Rapid Estimation for Logo Detection on Mobile Devices

Lusi Liao[1, a], Shuwu Zhang[2, b] and Shuqi Wang[1, c]

[1]School of Information Engineering, Communication University of China, Beijing 100011, China;

[2] Institute of Automation, Chinese Academy of Sciences, Beijing 100091, China.

[a]blose89@163.com, [b]shuwu.zhang@ia.ac.cn, [c]648127199@qq.com

Keywords: logo detection, geometric feature, best overlap, mobile devices.

Abstract. Along with the technological development, running complicated program on mobile phone become possible. Mobile device allows to easily capture pictures and do corresponding processing. Thus, this paper puts forward a rapid estimation method which is aimed at detecting large-scale logos in the natural environment by using mobile device. For this purpose, feature detectors and objectness measure are applied to rapid estimation method. By using the feature extraction results to evaluate the objectness results, the higher the evaluation is, the more possible the objectness result is a logo region. This paper take 1000 pictures in the natural environment for measurement, the experiment results show the effectiveness of the method.

1. Introduction

In the last few years, logo recognition in the reality environment has drawn more and more attention. Currently the methods on the logo detection are mainly training a classifier with the global feature or local feature which is extracted from certain logo, and finally get the model and classification equation. When using these methods to deal with a large number of logos in the natural environment, it needs lots of training images that will bring a huge amount of computation. Thus, this method cannot meet the demand for real-time logo detection at the present stage. Now with the development of intelligent handset technology, mobile phone has a good inner set camera and processor, such as human face detection on mobile phone is already a mature technology. Therefore, using mobile device to get to the logo images and do processing work to complete the function of logo detection has become possible.

This paper introduce a fast estimation method, which can estimate the possible logo region in an image. This method is divided into two parts. The first part consists of feature extraction and object detection. Second part is a screening process, the results of feature extraction are used to screening the target windows of the object detection results. In this paper, FlickrLogos-32 dataset and 1000 pictures taken in the natural environment are used for testing, the experimental results indicate that the method is effective.

2. Related work

Logo detection is a kind of object detection, common algorithm for object detection can be divided into two categories: salient region detection and objectness estimation.

The salient region detection method measures the saliency of the image sub region by computing the contrast ratio of each image sub region and its adjacent region in a certain range. Regional stability is one of the characteristics of logo, thus logo is a significant object that has high contrast degree in the real scene. Using saliency detection, the logo area can be easily detected in an image. At present many saliency detection algorithm can bring a good result. R.Achanta et al. proposed a Frequency-tuned Salient Region Detection [4] that outputs full resolution saliency maps with well-defined boundaries of salient objects. FT method achieves excellent performance of image segmentation and saliency detection. In 2014, Cheng M M et al. put forward Salient Shape method [5] which is a simple, fast, and effective algorithm for locating and segmenting salient objects by

analysing image collections. Salient Shape method uses group saliency to achieve superior unsupervised salient object segmentation by extracting salient objects that maximize between-image similarities and within-image distinctness. All these methods can display a good effect in detecting the most salient object from natural world. However, there are too many kinds of objects in the natural environment, some non-logo objects can be also detected in the process of testing, which causes relatively large interference. As the logo appears in the natural environment is more close to the text target, in this paper logo is considered as text target. MSER (Maximally Stable Extremal Region) [1] is a traditional method used in text detection to determine the connected regions. MSER can detect areas in different degrees and have achieved a good effect in practical applications. Moreover, OpenCV library already implement the MSER algorithm, makes it more convenient to use this method. Thus this paper uses MSER method to complete the logo connected region detection part.

The human visual system can perceive the target before identify it, inspired by this, some generic object detection method is put forward. Alexe et al [2] first presented the concept of "objectness", author uses the characteristics such as saliency, boundary, super pixel, color and position, calculating the possibility of a window containing arbitrary targets. Inspired by the research of Alexe et al, Uijlings et al [3] proposed a selective search method to obtain the possible target area of the image without relying on specific target class. BING [7] algorithm proposed by Cheng M M et al. is similar to selective search. Based on the closed regions of object, BING can predict almost all of the objects at high speed. However, BING method needs to learn a large number of samples, which makes it more suitable for the identification of a specific object. In the research of the paper, the efficiency of BING is not as good as selective search. Therefore, this paper combines the selective search method and the MSER method which mentioned above, to forecast the logo region in an image. By using the MSER results to evaluate the selective search results, the higher the evaluation is, the more possible the objectness result is a logo region.

3. Experiment

The overall framework for the rapid estimation method is shown in Fig. 1.
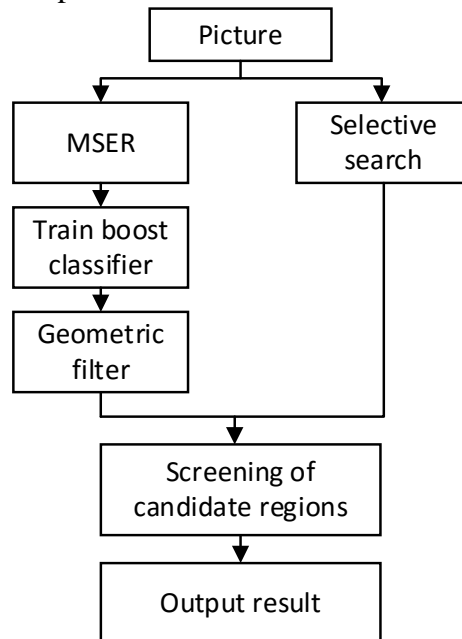

Fig. 1 Flow chart for rapid estimation method of the logo area

As shown in Fig. 1, the whole process is divided into two parts. The first part consist of two parallel methods, feature extraction and object detection. This paper uses MSER and selective search as feature extraction and object detection. The second part by screening the candidate regions to output the final results.

3.1 Feature extraction.

The basic principle of MSER method is to do binarization processing of a grayscale image by calculating threshold. With the increase of the threshold, some of the connected regions in the image changes very little, even without change. These regions are known as the most stable region. The mathematical definition of MSER is as follows:

$$q(i) = |Q_i - Q_{i-\triangle}| / |Q_{i-\triangle}| \tag{1}$$

Among them, Q_i represents a connected region with a threshold value of i, $\triangle$ is micro changes of threshold, $q(i)$ is the change rate of Q_i when threshold is i. Formula (1) shows that the judgment of whether a region is MSER region controlled by parameters $\triangle$. However, the number of connected regions can be extracted from the images with different complexity are not the same. If a fixed value of $\triangle$ is taken, it may cause too much interference or lose of logo regions. To avoid such problem, appropriate value for each image is selected according to the formula (2).

$$\triangle = 10 + 3 * N_{KpC} \tag{2}$$

N_{KpC} is the complexity of the image, the N_{KpC} expressed by key points on unit area, and the unit area is set to 900. $\triangle$ rises with the increasing complexity of the image to be detected, only the stable region will still be detected. If the image is complex, more key points will be detected, and vice versa. In this paper, FAST algorithm has been chosen for a fast running speed and a relatively large key points. On the other hand, the OpenCV4Android library already implements the FAST algorithm, makes the usage of FAST algorithm more convenient.

The MSER detection results of an image contains interference areas as well. After analyzing on the geometric features of logo areas, a boost classifier is designed to remove the interference area. The geometric features of the logo MSER region are as follows:

(1)Area. The MSER area of logo is shown by formula (3)

$$T (MinArea) < Feature (1) = AreaMSER < T (MaxArea) \tag{3}$$

(2)Duty ratio. The Duty ratio, namely ratio of MSER area and minimum external rectangle of MSER, such as formula (4) shown.

$$T (MinAreaRatio) < Feature (2) = AreaMSER/AreaMSERBox < T (MaxAreaRatio) \tag{4}$$

(3)Aspect ratio. The aspect ratio is aspect ratio of minimum external rectangle of MSER, as shown in formula (5).

$$T (MinAspectRatio) < Feature (3) = Width/Height < T (MaxAspectRatio) \tag{5}$$

(4)Regularity. The regularity of logo is represented by the ratio of the original boundary pixels value in the MSER area and the boundary pixels value after smoothing. As shown in formula (6), $N_{original}$ refers to the MSER boundary pixel value, N_{smooth} refers to the MSER boundary pixel value after do smoothing process to MSER.

$$T (MinRegular) < Feature (4) = N_{smooth}/N_{original} < T (MaxRegular) \tag{6}$$

(5)Non-boundary. When using a mobile phone to take picture, the main target is generally will not at the picture boundary, so that the MSER area near the border as the interference region. The non-boundary feature is shown by formula (7), N_{border} refers to pixel value at the border of MSER.

$$Feature (5) = N_{border}/ N_{original} < T (Border) \tag{7}$$

Boost classifier is designed based on the five features. The trained classifier can try to ensure obvious logo areas cannot be deleted by mistake, and effectively filter out the interference of non-logo regions at the same time.

3.2 Object detection.

In the natural environment, logo can maintain the color of the connected domain and the relative stability of the intensity under the influences of light and deformation. Thus, logo is very obvious in the surrounding background. In the algorithms based on the objectness estimation, the selective search algorithm can obtain the target area of the image accurately and quickly without relying on the specific target category.

In this algorithm, the image is divided into a number of rectangular regions. Initialize regions by [6] first, then calculate similarity of each neibouring region. Search the maximum similarity region until the similarity comparison is completed. There are only a small part of the selective search windows

contain logo. The higher the intersection-over-union (IOU) ratio of the candidate window and the logo region is, the more possible the window contains a logo. In order to keep the window that contains a logo best, it is necessary to filter the generated windows. Therefore, a window screening method based on the best overlap of feature region is used to filter windows to be tested.

3.3 Window screening method based on the best overlap of feature region.

In order to filter the candidate windows generated by selective search, the MSER regions generated by feature extraction step are used to calculate the overlap ratio of the windows to be tested. The higher the overlap ratio between the window and the MSER area is, the more possible the window contains logo area.

Each of logo MSER regions represents one sub connected area of logo. When using each MSER area to determine the windows, only the windows which have high IOU with MSER area will be found, and these windows may not be able to accurately locate the logo. However, in the practical application of logo detection, the type of logo cannot be known in advance, so it is not feasible for simply using a single MSER area to screening the candidate window.

This paper is based on the following two methods to carry out the screening of candidate windows.

(1)Maximum IOU of a single MSER on a single window.

For logos with few sub connected areas, only rely on a single MSER area can filter the candidate windows. Windows that get a high IOU with external rectangle of MSER will be retained.

(2)Global IOU of multiple MSER in a single window.

For logos with multiple sub connected areas, the MSER areas of logo are regarded as a whole, screening the candidate windows rely on this whole MSER. Windows that get a high global IOU with the multiple windows will be retained.

Screening Algorithm is as follows：

Algorithm: Screening Algorithm

Input: Selective Search boxes B = {b_1,…,b_n}
MSER regions R = {r_1,…,r_n}.
Output: Set of logo location boxes set L.
Initialise overlap set O(i)=0, threshold t_1,t_2;
Foreach selective search box b_i **do**
Calculate overlap O(i) = $b_i \cap r$*/ $b_i \cup r$*
If O(i)<t_1
Remove b_i
Else
L = L $\cup$ b_i

Foreach logo location box l_i **do**
Calculate overlap o(l_i,l_j)
If o(l_i,l_j)<t_2
Remove l_j

Finally
Extract logo location boxes L from all regions in R.

After screening windows, retain at most 5 windows that contain the logo best. Finally, with the least output to ensure the best detection effect, the burden of subsequent recognition is greatly reduced.

4. Experimental Results

4.1 Database.

Two databases are used in this paper, including the FlickrLogos-32 database and a self-built database for testing in the natural environment. The FlickrLogos-32 database is provided by Computing and Computer Vision Lab Multimedia, it consists of 4280 training images and 3960 test images. The self-built database consists of 1000 images collected by mobile phone camera.

4.2 Experimental Platform.

The experiments of this paper are based on the Android platform. The mobile phone is Xiaomi with 2S 4 core 1.7 GHz processor and 2GB memory, operating System is Android 4.1.1.

4.3 Experimental Design and Results.

The purpose of this paper is to realize the method of logo detection using mobile applications in the natural environment. In order to provide a better user experience, it needs the least candidate boxes to get the best localization result. Therefore, in this paper, the method's quality is evaluated by the detection accuracy and the numbers of output boxes.

(1) Analysis of Results

The output results of an image have the following cases: 1) Output only one window, and the window contains a complete log; 2) Output multiple windows, and there is one or more windows contain a complete logo; 3) The output window does not contain the logo or it contains incomplete logo; 4) No output window.

Case 1) and 2) regard as detect success. Case 3) and 4) are judged to detect failure. According to the above judgment thought, the overlap ratio of each detected region r_t and the ground truth regions r_s are calculated to evaluate the validity of the method. The mathematical expression of the detection results as shown in the formula (8)

$$\delta i(r_t, r_s) = \begin{cases} 1 & if\ o(r_t, r_s) \geq t \\ 0 & if\ o(r_t, r_s) < t \end{cases} \tag{8}$$

$o(r_t, r_s)$ is the overlap ratio of two regions, 1 stands for success, 0 stands for failure. In this paper, experiments are carried out on the FlickrLogos-32 database and self-built database. The detection results are shown in Table 1.

Table 1 Detection accuracy on different databases

Database	FlickrLogos-32	original	original-normal
Accuracy	66.7%	77.4%	86.4%

When logo detection in the FlickrLogos-32 database, the accuracy rate is 66.7%. Because some of the logos in the database are too large, too small or significant, MSER is difficult to extract geometric features of logos in such kind of picture, which causes detection failure. In the case of the pictures taken by the camera, the detection accuracy rate reached 77.4%. Some of the pictures influence by light or fuzzy will affect the test results. After removing the images with obviously not good quality from the database and test again, the detection accuracy can reach 86.4%.

(2) Comparison on different pairs

Salient Shape algorithm, FT algorithm and BING algorithm are implemented to compare with the rapid estimation method, the database is FlickrLogos-32. In order to evaluate the accuracy better, the evaluation criteria for the optimal coverage proposed in the selective search is used: ABO (Average Best Overlap) and MABO (Mean Average Best Overlap). Mathematical expressions of ABO as the formula (9) shown:

$$ABO = \frac{1}{|G^c|} \sum_{g_i^c \in G^c} \max_{l_j \in L} Overlap(g_i^c, l_j) \tag{9}$$

$Overlap(g_i^c, l_j)$ is the overlap ratio between the candidate region and the ground truth. MABO is the mean value of ABO for all test data.

Table 2 Comparison on MABO between different pairs

Method	MSER	SalientShape	FT
SelectiveSearch	0.6862	0.5602	0.5285
BING	0.5743	0.4732	0.4108

As observed from Table 2, feature detector combined with BING method has a lower MABO than combined with the selective search method. The BING method needs to learn a large number of samples. In cases when detection on a large scale logo, BING method is not as good as the selective search method. With the same object detection method, combined with MSER method has a higher MABO than Salient Shape and FT. Whether it is Salient Shape or FT, after extracting the significant object in the picture, it is not very good to distinguish between the logo and the non-logo. Thus, the

rapid estimation method which uses selective search and MSER has a well performance in logo detection.

Table 3 Effect comparison between SS and SS+MSER

Method	MABO	Average windows	MABO/Average windows
Selective Search	0.7948	50.06	0.158
Rapid Estimation	0.6862	1.08	0.6353

As shown in Table 3, compared with the selective search method, although rapid estimation method has relatively low MABO, it has a significant reduction in the numbers of the average windows to be measured. Each window has a high contribution to MABO, which makes the output window is more reliable than the results generated by the selective search. Besides, it can also make mobile applications achieve a better user experience.

Examples of the detection procedure are shown in Fig. 2. As example shows, the rapid estimation method can detect the logo region efficiently.

Fig. 2 Results of rapid estimation method on mobile phone

5. Summary

In this paper, a method for logo detection on mobile phone is implemented. Based on the algorithm of maximal stability region and selective search algorithm, a rapid estimation method is proposed. Logo regions are extracted according to the maximum stability region algorithm, and the object boxes are generated by selective search. Both of them can provide information to each other, completing the rapid estimation of logo under the large-scale natural scenes. Experimental results show that the method achieves a high level of window efficiency and detection accuracy.

References

[1]. Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions[J]. Image & Vision Computing, 2004, 22(10):761-767.
[2]. Alexe B, Deselaers T, Ferrari V. What is an object? [C]. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010:73-80.
[3]. Uijlings J R R, Sande K E A V D, Gevers T, et al. Selective Search for Object Recognition[J]. International Journal of Computer Vision, 2013, 104(2):154-171.
[4]. Achanta R, Hemami S, Estrada F, et al. Frequency-tuned Salient Region Detection[C]. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009). 2009:1597-1604.
[5]. Cheng M M, Mitra N J, Huang X, et al. SalientShape: group saliency in image collections[J]. Visual Computer, 2014, 30(4):443-453.
[6]. Felzenszwalb P F, Huttenlocher D P. Efficient Graph-Based Image Segmentation[J]. International Journal of Computer Vision, 2004, 59(2):167-181.
[7]. Cheng M M, Zhang Z, Lin W Y, et al. BING: Binarized Normed Gradients for Objectness Estimation at 300fps[C]. Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014:3286-3293.