Sketch-Based Image Retrieval by Salient Contour Reinforcement

Yuting Zhang, Xueming Qian, *Member, IEEE*, Xianglong Tan, Junwei Han, *Senior Member, IEEE*, and Yuanyan Tang, *Fellow, IEEE*

***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. To tackle this problem, we propose an SBIR-based approach by salient contour reinforcement. In our approach, we divide the image contour into two types. The first is the global contour map. The second, called the salient contour map, is helpful to find out the object in images similar to the query. In addition, based on the two contour maps, we propose a new descriptor, namely an angular radial orientation partitioning (AROP) feature. It fully utilizes the edge pixels' orientation information in contour maps to identify the spatial relationships. Our AROP feature based on the two candidate contour maps is both efficient and effective to discover false matches of local features between sketches and images, and can greatly improve the retrieval performance. The application of the retrieval system based on this algorithm is established. The experiments on the image dataset with 0.3 million images show the effectiveness of the proposed method and comparisons with other algorithms are also given. Compared to baseline performance, the proposed method achieves 10% higher precision in top 5.**

***Index Terms*—Contour matching, image retrieval, salient contour, sketch based image retrieval (SBIR).**

I. Introduction

DEVELOPMENTS in internet and mobile devices have increased the demand for powerful and efficient image retrieval tools. How to find the images we need from large scale datasets? In order to answer this question, content-based image retrieval (CBIR) develops rapidly and works well. The existing CBIR systems [2], [31], [32], [40] still mainly use keywords or images as the query, which still can't fulfill all user demands. It is often difficult to precisely describe the content of the desired images using plain text. When the user does not have the natural scene images and accurate textual descriptions to justify his/her search intention, there are some difficulties in obtaining relevant images. To avoid these problems, the sketch-based image retrieval (SBIR) system is generated. This system is more convenient for users, because the end user could simply draw a sketch and then use the sketch as the input for effective image retrieval. For this reason, SBIR technology has become an active research area [3], [7], [9], [24].

SBIR methods often use a hand-drawn sketch composed of rough black and white to retrieve the relevant images. Neuroscience research shows that the human brain is good at recognizing different objects based on the contour or shape [19]–[23]. In real life, we can draw a contour or a shape to represent the relevant image. In addition, sometimes a sketch produced by a user will look rough because of poor drawing skills or limited time for drawing. So it is possible that the sketches are different from natural scene images in some aspects. Considering this factor, an effective SBIR systems must be able to deal with the ambiguousness existed between the simple stroke and the natural scene images. This brings great challenges for solving the problem. Actually, in addition to the SBIR technology, many other edges related tasks, such as sketch classification [28]–[30], sketch recognition [33]–[36], and real-time user guide for freehand drawing [37]–[39] are also extensively studied.

Traditional draw-and-search systems [17], [18] require that the input sketch contains color information and looks similar to a natural scene image. Although the research of these systems has achieved great progress, it is still not based on line drawings. Supported input is actually relatively thick colored lines block. In addition, this approach converts SBIR technology to CBIR technology. The user has to draw the sketch carefully and provide colors to make the sketch visually similar to the natural scene images. Then, CBIR fuses different features (such as shape, color, and texture) together to perform a retrieval system. However, this method brings a lot of burdens to the user by requiring detailed drawings, and more importantly, it does not solve the core problem of SBIR, i.e., matching a line-formed sketch with images.

The critical problem in SBIR is to measure the relevance between an image and a query sketch. Intuitively, we can carry out sketch based retrieval by measuring the contour similarity for the input query sketch and image [4], [5], [24]. So the core problem of SBIR is converted to contour (or shape) matching. However, there is a huge gap between computer's shape recognition and the human brain's shape matching. Shape plays an

Manuscript received July 31, 2015; revised December 16, 2015 and February 29, 2016; accepted May 07, 2016. Date of publication May 12, 2016; date of current version July 15, 2016. This work was supported in part by the Program 973 under Grant 2012CB316400, in part by the National Natural Science Foundation of China under Grant 61373121, Grant 61173109, and Grant 61332018, and in part by Microsoft Research Asia. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Balakrishnan Prabhakaran. *(Corresponding author: Xueming Qian.)*

Y. Zhang and X. Tan are with the SMILES Laboratory, School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: zhangyuting@stu.xjtu.edu.cn; xjtuicemaple@sohu.com).

X. Qian was with the Ministry of Education Key Laboratory for Intelligent Networks and Network Security. He is now with the SMILES Laboratory, School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: qianxm@mail.xjtu.edu.cn).

J. Han is with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: jhan@nwpu.edu.cn).

Y. Tang is with the University of Macau, Taipa, China (e-mail: yytang@umac.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2016.2568138

important role in computer vision and image processing, and it has been widely used as a basic representation for object detection and recognition in images [41], [56]–[58]. To describe the shape information of the sketch, many descriptors are proposed. A shape descriptor, called shape contexts [11], consisting of a radial histogram of the relative coordinates has been applied to shape matching. To bridge the representational gap, many methods generate some intermediate descriptors, e.g., edge histogram [5], [25], [54], or directly extract representative contours, e.g., Canny edge [26], from natural scene images in the database. Other research efforts have been spent on the descriptor extraction [4]–[7].

When extracting the descriptor, some works focus on global descriptors, and other focus on local descriptors. Global features [1]–[3], [53] can be better used in image analysis, matching, and classification. However, global features are unreliable to deal with local affine variations. To overcome such drawback, many local descriptors are proposed [4], [5].

In a sketch-based image retrieval system, users care more about search precision than recall. The main directions in sketch retrieval are sketch querying [42]–[45] and fusion sketch and tag [27], [46], [47] querying. To find more relevant images in SBIR system, we need to overcome the following two challenges: representing and matching. Contour based retrieval is important because contour provides information about image structure and texture. In addition, orientation has been exploited widely in the computer vision community [24], [39], [48]–[52]. Orientation shows outperforming results in tasks such as object recognition and object categorization. In general, all these stages, including image segmentation, contour extraction, and image saliency, are difficult and extensively researched. Here, we do not claim to solve these challenging problems in general.

In this paper, we propose a SBIR based approach using the extracted salient contour feature, which makes full use of contour and orientation to improve the accuracy. To solve the representation gap, we use Berkeley detector [12] to extract the images contour maps and propose two contour maps (the global contour map and the salient contour map). The global contour map is defined to find the relevant image with a simple background. The salient contour map is defined to tackle the problem that an object is similar to the query. In order to solve the matching problem, we propose a novel angular radial orientation partitioning (AROP) feature. It is an enhanced angular radial partitioning (ARP) feature [3]. With the reinforcement of the salient contour, this feature uses contour and orientation to constrain the spatial information.

The main contributions of this paper are summarized as follows.

1) We propose a global contour map (GCM) to describe the background contour. The GCM has impact to the main object contour by adjusting its weight. Compared to the contour map, the proposed GCM is more reliable and better suited to SBIR. It demonstrates superior performance and is helpful for reducing the impact of complex background. Our experiments demonstrate that the proposed GCM plays a key role in significantly improving retrieval performance.
2) We propose a salient contour map (SCM) to increase the impact of the main object contour, corresponding to the object contour of all extracted edges. Our experiment demonstrates that SCM is helpful to filter out false matches and alleviate the impact of noisy edges.
3) Based on our candidate contour maps, we propose a new global descriptor, AROP feature, to describe the spatial information of contours within each block. Compared to image patch-based descriptor, our AROP feature contains more information, which makes the retrieval result more accurate and reliable.

Compared with our preliminary work [55], several enhancements have been made in this paper. We summarize them as the follows: 1) we enhance global and salient feature for feature extraction by reinforcing the saliency map to improve their robustness in SBIR; 2) we propose an effective approach to rank the feature, based on which we can reduce the computation time; and 3) more extensive experiments and comparisons are conducted.

The remainder of this paper is organized as follows. Works related to sketch-based retrieval are reviewed in Section II. We describe the proposed approach in Section III, our experiments are presented in Section IV, and the discussion is given 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. In the following, we briefly describe some approaches which are widely used in SBIR systems.

The query by visual example (QVE) is one of the earliest approaches in SBIR [6]. In this approach, they resize the query and the database images to $64 * 64$ pixels and then they use the proposed gradient operator to extract edges. The correlation is calculated by shifting these blocks. The edge histogram descriptor [2] and the histogram of oriented gradients (HoG) [1] are also used to establish the SBIR system [14]. They are both global features extracted from the edges of images. Eitz *et al.* [4], [5] use local descriptors to achieve state-of-the-art retrieval precision. And QVE [6] is a typical method by using blocks and local features. Cao *et al*. 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 our previous work [24], we propose a SBIR approach with re-ranking and relevance feedback schemes. We make full use of the semantics in query sketches and the top ranked images of the initial results to improve retrieval performance. We further apply relevance feedback to find more relevant images for the input query sketch. The integration of the re-ranking and relevance feedback results in mutual benefits and improves the performance of SBIR.

Chen *et al*. [46] present a system that composes a realistic picture from a simple freehand sketch annotated with text labels. First, they use the text label to search the relevance item and background. And then, they choose candidate images for each scene item and background. During filtering, each image is segmented to find the scenic elements corresponding to the

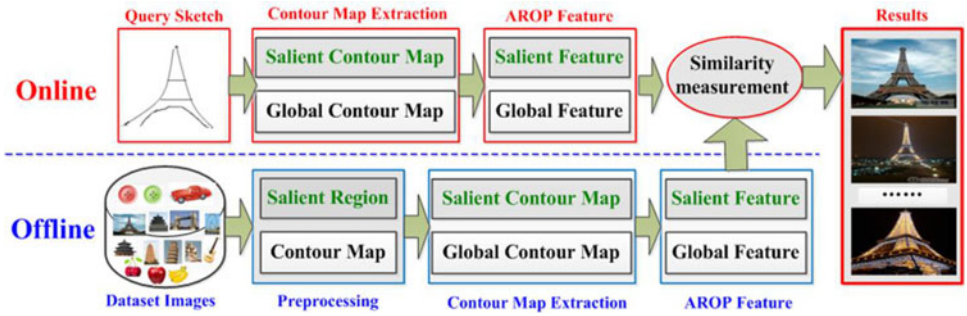


Fig. 1. Framework of our system.

sketch. Finally, they optimize the combination of filtered images into a picture by two steps. First, they optimize the blending boundary and assign each pixel within the boundary a set, indicating whether the texture and color of that pixel are consistent or not. Second, they compute the blending result by combining improved position blending and alpha blending.

The edgel index approach is a shape-based indexing method [7]. It solves the shape-to-image matching problem using pixel-level matching. Its Mind Finder system [7], [27] is a real-time image retrieval system. Oriented chamfer matching is used to compute the distance between contours to conveniently build the index structure [8]. Cao *et al.* use a binary similarity map (a hit map) instead of the distance map [7]. For each input sketch, a set of hit maps is created, which correspond to the number of orientations. They also design 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 a database image D (represents the contours of an image) and the query sketch Q. Then, they build an edgel index structure for fast retrieval, which records the value of position and channel of the hit map. And last they use two-way matching. The one-way $D \rightarrow Q$ often leads to trivial results. Unsatisfactory results could be filtered out by combining the opposite direction matching $Q \rightarrow D$. Then, they multiply these two similarity scores to obtain a final score that reduces the influence of trivial results.

The ARP method based SBIR approach is first proposed in [3]. It refines the angular partitioning feature [13] using radial partitioning. 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. The ARP feature is obtained by partitioning the image into $M \times N$ sectors, which uses the image center as the center of circles. N is the number of radius partitions and M is the number of angular partitions. The range of each angle is $\theta = 2\pi/M$ and the radius of successive concentric circles is $\varnothing = R/N$, where R is the radius of the surrounding circle of the image [3]. The contour is divided to $M = 8$ angulars and $N = 4$ 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 feature is with dimension $M \times N$.

Recent work [9] also utilizes two candidate maps based on the main region and the region of interest. However, there are significant differences between their method and our method. 1) The motivation is different. The proposed method focuses on alleviating the impact of background and makes full use of the salient contour, while their method focuses on localizing a region to improve performance under affine transformation. 2) The descriptor design is different. The proposed descriptor (AROP feature) captures the orientation and spatially relationships between local contour and all its neighbouring contours in every sector [9], while their descriptor only captures the relationship between two overlapping neighbouring. 3) The matching strategy is different. In the proposed work, matching is used to find the similar images in global and parts. However the work in [9] formulates matching using the first components of the feature as an index structure, with the result that key information might be overlooked.

Cheng *et al*. [10] propose a salient object detection algorithm which can be utilized to improve sketch retrieval performance. First, they propose a histogram-based contrast method to define salient value for each pixel using color statistics of the input image. Pixels with the same color have the same saliency. And then, they simultaneously evaluate global contrast differences and spatial weighted coherence score to obtain regional contrast (RC). They also introduce SaliencyCut based on automatic salient region extraction, which uses the computed saliency map to assist in automatic salient region extraction. The regional contrast is used in SBIR system. They rank the image by shape contexts [11] distances between their salient region outlines and user input sketch based on their SaliencyCut algorithm. As a result, their retrieval method is more effective.

III. The Proposed Approach

The framework of the proposed SBIR system is shown in Fig. 1, which consists of the offline part and the online part. In the offline part, for the dataset images, we sequentially execute three steps: 1) we first carry out preprocessing for dataset images to extract image salient regions and contour maps by RC [10] and Berkeley detector [12] respectively; 2) we use salient contour reinforcement method to extract candidate contour maps containing the global contour map and the salient contour map;

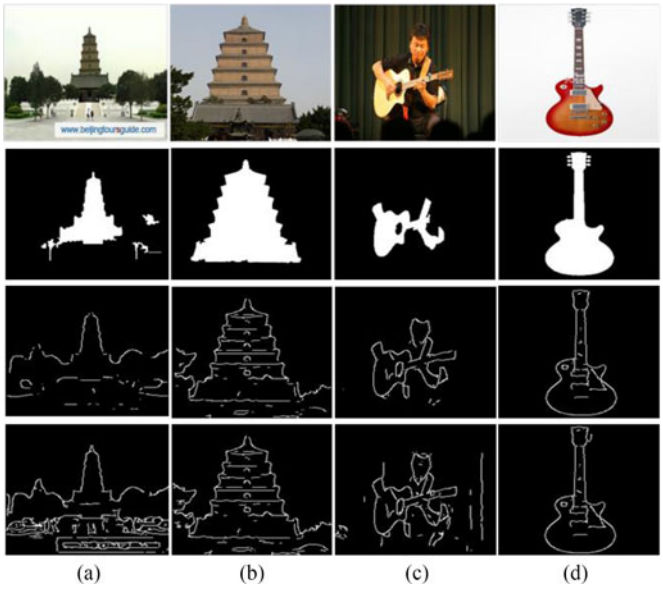


Fig. 2. Example of global contour maps. The first row is the natural images. The second row is the salient regions of the images. The third row is contour maps. The fourth row is GCM (x, y).

3) we extract AROP features based on the candidate maps named global features and salient features.

In the online part, for a given input query sketch, we extract the global contour map and the salient contour map based on the contour map. Then, similar to the offline, we extract AROP feature based on the two candidate maps.

After we extract these two types of features, we measure the similarity between the query sketch features and the dataset image features. Finally, we sort the similarity score to get the result. In the following sub-sections, we firstly introduce the offline system and then the online system step by step.

A. Salient Region and Contour Map Extraction

In our offline system, we use RC [10] method to extract the saliency map for each dataset image, and we use the Berkeley detector to extract contour map [12].

1) Salient Region Extraction: We apply region-based contrast (RC) method in [10] to get the RC saliency map. Cheng *et al*., initialize a segmentation obtained by binarizing the RC saliency map using a fixed threshold T_b. And then, the largest connected region is considered as the initial candidate region of the most dominant salient object. The other regions are labeled as background. In this paper, we utilize the suggested parameter $T_b = 70$ according to [10] to get the RC saliency map. The saliency map for an image is defined as follows:

$$SR\left(x,y\right) = \begin{cases} 1, & \text{if } T_b > 70 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $SR(x, y) = 1$ denotes the pixel (x, y) belonging to the salient region, otherwise belonging to the background region.

For the input natural scene images as showed in the first row of Fig. 2, the corresponding salient regions are shown in the second row. From Fig. 2, we can know that the RC method can extract the main object or scene well.

2) Contour Map Extraction: Common edge detection algorithm cannot meet the requirements for better generating contours that vividly reflect objects' shapes and suppress the influence by textures and noise in images. Therefore, researchers often adopt the Berkeley detector [12] to extract object contours. In the proposed method, we use Berkeley detector to extract contour map and edge's gradient.

For an image, we apply the Berkeley detector to each image (resize to 200×200) as that utilized in our previous work [24]. Thus we will get the true posterior probability (defined as $p(x, y)$) and orientation at every potential edge. We define $B_{\text{th}}(x, y)$ as the raw contour map under the cut-off threshold *th*

$$B_{\text{th}}\left(x,y\right) = \begin{cases} 1, & p\left(x,y\right) > \text{th} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

We define $B_g(x, y)$ as the raw orientation map and $B_o(x, y)$ as the quantized orientation map

$$B_o\left(x,y\right) \begin{cases} j, & \text{if } B_{\text{th}}\left(x,y\right) = 1 \text{ and } B_g\left(x,y\right) \in \left[\frac{(j-1)\pi}{O}, \frac{j\pi}{O}\right] \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

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

From the true posterior probability $p(x, y)$, we can know that when th < 0.5, there will be more edges in the contour map. However, the increased edges mostly are contributed by a complex background. When th > 0.5, there will be less contour information, which may not represent the object contours. In [7], [12], the authors choose $B_{0.5}(x, y)$ as the image contour map. Similarly, we set CM $(x, y) = B_{0.5}(x, y)$ and mark CM(x, y) as the contour map. From the CM $(x, y) = B_{0.5}(x, y)$,. the contour map can present the image contour accurately as shown in the third row in Fig. 2.

B. Global Contour Map and Salient Contour Map Extraction

We divide the contour of an image into two types: the salient contour map and the global contour map. The global contour map aims at alleviating the impact of background. Salient contour map aims at finding the similar object between the sketch and the corresponding image.

1) Global Contour Map Extraction: When users use the SBIR system, they mostly concentrate on finding the object in the sketch. In order to meet user requirements, the retrieval result of SBIR system should be the images that contain the same object or scene. In addition, they should be as simple as possible. For these purposes, we extract global contour map by reinforcing salient contour map.

In order to distinguish the complex background images, we should separate the complicated background from a contour map by decreasing the background's threshold *th*. In order to represent the contour of the main object (or scene) accurately, we extract the global contour map by salient contour reinforcement. We choose th < 0.5 to obtain the background contour map. In

this paper, we set th $= 0.3$, and we will discuss it in Section V. The global contour map is defined as follows:

$$\begin{aligned}\mathrm{GCM}\ (x,y) \ = \ & \mathrm{CM}\,(x,y) \times \mathrm{SR}\,(x,y) + B_{0.3}\,(x,y) \\ & \times (1 - \mathrm{SR}\,(x,y)) \end{aligned} \tag{4}$$

where $\mathrm{CM}(x,y)$ is the contour map, $B_{0.3}(x,y)$ is the contour map under th by (2), and $\mathrm{SR}(x,y)$ is the saliency map by (1).

The global contour maps are shown in the fourth row of Fig. 2. From Fig. 2(c), we can know that the contour map $\mathrm{CM}(x,y)$ represents the contour of the object. From Fig. 2 (d), we can find that the background contour map contains more detailed edges when the background is complex. From Fig. 2(c) and (d), we can know that the edges in background contour map will not increase when the background is simple. When the background is complex, the global contour map contains more edge pixels in the background region. In a word, the global contour map can be used to find out the images that contain the same object and scene.

2) Salient Contour Map Extraction: In fact, when we see an image, we usually look through the whole image for a short while, and then, focus our eyes on the salient part. In most cases, we are more concerned about whether the input sketch can be found in the image. In order to find out the images that contain the same object or scene, we propose to utilize the salient contour map (SCM) to refine the result. The reason is that the SCM mainly contains the contour of objects or scene. But we don't directly use the SaliencyCut algorithm (RCC) and RC method in [10]. In RC method, there will be more salient regions in an image, while only part of them belongs to object. In RCC algorithm, the refined region will not contain all part of the object. We obtain the salient contour map through the following steps.

First, we have obtained the RC saliency map from Eq.(1) in each image and we use bounding-box to obtain the minimum rectangle of each RC saliency map as shown in the second row of Fig. 3.

Second, we refine the saliency map and get the candidate rectangle. We filter some small rectangles as shown in the second row of Fig. 3(b) and (d), and get the refined saliency map (RSM). If there is just one connected region, we make the connected region as the candidate rectangle as shown in the third row of Fig. 3(a) and (d). If the number of connected region is more than two, we merge them and obtain their bounding box and view it as the candidate rectangle as shown in the third row of Fig. 3(b) and (c).

Third, in image candidate rectangle, we extract the salient contour map $\mathrm{SCM}(x,y)$ from the contour map $\mathrm{CM}(x,y)$. The salient contour map is defined as follows:

$$\mathrm{SCM}\,(x,y) = \mathrm{SR}\,(x,y) \times \mathrm{CM}\,(x,y) \tag{5}$$

where $\mathrm{CM}(x,y)$ is the contour map and $\mathrm{SR}(x,y)$ is the refined saliency map.

In the first row of Fig. 3, the yellow rectangles are the candidate rectangles by refining the saliency map. We can know that the refined saliency map can represent the object or scene in an image. From Fig. 3(c), we observe that the contour map contains the main object in the image even when the number of objects is two. From the third row in Fig. 3, we can know the salient contour map contains the object or scene. In a word, the salient contour map can be utilized to find similar objects in sketch based image retrieval.

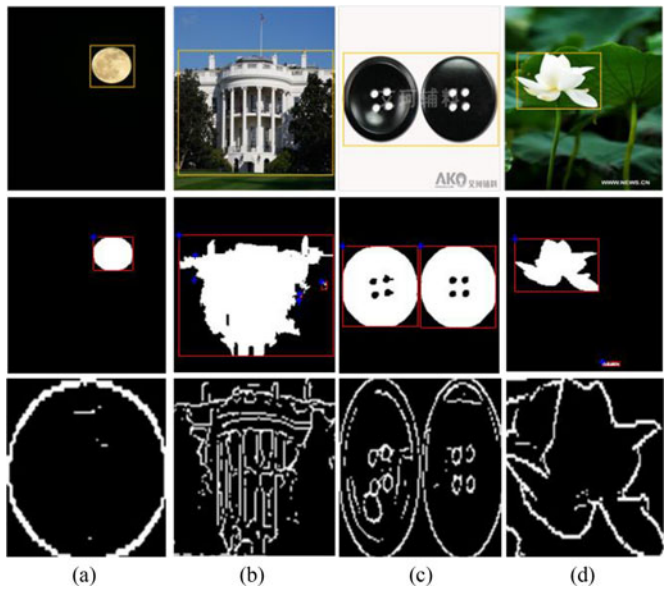


Fig. 3. Example of saliency maps. The first row is the image. The second row is the minimum rectangle (marked in red) of each saliency map. The third row is saliency contour map $\mathrm{CSM}(\mathrm{x},\mathrm{y})$, corresponding to the yellow border in the first row.

C. AROP Feature Extraction

AROP feature is an enhanced ARP feature [3]. ARP [3] is a coarse representation for the contour image. It just statistics the number of edge pixels in each sector, and they do not consider the edge pixel gradient orientation, which has been proved to be effective for matching [1], [51]. Thus, in the proposed method, we make full use of the gradient orientation.

In angular radial orientation partition, there are two methods for radius partition. So, we first introduce the method of radius partition and then we introduce the AROP feature extraction.

1) Radius Partition: In [3], there are generally two types of radial partitioning: uniform and non-uniform. Non-uniform method usually adopts the square root (in short sqrt) solution. For the kth radial partition, its radius is defined as $\rho_k = \sqrt{k}R/\sqrt{N},\ k = 1, 2, \ldots, N$, while k is the number of the slice from the inside out, R is the biggest radius, and N is the number of radius partition. Uniform method divides the radial partitions by equal radius. The uniform is defined as $\rho_k = kR/N$. In our method, we use the uniform partition and the reason will be discussed in the Section V.

2) AROP Feature: Similar to the ARP method, we divide the candidate contour maps into $M \times N$ sectors, where M is the number of angle partition and N is the number of radius partition. And then we count the number of edge pixel under different orientation maps $B_o(x,y)$ as shown in Fig. 4(b). That is to say, we represent each sector by an O dimensional orientation

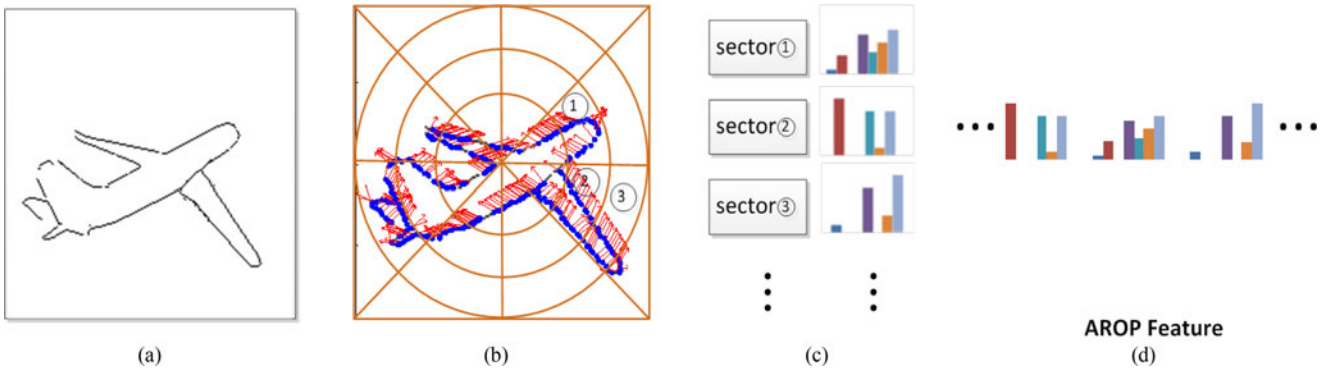


Fig. 4. AROP feature extraction. (a) is the contour map, (b) is the angle, radius and orientation partition (red line is the gradient orientation, which is quantized to 8 directions), (c) is the histogram of every sector, and (d) is the AROP feature.

vector, as shown in Fig. 4(c). Finally, we cascade the feature of the total O orientation channels to represent the image as shown in Fig. 4(d). By this means, the total dimension of AROP feature is $M \times N \times O$.

For each dataset image, we will extract the AROP features for the two types of contour maps: global contour map and the salient contour map. The AROP features are named as global AROP feature and salient AROP feature, respectively.

The global AROP feature is comparatively sparse and its contribution in feature matching is low. When we make statistics of the global AROP feature, the background region's AROP feature value will be much bigger, which can make the image more distinctive. The salient AROP feature can be used to find the images that contain the similar object with the input sketch.

For a dataset image, we define the global AROP feature as $f_t^1, t = 1, 2, \ldots, T$, where T is the number of dataset images. Similarly, we define the salient AROP feature as f_t^2.

Compared to the $M \times N$ dimensional ARP feature, the AROP feature contains much more local spatial information. This local spatial information can narrow the scope of the match and enhance the accuracy rate.

D. Feature Matching With Salient Contour Reinforcement

Now, we introduce the online part. For a query sketch, we first extract global and salient features as described in offline part. Then, we compute the similarity for the query sketch and the dataset image features according to their AROP features.

1) Feature Extraction for Query Sketch: As the fact that the query sketch only composes of clear lines on a clean background, we set the contour map $\mathrm{CM}(x, y)$ as the global contour map $\mathrm{GCM}q(x, y)$, i.e., $\mathrm{GCM}_q(x, y) = \mathrm{CM}(x, y)$. We choose the main object region which is the smallest rectangle that contains the largest pixel value. We obtain the salient contour map $SCM_q(x, y)$ for the sketch by detecting its bounding box of the query sketch.

We get the corresponding AROP features for the global contour f_q^1 and the salient contour f_q^2 for the query sketch q.

2) Similarity Measurement: In our offline and online systems, we represent each image by AROP feature. So, we can use Euclidean distance to measure their similarities. For the query sketch, let us denote the global contour similarity of query sketch and a dataset image t as $S_g(t)$ and salient contour similarity as $S_s(t)$

$$\begin{cases} S_g\,(t) = \|f_t^1 - f_q^1\|_2 \\ S_s\,(t) = \|f_t^2 - f_q^2\|_2 \end{cases} \tag{6}$$

where $\|\cdot\|_2$ denotes the Euclidean distance of two vectors.

The similarity between the query sketch and the dataset images can be determined by summing the weighted scores of $S_g(t)$ and $S_s(t)$ as follows:

$$S\,(t) = w \times S_g\,(t) + (1 - w) \times S_s\,(t) \tag{7}$$

where w is the weight, $w \epsilon [0, 1]$ and $t = 1, 2, \ldots, T$. In our experiments, we set $w = 0.8$.

Computing the $S_g(t)$ and $S_s(t)$ between the query sketch and all the T dataset images is time-consuming. So, in this paper, a fast salient contour reinforcement approach is utilized, which is expressed as follows.

1) We calculate the $S_g(t)$ and rank them in descending order and select the top L images, where $L \ll T$. In our experiment, we set $L = 5000$.
2) We calculate the $S_s(t)$ score for the top ranked L images, and then, we obtain the final ranked images.

IV. Experiments

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 SHoG method in [4], and the method in [9] on our crawled dataset. We also use the SBIR system in [10] as a comparison. They use shape contexts method [11] to compute the similarity of two outlines of the object (named RC-SC). All experiments were carried out in the same environment.

A. Datasets

1) SBIR_100K Dataset: This dataset was used in [4] (denoted as dataset_100k) and contains 101 240 images. There are 1 240 benchmarked images for 31 query sketches, and 100 000 noise images.

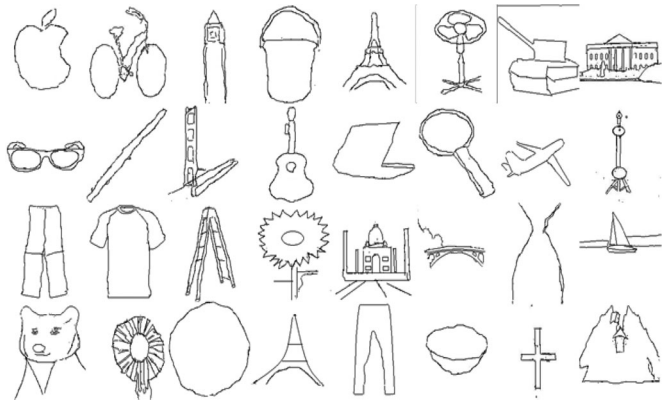


Fig. 5. Some examples of the input sketches.

2) Our Dataset: The experimental dataset consists of 293 215 images, and the storage cost is 119 GB. Our dataset consists of two parts. One is called Sketch-describable Dataset with 65 366 images collected from Google using keywords.

There are a total of 81 topics and each topic approximately consists of 1000 images. Another part is GOLD [2], [24], [40], [55] set, which is mainly about landmarks and landscapes.

We select 162 sketches which cover most of 81 topics in the Sketch-describable Dataset as queries to sketch retrieval systems. Some of them are shown in Fig. 5.

B. Performance Evaluation

We use the precision under depth n (denoted as Precion@n) to measure the objective performance, defined as follows:

$$Precion@n = \frac{1}{Z}\sum_{m=1}^{Z}\frac{1}{n}\sum_{i=1}^{n}R_m(i) \tag{8}$$

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

C. Objective Comparisons

MSF [9], HoG [1], SHoG [4], and ARP [3] methods are all proposed by calculating the histogram in each image contour partition. The RC-SC method [10] is the method using shape contexts [11] directly based on salient region. The Edgel method [7] compares edge pixel similarity and performs better in big dataset. Considering this, we select these methods as comparison methods. Correspondingly, Precision@n curves of other methods and the proposed method with the depth varying in the range [1], [50] are shown in Fig. 6(a) (our dataset) and Fig. 6(b) (SBIR_100K dataset). The curves are drawn by the average results of the query sketches on our database. For fair comparison, the parameters *M* and *N* are set to 8 and 4, respectively, for both ARP and AROP, and the partition of radius is uniform. The orientation channels in the proposed method and edgel method are both set to 8.

In our objective comparison, we find that the proposed algorithm is 10% more accurate than the other methods for the range from top-5 to top-40 results. For $n = 1$, our method is 5% more accurate than the edgel method and 10% more accurate than the other methods. Because we propose the AROP feature based on the global and local contour maps, our method makes the relevant image more similar and irrelevant image more different.

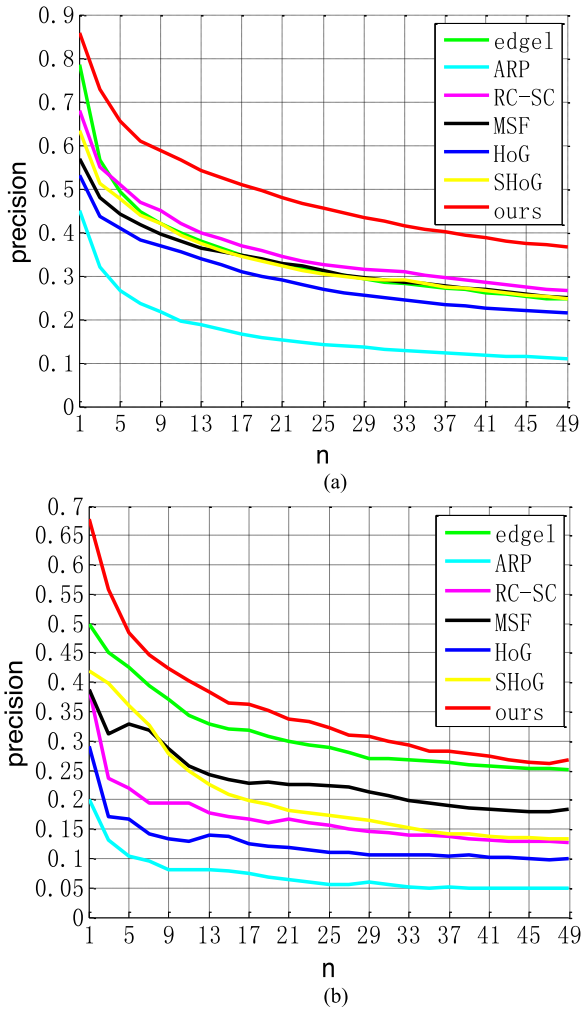


Fig. 6. Precision comparison. Edgel is the method in [7], ARP is the method in [3], RC-SC is the method in [10], MSF is multi-scale feature in [9], SHoG is the method in [4], and HoG is the method in [1]. (a). Precision comparison with other methods using our dataset. (b). Precision comparison with other methods using SBIR_100k dataset.

TABLE I
COMPARISON TO TIME COST TO PROCESS A QUERY ON AVERAGE IN DIFFERENT ALGORITHMS

Method	Edgel	ARP	RC-SC	MSF	HoG	SHoG	Ours
Times (s)	9.08	0.64	$5 * 10^4$	4.48	7.21	2.97	1.42

The average computational costs of the three methods are shown in Table I. The HoG and RC+SC based approaches are time consuming. These experiments were implemented using MATLAB on Linux, and the code was only optimized in MATLAB. But the relative computational costs are obviously different. Edgel method costs 9.077s and the MSF method costs 4.478s. The costs are all more than the proposed method. The

ARP method costs 0.635s, which is less than ours. The reason is that the AROP features dimensionality is *O* times than ARP features dimensionality.

V. Discussions

We now discuss the impacts of the parameters on the performance of our sketch-based retrieval system. In AROP feature extraction, the parameter M is the number of angle partition, *N* is the number of radius partition and *O* is the orientation channel. We discuss how each AROP feature affects the overall result. The parameter th is used to get the main contour threshold. The weight w, which computes the similarity score in matching.

A. Parameter M, N, and O in Feature Extraction

There are three parameters in AROP features. They are the number of angular partition *M*, the number of radial partition *N*, and the number of orientation partition *O*. The precision curves by varying the parameters are given in Fig. 7. Only the performance of AROP is given here for comparison.

For the number of angular partition *M*, when it is set too small, there will be less discriminative between different features. When it is set too big, the dimension of AROP feature is too larger to waste time. When fixing the parameters of *N* and *O*, we show the performance of our approach by selecting *M* in the range of [1], [12]. Other parameters are set to $N = 4$, $O = 8$. The precision curves are shown in Fig. 7 (a). $M = 12$ outperforms $M = 8$ in the end. The reason that the performance is better with increasing of *M* is summarized as follows: When *M* is small, it is too coarse to represent images' distinctiveness. Similarly, when *M* is large, the AROP feature will become more meticulous. In addition, it will bring many null partitions, which will increase the difficulty of judging the similarity between different images.

For the number of radial partition *N*, we provide the performances of $N = [1, 8]$ shown in Fig. 7(b), where we set $M = 8$, and $O = 8$. From the figure, we find that when *N* is increased, the performance is increased, but when *N* is larger, the performance is reduced. We also find that the performance is the best when *N* is approximately equal to 4. When *N* is smaller, there is less detailed information to describe the image. When *N* is larger, then the image is departed into many small sectors, which will increase the effect of the noise contour.

For the number of orientation partition *O*, $O = [1, 12]$ is discussed. Other parameters are set as $M = 8$, $N = 4$. The precision curves are shown in Fig. 7(c). When $O = 1$ the proposed method is identical to the ARP based approach. We can find that, $O = 4$ and $O = 10$ both have worse precision. But $O = 8$ performs better. The reason is summarized as follows: when *O* is set too small, every channel of orientation has many edge pixels and this will bring about worse discrimination. When *O* is set too large, every channel of orientation has little edge pixel, and the dimension of AROP feature is large. All of these will bring about too much error to make the precision lower. So we set *O* to 8 by considering the computational complexity and storage cost.

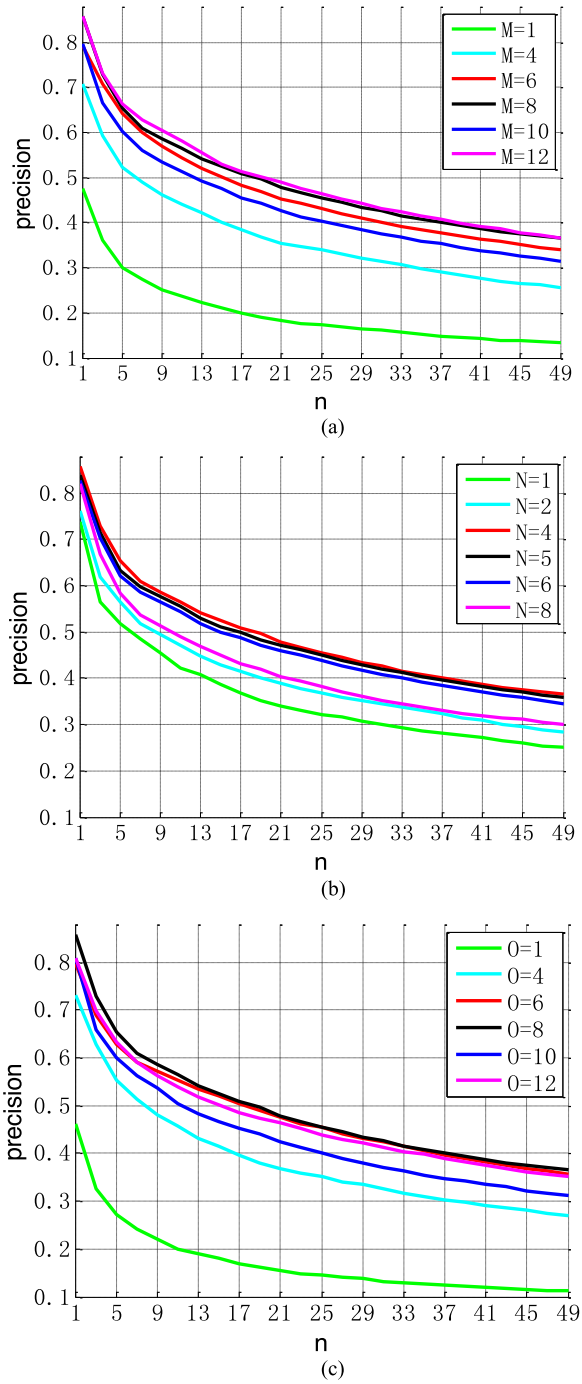


Fig. 7. Precision curves under different *M*, *N*, and *O*. (a) Precision curves when *M* varies, (b) precision curves when *N* varies, and (c) precision curves when *O* varies.

B. Two Types of Radius Partitioning

Fig. 8 shows the precision curves of the two means of radial partitioning types. From Fig. 8, for either the ARP method or the proposed method, the type of uniform partition performs well. In the uniform partition, the radius of any successive concentric circles is the same. This will make the comparison more effective. In Fig. 8, ARP-uniform means the ARP method using uniform radius partitioning. ARP-sqrt means the ARP

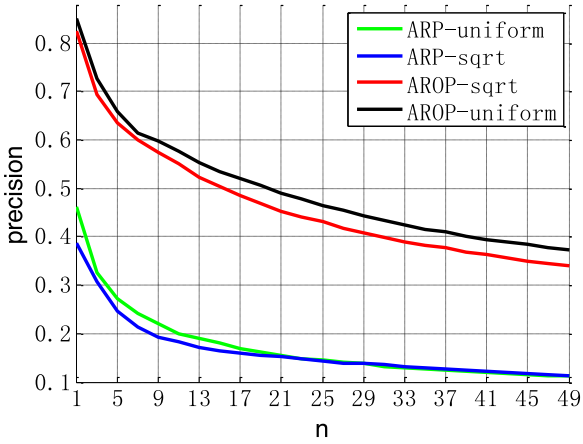


Fig. 8. Comparison of the different means of radial partitioning.

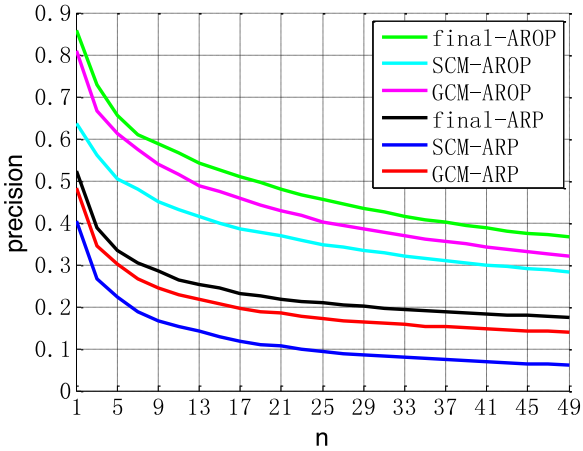


Fig. 9. *Precision@n* curves for two types features.

method using non-uniform radius partitioning. Similarly, AROP-uniform and AROP-sqrt represent the two types based on the proposed method. From Fig. 8, we can find that the uniform radial partitioning type is 3% more accurate than square root (in short sqrt) type for the top 50 results based on AROP method. For ARP method, we can find that the uniform radial partitioning type is 5% more accurate than sqrt type for the top 1 result. The accuracy of the two types is same.

C. Impact of Two Types of AROP Features

In the proposed method, we propose two contour maps: global contour map and salient contour map. Fig. 9 shows how exactly each part affects the overall result. In Fig. 9, final-AROP curve means the result of AROP features based on two contour maps. SCM-AROP curve means the result of AROP feature based on salient contour map. GCM-AROP curve means the result of AROP feature based on global contour map. The curves of ARP are similar to AROP method. From Fig. 9, we can find that the feature based on global contour map performs better than the feature based on salient contour map. However, the proposed method performs the best.

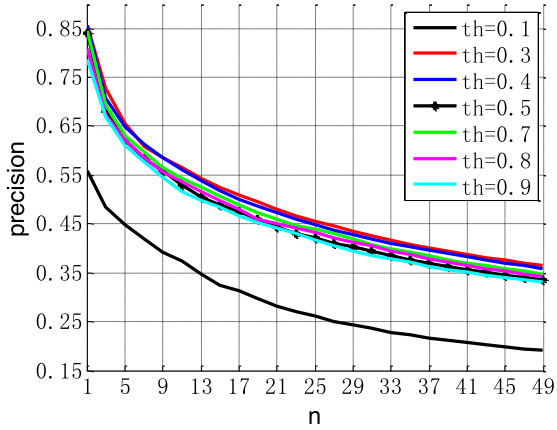


Fig. 10. *Precision@n* curves for various *th*.

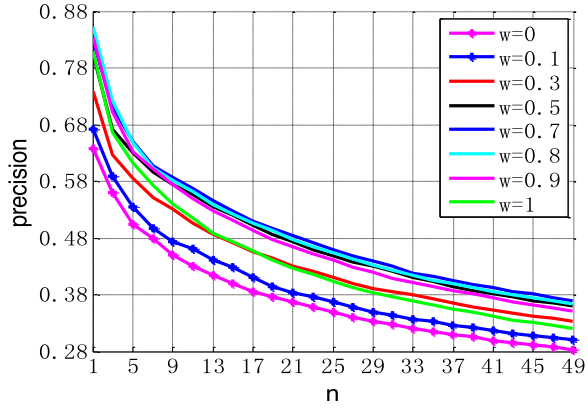


Fig. 11. *Precision@n* curves for various *w*.

D. Parameter th in Contour Map Extraction

In the global contour map GCM(x, y) extraction, we use $B_{0.5}(x, y)$ as the contour in salient region and $B_{0.3}(x, y)$ as the contour in surrounding salient region. In this section, we will discuss the parameter th in extracting contour of surrounding salient region. Fig. 10 shows the precision@n for various th. From Fig. 10, we can find that performance is the best when $\mathrm{th} = 0.3$ or $\mathrm{th} = 0.4$. $\mathrm{th} = 0.5$ is the meaning that the global contour map is the contour map $B_{0.5}(x, y)$, i.e., GCM $(x, y) = B_{0.5}(x, y)$. The precision when $\mathrm{th} = 0.2$ is a rapid decline. The reason is that the global contour map contains more detailed edges, which will make the value of AROP feature larger, i.e., background similarity score has a dominant role. From Fig. 10, we can know that the contour map can filter some irrelevant images. When $\mathrm{th} > 0.5$, the performance will be worse with the increase of th . The reason is that, when th becomes large, the edge pixel of contour will be less, which can't represent the background accurately. From Fig. 10, we find that the $\mathrm{th} = 0.3$ is 2% more accurate than th at the range [0.5, 0.9] for the top 50 results. $\mathrm{th} = 0.3$ is 30% more accurate than $\mathrm{th} = 0.1$ for the top 50 results.

E. Parameter w in Similarity Calculation

We now discuss the impacts of the parameters on the performance of our sketch-based retrieval system. In the proposed

Fig. 12. (a) and (b) are the sketch retrieval result. The first row contains the top-ranked result using proposed method in [7], the second row contains the results using ARP method [3], the third row contains the results using RC-SC method [10], the fourth row contains the results using MSF method [9], the fifth row contains the results using HoG method [1], the sixth row contains the results using SHoG method [4], and the last row contains the results using the proposed method.

method, the parameter w in (7) 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 map. Accordingly, w should range between 0 and 1. $w = 0$ denotes the AROP feature is extracted from the salient contour. $w = 1$ denotes the AROP feature is based on global contour. From Fig. 11, we can know that our method performs approximately 5% better than using salient contour map or global contour map.

As shown in Fig. 11, the method performed best when w is approximately 0.7. From Fig. 11, we find that the global contour map and the salient 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 value. So, we can use the global contour map to filter the image with complex background.
2) The saliency map contains the main image object. This information can make the image with common object more similar.

F. Subjective Comparisons

We compare the proposed methods to the MSF [9], edgel [7], ARP [3], and RC-SC [10], etc., using two input sketches.

Fig. 12(a) and (b) shows the retrieval results of the other methods and our method (the last row). As shown in Fig. 12 (a), our top 10 results are all correct and the other methods return several irrelevant images. In Fig. 12(b), the edgel method and the MSF method in [9] return more irrelevant images, but our top five results are all correct. As can be seen from Fig. 12, the proposed method performs best than other comparison methods. In addition, the proposed method has location diversity. Fig. 12 shows that our results also contain some incorrect images, but they are all similar to the queries in shape, and the results are better than those of the other methods.

VI. Conclusion

To address the SBIR, we first proposed two contour maps: global contour map and saliency map. The global contour map is used to filter the complex background and the saliency map is used to find the image of a common object. Then, we introduced an AROP feature that has higher performance between the sketch and profile based on two types of contours. In order to reduce searching time, firstly, we filtered the complicated images using contour segment. Secondly, we chose top 5000 in the result based on global contour map feature similarity score. Then, we computed the saliency map feature's similarity. In the experiments, the weight of AROP features has been discussed in detail. On our image dataset, AROP feature has certain advantages over the other methods in retrieval precision. Various experiments proved that sketch retrieval algorithm outperforms the other methods.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, Jun. 2005, vol. 1, pp. 886–893.

[2] 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*, vol. 15, no. 8, pp. 2058–2071, Dec. 2013

[3] A. Chalechale, G. Naghdy, and A. Mertins, "Edge image description using angular radial partitioning," *Proc. IEE*, vol. 151, no. 2, pp. 93–101, Apr. 2004.

[4] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," *IEEE Trans. Vis. Comput. Graph.*, vol. 7, no. 11, pp. 1624–1636, Nov. 2011.

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

[6] K. Hirat, and T. Kato, "Query by visual example," in *Advances in Database Technology*. Berlin, Germany: Springer, 1992, pp. 56–71.

[7] Y. Cao, C. Wang, L. Zhang, and L. Zhang, "Edgel index for large-scale sketch-based image search," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 761–768.

[8] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla, "Model-based hand tracking using a hierarchical Bayesian filter," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1372–1384, Sep. 2006.

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

[10] M. Cheng, N. Mitra, X. Huang, P. Torr, and S. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[11] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

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

[13] A. Chalechale, and A. Mertins, "Sketch-based image matching using angular partitioning," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 35, no. 1, pp. 28–41, Jan.2005.

[14] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "A descriptor for large scale image retrieval based on sketched feature lines," in *Proc. 6th Eurographics Symp. Sketch-Based Interfaces Model.*, 2009, pp. 29–36.

[15] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects?" *ACM Trans. Graph.*, vol. 31, no. 4, pp. 44-1–44-10, 2012.

[16] S. Ren, C. Jin, C. Sun, and Y. Zhang, "Sketch-based image retrieval via adaptive weighting," in *Proc. ACM Int. Conf. Multimedia Retrieval*, 2014, pp. 427–430.

[17] C. E. Jacobs, A. Finkelestein, and D. H. Salesin, "Fast multiresolution image querying," in *Proc. ACM SIGGRAPH*, 1995, pp. 277–286.

[18] J. R. Smith and S. F. Chang, "Visual SEEk: A fully automated content-based image query system," in *Proc. ACM Multimedia*, 1996, pp. 87–98.

[19] X. Qian, Y. Zhao, and J. Han, "Image location estimation by salient region matching," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4348–4358, Nov. 2015.

[20] J. Han *et al.*, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2015.

[21] L. Lei, S. Chunhua, Z. Chunyuan, and A. van den Hengel, "Shape similarity analysis by self-tuning locally constrained mixed-diffusion," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1174–1183, Aug. 2013.

[22] Z. Zhong, J. Zhu, and S. C. H. Hoi, "Fast object retrieval using direct spatial matching," *IEEE Trans. Multimedia*, vol. 17, no. 8, pp. 1391–1397, Aug. 2015.

[23] B. W. Hong and S. Soatto, "Shape matching using multiscale integral invariants," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 1, pp. 151–160, Jan. 2015.

[24] X. Qian, X. Tan, Y. Zhang, R. Hong, and M. Wang, "Enhancing sketch-based image retrieval by re-ranking and relevance feedback," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 195–208, Jan. 2016.

[25] X. M. Liu, C. Wang, H. Yao, and L. Zhang, "The scale of edges," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 462–469.

[26] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[27] Y. Cao *et al.*, "MindFinder: Interactive sketch-based image search on millions of images," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 1605–1608.

[28] E. Mathias, H. James, and A. Marc, "How do humans sketch objects?" *ACM Trans. Graph.*, vol. 31, no. 4, pp. 44-1–44-10, 2012.

[29] R. G. Schneider and T. Tuytelaars, "Sketch classification and classification-driven analysis using fisher vectors," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 174:1–174:9, 2014.

[30] X. Cao, H. Zhang, S. Liu, X. Guo, and L. Lin, "SYM-FISH: A symmetry-aware flip invariant sketch histogram shape descriptor," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 313–320.

[31] X. Yang, X. Qian, and Y. Xue, "Scalable mobile image retrieval by exploring contextual saliency," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1709–1721, Jun. 2015.

[32] T. Mei, Y. Rui, S. Li, and Q. Tian, "Multimedia search reranking: A literature survey," *ACM Comput. Surv.*, vol. 46, pp. 57–76, 2014.

[33] T. M. Sezgin, "Sketch based interfaces: Early processing for sketch understanding," in *Proc. of PUI-2001*, 2001, pp. 1–8.

[34] T. Hammond and R. Davis, "Ladder, a sketching language for user interface developers," in *Proc. ACM SIGGRAPH*, 2007, pp. 518–532.

[35] J. Laviola and R. Zeleznik, "Mathpad2: A system for the creation and exploration of mathematical sketches," *ACM Trans. Graph.* vol. 23, no. 3, pp. 432–440, Aug. 2004.

[36] N. Donmez and K. Singh, "Concepture: A regular language based framework for recognizing gestures with varying and repetitive patterns," in *Proc. Eurograph. Workshop Sketch-Based Interfaces Model.*, 2012, pp. 29–37.

[37] T. Y. Ouyang and R. Davis, "Chemink: A natural realtime recognition system for chemical drawings," in *Proc. 16th Int. Conf. Intell. User Interfaces*, 2011, pp. 267–276.

[38] Y. Lee, C. Zitnick, and M. Cohen, "Shadowdraw: real-time user guidance for freehand drawing," *ACM Trans. Graph.*, vol. 30, no. 4, Art. no. 27, 2011.

[39] D. Dixon, M. Prasad, and T. Hammond, "Icandraw: Using sketch recognition and corrective feedback to assist a user in drawing human faces," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2010, pp. 897–906.

[40] X. Qian *et al.*, "Landmark Summarization with Diverse Viewpoints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1857–1869, Nov. 2015.
[41] J. Han, *et al.*, "Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding," *ISPRS J. Photogramm. Remote Sens.*, vol. 89, pp. 37–48, 2014.
[42] A. M. Miguelena Bada, G. de Jesus Hoyos Rivera, and A. Marin Hernandez, "Garabato: A proposal of a sketch-based image retrieval system for the web," in *Poc. IEEE CONIELECOMP*, Feb. 2014, pp. 183–188.
[43] D. Pan, P. Shi, and C. Li, "Sketch-based image retrieval by using saliency," in *Proc. IEEE Int. Conf. Fuzzy Syst. Knowl. Discovery*, 2014, pp. 825–829.
[44] S. Liang, J. Luo, W. Liu, and Y. Wei, "Sketch matching on topology product graph," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1723–1729, Aug. 2015
[45] K. Tseng, Y. Lin, and Y. Chen, "Sketch-based image retrieval on mobile devices using compact hash bits," in *Proc. ACM Int. Conf. Multimedia*, 2012, pp. 913–916.
[46] T. Chen, M. Cheng, P. Tan, A. Shamir, and S. Hu, "Sketch2Photo: Internet image montage," *ACM Trans. Graph.*, vol. 28, pp. 89–97, 2009.
[47] T. Chen *et al.*, "PoseShop: Human image database construction and personalized content synthesis," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 5, pp. 824–837, May 2013.
[48] J.-L. Shih and L.-H. Chen, "A new system for trademark segmentation and retrieval," *Image Vis. Comput.*, vol. 19, pp. 1011–1018, 2001.
[49] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recog.*, vol. 29, pp. 1233–1244, 1996.
[50] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 1999, vol. 2, pp. 1150–1157.
[51] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.
[52] R. Hu, M. Barnard, and J. Collomosse, "Gradient field descriptor for sketch based retrieval and localization," *IEEE Int. Conf. Image Process.*, 2010, pp. 1025–1028.
[53] S. Wang, J. Zhang, T. Han, and Z. Miao, "Sketch-based image retrieval through hypothesis-driven object boundary selection with HLR descriptor," *IEEE Trans. Multimedia*, vol. 17, no. 7, pp. 1045–1057, Jul. 2015.
[54] X. Xue and X. Wu, "Directly operable image representation of multiscale primal sketch," *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 805–816, Oct. 2005.
[55] Y. Zhang, X. Qian, and X. Tan, "Sketch-based image retrieval using contour segments," in *Proc. IEEE Multimedia Signal Process.*, Oct. 2015, pp. 1–6.
[56] J. Han *et al.*, "Two-stage learning to predict human eye fixations via SDAEs," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 487–498, Feb. 2016.
[57] J. Han *et al.*, "An object-oriented visual saliency detection framework based on sparse coding representations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 12, pp. 2009–2021, Dec. 2013.
[58] J. Han *et al.*, "Learning computational models of video memorability from fMRI brain imaging," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1692–1703, Aug. 2015.

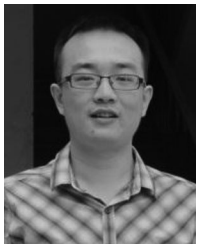
Yuting Zhang is currently working toward the MSD degree with the SMILES Laboratory, Xi'an Jiaotong University, Xi'an, China.

Her research interests include large scale sketch-based image retrieval and image content understanding.

Xueming Qian (M'09) received the B.S. and M.S. degrees from the Xi'an University of Technology, Xi'an, China, in 1999 and 2004, respectively, and the Ph.D. degree in electronics and information engineering from Xi'an Jiaotong University, Xi'an, China, in 2008.

From November 2011 to March 2014, he was an Associate Professor with Xi'an Jiaotong University, where he is currently a full Professor. He is also the Director of the SMILES Laboratory, Xían Jiaotong University. He was a Visiting Scholarar with Microsoft Research Asia, Beijing, China, from August 2010 to March 2011. His research interests include social media big data mining and search.

Prof. Qian was the recipient of a Microsoft Fellowship in 2006 and Outstanding Doctoral Dissertations of Xi'an Jiaotong University and Shaanxi Province in 2010 and 2011, respectively.



Xianglong Tan is currently working toward the MSD degree with the SMILES Laboratory, Xi'an Jiaotong University, Xi'an, China.

His research interests include large scale sketch-based image retrieval and image content understanding.

Junwei Han (M'10–SM'15) received the B.S. and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 1999 and 2003, respectively.

He has been a Research Fellow with Nanyang Technological University, Singapore; the Chinese University of Hong Kong, Hong Kong, China; Dublin City University, Dublin, Ireland; and the University of Dundee, Dundee, U.K. He was a Visiting Student with Microsoft Research Asia, Beijing, China, and a Visiting Researcher with the University of Surrey, Guildford. U.K. He is currently a Professor with Northwestern Polytechnical University. His research interests include computer vision and multimedia processing

Yuanyan Tang (F'04) received the B.E. degree in electrical and computer engineering from Chongqing University, Chongqing, China, the M.Eng. degree in electrical engineering from the Beijing Institute of Post and Telecommunications, Beijing, China, and the Ph.D. degree in computer science from Concordia University, Montreal, QC, Canada.

He is currently a Chair Professor with the Faculty of Science and Technology, University of Macau, Macau, China, and Professor/Adjunct Professor/Honorary Professor with several institutes including Chongqing University, Concordia University, and Hong Kong Baptist University, Hong Kong, China. His current research interests include wavelet theory and applications, pattern recognition, image processing, document processing, artificial intelligence, and Chinese computing.

Prof. Tang is a Fellow of the International Associate of Pattern Recognition.