Journal of Computational Information Systems 10: 14 (2014) 6093–6100
Available at http://www.Jofcis.com

Enhanced Maximally Stable Extremal Regions with Canny Detector and Application in Image Classification [⋆]

Shengsheng WANG, Weilie WANG, Dong LIU[*], Fangming GU,
Bolou Bolou DICKSON

College of Computer Science and Technology, Jilin University, Changchun 130012, China

Abstract

Maximally Stable Extremal Regions (MSER) is a commonly used region detector benefited from its affine-invariant and stability. It can find blob-like structures and regions of arbitrary shape in image. However, MSER has been shown to be sensitive with noise, and unnecessary pixels may be brought into the detected regions that directly affect the stability criterion and locality of fitted regions. In this paper, a novel approach to improve MSER by adding edge detection information is presented and applied to image classification. We first utilize the original MSER detector. Secondly, Canny detector is used to obtain edge information, and the dilate operation is employed to erase ambiguous edge in order to make the interest regions more representative. Finally, the image classification framework based on our improved MSER is given. Experiments on two datasets show that our method significantly improves MSER method, and gain a better performance than previous methods in image classification.

Keywords: Maximally Stable Extremal Regions; Canny Detector; Bag of Words; Image Classification

1 Introduction

In the field of computer vision such as image classification, representing an image based on local feature has become one of the most popular methods, e.g., the bag-of-word (BoW) model. The effective performance of interest region detector is a key factor in obtaining local feature. Local feature detector has been widely used as a robust image representation compared with global feature, and its characters become popular in the field of image recognition.

A considerable amount of approaches have been proposed for detecting regions of interest. For example, an affine-invariant detector referred to as Harris-Affine [1] starts with computing a Harris corner detector over scales as initialization to locate the position of features, the characteristic

[⋆] Project supported by the National Natural Science Foundation of China (61133011, 61303132, 61103091, 61202308), Science & Technology Development Project of Jilin Province (20140101201JC, 201201131), the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry of China and the Outstanding Youth Science Foundation of Jilin University.

[*] Corresponding author.
Email address: promisedong@163.com (Dong LIU).

1553–9105 / Copyright © 2014 Binary Information Press
DOI: 10.12733/jcis11079
July 15, 2014

scale for each point then determines a scale invariant region. Hessian-Affine detector is similar to spirit with Harris-Affine except that it starts from the determinant of the Hessian rather than Harris interest points. Heuristic technique has been used in [2] to exploit regions based on geometry of the edge, which can represent the image in a very compact way and allow fast comparison and feature matching with images. Compare to other detectors, Maximally Stable Extremal Regions (MSER) [3] can detect the regions stability over a range of scales, viewpoints and illumination changes so that many researchers have been attracted to do much work on it. For example, M. Donoser [4] and E. Murphy-Chutorian [5] extended MSER with the forest like data representation of watershed, making pixels grow the regions from tree stored information about extremal regions. Per-Erik Forssen and PerDavid G. Lowe [6] modify the MSER detector in a scale pyramid to achieve better scale invariance, then use the SIFT to computing the descriptor to do image matching. Huizhong Chen and Sam S.Tsai [7] proposed a novel MSER with edge-enhanced for text detection, they use geometric and stroke width information to exclude on-text objects.

For many affine invariant detectors, the output shape is an ellipse, so dose MSER detector. An example of the original regions detected (b) is given in Fig. 1. We find that these regions with arbitrary shape tend to be sensitive to image blur, they may merge the unnecessary edges that directly affects the stability criterion and the locality of fitted regions. One example is given in Fig. 1, the regions (c) deviated due to the unnecessary edge. To overcome this problem, we introduce a novel method called MSERC, which detect the region of interest as showed in Fig. 1(d). In this paper, novel approach to improve MSER by adding edge detection information is presented and applied to image classification.

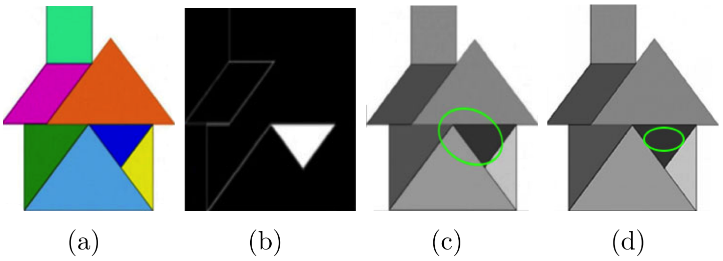


Fig. 1: An example showing the MSER and MSERC: (a) the original image, (b) one of the arbitrary regions detected by MSER, (c) the result of MSER, (d) the result of MSERC

The rest of the paper is organized as follows: Section 2 introduces the Maximally Stable Extemal Regions (MSER) detector. Section 3 shows our proposed method MSERC. Section 4 presents the framework of image classification based on MSERC. Experiments and results are shown in Section 5. Finally, in Section 6, the conclusion is presented.

2 Maximally Stable Extremal Regions

Maximally Stable Extremal Regions (MSER) [3] is a kind of local affine-invariant feature detector, which is based on the idea of the watershed algorithm. The regions detected by MSER have many desirable properties: invariance to adjacency preserving, stability to changes and multi-scale variability. Each extremal region is a connected component with appropriately threshold, all pixels in the MSER have either higher (bright-on-dark regions, MSER+) or lower (dark-on-bright regions, MSER-) intensity than the pixels on its outer boundary. As there is no global threshold in the detector, part of image may exist multiple nested regions. The formal definitions of MSER concept [3] is given in below.

Definition 1 *Image I is a mapping $I : D \subset \mathbb{Z}^2 \to S$. Extremal regions are defined on images where: 1. S is totally ordered. 2. An adjacency relation $A \subset D \times D$ is defined, i.e. $p, q \in D$ are adjacent (pAq) if $\sum_{i=1}^{n} |p_i - q_i| \leq 1$.*

Definition 2 *Region Q is a contiguous subset of D, i.e. for each $p, q \in Q$, there is a sequence $p, a_1, a_2, ..., a_n, q$ and $pAa_1, a_iAa_{i+1}, a_{i+1}Aq$.*

Definition 3 *Region Boundary $\partial Q = \{q \in D \backslash Q : \exists p \in Q : qAp\}$.*

Definition 4 *Extremal Region $Q \subset D$ is a region that for all $p \in Q$, $q \in \partial Q : I(p) > I(q)$ or $I(p) < I(q)$.*

The process of enumerating extremal regions is showed below. First, pixels are sorted by the value of intensity, place pixels in the image (decreasing or increasing), then growing and updating the pixel used the 4-neighbourhoods (Fig. 2 show the process of pixels merging), stored the connected components by components tree using the efficient union-find algorithm [8]. Relative regions on the same branch changing at a local minimum are seen as the "maximally stable" regions, it is measured by a function (1):

$$\mathrm{q}(i) = |Q_{i-\Delta} \backslash Q_{i+\Delta}| / |Q_i| \tag{1}$$

where $|\cdot|$ stands for the element number of a region, Δ is a parameter used in algorithm. In the other words, the stable one is the part of the image whose local binarization has little change over a range of threshold.

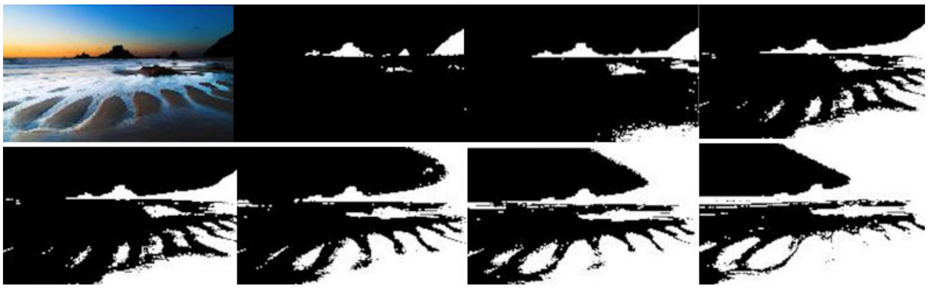


Fig. 2: The binary image under different threshold (left to right: 10 30 50 70 90 120 150)

3 Improve MSER by Canny Detector

Our proposed method is aimed at overcoming the weakness of MSER's sensitive to blurring edge, so we used the Canny operator [9] which has the good performance on noise immunity and detection accuracy, the properties of Canny detector just can make up the shortage of MSER. Our proposed method can be presented as follow:

Step 1 Utilize the original MSER proposed by J. Matas to detect the regions.

Step 2 Detect the edge using the Canny operator.

Step 3 In order to have the lower fault tolerance, we dilate the edge detected in Step 2.

Step 4 Erase the regions of Step 1 using the dilated edge template. Check connectivity of regions, if a region is separated into several pieces, they are regarded as the new forming regions. An example is showed in Fig. 3.

Step 5 Fit regions to ellipse.

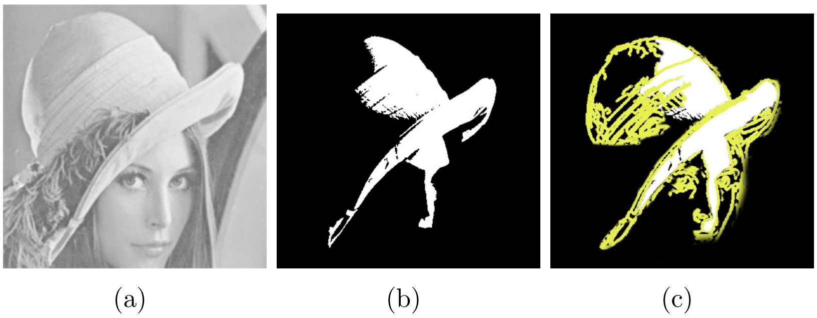


Fig. 3: A is the original image, b is an extremal region, c show the result of adding edges, the extremal region is divided into three parts, yellow lines are detected by Canny detector

We improve the MSER from the following two aspects: Firstly, our proposed method MSERC utilize the dilated edge to erase the unnecessary part which influence the locality of regions, making the descriptor contain more meaningful information. Secondly, we confirm that the number of sampling features also plays important role in image classification. Eric et al. [10] indicates that the more patches, the better performance for image classification. We present the improved method in accordance with above sampling strategy, because the region that tends to be affected by image blur can be divided into several parts by Canny detector, which increases the number of sampling features.

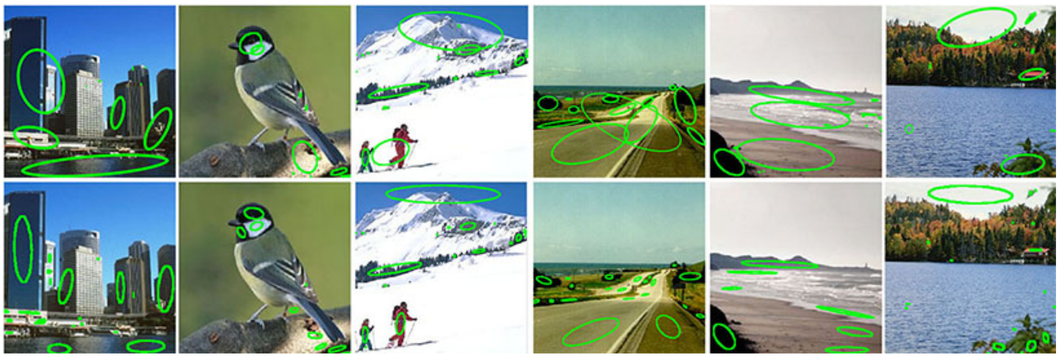


Fig. 4: The first row is the regions detected by MSER and the second row is detected by MSERC

We also make some comparison between MSER and MSERC shown in Fig. 4 (the picture only show the corrected regions, ignoring the other regions). It is indicated that MSERC tend to be more accurate in locality, they express better semantic information which can make more contribution to describe patches for image classification.

4 Image Classification Based on MSERC

Fig. 5 depicts the overall framework of our image classification method based on MSERC. In detection stage, each training image is detected a series of DRs (Distinguished Regions) by using the proposed MSERC method. Then SIFT features [11] with 128 dimension are extracted from each DR. Next, we employ Bag-of-word model to represent images due to its effectiveness. Based on SIFT features, we use K-means algorithm to cluster DRs which obtained from all the training images. Then, visual words are created according to the means of the clusters, which subsequently form a visual dictionary. After that, each DR in the training set is assigned to a visual word. And then, every image can be represented by a histogram of word frequency, which can be used for training a Support Vector Machine (SVM) classifier.

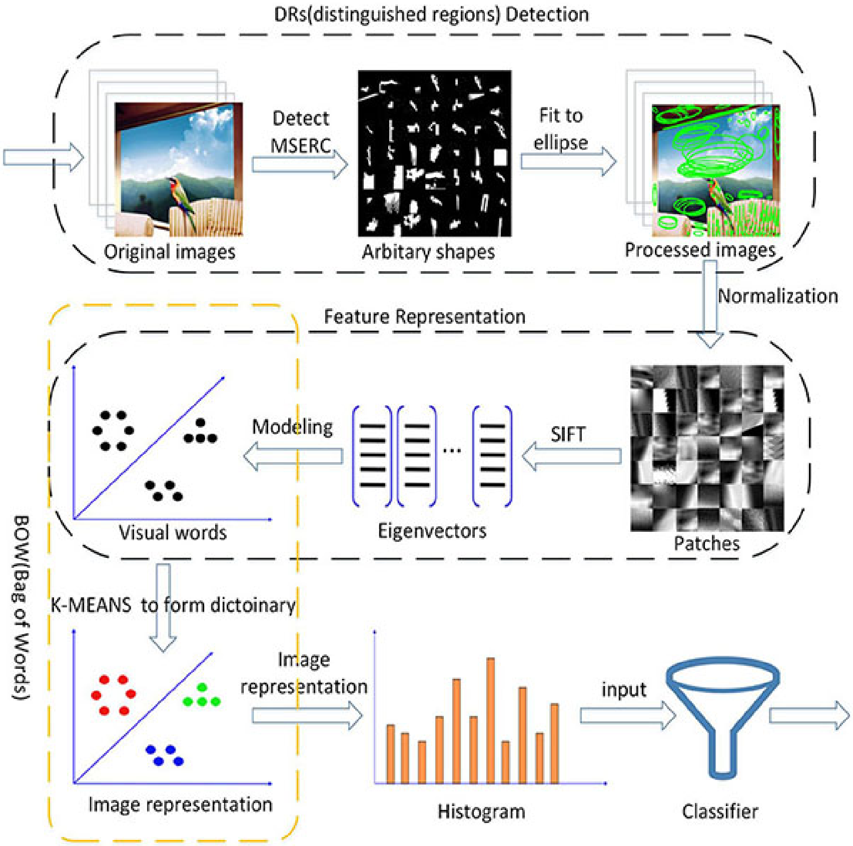
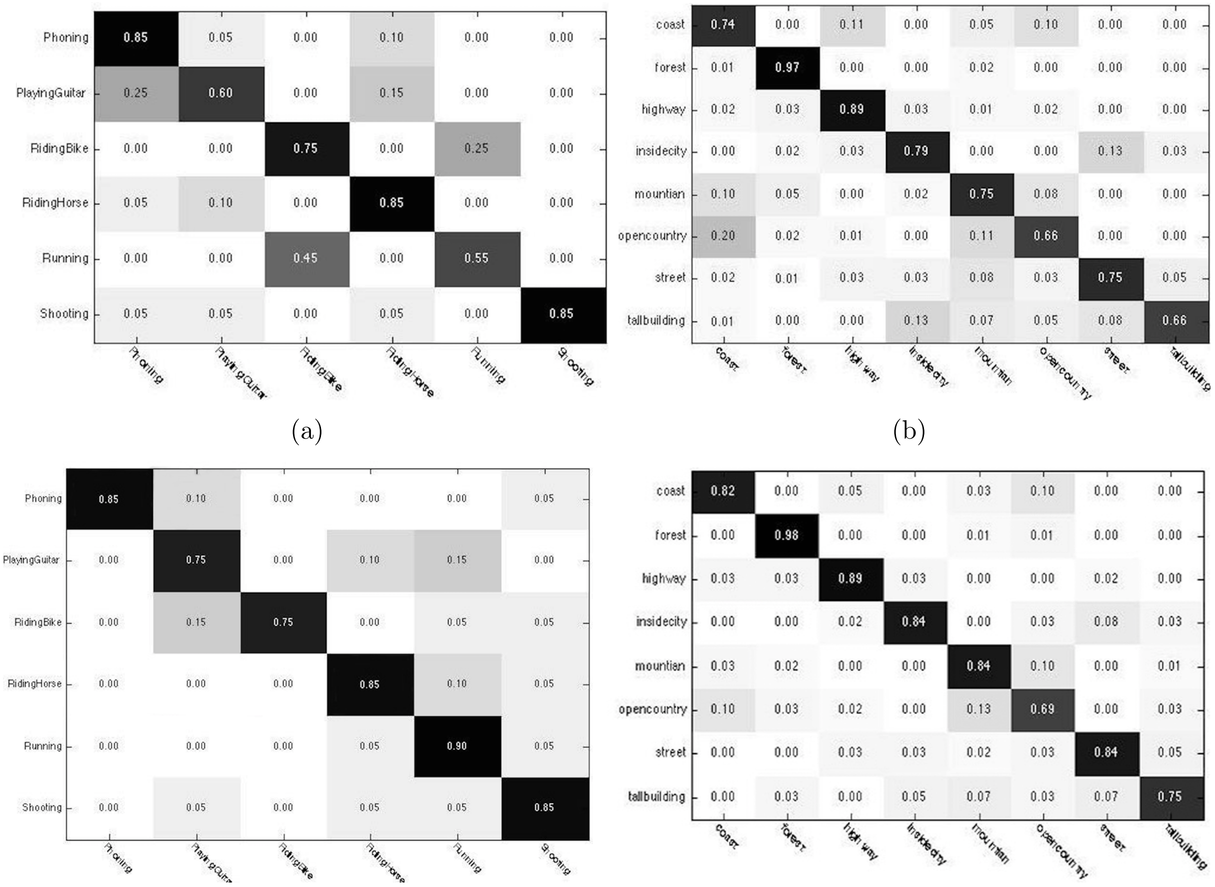


Fig. 5: The framework of the classification based BOW model

5 Experiments

In this section, we demonstrate the performance of our method using experiments on two datasets. More specifically, for SCENE-8, it contains 2688 color images from eight categories: coast(360), forest(328), highway(260), inside city(308), mountain(274), open country(410), street(292), tall building(356). The sizes of images are all 256 × 256, we randomly select 60 images to training and 60 images to testing for each category. For the dataset of ACTION-6 from actions in Still Web Images, it contains 360 color images from six categories: phoning(60), playing guitar(60), riding bike(60), riding house(60), running, shooting(60). The average size of images is 200 × 200, we select 40 images to training and 20 images to testing for each category. Some sample images are shown in Fig. 6.

Fig. 6: Sample images from two datasets



(a)

(b)

(c)

(d)

Fig. 7: A and b are confusion matrix of MSER from ACTION-6 and SCENE-8; c and d are matrix of MSERC

For extracting the BOW features after interest regions have been detected using our method, the SIFT descriptor is adopted to represent each interest region and a codebook was constructed by clustering the descriptors. The size of codebook is set to 500, which is a suitable choice in this case. Note that the codebook is constructed using a K-nearest neighbors (KNN) [12] algorithm with the Euclidean distances and 100 maximally iterations. For the classifier, we use the SVM with RBF kernel. In Fig. 7, we compare the classify performance between MSER and MSERC from two datasets.

Then we do the comparison among MSER, MSERC, Hessen Laplace and Harris Laplace. Table 1 show the average accuracy of classification and Fig. 8 presents a comparison of the four detectors in every category.

Table 1: The accuracy of classification using MSER, MSERC, Hessian Laplace and Harris Laplace

Dataset	MSER	MSERC	Hessian-Laplace	Harris-Laplace
ACTION-6	0.7417(89/120)	0.8250(99/120)	0.7750(93/120)	0.7500(90/120)
SCENE-8	0.7754(372/480)	0.8313(399/480)	0.7250(348/480)	0.7425(356/480)

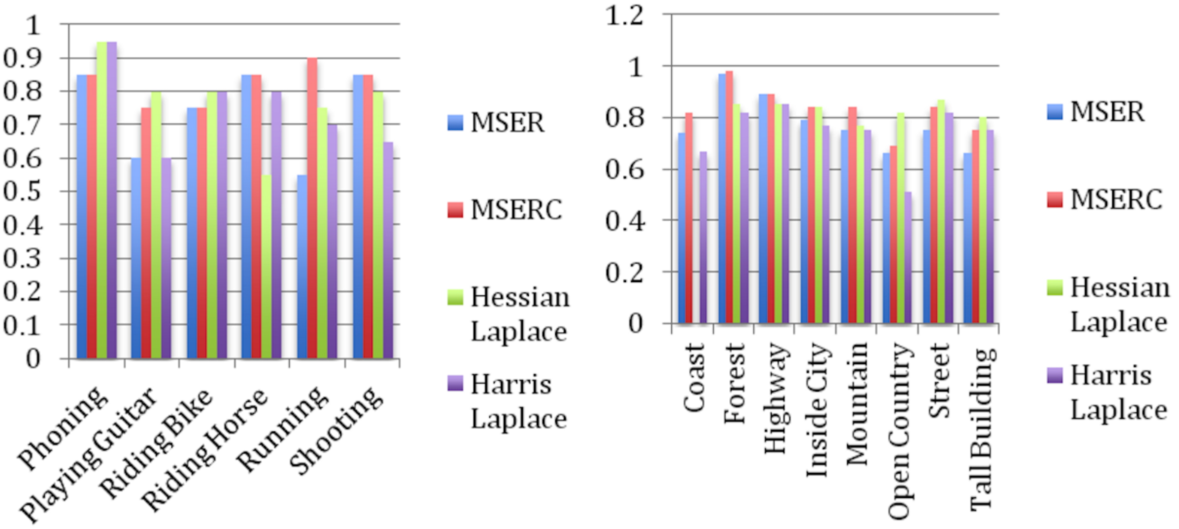


Fig. 8: Comparison of the four detectors in every category of ACTOIN (left) and SCENE (right) datasets

From the experiment results, we can see that our method consistently outperforms the MSER, no matter the accuracy of region locality or the performance for image classification on both two datasets. However, for the ACTION-6 of phoning, playing guitar and riding bike category, our method perform slightly poor than Hessian-Laplace, it may caused by the cases that our method could not catch the location of tiny action as good as Hessian-Laplace. For the SCENE-8 dataset of open country category, the complex condition of outdoor may affect the region detection, which may lead to lower accuracy. In generally, our proposed method has higher accuracy for image classification (Table 1) than other detectors, which is attribute to we filter out ambiguous edge information in our proposed method, making the detected regions achieve better representation for image.

6 Conclusion

In this paper, a novel DR detection method based on MSER is proposed, which is called MSERC. It overcomes the weakness of MSER, making it more suitable for image classification. We utilize the dilated Canny detector to divide and erase the regions, enhancing the locality of regions and the mount of features which is key to performance of image classification. Additionally we present the overall framework of our image classification method based on MSERC. Based on the experiments on Action-6 and Scene-8 databases, we can draw a conclusion that our MSERC method can effectively detect the regions of interest and gain better performance than previous method in image classification. The MSERC detector also can be applied to content-based image retrieval and recorded video analysis.

References

[1] Krystian Mikolajczy, Cordelia Schmid. Scale & affine invariant interest point detectors. International Journal of Computer Vision, 60: 63-86, 2004.

[2] Tinne Tuytelaars and Luc Van Gool. Content-based image retrieval based on local affinely invariant regions. Visual Information and Information Systems Lecture Notes in Computer Science, 1614: 493-500, 1999.

[3] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. In Proceedings of the British Machine Vision Conference, 22: 761-767, 2002.

[4] M. Donoser and H. Bischof. Efficient maximally stable extremal region (MSER) tracking. In Proceedings of the Conference on Computer Vision and Pattern Recognition, 1: 553-560, 2006.

[5] E. Murphy-Chutorian and M. Trivedi. N-tree disjoint-set forests for maximally stable extremal regions. In Proceedings of the British Machine Vision Conference, 739-748, 2006.

[6] P. E. Forssen and D. G. Lowe. Shape descriptors for maximally stable extremal regions. In Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV'07), pp. 1-8, 2007.

[7] Chen, H. Z., et al. Robust Text Detection in Natural Images with Edge-Enhanced Maximally Stable Extremal Regions. IEEE 18th International Conference on Image Processing (ICIP), 2609-2612, 2011.

[8] R. Sedgewick. Algorithms. Addison-Wesley, 2nd edition, 1988.

[9] J. Canny. A computational approach to edge detection. IEEE Transaction on Pattern Analysis and Machine Intelligence. Pattern Anal. Mach. Intell, 8: 679-698, 1986.

[10] Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification. In proceeding of European Conference on Computer Vision, 490-503, 2006.

[11] Forssen and Lowe. Shape descriptors for maximally stable extremal regions. IEEE 11th International Conference on Computer Vision, 1-8, 2007.

[12] Hengfei ZHANG, Zhiyuan ZENG, Xiaojun TAN. KNN Queries on Spatial Polygons Based on MR+-tree. Journal of Computational Information Systems, vol. 8(10): 4069-4077, 2012.