

A Method of Traffic Lights Detection Based on Visual Selective Attention

Xueling Wang, Youfu Wu^{a,*}, Peng Yang and Zusheng Chen

School of Science, Guizhou Minzu University, Guiyang 550025, China

^aYoufu_wu_64@aliyun.com

Keywords: Visual attention; Image enhancement; Region of interest; Region labeling.

Abstract. In this paper, the candidate target traffic lights under different backgrounds were automatically detected with the visual selective attention mechanism. The region of interest in color image was projected into the gray, and the gray image enhancement was done through top-hat algorithm. The binary image was obtained by using Otsu's method. And finally the logical 'and' operation was carried out for the binary image and the binary image which was obtained according to threshold segmentation in HSI space. The detection on depth information feature of the lights surrounded by the dark boundary was applied to the interested target candidate region. Experimental results show that this method was robust.

Introduction

Humans have begun to study intelligent traffic tools such as the unmanned aircraft, unmanned vehicle, etc. The previous focus on video image processing of traffic lights was static and fixed camera [1]. The present studies mainly focused on traffic lights detection with the dynamic and complex background. These methods were to extract color or shape features from the viewpoint of machine, or combine color and shape features for the target detection. These methods were only used for circular and arrow traffic lights detection. Hyun-Koo Kim [2] put forward to RGB color space conversion to YCbCr space and had circular lights candidate regions positioned through color threshold. Then transformed circular lights into binary, and removed noise using light shape feature. However, this method for the complex background and their own complex identifier of traffic lights, the detection rate and speed were yet to be improved. Jianwei Gong [3] transformed the image into HSV space, extracted circular light candidate region through color threshold segmentation and morphological operation. RGB color space conversion to HSI color space was proposed by Chunxiang Wang [4], combining color feature and circular light shape feature to remove noise and verify candidate regions of the traffic lights using the normalized template matching cross-correlation technique. This paper deals with human visual system characteristics to simulate visual attention. The experimental results show that this method can simply and effectively detect the traffic lights, and may meet the real-time requirement of detection system.

Visual selective attention

Visual selective attention study is interdisciplinary disciplines combining psychology with physiology. There are two mature selective attention mechanism of the human visual system at present: bottom-up and top-down analysis. The former belongs to the primary stage of cognitive process, while the latter belongs to the advanced stage of cognitive process. Two kinds of mechanisms are analysis of salient, and prominent parts of the scene can be interpreted as salient. Computational vision attention mechanism research is mainly focused on bottom-up integration of a variety of characteristics of salient regions detection. According to the salient feature of the region of interest, the candidate region image was selected to realize visual attention. In fact, human visual attention is affected by a lot of factors such as movement, color and shape. We determined a rough region of interest containing objects to be measured, and extracted salient features as possible, such as contrast, size, shape, color, so as to realize the rapid detection of target.

Determination of the probable region of interest

Human eyes can unconsciously have interest in a region or a few regions of computer graphics at the time of observation under the interaction with visual selective attention mechanism. These regions are called region of interest (ROI). Usually, the visual quality of the whole image depends largely on ROI and it is often difficult to be aware of the degraded uninterested region, because of less impact on the visual quality of the whole image. If we only find the treatment of these areas, it can greatly improve the efficiency and accuracy of image processing and analysis, so that we will achieve the realization of coarse-to-fine attention regions. In this paper, camera was fixed into the angle of 10~15 degrees with the horizontal direction on the roof. ROI was locked at 1/4~3/4 of the scene from bottom to top, marked with blue framework in the origin image.

Detecting the salient of gray scale contrast

The following work will be on the target of interest in ROI to further lock so as to realize attention regions from course to fine. The main step was using Top-hat algorithm to extract gray scale contrast feature in ROI, and Top-hat operation was a kind of non-linear spatial domain filter based on local image structure characteristics. It can reflect differences of potentially salient objects and surrounding backgrounds at two aspects of local spatial distribution and gray scale difference, in line with the characteristics of early stage visual attention conversion to visual stimulation. The detailed algorithm was dilation after erosion for the original image f , removal of bright features of structural elements, original image subtraction image with open operation.

$$\begin{aligned} \text{Tophat}(f) &= f - ((f \oplus s) \ominus s) \\ f \oplus s &= \max \{ f(x-x_0, y-y_0) - s(x_0-y_0) \mid (x_0, y_0) \in D_s \} \\ f \ominus s &= \max \{ f(x-x_0, y-y_0) + s(x_0-y_0) \mid (x_0, y_0) \in D_s \} \end{aligned} \quad (1)$$

Where x, y is pixel position of input image, x_0, y_0 is coordination of structural elements, D_s is definition domain of structural elements. $(f \oplus s) \ominus s$ is s do open operation with f , namely, erosion after dilation. $f \oplus s$ denotes s dilation operation for f , $f \ominus s$ denotes s erosion operation for f .

After the aforementioned process salient bright region in the dark background was extracted, Otsu's^[5] algorithm was used to get the binary image S_0 . This algorithm was an adaptive threshold ascertain. Firstly set threshold as T , image was divided into two R_1 and R_2 , have $R_1 = \{f_1 \mid f_{\min} \leq f \leq T\}$ (where f is gray scale pixel value). If N_i is number of the gray scale I , then total $R_2 = \{f_2 \mid T < f \leq f_{\max}\}$

pixel number $N = \sum_{i \in [f_{\min}, f_{\max}]} N_i$, the probability of every gray scale is $P_i = N_i / N$, so the total probability of

gray scale in R_1 range is $\rho_1 = \sum_{i \in [f_{\min}, T]} P_i$, and its expectation is $E_1 = \sum_{i \in [f_{\min}, T]} i \times P_i / \rho_1$. The total probability of

gray scale in R_2 range is $\rho_2 = \sum_{i \in [T+1, f_{\max}]} P_i$, and its expectation is $E_2 = \sum_{i \in [T+1, f_{\max}]} i \times P_i / \rho_2$. Expectation of the

whole image is $E = \sum_{i \in [f_{\min}, f_{\max}]} i \times P_i = \sum_{i \in [f_{\min}, T]} i \times P_i + \sum_{i \in [T+1, f_{\max}]} i \times P_i = \rho_1 E_1 + \rho_2 E_2$, and variance between R_1 and R_2 range

can be defined as: $\sigma^2(T) = \rho_1(T)(E - E_1)^2 + \rho_2(T)(E - E_2)^2$, when σ^2 is max, optimum segmentation can be attained. Fig.1 (a) is the input original image, wherein the blue rectangle for the rough region of interest set, Fi. 1 (b) salient gray contrast effect after top-hat transformation and Otsu's process for the traffic lights.

The input color RGB image consisting of hue, saturation and intensity three elements was converted to HSI color space model. The usage of separable HSI space characteristics of intensity, hue and saturation was applied to extrude their salient of candidate target. Detailed realization HIS color model using Eq. (2) and Eq. (3).

Let's

$$\theta = \cos^{-1} \left\{ \frac{[(R-G)+(R-B)]/2}{\sqrt{(R-G)^2+(R-B)(G-B)}} \right\} \quad (2)$$

Then H, S, I components were

$$H = \begin{cases} \theta & B \leq G, \\ 360 - \theta & B > G \end{cases}, S = 1 - \frac{3 \cdot \min(R, G, B)}{R + G + B}, I = \frac{R + G + B}{3} \quad (3)$$

Experience threshold was drawn by characterizing each component as histogram statistics collection color samples of traffic light color feature respectively. For example, we fix the threshold value, TH_1, TH_2, TH_3, TH_4 is hue threshold, TS is saturation threshold, TI is intensity threshold.

$$f = \begin{cases} f_R & H_R \leq TH_1 \quad \text{or} \quad H_R \geq TH_2 \quad \text{and} \quad S_R > TS \quad \text{and} \quad I_R > TI \\ f_G & H_R > TH_3 \quad \text{and} \quad H_R < TH_4 \quad \text{and} \quad S_R > TS \quad \text{and} \quad I_R > TI \\ f_Y & H_R > TH_1 \quad \text{and} \quad H_R < TH_3 \quad \text{and} \quad S_R > TS \quad \text{and} \quad I_R > TI \end{cases} \quad (4)$$

Among them, f_R, f_G, f_Y is red, green and yellow traffic lights candidate region.

Detecting the salient of hue, intensity and saturation

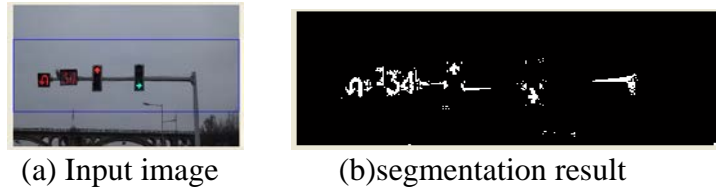


Fig. 1: An experiment result.

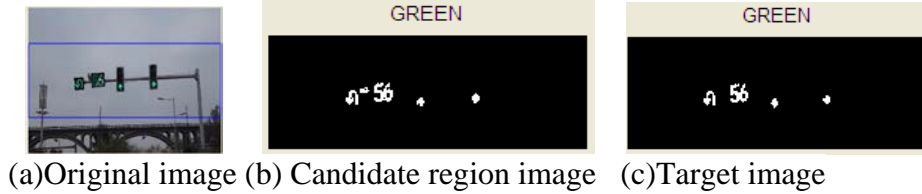


Fig. 2: An experiment result with noise.

Combination and decision

Based on the Itti's ^[6] proposed computational model of visual attention, we present algorithm combining with the contrast, hue, intensity and saturation, and fusing local depth information. The combination method is f_R, f_G, f_Y with $S0$ making logic 'and' operation respectively.

$$\begin{cases} F_R = f_R(x, y) \cdot S0(x, y) \\ F_G = f_G(x, y) \cdot S0(x, y) \\ F_Y = f_Y(x, y) \cdot S0(x, y) \end{cases}$$

Among which x, y is corresponding coordination of two sets of images. Because of binary value image, its pixel value is 0 and 1. Logic and operation rules were $0 \wedge 0 = 0, 0 \wedge 1 = 0, 1 \wedge 0 = 0$, and $1 \wedge 1 = 1$. Therefore, through multiplying the two images corresponding points can be get rid of a majority of noise. The noise removal image was obtained after logic and operation, as shown in Fig.3, it can be seen that the test results are satisfactory. If the candidate region contains a similar light or noise block as shown in Fig.2, we selected the local depth information that the target in original image surrounded by black characters, excluded non-target area and to realize attention focus rapid detection.

With the collection of Huaxi Avenue traffic light video and simulation of unmanned vehicle to collect video with six million pixel camera on the roof, a PC machine was used to test algorithm performance by cutting into the video with size 640x480. Kim's and Wang's algorithms were based on detection of color feature and circular light shape feature for candidate target in the entire scene, and the noise was removed according to the differences between the circular light shape and noise shape, such as the use of target area, the ratio of length to width, and the outer rectangle filling ratio and boundary characteristics. Due to complex shape of digital traffic lights, their methods cannot be

applied to remove noise for digital and arrow lights. But our method can be feasible for digital and arrow lights with higher detection rate. The detection time can meet the real-time requirements compared with the methods of Kim's algorithms as Fig.4 and Wang's algorithms as Fig.5.

Summary and conclusion

We have presented a visual selective attention mechanism. Firstly we extracted gray scale intensity contrast feature using top-hat algorithm, and then combined with the salient feature of HSI color space of hue, intensity and saturation, fusion of local depth information. Object in the vicinity of original image was used as the feature of black pixel in order to realize accurate detection of target traffic light. Finally statistical methods were used to verify the output of every frame. The experimental results show that this method can accurately detect the traffic light. But for the target and noise of adhesion, the algorithm cannot remove adhesion noise, and not detect ROI and objects with similar color and shape feature in dark scene. How to make better use of local depth information and to reduce the error rate is the focus of future research.

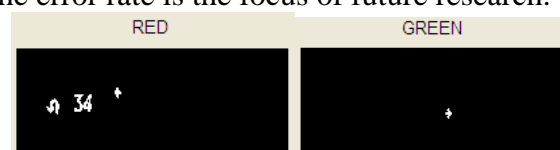


Fig. 3 Test results of Fig. 1



Fig.4: Test results of Fig.1 using Kim's method



Fig.5:Test results of Fig. 1 using Wang's method

Acknowledgment

Supported by Guiyang Science Technique Department (2010, 2012), Science Technique Department of Guizhou (2007, 2009, 2012) and Guizhou Education Department (2011, 2013["125project"]).

References

- [1] Y.C. Chung, J.M. Wang, S.W. Chen. A vision-based traffic light detection system at intersections. *J. of Taiwan Normal Univ.: Math. Sci. & Tech.* 47 (2002) 67-86.
- [2] H.K. Kim, J. H. Park, H.Y. Jung. Effective traffic lights recognition method for real time driving assistance system in the daytime. *World and Acad. of Sci. Eng. and Tech.*, 11(2011)311-314.
- [3] J.W.Gong, Y.H.Jiang, G.M.Xiong, C.H.Guan, G.Tao, H.Y. Chen. The recognition and tracking of traffic lights based on color segmentation and CAMSHIFT for intelligent vehicles. 2010 IEEE Intelligent Vehicles Symposium (IV), USA, (2010) 431-435.
- [4] C.X. Wang, T. Jin, M. Yang, B. Wang. Robust and real-time traffic lights recognition in complex urban environments. *Int. J. of Comput. Intelligence Sys.*, 4(2012)1383-1390.
- [5] N. Otsu. A threshold selection method from gray-level histogram. *IEEE Trans. on Sys. Man and Cybernetics*, 9(1979)62-66.
- [6] L. Itti, C. Koch, E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Anal. & Mach. Intelligence*, 20(1998) 1254-1259.