A LOCAL SHAPE DESCRIPTOR FOR MOBILE LINEDRAWING RETRIEVAL

Yucong Xuan[‡] *Lingyu Duan*[‡⋆] *Tiejun Huang*[‡]

[‡]The Institute of Digital Media,School of EE&CS, Peking University, Beijing, 100871, China
{xuanyucong, lingyu, tjhuang}@pku.edu.cn

ABSTRACT

Coming with the rapid spread of Intelligent terminals with camera, mobile visual search techniques have undergone a revolution, where visual information can be easily browsed and retrieved upon simply capturing a query photo. However, most existing work targets at compact description of natural scene image statistics, while dealing with line drawing images retains an open problem. This paper presents a unified framework of line drawing problems in mobile visual search. We propose a compact description of line drawing image named Local Inner-Distance Shape Context (LISC) which is robust to the distortion and occlusion and enjoys scale and rotation invariance. Together with an innovative compression scheme using JBIG2 to reduce query delivery latency, our framework works well on both a self-built dataset and MPEG- 7 CE Shape-1 dataset. Promising results on both datasets show significant improvement over state-of-the-art algorithms.

Index Terms— Mobile Visual Search, linedrawing, inner-distance, shape context, JBIG2 compression

1. INTRODUCTION

Nowadays, using a mobile phone to retrieve information through wireless network is popular. There is much previous work [1] on Mobile Visual Search. Most of those work is designed for natural scene images like landmark, however, linedrawing, which is widely used in engineering design, architecture and cartoon and play an important role for digital library and sketch retrieval, are seldom studied.

Considering the difference between natrual scene images and linedrawings, there are two obstacles of the mobile linedrawing retrieval. The first one is the lack of a distinctive descriptor. The difficulty results from the following 2 facts: a) Since a drawing is usually a binary image composed of several lines, there is no color information and little texture information to use. To this end, color-based and gradient-based visual features, e.g. SIFT [2] and CHoG [3] have poor description capability for linedrawings; b) Drawings are much more complex than a silhouette because there are many other curves besides the contour, therefore most shape-based features which are designed for shapes containing only silhouettes fail in this situation. Although some representations (shock graph [4], shape context[5], etc.) take the interior content into consideration, they just regard the interior content as auxiliary information in describing the silhouette. The second obstacle is the lack of an efficient scheme of low bit transmission for linedrawing retrieval through wireless network.

In this paper, we propose a visual descriptor and an integrated framework for mobile linedrawing retrieval. The contribution of the paper is twofold. First , we present a compact visual descriptor to convey linedrawings. The overall framework towards compact descriptor for linedrawing is shown in Figure 1.We propose to use inner-distance shape context [6] to capture the line structure around each corner of a drawing. Second, we present a JBIG2-based compression scheme to reduce the querys size. JBIG2[1] is an image compression standard for bi-level images. Since the linedrawings mainly contain only black and white pixels, naturally, we choose to compress the taken images into JBIG2 format. This produces 1:25 compression rate comparing to the solely JPEG based compression. These two dedication are combined to form a low bit rat Mobile Visual Search framework for linedrawings and illustrated in Figure 1.

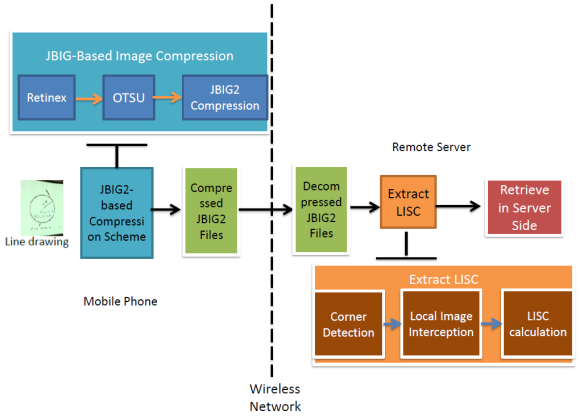

Fig. 1. The proposed linedrawing retrieval framework.

The rest of the paper is organized as follows: In section 2 we reviews some related work. Section 3 define the LISC linedrawing descriptor. Section 4 details our JBIG2-based

⋆ Corresponding Author

[1]http://en.wikipedia.org/wiki/JBIG2

compression scheme. Experiment results are shown in Section 5. Finally Section 6 concludes this paper.

2. RELATED WORK

Shape Descriptor: Shape Describing has been an active research topic in computer vision for many years.Roughly speaking, work handling shape can be classified into two categories The first category is gradient based shape descriptor. This is a kind of more general approaches as it is suitable natural images. For instance, the EHD[7]and HOG[8] approaches split the picture into small blocks and calculate the gradient of every block. However, this kind of approaches is not suitable for linedrawing retrieval as it cannot handle distortion . What is more, it is involved with too much noise by densely extracted features from images, so it is usually used with a SVM to do a shape classification or verification.

The second category is contour based shape descriptor(e.g [5] [9], etc.). As for this category, there are several typical approaches like Shape Context[5] and stroke[10]. Shape Context, which is most related to ours work models the problem of shape description as the procedure of calculating the shape pixel distribution in a polar coordinate system. Although it is widely used in shape retrieval, Shape Context proposed by MPEG-7 is based on the use of shape boundary points as opposed to shape interior points[11]. As for stroke feature, it limits the interior content into only several primitives. To cope with interior lines, LISC takes the idea of local innerdistance to extend Shape Context. LISC handles the interior lines by spliting the linedrawing into several shape patches. Inside the shape patches,there are fewer interior lines, so it is possible to handle the shape patch using Shape Context. Also, since the patches are independent with each other, LISC can still preserve enough information when there is certain occlusion. In addition, LISC also enjoys rotation and scale invariance.

Compact Descriptor for Visual Search: With the ever growing computation power of the mobile devices, recent works[12] have proposed to directly extract and send image descriptors on the mobile end to achieve a low cost wireless transmission. To this end, local descriptors in the literature like SIFT [2], SURF[13] or PCA-SIFT [14] are over size, e.g. sending hundreds of such descriptors per image typically costs more data throughput comparing to sending the original image. Taking the advantage of our images being most black and while, we present a JBIG2-based compression scheme to reduce the querys size, which produces 1:25 compression rate comparing to the solely JPEG based compression and certainly outperforms compression methods of over size features.

3. LOCAL INNER-DISTANCE SHAPE CONTEXT

In practical, since changes in scale and angle of taken photo are unavoidable, the extracted feature should be invariant to scale and rotation. To handle this issue, our Local Inner- Distance Shape Context(LISC) is inspired by local visual features [2][3]. Here we use corners as interest points and redefined inner-distance shape context is applied to formulate the line structure around each corner. As shown in Figure 2, calculation of our LISC includes three steps:

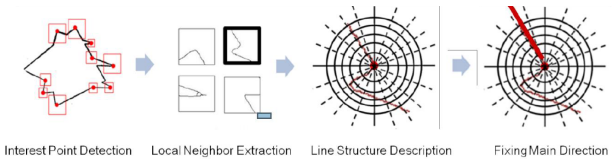

Fig. 2. Pipeline of calculating LISC. In the left image, the corners share a same scale, different sizes of red rectangles are just for better illustration.

3.1. LISC Extraction

Corner Detectors: Since most information that indicates the intrinsic properties of a sketch distribute around corners, an accurate corner detector is necessary in extracting visual characteristics. The state-of-the-art corner detectors have been widely applied in many fields. However, considering the sketch consists of planar curves, some detectors like SUSAN, COP, Harris and CSS suffer from the high computation cost or being sensitive to noise. In this paper, we utilize a detector which derives from CSS but improve the accuracy [15]. The method takes the local maxima curvature as the initial corner candidates and verifies the candidates by using an adaptively calculated threshold and angel tangent.Some detected corner are shown in Figure2.

Neighboring region interception: For each detected corners, we intercept a square neighboring region whose side length is λ. Since these regions contain distinctive information, the drawing could be represented as a set of these regions:

$$D = \{R_1, R_2, \cdots, R_N\}$$

where D represents a drawing, and R_i represents one of the N regions.

To cope with the scale changes in taken photos, we adaptively assign the value of λ by:

$$\lambda = \frac{\sum length_{C_i,C_j}}{N \times (N-1)} \quad (1)$$

where $length_{C_i,C_j}$ denotes the Euclidean distance between detected corners C_i and C_j, and $N \times (N-1)$ is the number of pairs of C_i and C_j. So the value of λ would naturally reflect the scale change of drawings.

Definition of Inner Distance: The original definition of inner-distance in [6] is the length of the shortest path between two points within an object. Given an articulated object $O \subset \mathbf{R}^2$, assume that O has n parts and each of them is convex. In this way, O can be defined as:

$$O = \{\bigcup_{i=1}^{n} O_i\} \bigcup \{\bigcup_{i=1} J_{ij}\} \tag{2}$$

where,$O_i \subset \mathbf{R}^2$ is connected and closed convex,$J_ij \subset \mathbf{R}^2$,connected and closed, is the junction betweenO_i and O_j.

Let $\Gamma(x, y; O)$ denotes the shortest path form $x \in O$ to $y \in O$. Assumes that $\Gamma(x, y; O)$ goes through m different junctions in O. Then the inner-distance $d(x, y; O)$ is the length of $\Gamma(x, y; O)$. We assume that $\Gamma(x, y; O)$ can be decomposed into segments. Each segment is either within a part or across a junction. Suppose that there are l segments in $\Gamma(x, y; O)$ We can express $\Gamma(x, y; O)$ as:

$$\{\Gamma(p_0, p_1; R_1), \Gamma(p_1, p_2; R_2), \ldots, \Gamma(p_{l-1}, p_l; R_l)\} \tag{3}$$

Therefore, with the definitions above, the inner-distance between x and y,$\forall x, y \in O$ is defined as:

$$d(x, y; O) = \sum_{i=1}^{l} d(p_{i-1}, p_i; R_i) \tag{4}$$

Where,R_i is corresponding to O_i in equation(2). Especially when O_i is convex, the $d(p_{i-1}, p_i; R_i$ reduces to the Euclidean distance. For detail, please refer to [5].

Its obvious that Compared with the Euclidean distance, inner-distance considers the articulation configuration for objects that are not convex, therefore the LISC convey the properties of curves structure more precisely than original shape context. In this paper we bring in the idea of inner-distance, however, we focus on the structural of curves in a drawing, but not the articulation of parts of an object. So we redefine the original inner-distance as drawing-based inner-distance, which means the shortest distance along the curves between two points.

Given an drawing $D, \forall x, y \in D$, then the drawing-based inner-distance can be defined as:

$$d_{drawing}(x, y; D) = |\Gamma(x, y; D)| \tag{5}$$

Where $\Gamma(x, y; D)$ is the path from x to y along the curves in D.

To compute this drawing-based inner-distance, a shortest path algorithm is applied. It consist two steps:

First, let G denotes a graph that is built with the points in a part P of drawing, $V(G) = \{v_i \in P | i = 1, 2, \ldots, N\}$ is the set of vertexes of G where N is the number of the points in P, $E(G) = \{e_{ij} | v_i$ is adjacent with v_i in P, where $i, j = 1, 2, \ldots, N\}$ and matrix $dist$ stores the drawing-based distance between each pairs of points in $V(G)$.Initially,

$$dist[x][y] = \begin{cases} \text{Euclidean distance from } v_i \text{ to } v_j, & e_{ij} \in E(G) \\ \infty, & \text{otherwise.} \end{cases} \tag{6}$$

Here, the x, y is points in the image corresponding to the vertexes v_i, v_j in the graph. Then, a standard shortest path algorithm is applied to the G . A baseline of the time complexity is $O(N^3)$ by choosing Floyd-Warshalls algorithm. There are several accumulating algorithm like Shortest Path Faster Algorithm(SPFA) which can decrease the time complexity to $O(N^2.5)$ complexity. Finally, the drawing-based inner-distance from v_i, v_j:

$$d_{drawing}(x_i, y_j; D) = dist[x][y] \tag{7}$$

Calculating the inner-distance shape context: By considering the vectors originating from a reference point to all other points on the shapes, the original shape context at a reference point is defined as a histogram of the relative polar coordinates of all other points:

$$h_i(k) = \#\{x_j \neq x_i : (x_j - x_i) \in bin(k)\} \tag{8}$$

Where the bins uniformly divided the log-polar space. And x_i is the reference point, x_j is any other point on shapes, so $x_j - x_i$ is the vector that originates from x_i to x_j

Following the [6], we replace the Euclidean distance in original shape context with the drawing-based inner-distance. It means that d-IDSC of a part centered in x_i can be defined as a histogram of relative polar coordinates of all other points x_j:

$$h'_i(k) = \#\{x_j \neq x_i : (x'_j - x'_i) \in bin(k)\} \tag{9}$$

Where x_i is reference point, and x_p in a drawing D maps to x'_p in the polar coordinates, where $p = 1, 2, \ldots, N$.Assumes that $x'_j = (\rho, \theta)$, therefore $\rho = |x'_j| = d_{drawing}(x_i, x_j; D)$ is the relative distance, and the relative angle θ is defined as:

$$\theta = arctan(\frac{\Delta Y}{\Delta X}) \tag{10}$$

Where the ΔX denote the difference of the projections of x_i and x_j on x-axis, while the ΔY on y-axis.

Different with original Shape Context and original LISC that would take each point in contour as reference point to construct rich features for an image, for each part of drawing, only the detected corner is taken as reference point, as a result the number of LISC features of a drawing equals to the number of detected corners.

3.2. Scale and Rotation Invariance

A good visual feature should present some essential properties like transform, scaling and rotation invariance. Inheriting from original shape context, LISC is invariant to transformation, however, as for the rotation invariance, Belongie [5] suggested rotating the coordinate system at each reference point so that the positive x-axis is aligned with tangent vector. But this method cannot work on LISC since most detected corners, usually the intersection of several curves, would not have well defined tangents.

To further achieve rotation invariance, we allocate the main directions as shown in Figure2. To allocate the main direction, we map the pixels in the spatial neighborhood into the polar coordinate space, then calculate the distribution along different angles. The angle with maximal distribution is selected as the main direction to rotate the region correspondingly.

As for the scale invariance, we propose a region adaptive division scheme. As described in Section 3.1,we calculate the average distance among all interesting points in a given linedrawing, then assign the scale of region using equation 1. The linedrawings is re-scaled by the relative distances between interest points, which ensures the query and reference linedrawings are in the same distance scale.

3.3. Similarity Measurement

As LISC is the distribution represented as histograms, it is natural to use x^2 test statistic:

$$C_{ij} \equiv C(x_i, x_j) = \frac{1}{2} \sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \qquad (11)$$

Where $C(x_i, x_j)$ denote the cost of matching these two points, and $h_i(k)$ and $h_j(k)$ denote the K-bin normalized histogram at x_i and x_j , respectively.

Given the set of costs C_{ij} between all pairs of corners x_i on the first drawing and y_j on the second drawing. As mentioned in [16], the total cost of matching these two drawings is calculated by minimizing:$\sum_i C(x_i, y_{\pi(i)})$. Where the π is a permutation. This is an instance of the square assignment and can be solved by using TPS.

BoW model can be also applied to measure the similarity between two drawings [4]. In practice, the d-IDSCs of a drawing are indexed by an inverted index and the similarity between each pair of drawings is measured by *tf-idf*. In our experiments, we compare the retrieving performance of these two measurements.

4. JBIG2 COMPRESSION SCHEME

Besides the effectiveness of visual descriptor, the compactness of query sent via 3G or wifi is another critical factor on which the user experience depend. The existing works extract compact features directly [1] or compress the query image firstly. Since JBIG2 is widely used in compressing the bi-level images, e.g., scanned document image, we propose compressing the query drawings into JBIG2 format files on mobile phones.

As the high compression ratio of JBIG2 can be achieved only when the input is a bi-level image, the taken photo must be binarized before compression. However, as the Figure 3. shows, the classical binarizing method(OTSU)[17] can not generate perfect bi-level images for the photos taken by mobile phones. To solve this problem, we propose augmenting the photos before binarization. As shown in the Figure 1, our JBIG2-based compression scheme includes three steps:

Augmentation In this step, we apply image augmentation on initial taken photos to remove the illumination change and shadows on them. The augmentation algorithm we apply here is Retinex [18].

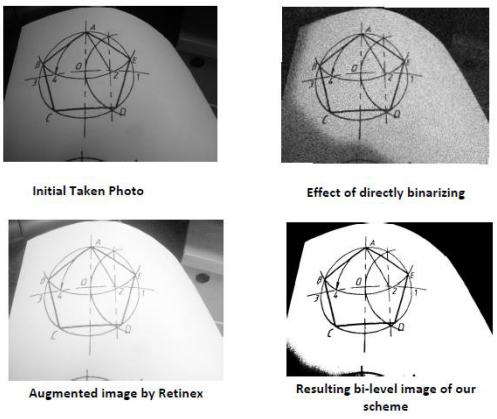

Fig. 3. Effect of binarization.

Binarizing We use OTSU to binarize the augmented images in this step. Since the adverse effect of illumination and shadows are removed, the resulting bi-level images are more suitable for JBIG2 compression algorithm, the Figure x. shows the comparing results of directly binarized query image and the results of our scheme.

JBIG2 Compression Finally, we use the official JBIG2 algorithm to compress the bi-level query image. Empirical experience shows that this compression scheme produces 1:25 compression rate comparing to the solely JPEG based compression.

5. EXPERIMENTS

In this section, we conduct two groups experiments. The first group is carried to validate the effectiveness of LISC, comparing other visual features. And the second group is designed for comparing the retrieval performance of different compression techniques.

5.1. Dataset

Our experiments are conducted on two data set. One is the standard MPEG-7 CE shape dataset, which contains 1400 shapes and is consist of 70 categories. Among each category, there are 20 similar shapes. Another dataset contains 13000 line-drawings, which are collected from industrial books and geometry books. For each line-drawing, there are 10 similar drawings that can serve as references.

For each reference dataset, we we randomly pick out 100 drawings to build the corresponding query set. We use mobile phone to take 5 photos of each query from different angles. Consequently, both of these two resulting query datasets include 500 photos.

5.2. Effectiveness of LISC

Evaluation: We use the precision-recall curve to evaluate the retrieval performance. Where precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved.

Queries: All the queries in query dataset are tried. Since JBIG2 is loseless compression, the query photos in this experiment are not compressed.

Baselines:(1) SIFT. (2) Shape context (3) Leungs Stroke Feature[10].

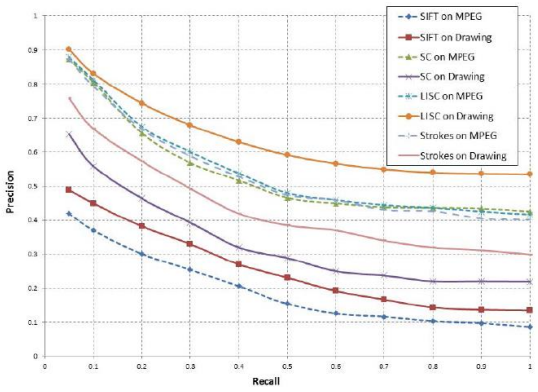

Fig. 4. Retrieval performance.

As shown in Figure 4. On both self-built drawing reference set and MPEG CE shape-1 dataset, the proposed LISC outperforms the other three methods. Although the retrieval performance of LISC on MPEG CE-1 shape set is almost the same with Shape Context and strokes, the poor performance of the latter two features on drawing dataset prove the Shape Contexts drawback of neglecting the interior content, as well as the fact that the limited shape primitives defined in [10] can not correctly convey the complex drawings. Besides, the lowest retrieval performance of SIFT demonstrate that the existing MVS applications are inappropriate for searching line-drawings. An example of retrieval result on dataset CADAL is shown in Figure 5.

5.3. Comparison of different low bit rate methods

Evaluation: We use mean Average Performance (mAP) to evaluate the image retrieval performance. Given in total N queries, mAP is defined as follows:

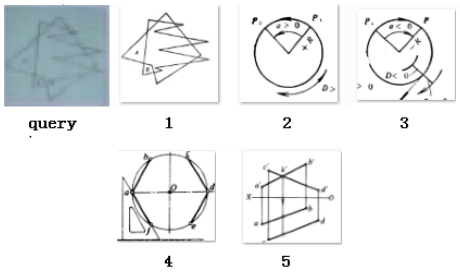

Fig. 5. Retrieval example using the proposed framework.

$$mAP = \frac{1}{N_q}\sum_{i=1}^{N_q}\frac{\sum_{r=1}^{N}P(r)}{\#-of-relevant-images}$$

where N is the number of relevant images of the ith query; $P(r)$ is the precision at rank r.

Queries: We adjust the parameters of compression methods to generate queries in different compression ratio, in order to record the retrieval performance of varied query size.

Baselines: (1) Product Quantized-LISC: PQ-LISC follows the idea of PQ-SIFT, quantize the raw LISC descriptors by a standard quantization technique. (2) VSQT: Compressed the query images by learning a optimized quantization table of JPEG which retain the information for visual search rather than information for visual perception.

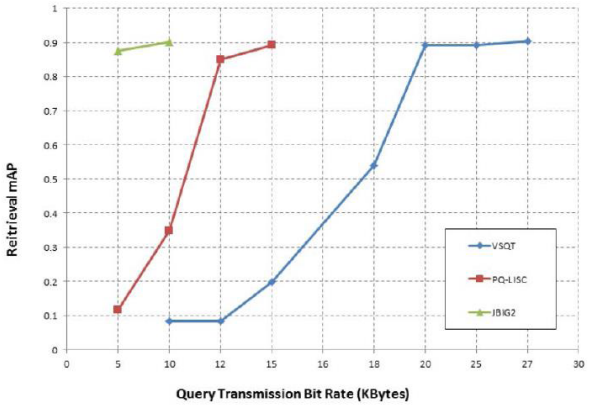

Fig. 6. Comparison of different low bit rate methods.

As shown in Figure 6. At the same mAP, JBIG2-based compression outperform VSQT by further 20KB reduction, and the average compressing ratio of VSQT on drawings is about 3.61performance with JBIG2, its heave computational cost and battery consumption would probably make compact descriptor less competitive.

5.4. Time Cost

Table 1 compares the computational complexity between JBIG2-based compression scheme and calculation of LISC

Routine	Energy Consumption	Time Cost
Compress JBIG2	89mA	91s
Extract LISC	102mA	105s

Table 1. Through three rounds at the Muni.

on iPhone4. The results show that directly sending JBIG2 compressed query has huge advantages in terms of computational complexity over extracting features on a mobile devices.

6. CONCLUSION AND FUTURE WORK

This paper presents an unified framework of a Mobile Visual Search application that takes linedrawings as query to search relevant information . Our achievements improves state-ofthe- art mobile visual search technologies in both (1) effectiveness of visual descriptor and (2) compactness of the query towards linedrawings. Extensive experiments demonstrate high retrieval performance of our proposed LISC and this frameworks advantages at low bit rates. And he efficient algorithm for calculating LISC will be included in future work.

Acknowledgement

This work was supported by the Chinese Natural Science Foundation under Contract No. 61271311 and No. 61121002, and in part by the Research Fund of ZTE Corporation.

7. REFERENCES

[1] Jie Chen Rongrong Ji, Ling-Yu Duan, "Towards low bit rate mobile visual search with multiple channel coding," *ACM Multimedia(MM)*, 2011.

[2] David Lowe, "Distinctive images features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, November 2004.

[3] David Chen Vijay Chandrasekhar, Gabriel Takacs, "Chog: Compressed histogram of gradients a low bit-rate feature descriptor," *CVPR*, 2009.

[4] Benjamin B Kimia, Allen R Tannenbaum, and Steven W Zucker, "Shapes, shocks, and deformations i: the components of two-dimensional shape and the reaction-diffusion space," *International journal of computer vision*, vol. 15, no. 3, pp. 189–224, 1995.

[5] Jitendra Malik Serge Belongie and Jan Puzicha, "Shape context: A new descriptor for shape matching and object recognition," *Advances in neural information*, vol. 15, no. 5, pp. 745–770, November 2001.

[6] David W. Jacob Haibin Ling, "Shape classification using the inner-distance," *PAMI*, vol. 15, no. 5, pp. 745–770, Feb 2007.

[7] Chee Sun Won, Dong Kwon Park, and Soo-Jun Park, "Efficient use of mpeg-7 edge histogram descriptor," *Etri Journal*, vol. 24, no. 1, pp. 23–30, 2002.

[8] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[9] Christian Bauckhage and John K Tsotsos, "Bounding box splitting for robust shape classification," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. IEEE, 2005, vol. 2, pp. II–478.

[10] Chen T Leung WH, "Hierarchical matching for retrieval of hand-drawn sketches," *Proceedings of the IEEE international conference on multimedia and exposition*, vol. 2, no. 2, pp. 29–32, June 2003.

[11] Mingqiang Yang, Kidiyo Kpalma, Joseph Ronsin, et al., "A survey of shape feature extraction techniques," *Pattern recognition*, pp. 43–90, 2008.

[12] David M Chen, Sam S Tsai, Vijay Chandrasekhar, Gabriel Takacs, Jatinder Singh, and Bernd Girod, "Tree histogram coding for mobile image matching," in *Data Compression Conference, 2009. DCC'09.* IEEE, 2009, pp. 143–152.

[13] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in *Computer Vision–ECCV 2006*, pp. 404–417. Springer, 2006.

[14] Yan Ke and Rahul Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. IEEE, 2004, vol. 2, pp. II–506.

[15] H. Nelson X. Chen and H. C. Yung., "Corner detector based on global and localcurvature properties," *Optical Engineering*, 2008.

[16] Yang He and Amlan Kundu, "Shape classification using hidden markov model," in *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on*. IEEE, 1991, pp. 2373–2376.

[17] N Otsu, "A threshold selection method from gray-level histograms," *Automatica*, 1975.

[18] DJ Jobson Z Rahman, "Retinex processing for automatic image enhancement," *Journal of Electronic Imaging*, 2004.