Visual Attention Saliency Model for License Plate Location

Boyu Gu

Changchun University of Science and Technology
guboyu1101@163.com

Abstract

In order to locate the license plate in noisy background or various illumination conditions, a license plate location model based on visual attention is proposed, which imitates the biological mechanism of human visual. Four major visual sensitive characteristics of Chinese license plate are applied. First, the features of intensity, color and texture are extracted and integrated to structure the saliency map; second, the obtaining the optimal threshold based on the shape feature for foreground segmentation; at last, support vector machine is applied to eliminate the false plates in the image. Experimental results show that the proposed model is adaptive to various illumination conditions, complicated background, and resistant to noise.

Keywords: visual attention, saliency model, license plate location, various illumination conditions, noisy background

1. Introduction

Automatic license plate recognition technology has numerous important applications in people's daily life [1]. License plate location is a crucial stage in a license plate recognition system, which directly influences the performance of the subsequent process [2].

Now, morphology, edge detection, and statistical learning are three mainstream methods for license plate location. However there are two main issues in aforementioned methods: various illumination conditions and noisy background. Human and other primates have a remarkable ability to interpret complex scenes in real time, which is invariant to illumination and noisy.

Vision bionics has broad application prospects in the field of machine vision and image understanding [3]. In human vision system, intermediate and higher visual processes appear to select a subset of the available sensory information before further processing [4], this selection appears to be implemented in the form of a spatially circumscribed region of the visual field, which is called "focus of attention". Information from only a small region of the whole visual field can progress through the cortical visual hierarchy. The attended region is selected through dynamic modifications of cortical connectivity or through the establishment of specific temporal patterns of activity [5]. Itti model [6] is a visual saliency model proposed by Itti, *et al.,* which is similar to the biological mechanism of human visual attention. The model builds on the neuronal architecture of primate visual system proposed by Koch, *et al.*, [7], it is related to the so-called "feature integration theory", explaining human visual attention strategies [8].

In this paper, a license plate location adaptive to various illumination conditions and noisy background is proposed, which is proposed based on Itti saliency model. The interested objectives are license plates; however, the Itti model can only extract the salient objectives regardless of whether they are the interested objectives. Therefore, the proposed model builds on the biological mechanism of human searching for the license plates in image; it structures the saliency map according to four major characteristics of Chinese license plate which are intensity, color, texture

and shape. In primates, the saliency map is believed to be located in the posterior parietal cortex [9] as well as in the various visual maps in the pulvinar nuclei of the thalamus [10]. Experimental results show that the proposed model is not constrained by the amount and position of license plates, adaptive to complicated background, resistant to noise.

2. License Plate Location Model

The frame of the proposed license plate location model is shown in Figure 1, there are four major stages in the proposed model, which are 1) early visual features extraction, 2) visual saliency map structuring, 3) candidate regions segmentation, and 4) license plate selection
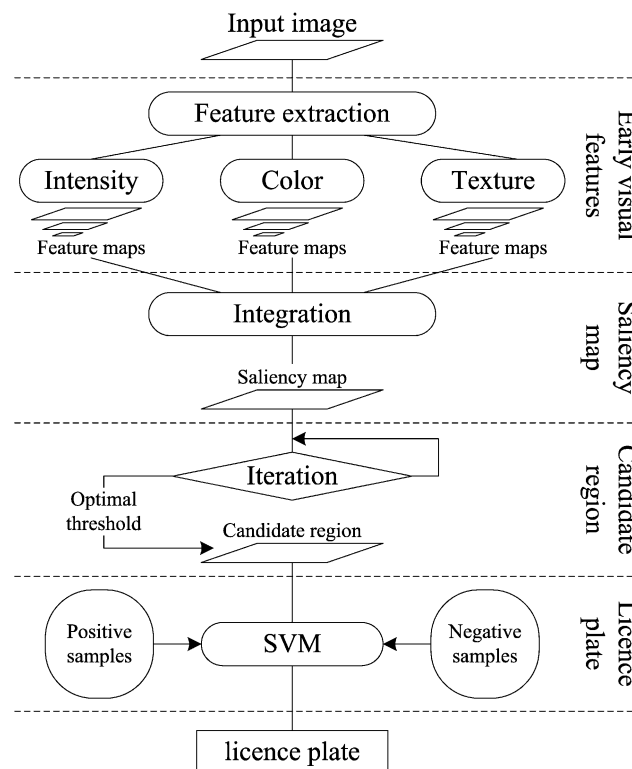
Figure 1. Frame of the License Plate Location Model

It can be known from Figure 1, the first stage in the model is the "pre-attention" phase, which is located in the bottom hierarchy of the visual attention mechanism; the second and third stages describe the "attention" phase; and the last stage is the "post-attention" phase at the top hierarchy, the positive and negative samples are akin to the memories and knowledge for information comprehension.

In the process of constructing the saliency map, some operations and definitions are introduced to mathematically describe the above-mentioned mechanisms. 1) Feature pyramid: progressively subsample the input image to create a 9-level image pyramid (1:1~1:256, the level $\sigma \in \{0...8\}$), then, the features of visual sensitive characteristics are extracted in every level of the image pyramid to form the feature pyramids. 2) Center-surround operator Θ: Interpolate images to the finer scale and point-by-point subtraction. And for a feature F, the feature map is defined as $F(c,s)=|F(c)\Theta F(s)|$, where the center and surround pixel are at the cth and sth level of the feature pyramids, $c \in \{2,3,4\}$, $s=c+\delta$, $\delta \in \{3,4\}$. 3) Normalization function $N(\cdot)$: normalize the values in the map to a fixed range $[0,M]$, then find the location of the global maximum M of the map and compute the

average m' of all its other local maxima; and globally multiplying the map by $(M\text{-}m')^2$. 4) across-scale addition operator $\oplus$: scale the maps to the size of 4th level in pyramid, and add together point-by-point.

2.1. Early Visual Features

The early visual features are the most visual sensitive characteristics of license plate. First, form the feature pyramids, the level $\sigma \in \{0\ldots8\}$; then, obtain the feature maps by using the center-surround operator.

2.1.1. Intensity Feature Map: suppose that r, g, b are the input value of each color channel of RGB image, according to the well known psychology formula, the Intensity feature is:

$$I = 0.2989r + 0.5870g + 0.1140b \tag{1}$$

Since high contrast regions (such as a white plate on a black vehicle) in the image are easily captured by human eyes, the intensity feature map is:

$$I(c,s) = |I(c)\Theta I(s)| \tag{2}$$

Where the level of center pixels $c \in \{2,3,4\}$, and the level of surround pixels $s=c+\delta$, $\delta \in \{3,4\}$.

2.1.2. Color Feature Map: first, build 4 new color channels RGBY (red, green, blue, yellow):

$$\begin{cases} R = r - (g+b)/2 \\ G = g - (r+b)/2 \\ B = b - (r+g)/2 \\ Y = (r+g)/2 - |r-g|/2 - b \end{cases} \tag{3}$$

The color feature map is formed by the extended blue and yellow color-opponents, which can be described as:

$$\begin{cases} B(\Gamma,c,s) = |(B(c) - \Gamma(c))\Theta(B(s) - \Gamma(s))| \\ Y(\Gamma,c,s) = |(B(c) - \Gamma(c))\Theta(B(s) - \Gamma(s))| \end{cases} \tag{4}$$

Where Γ is the color channel, for the blue color-opponents $B(\Gamma,c,s)$, $\Gamma \in \{R,G,Y\}$; and for the yellow color-opponents $Y(\Gamma,c,s)$, $\Gamma \in \{R,G,B\}$.

2.1.3. Texture Feature Map: texture is an important feature for license plate location, especially in the conditions that the ability of visual perception for color is decreased, such as night or low light circumstances. The Gabor filter is akin to the way of image texture processed by human visual system. First, the color channels are Gabor filtered:

$$\Gamma'(\theta) = \gamma(\theta) * \Gamma \tag{5}$$

Where $\Gamma \in \{B,Y\}$, $\gamma(\theta)$ is the Gabor filter in the direction of θ; human eyes are most sensitive to the texture information in the directions of horizontal and vertical, so $\theta \in \{0°,90°\}$.

The energy concentration of texture feature can be analyzed by image signature [11], which exists in the discrete cosine transform (DCT) of an image. It is similar to the well-known K-L transform. The saliency regions can be extracted by DCT since the visual sensitive information is distributed concentrated in frequency domain. First, the image signature is:

$$\text{ImgSig}(\Gamma') = \text{Sign}[\text{DCT}(\Gamma')] \tag{6}$$

Where Sign is the sign operator, DCT is discrete cosine transformation. Suppose that IDCT is inverse discrete cosine transform, the image can be reconstructed:

$$\hat{\Gamma} = \mathrm{IDCT}[\mathrm{ImgSig}(\Gamma')] \tag{7}$$

The region of energy concentration can be extracted by image signature and reconstruction. To ensure the spatial continuity of the saliency objectives, the reconstructed image need to be smoothed:

$$\widetilde{\Gamma} = g * (\hat{\Gamma} \circ \hat{\Gamma}) \tag{8}$$

Where g is Gaussian filter, $\circ$ is Hadamard product operator. And the texture feature map is:

$$T(\Gamma, \theta, c, s) = | \widetilde{\Gamma}(\theta, c) \Theta \widetilde{\Gamma}(\theta, s) | \tag{9}$$

2.2. Visual Saliency Map

To extract the saliency regions, three conspicuity maps need to be created by the feature maps, and then a saliency can be structured by integrating the conspicuity maps.

2.2.1. Conspicuity Maps: after the early visual features extraction, there are 6 intensity feature maps, 36 color feature maps and 24 texture feature maps. Combine the feature maps, and the intensity conspicuity map is:

$$\bar{I} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(I(c,s)) \tag{10}$$

The color conspicuity map is:

$$\overline{C} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [\sum_{\Gamma} N(B(\Gamma,c,s)) + \sum_{\Gamma} N(Y(\Gamma,c,s))] \tag{11}$$

And the texture conspicuity map is:

$$\overline{T} = \sum_{\theta} N[\sum_{\Gamma} \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(T(\Gamma,\theta,c,s))] \tag{12}$$

2.2.2. Saliency Map: the confidence coefficients of the features in various circumstances are obtained by a sample set containing 600 images (376 in day, 224 in night). First, achieve the accuracy rates r_I, r_C, and r_T, by means of only one kind of early visual feature. The confidence coefficient of feature F ($F \in \{I,C,T\}$) is:

$$R_F = \frac{r_F}{r_I + r_C + r_T} \tag{13}$$

The saliency map is:

$$S = R_I \bar{I} + R_C \overline{C} + R_T \overline{T} \tag{14}$$

The confidence coefficients in day are R_I=0.143. R_C=0.416, R_T=0.441; and in night are R_I=1.97, R_C=0.319, R_T=0.484. in the day, the light condition is ideal, the color is easy to discriminated by eyes; and in night, since human eyes are less sensitive to colors, the confidence coefficient of color is reduced.

2.3. Candidate Regions

Segment the saliency map, and each connected area is a candidate region. The threshold for foreground segmentation is obtained by an iterative process based on shape context [12] according to the shape feature of license plate.

Step 1: initialize the threshold φ and its maximum and minimum. φ_{min}=0.1, φ_{max}=0.9, Increment τ=0.01.

Step 2: obtain a point set α by sampling the boundary of a connected area. Then, create a rectangle circumscribed to the connected area, the length-width ratio of which is 22:7. Also, a point set β is obtained by sampling the rectangle.

Step 3: place α and β to polar coordinates, divide the radius r to 5 areas, angle θ to 12 areas, and relative vectors are formed by a point and all neighboring points. Then, compute the histogram h of relative vectors in all 60 (r,θ) areas, that is the shape context of the point.

Step 4: compute the cost of the two points α_i and β_j:

$$C_{ij} = C(\alpha_i, \beta_j) = \frac{1}{2} \sum_{k=1}^{60} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \qquad (15)$$

Where $h_i(k)$ and $h_j(k)$ are the histogram of α_i and β_j in kth (r,θ) area, $\alpha_i \in \boldsymbol{\alpha}$, $\beta_j \in \boldsymbol{\beta}$. The costs of all points form a cost matrix, and run a maximal-matching algorithm (e.g. Hungary algorithm) and find the maximal-matching $\pi(i)$ to minimizing the overall cost:

$$H(\pi) = \sum_i C(\alpha_i, \beta_{\pi(i)}) \qquad (16)$$

Then, the cost of two shapes $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is $D_{\alpha\beta}=[H(\pi)|\pi(i)=\text{argmin}(H(\pi))]$.

Step 5: find the minimum cost $D_{min}(t)$ of all connected areas, where t is the present iterations. If $|D_{min}(t)-D_{min}(t-1)|>\eta$, set $\varphi=\varphi+\tau$, and jump to Step 2 (here $\eta=10^{-5}$); or else, stop the iteration and jump to Step 6.

Step 6: there are saliency regions match the shape of license plate, and every connected area is considered as a candidate region.

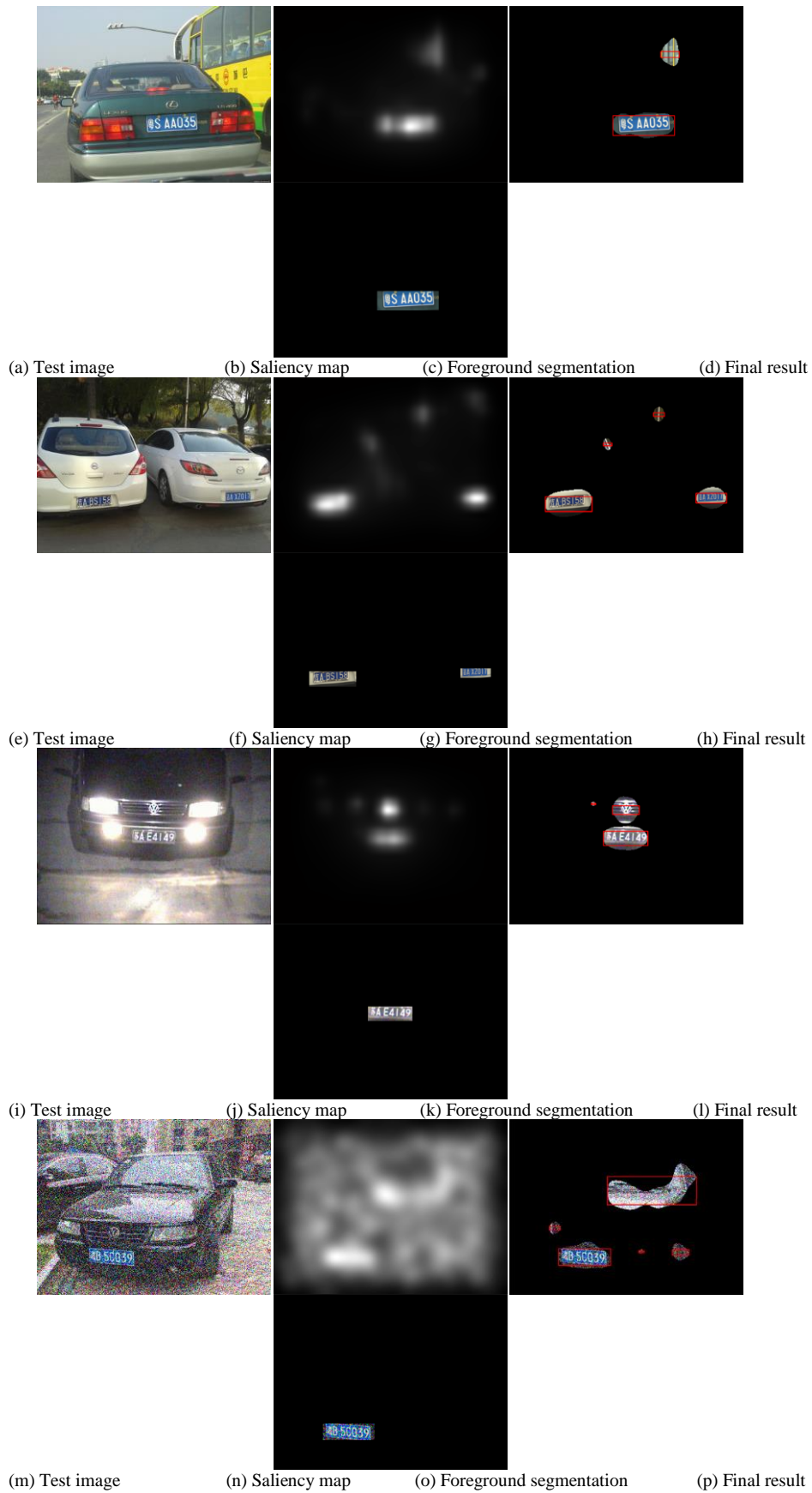### 2.4. License Plate Regions

The support vector machine (SVM) classifier is applied to differentiate the positive and negative samples, and eliminate the false plates. In the proposed model, the RBF kernel function is applied to compute the inner product. The training set is consisted of 600 positive samples and 600 negative samples, the positive samples contain license plates in various circumstances, and the negative samples contain the front radiators, car lights, logo of vehicle manufacturers, and advertisements on automobile, *etc*. First, all candidate regions are scaled to 220×70; then, input the candidate regions to the SVM classifier; at last, every region classified as a positive sample is considered as a license plate.

3. Experimental Results

The proposed model is tested with images collected from toll stations, parking lots, gasoline stations, and internet. Figure 2 shows the results of some representative images.

As shown in Figure 2, (a)(e)(i)(m) are test images, (a) is an image with ideal circumstance; (e) contains 2 license plates; (i) is captured in night; and Gaussian noise (mean is 0, variance is 0.1) is added in (m). (b)(f)(j)(n) are saliency maps structured by the proposed model. (c)(g)(k)(o) are the results of foreground segmentation, the black area is background, and the areas marked by red rectangles are candidate regions. And (d)(h)(l)(p) are the final results of license plate location.

In Figure 2, the (a) is a relatively ideal image; (e) is an image contains two license plates; (i) is a low light image; and in (m), Gaussian noise is added in the image to simulate the noisy and complicated background. It can be known that the proposed model is not constrained by the amount and position of license plates, adaptive to variant illumination conditions, complicated background, and resistant to noise.

(a) Test image   (b) Saliency map   (c) Foreground segmentation   (d) Final result

(e) Test image   (f) Saliency map   (g) Foreground segmentation   (h) Final result

(i) Test image   (j) Saliency map   (k) Foreground segmentation   (l) Final result

(m) Test image   (n) Saliency map   (o) Foreground segmentation   (p) Final result

Figure 2. Results under Different Conditions

Table 1 shows the accuracy rate comparison, the proposed algorithm is compared with three wildly used efficient algorithms. The test set contains 1549 640×480 images, which is divided into 2 categories, 932 in day, and 617 in night.

Table 1. Accuracy Rate

Method	Accuracy rate（%）		
	Day	Night	Total
SVM classifier	89.59	85.25	87.86
Adaboost classifier	93.13	91.73	92.58
BP neural network	91.31	88.98	90.34
Proposed algorithm	94.10	92.54	93.48

And also, in order to confirm the robustness in noisy background, the Gaussian noise is added in all test images. The result is shown in Table 2.

Table 2. Accuracy Rate of Noise Images

Method	Accuracy rate（%）		
	Day	Night	Total
SVM classifier	82.30	81.04	81.79
Adaboost classifier	88.95	85.58	87.60
BP neural network	85.73	82.50	84.44
Proposed algorithm	89.81	86.22	88.38

From Table 1 and Table 2, it can be known that the proposed approach can improve the accuracy rate. Since the proposed model is similar to the biological mechanism of human visual attention, and the visual sensitive features of license plates are emphasized in the process of structuring the saliency map, it performs more robust especially in circumstances such as night, low light, low contrast, and complicated background. The robustness is improved in both variant illumination conditions and noisy images.

4. Conclusions

A bionic saliency model for license plate location is proposed, which imitates the biological mechanism of human searching for the license plates in image. The proposed model extracts visual sensitive characteristics of Chinese license plates since the interested objectives are only license plates. Experimental results show that the proposed model is adaptive to complicated background, resistant to noise, and performs robust in variant illumination conditions.

References

[1]. R. Cheng and Y. Bai, "A novel approach for license plate slant correction, character segmentation and chinese character recognition", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 7, no. 1, **(2014)**.

[2]. C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, I. D. Psoroulas, V. Loumos and E. Kayafas, "License plate recognition from still images and video sequences: A survey", Intelligent Transportation Systems, IEEE Transactions on, vol. 9, no. 3, **(2008)**.

[3]. Y. Zhou, "Fast and Robust Stereo Vision Algorithm for Obstacle Detection", Journal of Bionic Engineering, vol. 5, no. 3, **(2008)**.

[4]. J. K. Tsotsos, S. M. Culhane, W. Y. Kei Wai, Y. Lai, N. Davis, F. Nuflo, "Modeling visual attention via selective tuning", Artificial intelligence, vol. 78, no. 1, **(1995)**.

[5]. B. A. Olshausen, C. H. Anderson and D. C. Van Essen, "A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information", The Journal of Neuroscience, vol. 13, no. 11, **(1993).**

[6]. L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis.", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 20, no. 11, **(1998)**.

[7]. C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry. Matters of Intelligence", Springer Netherlands, vol. 4, **(1987)**, pp. 115-141.

[8]. A. M. Treisman and G. Gelade, "A feature-integration theory of attention. Cognitive psychology", vol. 12, no. 1, **(1980)**.

[9]. J. P. Gottlieb, M. Kusunoki and M. E. Goldberg, "The representation of visual salience in monkey parietal cortex", Nature, vol. 391, no. 6666, **(1998)**.

[10]. D. L. Robinson, and S. E. Petersen, "The pulvinar and visual salience", Trends in neurosciences, vol. 15, no. 4, **(1992)**.

[11]. X. Hou, J. Harel and C. Koch, "Image signature: Highlighting sparse salient regions", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 1, **(2012)**.

[12]. S. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts.", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 4, **(2002)**.

Author

Boyu Gu He received B.Sc. and M.Sc. degrees in Information and Communication Engineering from Changchun University of Science and Technology, Changchun, China in 2007 and 2011, respectively. He is now pursuiting Ph.D. degree in the School of Electronics and Information Engineering, Changchun University of Science and Technology. His research interests are pattern recognition and intelligent system.