


Fig. 1. Illustration for feature extraction.

The edgel index approach is a shape-based indexing method [7]. It solves the shape-to-image matching problem using pixel-level matching. Oriented chamfer matching [8] is used to compute the distance between contours to conveniently build the index structure, Wang et al. [7] used a binary similarity map (a hit map) instead of the distance map [7]. For each input sketch, N hit maps are created, which correspond to the N orientations. They also designed a simple hit function. Specifically, if a point falls in the valid region on a hit map in the same channel, it is considered as one hit. The sum of all the hits is the similarity between the image D (represented by its contours) and the query sketch Q. Then, they build an edgel index structure for fast retrieval.

The ARP method based SBIR approach is first proposed by Chalechale et al. in [3]. In ARP, the edge is firstly extracted by the Canny operator and Gaussian mask, and then the edge is thinned to obtain the abstract image. Finally they define angular partitions in the surrounding circle of the abstract image. Each number of pixels would be counted to form a histogram defined as ARP (Angular Radius Partitioning) feature.

Roman-Rangel et al. [17] propose a shape descriptor Histogram of Orientation Shape Context (HOOSC), which extends the Generalized Shape Context (GSC) [19] using a histogram of orientations and distance-based normalization. Different from the HOOSC feature, which is based on double sides of wide contours, we propose the AROP feature based on two candidate contours. Besides, the type of similarity is different. They use bag-of-shapemes, however, we compute the two AROP feature with weight.

Zhou et al. [9] use the human perception mechanism to identify two types of regions in one image: the first type of region is defined by a weighted center of image features, and it is used to retrieve objects in images regardless of their size and positions. The second type of region is to find the most salient part of an image. They firstly extract orientation features and then organize them in a hierarchal way to generate global-to-local features by a series of hierarchical and overlapping patches. Finally, a hierarchical database index structure is built.

Cheng et al. [10] propose a simple, efficient, naturally multi-scale, and produces full-resolution, high-quality saliency maps. Firstly, they propose histogram-based contrast method to define saliency value for image pixels using color statistics of the input image (i.e., pixels with the same color have the same saliency). And then they use histogram comparison and spatial weight to obtain region contrast (RC). They also introduce *SaliencyCut*, which uses the computed saliency map to assist in automatic salient object segmentation to automate salient region extraction. Used in SBIR system, the author rank the images by SC [11] distance between their salient region outlines and user input sketches based on their *SaliencyCut* algorithm. As a result, their retrieval method is more effective.

Berkeley detector [12] to employ Brightness Gradient (BG), Color Gradient (CG), Texture Gradient (TG), and combine information from these features in an optimal way. Berkeley detector can accurately detect the localize boundaries in natural scenes than that of Canny.

III. The Retrieval System Based on AROP Feature

The framework of the proposed SBIR system is shown in Fig.1. It consists of two parts: the offline part and the online part. In the offline part, we obtain global contour and saliency contour based on contour map use Berkeley detector [12] and saliency region use RC [10]. Then, we extract AROP feature from every contour. In the online part, for a given input query sketch, we extract the global contour and saliency contour based on the contour map. Then we extract the AROP feature for the two contours. Finally, we calculate the similarity between the input and the dataset images.

A. Contour Map and Orientation Map Extraction

The Berkeley detector [12] extracts contours. For an image, we apply the Berkeley detector to each image (resized to 200×200). We will get the true posterior probability (defined as $P(x, y)$) and orientation. When $P(x, y) > th$, we define $B_{th}(x,y)$ as the raw contour map, where th is the threshold value, $B_g(x,y)$ as the raw orientation map and $B_o(x,y)$ as the quantization orientation map.

$$B_{th}(x,y)=\begin{cases}1, & P(x, y) > th \\ 0, & otherwise\end{cases} \quad (1)$$

$$B_o(x,y)=\begin{cases} j, \text{ if } B_{th}(x,y)=1 \,\&\, B_g(x,y)\in\left[\dfrac{(j-1)\pi}{O},\dfrac{j\pi}{O}\right] \\ 0, \qquad \text{otherwise} \end{cases} \quad (2)$$

where j is the orientation channel, O is the number of quantized orientations. $B_o(x,y)$ is obtained by quantitating $B_g(x,y)$ into O orientation channels.

B. Candidate Contours Extraction

We divide the contour of image into two types: the first type is the global contour, suggesting that we can use it to reduce the similarity between the images with complex background. The second type is the salient contour, is helpful to retrieve images with objects similar to the query.

1) Offline Candidate Contours Extraction

We use contour map and the saliency region [10] to obtain the global contour and saliency contour. The global contour is used to present the contour of image accurately. The saliency contour is used to present the main object in the image and is used to find the relevant images containing a common object.

We apply RC [10] to get the saliency region in one image. The saliency map is defined as

$$RC(x,y)=\begin{cases} 1, & \text{if } B_{th}(x,y) \text{ in the saliency region} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

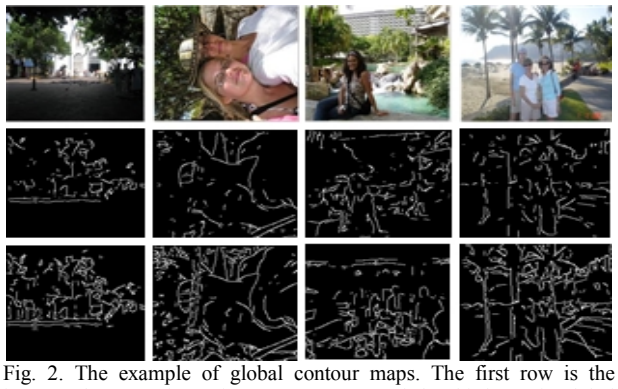


Fig. 2. The example of global contour maps. The first row is the images, the second row is contour maps, and the third row is global contour maps.

(a) Global contour map extraction

As previously mentioned, we obtain the contour map and saliency region in one image. We obtain the saliency map RC from (3) and the contour map from (1). From [7], [12], we can know that when $th=0.5$, the contour map can present the contour of the image and we choose this as the contour map of image. When $th<0.5$, there will be more small edge. When $th>0.5$, there will lose more contour information. In order to remove complex background images, we must introduce more detail edges but not all. So we choose $th=0.3$ for the saliency region. The global contour map is defined as

$$MCM(x,y)=\begin{cases} 1, & \text{if } B_{0.5}(x,y)=1 \,\&\, RC=1 \\ 1, & \text{if } B_{0.3}(x,y)=1 \,\&\, RC=0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

The global contour map is shown in Fig.2. When the background is complex, the global contour map contains more edge pixels in the background region. This will makes the feature value very big. So the global contour has discrimination for complex background images.

(b) Salient contour map extraction

In most cases, users are more concerned about whether the input sketch can be found mainly in the image. So we can obtain the saliency contour map through the following steps.

Firstly, we obtain the saliency region in each image and we use bounding-box to obtain the minimum rectangle of each saliency region.

Secondly, we obtain the saliency image region with main object. That is to say, if there is a saliency region, we make the region as the saliency image region. If the number of saliency region is more than 2, we make the maximum rectangle (defined as *r1*), which contains a maximum of edge pixels, as the main rectangle. Then we find the nearest rectangle (defined as *r*2). Then we make a new rectangle (defined as *r*) witch contain the two rectangles as the saliency image region resized to 100×100.

Thirdly, we extract the saliency image region's contour map using Berkeley detector and make the contour within the saliency region as the contour map $B_{th}(x,y)$. In order to make the object more clearly, we set *th=0.3* and the $B_{0.3}(x,y)$ is the saliency contour map.

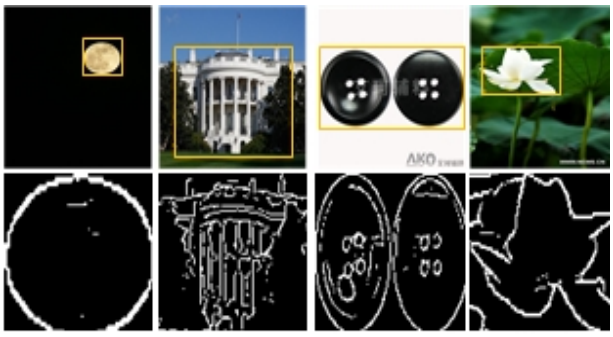


Fig. 3. The example of saliency contour maps. The yellow border is the saliency image region in the first row. The second row is saliency contour maps.

From Fig. 3, we can find the contour map just contain the main object in the image even when the number of object is two. So our saliency contour map can be used to find more relevance images.

2) Online Candidate Contours Extraction

Different from the dataset images processing, we apply Berkeley detector to obtain the input sketch contour map, and then we take the contour map as the global contour map. We choose the image region which is the smallest rectangle that contains the largest pixel value as the saliency region resized to 100×100. Finally, we take the contour map as the saliency contour map.

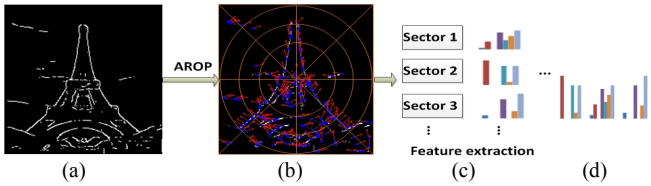


Fig. 4. AROP feature extraction. (a) is the contour map, (b) is the angle, radius and orientation partition (red line is the gradient orientation, which is quantized to 8 directions), (c) is the histogram of every sector and (d) is the AROP feature.

C. AROP Extraction

In order to introduce the AROP feature, we first give brief recommendation for ARP.

1) The Angular Radial Partitioning Feature Extraction

The ARP [3] -based SBIR approach refines the angular partitioning (AP) feature [13] using radial partitioning. The ARP feature is obtained by partition the image into $M \times N$ sectors uses the image center as the center of circles. M is the number of radius partitions and N is the number of angular partitions. The range of each angle $\theta = 2\pi / N$ and the radius of successive concentric circles is $\phi = R / M$ where R is the radius of the surrounding circle of the image [3]. The contour is divided to $N = 8$ angular and $M = 4$ radials. Based on the obtained contour map of the original image, the corresponding edge pixel number in each sector is utilized to represent each sector. Then, for the total $M \times N$ sectors, the final ARP vector is with dimension $M \times N$.

2) Angular Radial Orientation Partitioning Feature

ARP [3] is a coarse representation for the contour image. It just count the number of edge pixels in each sector, and they don't consider the gradient orientation of edge pixel, which is proved to be effective for matching. Thus, in the proposed method, we make full use of the gradient orientation.

Same with the ARP method, we divided the contour map into $M \times N$ sectors, where M is the number of angle partition, N is the number of radius partition. And then we use the number of edge pixel number under in different orientation maps $B_o(x, y)$. That is to say, we represent each sector by an O dimensional orientation vector, as shown in Fig.4 (b). Finally, we cascade the feature of the total O orientation channel to represent the image. By this means, the total dimension of AROP feature is $M \times N \times O$. For each dataset image and the input sketch, we will extract the AROP features based on the two type contour maps.

Compared to the $M \times N$ dimensional ARP feature, the AROP feature can represent more local spatial information. This local spatial information can narrow the scope of match and also enhance accuracy rate. In a word, AROP captures certain local spatial information of the image.

D. AROP Feature Matching

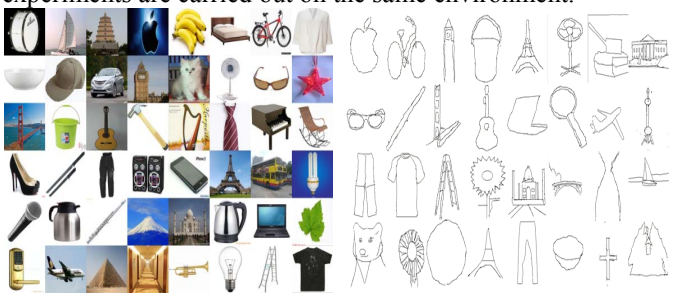
In the offline, we extract dataset images' AROP features and we define them as $f_t^k, k = 1,2$ *and* $t = 1,2,\cdots,T$, where T is the number of dataset images, k denotes in the case of different contours. f_t^1 is the AROP features based on the global contour map and f_t^2 is the AROP features based on saliency contour map. In the online, we extract AROP feature of input query sketch and define it as f^k. Let $sim1(t)$ and $sim2(t)$ denote the similarity of the global contours f_t^1 *and* f^1 and the similarity of the salient contours f_t^2 *and* f^2. Then we can measure the similarity of the input query sketch and the dataset images by taking the global contour similarity and salient contour similarity as follows:

$$sim(t) = w \times sim1(t) + (1-w) \times sim2(t), \quad t-1,2,\ldots,T \tag{5}$$

Finally, we sort the similarity scores in descending order, and determine the results (the top-n ranked).

IV. Experiments and Discussion

In order to show the effectiveness of the proposed approach, we compare our algorithm AROP with the method Edgel [7], the method ARP [3], the method in [9] and the proposed method on our crawled dataset. Besides, we extract the AROP feature on the contour map corresponding to RC region and use this method (named RC-AROP) as a compare. All experiments are carried out on the same environment.



(a). Example of dataset image. (b). Example of sketch

Fig. 5. The example of dataset image and the input sketch.

A. Dataset

The experiment dataset consists of 433,790 images, and the storage cost is 130 GB. Our dataset is crawled from Google using key words. We select 81 topics that can be easily described by sketches, so that there could be sufficient similar images for a query sketch. And there are approximately 1,000 images in each topic. This dataset also contains the GOLD set [20, 21], which mainly contains landmarks and landscapes. Topics mainly include living goods, fruits, animals, and some landmarks easy to be described by sketches. Some examples are shown in Fig.5 (a).

We draw 162 query sketches which cover most of the 81 topics in the Sketch-describable Dataset, and then set them as queries to sketch retrieval systems. Some of them are shown in Fig.5(b). In the following parts of this paper, experiments used the 162 sketches.

B. Performance Evaluation

We used the precision under depth n (denoted as *Precion@n*) to measure the objective performance, defined as

$$Precion@n = \frac{1}{Z}\sum_{m=1}^{Z}\frac{1}{n}\sum_{i=1}^{n}R_m(i) \tag{6}$$

where $R_m(i)$ is the relevance of the *i-th* result for query *m*, $i \in [1,2,\ldots,n]$, and $m \in [1,2,\ldots,Z]$. *Z*=162 for our dataset. If it is relevant to the query sketch then $R_m(i)=1$, otherwise $R_m(i)=0$.

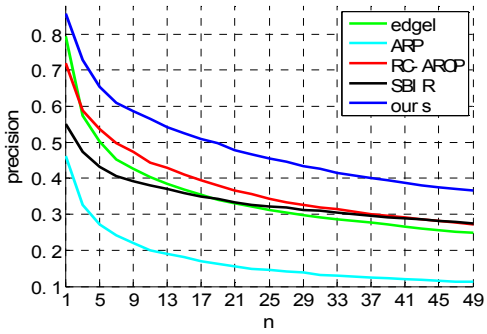


Fig. 6. Precision comparison with other methods. Edgel is the method in [7], ARP is the method in [3], RC-AROP is the AROP feature based on RC [10] method and SBIR is the method in [9].

C. Objective Comparisons

Correspondingly, the *Precision@n* curves of other methods and the proposed method with the depth varying in the range [1, 50] are shown in Fig.6. The curve is drawn by the average results of 162 queries on our database. For fair comparison, the parameters M and N are set to be 8 and 4 respectively for ARP and AROP. The partition of radius is uniform. The orientation channel (O) of AROP and edgel methods are both set to be 8 and in our method, $w=0.8$ (the weight of the similarity).

In our objective comparison, we find that the proposed algorithm improve 10% than the other methods in the top 10 results. For $n=1$, our method is also more accurate than the edgel method and the RC-AROP method. The proposed method makes the image comprising a common object more similarity and irrelevant image more different.

These experiments were implemented using Matlab. The average computational cost of the proposed method is 1.423 s. When the orientation channel *O* is increase by *O'*, the dimension of AROP feature will increase by $M \times N \times O'$. This can increase the computational complexity. However, the increased time can be ignored.

D. Discussions

We now discuss the impacts of the parameters on the performance of our sketch-based retrieval system. In the proposed method, the parameter *w* in (5) is used to compute the score of similarity. We set $w=0.8$ in our baseline experiments. This parameter determines the contributions of the global contour map and the saliency contour map. Accordingly, *w* should range between 0 and 1. When $w=0$, it means no AROP feature of global contour map. When $w=1$, it means no AROP feature of saliency contour map. As shown in Fig.7, the method performed best when *w* was approximately 0.7. From Fig.7, we find that the global contour map and the saliency contour are more important to the final performance for the following reasons.

1) The global contour map contains more edge information and presents the content of the image clearly. Besides it makes the image with complex background have more edge information. So we can use the global contour map to filter the image with complex background.

2) The saliency contour map contains the main object in the image. This information can make the image with common object more similarity.

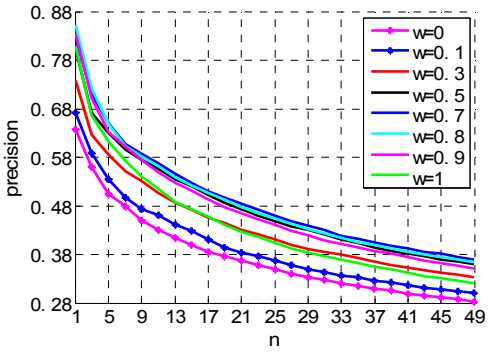


Fig. 7. Precision@n curves for various *w*.

E. Subjective Comparisons

In Fig.8, compared to the SBIR [9], edgel method and the proposed methods use three input sketches. Fig. 8 shows the retrieval results of the method in [9] (the first row), the edgel method (the second row) and our method (the third row). As shown in Fig. 8 (a), our top 10 results were all correct and the other methods returned several irrelevant images. In Fig. 8 (b, c), the edgel method and the method in [9] returned more irrelevant images, but our top five results were all correct. Fig. 8 shows that our results also contained some incorrect images, but they are all similar in shape to the queries, and the results were better than those of the other methods.

V. Conclusion

We have introduced a novel approach for image representation based on contour segments. Considering that false matches could degrade retrieval performance, we propose the global contour map and the saliency contour map. The proposed method can find the image with simple background and find the image with the common object. This is very important for the chaotic dataset. The experimental results show that the AROP features have certain advantages over the other methods in retrieval precision. Various experiments proved that sketch retrieval algorithm is beyond the other methods.

Acknowledgment

This work is supported in part by the Program 973 No.2012CB316400, by NSFC No.60903121, 61173109, 61332018, and Microsoft Research Asia.

References

[1] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "A descriptor for large scale image retrieval based on sketched feature lines," *Sketch Based Interfaces and Modeling*, 2009, pp. 29-36.

[2] P. Salembier, T. Sikora and B. S. Manjunath. "Introduction to MPEG-7: multimedia content description interface," John Wiley & Sons, Inc., 2002.

[3] A. Chalechale, G. Naghdy, and A. Mertins, "Edge image description using angular radial partitioning," *IEEE Proceedings-Vision, Image and Signal Processing,* vol. 151(2): 93–101, April, 2004.

[4] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," *IEEE Trans. Visualization and Computer Graphics*, 2011, pp.1624-1636.

[5] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "An evaluation of descriptors for large-scale image retrieval from sketched feature lines," *IEEE Trans. Visualization and Computer Graphics*, vol. 34, no. 5, pp. 482-498, 2010.

[6] K. Hirat, and T. Kato, "Query by visual example," in *EDBT'92*. Springer Berlin Heidelberg, 1992, pp. 56-71.

[7] Y. Cao, C. Wang, L. Zhang, and L. Zhang, "edgel index for large-scale sketch-based image search," *IEEE CVPR,* 2011, pp. 761-768.

[8] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla. "Model-based hand tracking using a hierarchical bayesian filter," *IEEE TPAMI*, vol. 28, no. 9, pp. 1372-1384, 2006.

[9] R. Zhou, L. Chen, and L. Zhang, "Sketch-based image retrieval on a large scale database," *ACM MM*, 2012, pp. 973-976.

[10] M. Cheng, N. Mitra, X. Huang, P. Torr, and S. Hu, "Global Contrast Based Salient Region Detection," *IEEE TPAMI.*, vol. 37, no. 3, pp. 569-582, 2014.

[11] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE* Trans. *Pattern Anal. Mach. Intell.*, 2002, pp. 509-522.

[12] D. R. Martin, C. C. Fowlkes, and J. Malik. "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE TPAMI.*, vol. 26, no. 5, pp. 530-549, 2004.

[13] A. Chalechale, and A. Mertins, "Sketch-based image matching using angular partitioning," *IEEE TSMCA*, vol.35, no. 1, pp. 28-41, 2005.

[14] T. Chen, M. Cheng, P. Tan, A. Shamir, and S. Hu, "Sketch2Photo: internet image montage," *ACM Trans. Graph.*, vol.28, no. 124, article 124, 2009.

[15] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects?" *ACM Trans. Graph*, vol. 31, no. 4, article. 44, 2012.

[16] S. Ren, C. Jin, C. Sun, and Y. Zhang, "Sketch-Based Image Retrieval via Adaptive Weighting," *ACM ICMR*, 2014, pp. 427-430.

[17] E. Roman-Rangel, C. Pallan, J. Odobez, and D. Gatica-Perez. "Analyzing Ancient Maya Glyph Collections with Contextual Shape Descriptors," *IJCV* 2011.

[18] Y. Rui, T. S. Huang, and S. Chang, "Image Retrieval: current techniques, promising directions, and open issues," *JVCIR*, 1999.

[19] Mori, G., Belongie, S., & Malik, J. (2005). Efficient shape matchingusing shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11), 1832–1837.

[20] J. Li, X. Qian, Y. Tang, L. Yang, and T. Mei, "GPS estimation for places of interest from social users' uploaded photos," *IEEE Trans. Multimedia* 2013, vol.15, no.8,pp.2058–2071.

[21] J. Li, X. Qian, Y. Tang, L. Yang, and C. Liu, "GPS estimation from users' photos," *in Proc. MMM* 2013, pp.118–129.

(a)

(b)

(c)

Fig.8. Sketch retrieval performances, the first row contains the top- ranked result using proposed method in [9], and the second row contains the results using edgel method, and the third row contains the results using the proposed method