Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008

BOOSTED ROAD SIGN DETECTION AND RECOGNITION

SIN-YU CHEN AND JUN-WEI HSIEH

Department of Electrical Engineering, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li 320, Taiwan
E-MAIL:shieh@saturn.yzu.edu.tw

Abstract

This paper presents a boosted system to detect and recognize roads signs from videos. The system first uses the Adaboost algorithm to learn the visual characteristics of road sign. Then, a cascaded structure is then used to detect road signs from videos in real time. After detection, a rectification process is then applied for rectifying different skewed road signs into a normal one. Then, its all embedded texts can be more accurately recognized using their distance maps. On the map, a weighting function is used to balance the importance between a road sign's inner and outer feature so that its embedded characters can be more accurately recognized. Experimental results have proved the superiority of the proposed method in road sign recognition.

1. Introduction

Road sign recognition is an important and essential task in an intelligent driver support system. The texts embedded in a road sign usually carry much useful information like limited speed, guided direction, and current traffic situations for helping the drivers drive safely and comfortably. However, it is difficult to detect road signs directly from videos due to different changes of environmental condition. For example, a road sign will have different appearance changes including its lightings, colors, or shadows under different days, seasons, and weathers. In addition, for the camera mounted in front of a moving car, its perspective effects will make a road sign have different sizes, shapes, contrasts, and motion blurs. In some cases, it would be occluded with other objects like trees. To tackle the above problems, there have been many works [1]-[6] proposed for automatic road sign detection and recognition via a vision-based technique. Since a road sign's color is highly contrast to its background and its shape is often specially designed for message showing, we can divide these approaches into two categories, i.e., color-based and shape-based. For the color-based approach, in [1], Bénallal and Meunier found that the difference between R and G, and the difference between R and B channels can form two stable features in road sign detection. Escalera *et al.* [2] used a color threshoding technique to separate road sign regions from their backgrounds on the RGB space. In addition to the RGB space, other color spaces like YIQ and HSV are also useful for road sign detection. For example, in [3], Kehtarnavaz and Ahmad used a discriminant analysis on the YIQ color space for detecting various road signs from videos. Fleyeh [10] used an improved HLS (Hue, Lightness, Saturation) color space for detecting color road signs from road scenes. For the color-based approach, since a road sign uses different colors (like red, blue, or green) to demonstrate its functionalities like warning or direction, different detectors should be designed for tackling its color variations.

In addition to color, shape is another important feature for road sign detection. In [5], Barnes and Zelinsky adopted the radial symmetry feature for locating possible road signs and then verified them using a correlation technique. In [6], Piccioli *et al.* proposed a template matching scheme to search all possible road signs from images. In addition, Wu *et al.* [4] used a corner feature and a vertical plane criterion to cluster image data to different categories so that each road sign can be found. Moreover, Garcia-Garrido *et al.* [11] extended the Hough transform to find any curves in an image and thus circular and stop signs can be detected. Usually, a road sign's shape will change according to its message types. For a good shape-based approach, it should have good abilities to overcome the shape variations and occlusions especially when a moving camera is used for video acquisition.

Fig. 1 Flowchart of our proposed system.

This paper presents a novel system to detect and recognize roads signs from videos using the Adaboost algorithm and edge maps. Fig. 1 illustrates the flowchart of the proposed system to recognize road signs. First of all, this paper uses the Adaboost algorithm to learn the visual characteristics of road sign. Then, based on the learned integral features, a cascaded structure is then used to quickly detect road signs from videos. After detection, since a skewed road sign will be found, a rectification process is also applied to rectifying each skewed road sign so that its all embedded texts can be well recognized. To recognize the extracted road sign, the distance transform is used for converting it into a distance map. On the map, a weighting function is used to balance the importance between a road sign's inner and outer shapes and thus it can be more accurately recognized. Experimental results have proved the effectiveness and efficiency of the proposed method in road sign recognition.

2. Boosted Traffic Sign Detection

This section presents a learning method to learn a strong classifier from a set of training images for road sign detection. The learning method we use is the Adaboost algorithm which combines a set of 'important' weak classifier to form a strong classifier for object detection.

2.1 Integral Images and Features

978-1-4244-2096-4/08/$25.00 ©2008 IEEE

For the task of road sign detection, it is better to use region feature rather than pixel-level feature since region feature has more robustness than pixels. In this paper, we use the integral image to generate a bank of rectangle features for representing a road sign. Given an image I, its integral image $S(x, y)$ contains the sum of intensity values of pixels in I accumulated from the original (0, 0) to the pixel (x, y), i.e.,

$$S(x,y)=\sum_{i=0}^{x}\sum_{j=0}^{y} I(i,j) .$$

The integral image can be computed recursively, by $S(x,y)= S(x, y-1)+S(x-1, y)+I(x, y)-S(x-1, y-1)$, with the boundary condition: $I(-1, y) = I(x,-1) = I(-1,-1) = 0$. Clearly, the computation of $S(x,y)$ can be finished using only one scan over the input image I. Given a rectangle R bounded by (l, t, r, b), its sum of pixel intensity value can be very efficiently achieved by taking advantages of the integral image S. Like Fig. 2, the sum $F(R)$ of pixel intensities in R can be easily calculated with the form

$$\begin{aligned} F(R) &= (A+B+C+R)+A-(A+B)-(A+C) \\ &= S(r,b)+S(l,t)-S(l,b)-S(r,t). \end{aligned}$$

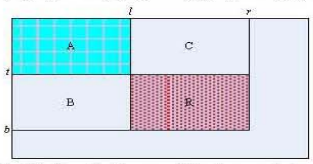
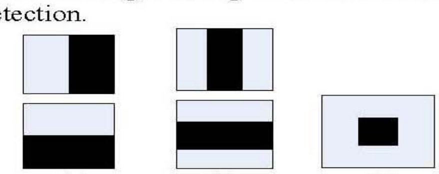


Fig. 2: Calculation of integral image.

Then, given a base training region, different combinations of rectangle can generate a set of simple rectangle features. Fig. 3 shows three kinds of rectangle features generated for road sign detection. Then, in what follows, the Adaboost algorithm is applied to learning a strong classifier for robust road sign detection.



(a) (b) (c)

Fig. 3: Different kinds of rectangle features.

2.2 Adaboost Algorithm

"Boosting"[12] is a learning algorithm to iteratively learn a strong classifier from a set of weak classifiers. A weak classifier uses a simple feature for determining positive samples from negative samples and is only required to be slightly better than chance. At each iteration, a "good" weak classifier is selected and added in turn to form a strong classifier which is a weighted sum of individual selected weak classifier. Assume we have m features extracted from a training sample. The selected weak classifier $h_t(x)$ is the one which has the minimum classification error ε_t when the feature f_t is selected from the m features. The weak classifier $h_t(x)$ determines whether a sample x is positive according to the following rule:

$$h_t(x): \quad x \cdot f_t > \theta_t ,$$

where $x \cdot f_t$ is a convolution result between x and f_t, and θ_t a threshold determined by minimizing the number of misclassified samples. At the tth iteration, the algorithm combines the weak classifiers h_1, ..., h_t to form the strong classifier H_t using weighted voting where h_t has the weight α_t. Thus, the tth strong classifier H_t is defined as

$$H_t = \sum_{i=1}^{t} \alpha_i h_i = H_{t-1} + \alpha_t h_t ,$$

where $\alpha_t = 0.5 \ln((1-\varepsilon_t)/\varepsilon_t)$. Assume that there are N training samples: $\{(x_1,y_1),...,(x_N, y_N)\}$, where $y_i = 1$ for positive examples, $y_i = -1$ for negative examples, and w_i for weighting the sample (x_i, y_i). Details of the Adaboost algorithm can be described as follows.

Adaboost Algorithm

Initially assign uniform weights $w_i^0 = 1/N$ for all x, and let T denote the maximum number of iteration.

At each iteration t:

1. Find the best weak classifier $h_t(x)$ which has a lowest error ε_t based on W^t;
2. $\alpha_t = 0.5\ln((1-\varepsilon_t)/\varepsilon_t)$;
3. Get the hypothesis $H_t = H_{t-1} + \alpha_t h_t$;
4. $w_i^{t+1} = \exp(-H_t(x_i)y_i)$ and normalize it such that $\sum_{i=1}^{N} w_i^{t+1} = 1$;

After training, $H(x_i) = sign\,[\sum_{t=1}^{T} \alpha_t\, h_t(x_i) - \sum_{t=1}^{T} \alpha_t\,]$ is the final hypothesis. Fig. 4 shows some of the important weak classifiers.

Fig. 4 Training results using Adaboost algorithm.

2.3 Cascade Structure of Strong Classifiers

In the previous section, the Adaboost algorithm improves the classification performance by iteratively adding features to the strong classifier. When more features are added, the adding technique will directly increase lots of computation time for the designed classifier to detect objects. However, if a cascaded structure is used for training the strong classifier, its efficiency will be significantly improved. The cascaded structure is introduced by Viola and Jones [13] and shown in Fig. 5. With the cascade, the original classification process will become a series of stage-wise rejections of non-object patterns. Usually, the number of non-object patterns is much larger than the number of object patterns. The cascade can quickly reject most of non-object regions but pass all of desired object patterns stage by stage. Each classifier in the cascade is required having a very high detection rate, but only a moderate false positive rate (e.g. 50%). An input region is passed from H_t to H_{t+1} if it is classified as an object, otherwise it is rejected. If the false-alarm rate of each classifier H_t is η_t (less than 50%) and independent to other classifiers, the accumulated false-alarm will become $\prod_{t=1}^{T}\eta_t$ after T weaker classifiers are applied. The detection error will be significantly reduced if more classifiers H_t are used. Since most non-object patterns are rejected quickly, the designed detector can extremely fast detect all desired objects.

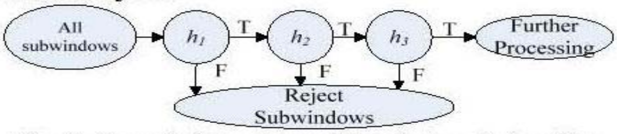
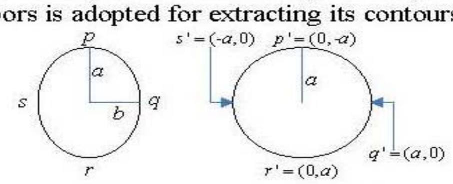


Fig. 5. Cascaded structure of the designed classifier.

3. Road Sign Rectification

In order to handle skewed road signs, a rectification procedure should be further applied for recognizing its embedded texts more accurately. Assume that R is the detected road sign. First of all, the Canny edge operator [7] is utilized for getting all its edge pixels. Then, the chain coding technique with 8 neighbors is adopted for extracting its contours.



(a) (b)

Fig. 6: Rectification of a circular road sign. (a) Input road sign. (b) Rectification result of (a).

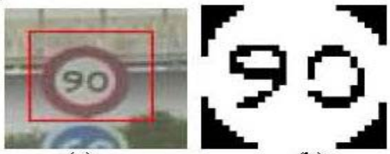
If R is a circular type, four control points will be selected for rectification. Like Fig. 6(a), p, q, r, and s are the most top, right, bottom, and left points of R, respectively. Considering them as control points, we can get the longest axis of R. Assume that its length is $2a$. Then, a projective transformation M can be found for rectifying R into a normal shape R' (see Fig. 6(b)). The relationship between R and R' can be defined as follows

$$x'=\frac{m_0x+m_1y+m_2}{m_6x+m_7y+1} \text{ and } y'=\frac{m_3x+m_4y+m_5}{m_6x+m_7y+1}, \quad (1)$$

where (x, y) is the coordinate of a pixel in R, (x',y') the coordinate of its corresponding point in R', and $(m_0,m_1,...,m_7)$ the parameters of the projective model M. Let p', q', r', and s' be the four corresponding points of p, q, r, and s in R', respectively, and have the following coordinates: $p'=(0,-a)$, $q'=(a,0)$, $r'=(0,a)$, and $s'=(-a,0)$. Given the above four pairs of correspondence, M can be solved by a linear method [8]. Once M is obtained, all the points in R can be transformed into R' using Eq.(1). The similar scheme is adopted for rectifying the rectangular and triangular Road Signs.

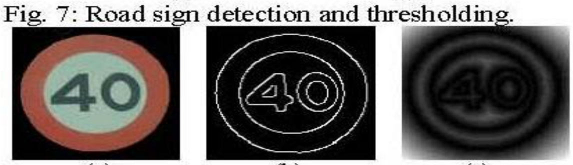
4. Traffic Sign Recognition

Once a road sign R is extracted, to recognize the texts in R, this paper uses a moment-based thresholding approach [9] to binarize R. After thresholding, the texts in R can be easily extracted using a connected component analysis. Like Fig. 7, (b) is the thresholding result of (a). Then, we can use the connected component analysis to extract different inner characters from (b). This section will the distance transform for recognizing all the extracted characters.



(a) (b)

Fig. 7: Road sign detection and thresholding.



(a) (b) (c)

Fig. 8: Result of distance transform. (a) Original Image. (b) Edge map. (c) Distance transform of (b).

Assume that B_R is a set of boundary pixels extracted from R. Then, the distance transform of a pixel p in R is defined as

$$DT_R(p)=\min_{q\in B_R} d(p,q), \quad (2)$$

where $d(p,q)$ is the Euclidian distance between p and q. In order to enhance the strength of distance changes, Eq.(2) is further modified as follows

$$\overline{DT}_R(p)=\min_{q\in B_R} d(p,q)\times\exp(\kappa d(p,q)), \quad (3)$$

where $\kappa=0.1$. Fig. 8 shows the result of the distance transform. (a) is an image R of road sign and (b) is its edge map. Fig. 8(c) shows the result of its distance transform. Thus, according to Eq.(3), a set F_R of contour features can be extracted from R. If we scan all pixels of R in a row major order, F_R can be then represented as a vector, i.e.,

$$F_R=[\overline{DT}_R(p_0),....,\overline{DT}_R(p_i),....], \quad (4)$$

where all p_i belong to R and i is the scanning index. To recognize the texts in a road sign more accurately, due to noise, their inner shapes play more important roles than their outer text patterns in road sign recognition. Thus, a new weight w_i which increases according to the distance between p_i and the center of R is included. Assume that O is the center of R and r_i is the distance between p_i and O, and the circumcircle of R has the radius z. Then, w_i is defined by the form:

$$w_i = \begin{cases} \exp(-|r_i|^2), \text{ if } r_i \le z; \\ 0, \quad \text{otherwise.} \end{cases} \quad (5)$$

With the help of w_i, Eq.(4) can be rewritten

$$\overline{F}_R = [w_0 \overline{DT}_R(p_0), \ldots, w_i \overline{DT}_R(p_i), \ldots]. \quad (6)$$

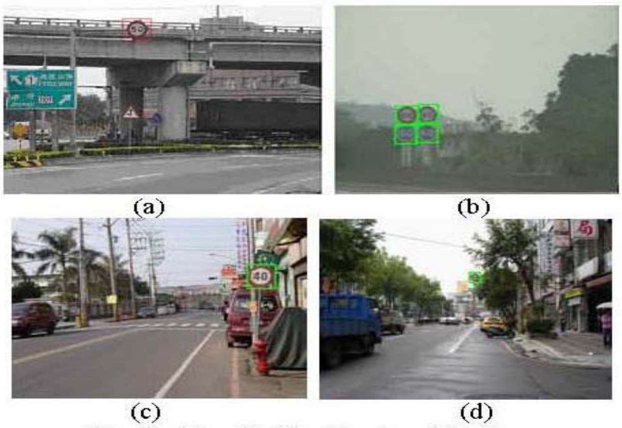
To recognize the type R_i, we extract its shape characteristics from a set of training samples. Assume there are N_i training templates collected for recognizing R_i. From the N_i samples, the mean μ_i and variance Σ_i of $\overline{F}_R$ can be then calculated. With μ_i and Σ_i, the similarity between a road sign candidate H and R_i can be measured by:

$$S(H, R_i) = \exp(-(\overline{F}_H - \mu_i)\Sigma_i^{-1}(\overline{F}_H - \mu_i)^t). \quad (7)$$

Then, H can be well classified into its corresponding category with the equation:

$$type(H) = \arg\min_{R_i} S(H, R_i). \quad (8)$$

5. Experimental Results



(a) (b)

(c) (d)

Fig. 9 Result of traffic sign detection.

Fig. 10 Result of rectangular sign detection.

To examine the performances of our proposed method, several video sequences collected from different highways and roads were used. The sequences were captured under different road and weather conditions (like sunny, cloudy) for examining the robustness of the proposed system. A database including more than four thousand images was collected and constructed for performance analysis. Fig. 9 shows the results of circular road sign detection. Fig. 10 shows the cases of rectangular road sign detection. Fig. 11 shows the results of road sign recognition. All the above results have proved that the proposed method is a robust, accurate, and powerful tool for road sign recognition.

Fig. 11 Result of traffic recognition.

References

[1] M. Bénallal and J. Meunier, "Real-time color segmentation of road signs," in *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering*, Canadian, vol. 3, pp.1823–1826, May 2003.

[2] A. D. L. Escalera, *et al.*, "Road Traffic Sign Detection and Classification," *IEEE Transaction on Industrial Electronics*, vol. 44, no. 6, pp.848-859, Dec.1997.

[3] N. Kehtarnavaz and A, Ahmad, "Traffic sign recognition in noisy outdoor scenes," *Proceedings of Intelligent Vehicles '95 Symposium*, pp.460-465, Sep. 1995.

[4] W. Wu, X. Chen, and J. Yang, "Detection of Text on Road Signs From Video," *IEEE Transactions on ITS*, vol. 6, no. 4, pp. 378- 390, Dec. 2005.

[5] N. Barnes and A. Zelinsky, "Real-time radial symmetry for speed sign detection," *Proceedings of IEEE Intelligent Vehicles Symposium*, Italy, pp. 566- 571, June 2004.

[6] G Piccioli et al., "Robust method for road sign detection and recognition," *Image and Vision Computing*, vol.14, no. 4, pp.209-223, 1996.

[7] J. Canny, "A computational approach to edge detection," *IEEE Transactions on PAMI*, vol. 8, no. 6, pp. 679-698, Nov. 1986.

[8] J.-W. Hsieh, "Fast Stitching Algorithm for Moving Object Detection and Mosaic Construction," *Image Vision and Computing*, vol. 22, no. 4, pp. 291-306, April 2004.

[9] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine vision*, Brooks/Cole Publishing Company, 1999.

[10] H. Fleyeh, "Color detection and segmentation for road and traffic signs," *IEEE Conference on Cybernetics and Intelligent Systems*, vol. 2, pp.809-814, 2004.

[11] M. A. Garcia-Garrido, M. A.Sotelo, and E. Martin-Gorostiza, "Fast traffic sign detection and recognition under changing lighting conditions," *Proceedings of the IEEE Conference on ITS*, pp.811-816, 2006.

[12] Freund, Y. Schapire, "Experiments with a new boosting algorithm," *Proceedings of the Thirteenth International Conference on Machine Learning*, pp.148-156, Bari, Italy, 1996.

[13] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp.137-154, May 2004.