

Automatic Detection of Reference Region in Label Images

Yuheng Wang^a, Chune Zhang^{b*}

Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

^aemail:14120336@bjtu.edu.cn, ^b email:chezhang@bjtu.edu.cn

Keywords: Reference Region; MSER; Shape-Matching; Similarity Measure.

Abstract. In regard to automatic quality inspection for label printings, a reference region is necessary for finding the region of interest. With such a reference region, the defects of the region of interest can be analysed. The reference region should be unique, stable and have a high contrast against the background. This paper proposes a method to automatically locate such a reference region. First, the maximally stable extremal regions method is used to obtain a series of regions, including reference regions and noisy regions. Next, a constrained model is established to filter out regions that are undesirable for use as reference regions. After that, to ensure all filtered regions are unique, a method to detect similar regions is applied based on shape-matching. Compared with other similarity measure methods, the shape-matching method shows better performance. Experiments on label images show the effectiveness and stability of our algorithm. The proposed method provides a new way for quality inspection of label printings in an industrial setting.

Introduction

The rapid growth of the retails has accelerated the expansion of the label printing industry. The printed content needs to be not only exquisite but also correct, such as numbers, Chinese and English characters, bar codes, graphics and other variable information. In others words, no errors are allowed.

During an industrial inspection, owing to the inherent deviations in mechanical conveyor systems, the absolute position of the label is not the same in the whole captured images every time. While the position of the region of interest (ROI) is fixed relative to the label image content. To find ROI for inspection, a reference region is first confirmed in the template image. Then, the relative location between the reference region and ROI is saved. Consequently, in a new label image, ROI for inspection is located depending on the reference region. In practice, owing to a relatively low understanding of visual inspection systems, unfittable reference regions are often chosen by workers. It is disadvantageous for automatic inspection. Moreover, training workers to confirm a reference region increases the cost of using such system. Here, we propose a way to automatically determine the reference region in label images.

Researchers have explored various ways to detect desired regions. Santella and DeCarlo [1] proposed a method to record the eye focus points of a person being tested and to obtain visual interest regions with a purely data-driven approach. The method could represent a human's initial interest in the image, but it had a low degree of automation because participants need to wear the test equipment throughout the experiment. Furthermore, that method mainly focuses on the analysis and understanding of human visual behaviour. Frejlichowski and Grzegorzewicz [2] presented an approach for automatic detection and extraction of salient regions based on Ullman and Koch's model of the bottom-up attention. The experiments with the proposed approach gave very promising results. Liang and Yuen [3] used linear regression to combine the automatic scale selection attention model with the coherent visual attention model. It is biologically more plausible because two important properties of human vision were taken into account. Zhang [4] designed a connected component descriptor for rapid logo detection. The descriptor combined the connected component pixel distribution histogram with the anchor-connected component features.

In practice, owing to the incredible diversity of the target images, detection of a reference region is extremely complex. A region for reference must be unique and have strong contrast against the background. According to the characteristics of reference regions, we propose the following algorithm to extract reference regions from the label images. First, we obtain the maximally stable extremal regions (MSERs) [5] from the input image, and presents a constrained model to enhance the MSERs' results. This process can filter unfittable regions. Finally, we remove regions that are similar based on a shape-matching measurement. This step is necessary to avoid possible location errors and finally one of the similar regions is set to be the reference region.

The rest of the paper is organized as follows. We propose our method in section II. In section III, the effectiveness and stability of the proposed method are evaluated in several experiments. Finally, the conclusions are given in section IV.

The Proposed Model

As mentioned above, the reference region is significant for ROI location. The centroid of the reference region can be regarded as a marking point. To be a reference region, it must be visually significant against the background and unique in the image. On the basis of these features, we present a hybrid algorithm for detecting the reference regions in the label images. It comprises the following modules: application of the MSER method, region refining and region merging, finally detection of similar regions. The general architecture of the proposed algorithm is presented below in Fig. 1.



Fig.1. The general architecture of the proposed algorithm

Application of MSER method. This approach extracts the maximally stable extremal regions from label images. Label images contain a wealth of information, such as numbers, letters, characters and special logos. Within a region, the contrast is comparatively small, but between the region and the background, the contrast is typically high. The MSER method feeds more effective and stably connected components to traditional segmentation methods by only recognizing stable extremal regions [6] [7]. At the same time, this process will successfully ignore the disturbance of noise [8]. MSERs contain both desirable and undesirable regions. Fig. 2 shows the original image; meanwhile, Fig. 3 shows the MSERs, which are marked by ellipses.



Fig.2. Original label image

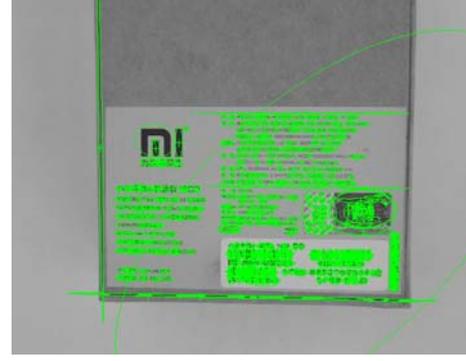


Fig.3. Result showing the maximally stable

stable

extremal regions (marked with

ellipses)

Region refining and region merging. Region refining: Not all the regions extracted after the application of the MSER method are suitable to be reference regions. According to the characteristics of a reference region, we establish a constrained model to refine the binarized MSERs to generate as few and precise regions as possible to be candidate reference regions. Here are the features we used to assess binarized MSERs.

1)Area: From among the regions we obtained after the application of the MSER method, the shape of a large region is much easier to recognize than that of a small region. Meanwhile, large regions are more accordant with the visual attention characteristics of human. So we retain the larger regions and remove the smaller ones, depending on the following criteria:

$$A_{MR}(i) > T_1 \quad (1)$$

where A_{MR} is the area of the binarized MSERs, where $i = 1, \dots, n$; n is the total number of MSERs, and T_1 is the threshold we set for the minimum desired area size. The areas are sorted in ascending order, and the corresponding value of the total area in the first 2/3 of the list is used as the threshold value T_1 .

2)Deviation: Texture is an extremely important visual feature that reflects the phenomenon of homogenization in a label image. Rich texture features can describe an image in a more accurate manner. However, rich texture within a region will be mistaken for an edge after Canny edge detection, and it is not conducive in the shape-matching measurement. Therefore, we set a maximum for the variance of the inner MSERs depending on the following criteria:

$$Dev_{MR}(i) < T_2 \quad (2)$$

where Dev_{MR} is the deviation of the binarized MSER and T_2 is the threshold we set for the maximum desired deviation.

3)Aspect Ratio: Regions that are too long or too wide may be the boundaries of the label images or other noisy areas. Hence, we set a scope for the aspect ratio as follows:

$$T_3 < H_{MR}(i) / W_{MR}(i) < T_4 \quad (3)$$

where H_{MR} is the height of the binarized MSER and W_{MR} is the width of the binarized MSER. T_3 and T_4 are the threshold we set for the lower and upper limit of aspect ratio.

4)Boundary Distance: In label images, MSERs may lie close to the image border. However, because of the slight deviations of the camera, these regions may be partially missing in the next shot. Thus, we set a restriction for the boundary distance as follows:

$$Dis_{MR}(i) < T_5 \quad (4)$$

where Dis_{MR} is the distance between the centroid of the binarized MSER and edge of the image. T_5 is the extremal distance. The value of T_5 is set to half of the length or width of the region.

Before refining the MSERs, we perform a morphological operation to eliminate noise and joint disconnected regions. Additionally, all thresholds are generous to retain all potentially possible regions.

Region Merging: Some regions obtained after the MSER process may be over-segmented, like boxes 5 and 6 in Fig. 4; these boxes overlap and should actually be assigned as one pattern. Now, however, they are treated as two separate regions. To restore the original response of the human visual system to images, we combine two such regions if the overlap area is over half the area of the smaller region.



Fig.4. Regions after refining

Detecting similar regions. Not all regions are suitable to be used as reference regions even after region refining and region merging. In our proposed method for detecting reference regions, a reference region must be unique within a picture. If two regions are similar to each other, it will be disastrous for localization. On the one hand, if one of the regions mentioned above is not a localization reference, there will not be a problem. On the other hand, if any of these regions is set to be a localization reference in the template image and when a new image comes up and the shape-matching algorithm finds a similar region as the localization reference, this will lead to a failure in localization. To avoid such a failure, we need to find similar regions and delete them. Here, we use a shape-matching algorithm to detect similar regions based on edge information of a selected region [9]. This approach involves three steps.

1) Create an edge based shape model of the template region: Select any one of the regions after region merging as the template region. Then, a variation of the Canny edge detection method is performed to find the edges, including finding the intensity gradient of the image, applying non-maximum suppression and doing hysteresis threshold. After extracting the edges, we save the X and Y derivatives of the selected edges along with the coordinate information $d_i = (t_i, u_i)^T$ in the template region.

2) Search for the best similar region: The template shape we created from the template region contains a set of points: $p_i = (r_i, c_i)^T$ and its gradients $d_i = (t_i, u_i)^T$ in the X and Y direction, where $i = 1, \dots, m$; here, m is the number of elements in the edge data set. Then, we calculate the gradients $e_{r,c} = (v_{r,c}, w_{r,c})^T$ in the search region, where $r = 1, \dots, a$; a is the number of rows in the search region and $c = 1, \dots, b$; b is the number of columns in the search region. In the following process, the template shape is compared at all positions in the search region using a similarity measurement. The idea behind the similarity measurement is to take the sum of all normalized dot products of the gradient vectors from the template shape and the corresponding points in the search image. We obtain a score at each point in the search image. We retain the highest score among all points as the similarity value between these two regions. The similarity can be formulated as follows:

$$s = \frac{1}{n} \sum_{i=1}^n \frac{d_i'^T e_{q+p'}}{\|d_i'\| \|e_{q+p'}\|} = \frac{1}{n} \sum_{i=1}^n \frac{t_i' v_{r+r_i', c+c_i'} + u_i' w_{r+r_i', c+c_i'}}{\sqrt{t_i'^2 + u_i'^2} \sqrt{v_{r+r_i', c+c_i'}^2 + w_{r+r_i', c+c_i'}^2}} \quad (5)$$

3)Delete similar regions: The similarity value between any of the two regions is obtained. We set an upper limit to determine similarity. Two regions are regarded as the same one if the similarity is larger than the threshold and the difference between the areas of two regions is below 30%. Finally, all the marked same regions are deleted and only the regions to be used for reference are retained. The final results are displayed in Fig. 5.



Fig.5. Final regions for reference in the label image

Experimental results

In this section, we first compare the performance of the last step of our proposed method with other similarity detection methods such as correlation, histogram matching and moment invariants. Fig. 6 shows the similarity between any two of the regions to be checked for similarity. Each block in the figure represents a similarity value; the lighter the grey level, the higher the similarity. The blocks in the diagonal of the figure are white, which means the similarity between the selected region and itself is 1. Comparing the result of the four different similarity measurement methods, our proposed method is better at expressing similarity. The difference between different regions is more obvious in our shape-matching measurement than in the other measurements.

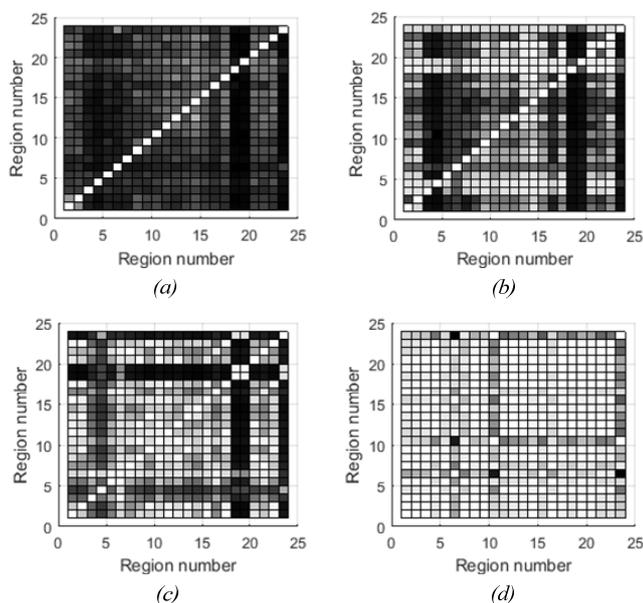


Fig.6. Similarity graph: (a) is a similarity graph based on shape- matching; (b) is a similarity graph based on correlation; (c) is a similarity graph based on histogram matching; and (d) is a similarity graph based on moment invariants.

To evaluate the stability of the proposed method, several label images that belong to different types of labels were tested. As is shown in Fig. 7, the method can extract final regions from different label images. All of the similar regions in Fig. 7(h) were deleted, and the final result is shown in Fig. 7(i).

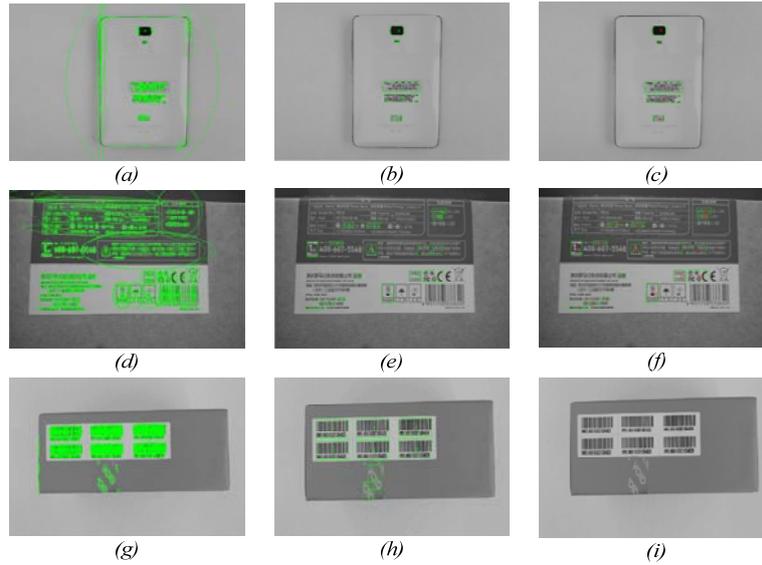


Fig.7. Final results: (a), (d), and (g) show the results after determining the maximally stable extremal regions; (b), (e), and (h) present the regions results after refining; finally, (c), (f), and (i) show the results of final regions for use as reference regions.

To check the effectiveness of the proposed method, we performed a test on three different types of label images. Each type contained eight images. The total number of reference regions was 275. Due to the vibration of the mechanical system, labels in images taken by industrial cameras usually show tiny deviations in both position or scale. Thus, the label image was placed at a slight angle and slightly scaled so as to simulate the real state of the mechanical system.

Three methods are considered and implemented for comparative evaluation. Method (a) uses a multiscale algorithm for the selection of salient regions of an image [10]. Method (b) uses dynamic threshold segmentation with region refining and region merging, followed by shape-matching measurement. Method (c) represents for our method. All the final regions were examined by experienced experts, and the results are displayed in TABLE I.

TABLE I. DETECTION RATES

Method	Image group1	Image group2	Image group3
(a)	43.1%	55.7%	51.5%
(b)	77.9%	63.6%	56.2%
(c)	95%	92.5%	75.4%

Compared to previous work in saliency, the proposed method has a higher detection rate. Salient region is not the same as reference region. Regions may be salient but not unique in the image.

Conclusion

In this paper, a MSER and shape-matching based hybrid algorithm is proposed for detecting reference regions in label images. The MSER method is applied to obtain all possible regions that could be used as reference regions. Then, a constrained model is established to enhance the results of the MSER method and filter out the unfittable regions. Finally, a shape-matching measurement method is applied to detect similar regions and delete them. As we can see from the experiments, our proposed method has a better performance than other detection methods. The algorithm is simple and effective. In addition, it is not sensitive to mechanical system deviations.

References

[1] A Santella, D DeCarlo. Robust clustering of eye movement recordings for quantification of visual interest[C]. Proceedings of the 2004 symposium on eye tracking research & applications.

New York: ACM,2004: 27-34.

[2] Frejlichowski D, Grzegorzewicz K. An approach to automatic detection and extraction of regions of interest in still images[M]//Image Processing and Communications Challenges 4. Springer Berlin Heidelberg, 2013:3-10.

[3] Liang J, Yuen S Y. An edge detection with automatic scale selection approach to improve coherent visual attention model[J]. Pattern Recognition Letters, 2013, 34(13):1519-1524.

[4] Zhang Y, Zhang S, Liang W, et al. Spatial connected component pre-locating algorithm for rapid logo detection[C]// IEEE International Conference on Acoustics, Speech & Signal Processing. 2012:1297-1300.

[5] 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.

[6] Kimmel R, Zhang C, Bronstein A M, et al. Are MSER Features Really Interesting?[J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2011, 33(11):2316-20.

[7] Gomez L, Karatzas D. MSER-Based Real-Time Text Detection and Tracking[C]// International Conference on Pattern Recognition. IEEE Computer Society, 2014:3110-3115.

[8] Mammeri A, Boukerche A, Lu G. Lane detection and tracking system based on the MSER algorithm, hough transform and kalman filter[C]// MSWIM. 2014:259-266.

[9] Steger C, Ulrich M, Wiedemann C. Machine Vision Algorithms and Applications[M]// Machine vision algorithms and applications. Wiley- VCH, 2007:339-346.

[10] Kadir T, Brady M. Saliency, scale and image description[J]. International Journal of Computer Vision, 2001, 45(2):83-105.