ACADEMIC World Academic Union

ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 9, No. 3, 2014, *pp*. 233-240

The Traffic Sign Detection based on the visual saliency

Yan Zhang[1] ,Shangbing Gao[1+], Shiliang Xu[1], Yue Zhang[1]

[1] Faculty of Computer Engineering, Huaiyin Institute of Technology, Huai'an, 223003, P.R.China
(*Received November 10, 2013, accepted January 12, 2014*)

Abstract. In this paper, we propose a new approach to detect salient traffic signs, which is based on visual saliency and the auto segmentation method based on region merging. The proposed algorithm introduces the visual saliency model, and puts forward the top-down and bottom-up two-way fusion mechanism, tried to consciously take the initiative to find and simulate human visual unconscious passive visual process by attracting interact. Then the segmentation method based on region merging is used to extract the traffic sign. Experiments show that the two-way fusion mechanism in target detection has fast processing speed and high accuracy. Extensive experiments on public datasets show that our approach outperforms state-of-the-art methods remarkably in salient traffic sign detection. Moreover, the proposed detection method has higher accurate rate and robustness to different natural scenes.

Keywords: visual saliency; Image segmentation; region merging; traffic sign

1. Introduction

Automatic extraction of traffic sign is a hot topic of intelligent transportation systems [1]. It has been widely applied in many applications such as: driving safety and automatic vehicle guidance etc. Traffic sign detection has two key goals: location and extraction. Because of complex traffic sign images, color-based and shape-based methods can not rapidly find the object and deal with illumination, viewpoint change. For example, Loy et al. [2] applied the radial symmetry algorithm [3] to detect regular polygons, which needs to set parameters in advance and unfit for all the signs. Moreover, the fully automatic segmentation of traffic sign from the background is very difficult; and there is not a mature and integrated method for traffic sign detection system until now. Therefore, to develop a real-time and accurate road traffic sign detection system has been a challenging task in computer vision.

Humans can identify salient areas in their visual fields with surprising speed and accuracy before performing actual recognition. Computationally detecting such salient image regions remains a significant goal, as it allows preferential allocation of computational resources in subsequent image analysis and synthesis. A number of very inspiring and mature saliency models have been recently introduced in the literature. Itti et al. [4] introduced a saliency model which was biologically inspired. Specifically, they proposed the use of a set of feature maps from three complementary channels as intensity, color, and orientation. The normalized feature maps from each channel were then linearly combined to generate the overall saliency map. Based on Itti's algorithm, many saliency models have appeared, such as, SR [5].

Our work is based on our previous work [6]. The extraction method based on MPCA [6] surprisingly has a performance superior to the other popular methods, but is not the most effective. To remedy such shortcoming, bi-directional integration mechanisms have been proposed. First of all, bi-directional integration mechanism is used to quickly locate the target in the image, and then the segmentation method based on region merging by maximal similarity was adopted in order to completely detect the traffic signs.
We use first two information to measuring the patch's saliency value. The central bias, which based on the principle that dominant objects often raise to the center of the image, is proposed by [7]. This underlying hypothesis brings two problems. First, background near the center of image maybe more salient than the foreground which is far away from the center. Second, for a salient object, the part near the center is more salient than that far away from the center. To diminish this effect, we give up the central bias and use the multiple scales to decrease the saliency of background patches, improving the contrast between salient and non-salient.

We get traffic sign's location using the proposed algorithm. Based on an interactive segmentation method

+ Corresponding author. Tel.: +86 25 83591046
E-mail address: luxiaofen_2002@126.com.

proposed by Ning et al. [8], which is semi-automatic image segmentation. In order to extract traffic sign region, we achieve automatic segmentation by auto-generating strokes after location.

This paper is organized as follows: section 2 describes the proposed traffic sign detection method; section 3 performs extensive experiments to verify the proposed method; and section 4 shows the conclusion of this paper.

2. THE PROPOSED TRAFFIC SIGNS DETECTION METHOD

Bi - directional integration mechanism is actually the fusion of two methods the "top-down" active search and "bottom-up" salient area detection. The "top-down" active search uses partition - first search strategy, that is, the most likely object area is the first class search area. If the object is not searched, then turn to the next level may be the regional search, or to a centre for the search area to expand your search scope

2.1. Multi-scale PCA method

Given an image I with dimension $H \times W$, non-overlapping patches with the size of $n \times n$ pixels are drawn from it. The total number of patches is $L = \lfloor H/n \rfloor \cdot \lfloor W/n \rfloor$. Denote the patch as $p_i, i = 1,2,\cdots,L$. Then each patch is represented as a column vector x_i of pixel values. The length of the vector is $3k^2$ since the color space has three components. Finally, we get a sample matrix $X = [x_1, x_2, \cdots, x_L]$, L is the total number of patches as stated above.

To effectively describe patches in a relatively low dimensional space, we used an equivalent method to PCA to reduce data dimension. Each column in the matrix X subtracts the average along the columns. Then, we calculated the co-similarity matrix $A = (X^T X)/L^2$, so the size of the matrix A is $L \times L$. The eigenvalues and eigenvectors were calculated based on the matrix A selected with their eigenvector $U = [u_1, u_2, \cdots, u_d]^T$ according to the biggest d eigenvalues, where u_i is an eigenvector. The size of the matrix U is $d \times L$.

New algorithm considers two factors for evaluating the saliency: the dissimilarities of color between image patches in a reduced dimensional space, and their spatial distance.

A patch is salient if the color of its pixels is unique. We should not, however, look at an isolated patch, but rather at its surrounding patches, which lead to a center-surrounding contrast. Thus, a patch p_i is considered salient if the appearance of the patch p_i is distinctive with respect to all other image patches.

Specifically, let $dist_{color}(p_i, p_j)$ be the distance between the patches p_i and p_j in the reduced dimensional space. Patch p_i is considered salient when $dist_{color}(p_i, p_j)$ is high for $\forall j$.

$$dist_{color}(p_i, p_j) = \sum_{n=1}^{d} \left| u_{ni} - u_{nj} \right| \quad (1)$$

The positional distance between patches is also an important factor. Generally speaking, background patches are likely to have many similar patches both near and far-away in the image. It is in contrast to salient patches that the latter tend to be grouped together. This implies that a patch p_i is salient when the patches similar to it are nearby, and it is less salient when the resembling patches are far away.

Let $dist(p_i, p_j)$ be the Euclidean distance between the positions of patches p_i and p_j, which is represented by the two centers of patches p_i and p_j in the image, normalized by the larger image dimension. Based on the observations above we define a dissimilarity measure between a pair of patches p_i and p_j as:

$$dissimilarity(p_i, p_j) = \frac{dist_{color}(p_i, p_j)}{1 + dist(p_i, p_j)} \quad (2)$$

This dissimilarity measure is proportional to the difference in appearance and inverse proportional to the positional distance.

To evaluate a patch's uniqueness, we can compute the dissimilarity between the patch and all of other patches and take the sum of these dissimilarities as the saliency of related patch. In practice, there is no need to incorporate its dissimilarity to all other image patches. It suffices to consider the K most similar patches that if the most similar patches are highly different from p_i , then clearly all image patches are highly different from p_i . Hence, for every patch p_i , we search for the K most similar patches $\{q_i\}, i = 1,2,\cdots,K$ in the image, according to (2).

$$S_i = 1 - \exp\{-\frac{1}{K}\sum_{k=1}^{K} dissimilarity(p_i, q_k)\} \quad (3)$$

In addition, the patch with large scale can not describe the boundary of small salient object. So we hope to use different scales that large scale to detect the whole information and the small scale to describe the salient object in details. Last, we compile all saliency value into final saliency.

For a patch p_i of scale r , the saliency value according to (3) is defined as

$$S_i^r = 1 - \exp\{-\frac{1}{K}\sum_{k=1}^{K} dissimilarity(p_i^r, q_k^r)\} \quad (4)$$

We consider the scales $R_c = \{r_1, r_2, \cdots, r_M\}$, using (4) to calculate the saliency of patch i as $\{S_i^{r_1}, S_i^{r_2}, \cdots, S_i^{r_M}\}$. The final saliency is computing as

$$S_i = \frac{1}{M}\sum_{r \in R_c} S_i^r \quad (5)$$

The results of different scales and the final result illustrated in Fig.1. The number of PCs set to 4.

Partition-first search strategy, in particular, due to the establishment of the height and location of traffic signs have a certain standard, you will be "guided" detection without blindly taken from top to bottom, from left to right on the "Dragnet" followed by search. That is, the destination most likely area for, if not then turn to the next possible areas, or to the most likely area for the Center to expand the search scope (as shown in Figure 3.1). Under normal circumstances, set the standard traffic signs usually appear in the upper middle or top right corner of the image. That is, in many cases systems deal only with the full image area of three-eighths, a small number of cases also need to deal with eleven-sixteenths of the chart area, significant savings in processing time of the system.

2.2. Traffic signs extraction based on the region merging by maximal similarity

Saliency map only provides coarse information about where traffic signs locate in the original image. Traffic sign extraction is also an important task and necessary for road traffic sign recognition work. Image segmentation can be applied to extract the prominent regions. In very complex images the fully automatic segmentation of the object from the background is very difficult.

In [8], Ning et.al proposed a natural framework that a new region merging based interactive image segmentation method. Considering the general characteristics that interested objects are mostly concentrated at the center of images and set from the top bottom, some of the background regions are expected to be at the left and right side of images, and saliency areas here are usually noise. We adopt the object map areas within a proportion of photo size as our object labels $L_1^O(x)$, and set background labels $L_1^b(x)$ at the left and right side and bypass noisy saliency areas:

$$L_1^O(x) = \left\{x : S^O(x) = 1, r(x) < k_1 w_1\right\} \quad (6)$$

$$L_1^b(x) = \{x : S^O(x) = 0, k_2 w_1 < r(x) < \frac{w_1}{2}\} \quad (7)$$

where r(x) is the horizontal distance to photo center, w_1 is the width of the low resolution image, and k1 and k2 are proportion coefficients

The object labels in I1 by saliency detection are usually isolated points or small areas centralized around the object boundary, we implement graph cuts and only get a coarse object segmentation C1(x) with inexact boundaries.

(a) The original images (b)The corresponding saliency maps

Figure 1.The saliency maps using our method

Figure 2. The example of traffic sign detection

However, with the aid of C1(x), "Professional Labels" can be formed for I: let the coarse object region in I derived from I1 is O1. We shrink the boundary of O1(x) to avoid inexact boundaries and form a more accurate object labels Lo; expand it to form a ring region, accomplished background labels Lb are the summation of the ring region and Lb

$$L^{O}(x) = \left\{ x : \varepsilon_{r_1}(O) \right\} \tag{8}$$

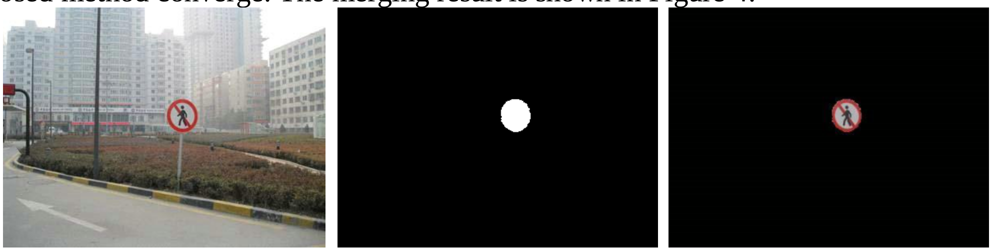
$$L^{b}(x) = \{ x : (\rho_{r_2}(O) - \rho_{r_3}(O)) \cup L_1^{b} \} \tag{9}$$

where $r_1 = l_1 R(O)$, $r_2 = l_2 R(O)$, $r_3 = l_3 R(O)$, $R(O)$ is the minimal value of the width and length of the circumscribed rectangular of O1, and l_1, l_2, l_3 are proportion coefficients; $\varepsilon_{r_1}(O)$ is an erosion operator indicating shrinking region O for r_1 pixels; $\rho_{r_2}(O)$ and $\rho_{r_3}(O)$ are the expansion operators denoting expanding region O for r_2 and r_3 pixels respectively.

(a) Artificial brushes (b) Automatically generated brushes

Figure 3. Interactive segmentation

The proposed MSRM algorithm is an iterative method. It will progressively assign the non-marker background regions in N to MB, and then all the left regions in N are assigned to MO. It can be easily seen that the proposed method converge. The merging result is shown in Figure 4.



(a) original image (b) the segmentation result (c) the extraction result

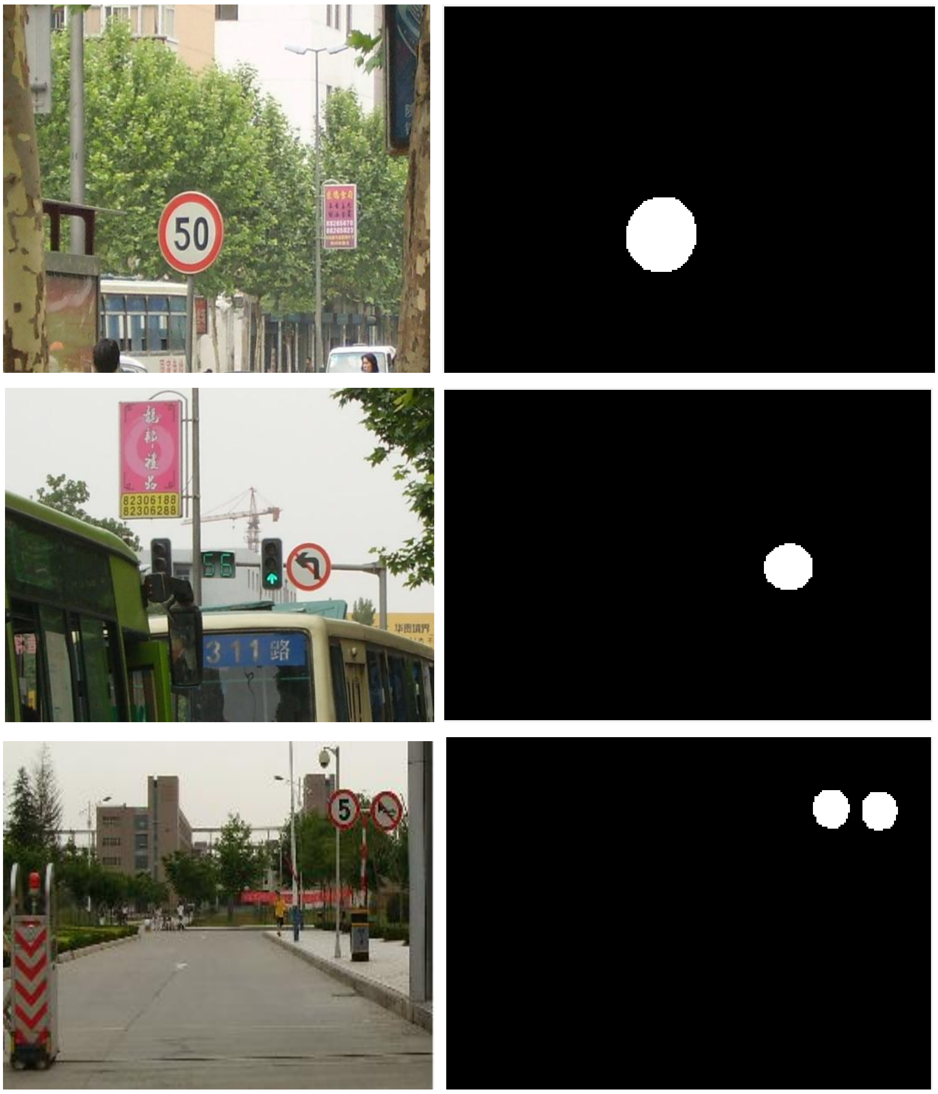
Figure 4. Region merging result

3. Experimental results

Our algorithm is implemented in Matlab 7.0 on a 2.8-GHz Intel Pentium IV PC. Extensive experiments are performed to validate the proposed method. We test the proposed algorithm under different surroundings, as shown Fig.4. All the images for test are gained from website and public datasets in our experiments, including different scene conditions, different colors and shapes.

We apply our method to test its performance on some adverse conditions, such as some traffic signs affected by illumination changes, occlusion, rotation, and shadow, different sizes.

Fig.4 (a) shows the detection results using the proposed model on some unfavorable conditions. Experimental results show the proposed method can make traffic sign region “popup” in an image of even bad environment and robust to viewpoint change and insufficient illumination. Fig.4 (b) shows the experimental results of the proposed traffic sign detection model in various suburban and highway scenes. The simple background on highway scene can make the detection more convenient. However, in urban areas and campus as shown Fig.4 (c), there may be some complex visual objects, such as trees and buildings, pedestrians and motor vehicles. Traffic sign detection becomes more difficult if only using color-based and shape-based detection method. From experimental results, we can see the proposed model is also invariant to traffic sign scale change, and our model does not rely on the size and shape of traffic sign, and can still accurately detect traffic sign area in complex scene.



(a) The original images (b) The final extraction results

Figure 4. Results of traffic sign detection using the proposed model in different conditions

4. CONCLUSIONS

In this paper we introduce a novel traffic sign detection method. According to the law of the traffic signs in the actual image position, put forward the partition first search strategy based on human visual habits, namely, the image is divided into primary, secondary search area and not the search area. Firstly, we obtain multiscale images using PCA, then compute saliency maps and label in original images, finally extract traffic signs area by automatic image segmentation. Through computing auto-generated strokes image in advance, we can realize automatic segmentation for traffic signs. The proposed algorithm could deal with traffic signs in the presence of light, scale and viewpoint change, experimental results test our model can achieve a high detection rate.

Acknowledgements

This work was supported by the grant from Jiangsu university students' innovative undertaking projects No. 201411049062X, sponsored by the Jiang Su Qing Lan project.

5. References

[1] Kastrinaki, V. and Zervakis, M. and Kalaitzakis, K, *A survey of video processing techniques for traffic applications*, Journal Image and Vision Computing, 21(4)(2004), pp.359-381.

[2] G. Loy and N. Barnes, *Fast shape-based road sign detection for a driver assistance system*, in Pro. IEEE/RSJ Intelligent Robots and Systems, 1(2004) ,pp.70-75.

[3] G. Loy and A. Zelinsky, *Fast radial symmetry for detecting points of interest*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(8) (2003), pp. 959–973.

[4] L. Itti, C. Koch, and E. Niebur, *A model of saliency-based visual attention for rapid scene analysis*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11)(1998), pp.1254–1259.

[5] Hou X.D, Zhang L.Q, *Saliency detection: A spectral residual approach*, In IEEE Conf Computer Vision and Pattern Recognition, (2007),pp.1-8.

[6] Gao S.B, Yan Y.Y. *Saliency and Active Contour based Traffic Sign Detection*, Journal of Information and Computing Science, 2012,7(3):235-240

[7] B. W. Tatler, *The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions*, Journal of Vision, 7(14)(2007), pages 1-17.

[8] J. Ning, L. Zhang, D. Zhang and C. Wub, *Interactive image segmentation by maximal similarity based region merging*. Pattern Recognition, 2010,43(2): 445-456.