Automatic Vehicle Detection using Dynamic Bayesian Networks for Aerial Surveillance

Sreepathi B.[1] & Sheela B. P.[2]

Abstract: **The advancement of computer technology and increasing needs of social security, studies of target object detection in aerial surveillance using image processing techniques are growing more and more important. These technologies can be employed in various applications, such as gathering enemy information for military purpose and searching for missing people in mountain areas. In an automatic vehicle detection system for aerial surveillance background colors are eliminated and then features are extracted. This system considers features including vehicle colors, edges and local feature points. For vehicle color extraction, system utilizes color transform to separate vehicle colors and non-vehicle colors effectively. For edge detection, system applies moment-preserving method to adjust the thresholds for canny edge detector automatically, which increases the adaptability and accuracy. A Dynamic Bayesian Network (DBN) is constructed for classification purpose. Based on the features extracted, a well trained DBN can estimate the probability of a pixel belonging to a vehicle or not. However, the features are extracted in a neighborhood region of each pixel. Therefore, the extracted features comprise not only pixel-level information but also relationship among neighboring pixels in a region.**

Key words: **vehicle detection, Aerial Surveillance, Dynamic Bayesian networks(DBNs),support vector machine (SVM)**

1. INTRODUCTION

The recent growth in the number of vehicles on the roadway network has forced the transport management agencies to rely on advanced technologies to take better decisions. In this perspective aerial surveillance has better place nowadays. Aerial surveillance provides increased monitoring results in case of fast-moving targets because spatial area coverage is greater. Thus aerial surveillance is supplement for ground-plane surveillance systems. One of the main topics in intelligent aerial surveillance is vehicle detection and tracking. The difficulties involved in the aerial Surveillance includes the camera motions such as panning, tilting and rotation. Also the different camera heights largely affect the detection results. Vision based techniques is one of the most common approach to analyze vehicles from images or videos.

1. *Information Science & Engineering Dept ,VTU*
sreepathib@gmail.com

2. *RYMEC, Cantonment Bellary*
sheelabp@gmail.com

The view of vehicles will vary according to the camera positions, lighting conditions and background situations.

The existing vehicle detection techniques are based on a large variety of techniques. The system, proposed by Hsu-Yung Cheng [1] escaped from the stereotype and existing frameworks of vehicle detection in aerial surveillance, which are either region based or sliding window based. We design a pixelwise classification method for vehicle detection. Hsu-Yung Cheng proposed Hierarchical model proposed by Hinz and Baumgartner [2] which describes different levels of details of vehicle features and detection method based on cascade classifiers has the disadvantage of lots of miss detections. Vehicle detection algorithm based on symmetric property [3] of car shapes is prone to false detections. The high computational complexity of mean-shift segmentation algorithm is a major concern in the existing methods. This technical report provides a survey on the existing methods which to an extent overcomes the disadvantages mentioned earlier. One method utilizes color transformation in case of still images and an approach tends to utilize wide area motion imagery.

Lin et al.[4] proposed a method by subtracting background colors of each frame and then refined vehicle candidate regions by enforcing size constraints of vehicles. However, they assumed too many parameters such as the largest and smallest sizes of vehicles, and the height and the focus of the airborne camera. Assuming these parameters as known priors might not be realistic in real applications. In [4], the authors proposed a moving-vehicle detection method based on cascade classifiers.

A large number of positive and negative training samples need to be collected for the training purpose. Moreover, multi scale sliding windows are generated at the detection stage. The main disadvantage of this method is that there are a lot of miss detections on rotated vehicles. Such results are not surprising from the experiences of face detection using cascade classifiers. If only frontal faces are trained, then faces with poses are easily missed. However, if faces with poses are added as positive samples, the number of false alarms would surge. Choi and Yang [5] proposed a vehicle detection algorithm using the symmetric property of car shapes. However, this cue is prone to false detections such as symmetrical details of buildings or road markings. Therefore, they applied a log-polar histogram shape descriptor to verify the shape of the candidates. Unfortunately, the shape descriptor is obtained

from a fixed vehicle model, making the algorithm inflexible. Moreover, similar to [7], the algorithm in [6] relied on mean-shift clustering algorithm for image color segmentation. The major drawback is that a vehicle tends to be separated as many regions since car roofs and windshields usually have different colors. Moreover, nearby vehicles might be clustered as one region if they have similar colors. The high computational complexity of mean-shift segmentation algorithm is another concern.

2. PROPOSED VEHICLE DETECTION FRAMEWORK

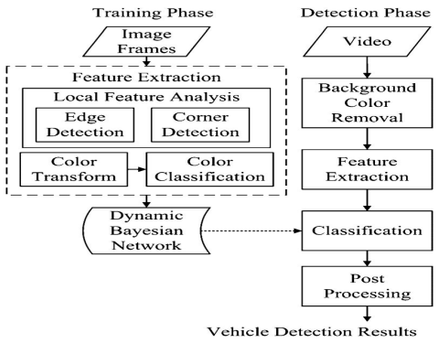


Fig 1: Proposed System Framework

In this paper, we design a new vehicle detection framework that preserves the advantages of the existing works and avoids their drawbacks. The framework can be divided into the training phase and the detection phase. In the training phase, we extract multiple features including local edge and corner features, as well as vehicle colors to train a dynamic Bayesian network (DBN). In the detection phase, we first perform background color removal. Afterward, the same feature extraction procedure is performed as in the training phase. The extracted features serve as the evidence to infer the unknown state of the trained DBN, which indicates whether a pixel belongs to a vehicle or not. In this paper, we do not perform region-based classification, which would highly depend on results of color segmentation algorithms such as mean shift. There is no need to generate multi-scale sliding windows either. The distinguishing feature of the proposed framework is that the detection task is based on pixel wise classification. However, the features are extracted in a neighborhood region of each pixel. Therefore, the extracted features comprise not only pixel-level information but also relationship among neighboring pixels in a region. Such design is more effective and efficient than region-based or multi scale sliding window detection methods.

A. Frame Extraction

In module we read the input video and extract the number of frames from that video.

Fig 2: Frame extraction

1). Background color removal

In this module we construct the color histogram of each frame and remove the colors that appear most frequently in the scene. These removed pixels do not need to be considered in subsequent detection processes. Performing background color removal cannot only reduce false alarms but also speed up the detection process.

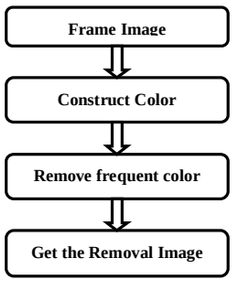


Fig 3: Background color removal

Since non vehicle regions cover most parts of the entire scene in aerial images, we construct the color histogram of each frame and remove the colors that appear most frequently in the scene. Take Fig. 4 for example, the colors are quantized into 48 histogram bins. Among all histogram bins, the 12th, 21st, and 6^{th} bins are the highest and are thus regarded as background colors and removed. These removed pixels do not need to be considered in subsequent detection processes. Performing background color removal cannot only reduce false alarms but also speed up the detection process.

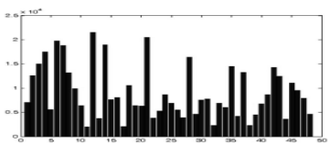


Fig 4: color histogram for frame

B. Feature Extraction

In this module we extract the feature from the image frame. In this module we do the following Edge Detection, Corner Detection, color Transformation and color classification as shown in fig 5. Feature extraction is performed in both the training phase and the detection

phase. We consider local features and color features in this paper.

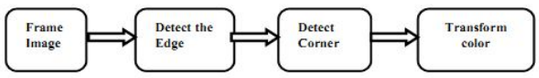


Fig 5: Local feature analysis and color transform

1) Local Feature Analysis: Corners and edges are usually located in pixels with more information .We use the Harris corner detector [8] to detect corners. To detect edges, we apply moment-preserving thresholding method on the classical Canny edge detector [9] to select thresholds adaptively according to different scenes. In the Canny edge detector, there are two important thresholds, i.e., the lower threshold T_{low} the higher threshold T_{high} As the illumination in every aerial image differs, the desired thresholds vary and adaptive thresholds are required. The computation of Tsai's moment-preserving method [11] is deterministic without iterations for L- level thresholding with L<5. Its derivation of thresholds is described as follows.

Let f be an image with n pixels and denote the gray value at pixel (x,y) .The ith moment m_i is defined as

$$m_i = \left(\frac{1}{n}\right)\sum_j n_j(z_j)^i = \sum_j p_j(z_j)^i, \qquad i = 1, 2, 3, \ldots \tag{1}$$

Where n_j is the total number of pixels in image f with gray value z_i and $p_j=n_j/n$. For bilevel thresholding, we would like to select threshold T such that the first three moments of image f are preserved in the resulting bilevel image g . Let all the below-threshold gray values in f be replaced by z_0 and all the above threshold gray values be replaced by z_1; we can solve for and based on the moment-preserving principle [8]. After obtaining p_0 and p_1 , the desired threshold T is computed using

$$p_0 = (1/n)\sum_{1}^{T} n_j. \tag{2}$$

In order to detect edges, we use the gradient magnitude, we use the gradient magnitude G(x,y) of each pixel to replace the gray value f(x,y) in Tsai's method.Then , the adaptive threshold found by(2) is used as higher threshold T_{high} in the canny edge detector.We set the lower threshold $T_{low}=0.1X(G_{max}-G_{min})+G_{min}$, where G_{max} and G_{min} where and represent the maximum and minimum gradient magnitudes in the image. Thresholds automatically and dynamically selected by our method give better performance on the edge detection. We will demonstrate the performance improvement on the edge detection with adaptive thresholds and the corresponding impact on final vehicle detection results in Section III.

2) Color Transform and Color Classification : In [13], the authors proposed a new color model to separate vehicle colors from non vehicle colors effectively. This color model transforms(R,G,B)color components into the color domain (u,v), i.e. as proposed by [1],

$$u_p = \frac{2Z_p - G_p - B_p}{Z_p} \tag{3}$$

$$v_p = \mathrm{Max}\left\{\frac{B_p - G_p}{Z_p}, \frac{R_p - B_p}{Z_p}\right\} \tag{4}$$

Where (R_p,G_p,B_p) is the R,G ,and B color components of pixel p and $Z_p=(R_p+G_p+B_p)$. It has been shown in [12] that all the vehicle colors are concentrated in a much smaller area on the plane u-v than in other color spaces and are therefore easier to be separated from non vehicle colors. Although the color transform proposed in [12] did not aim for aerial images, we have found that the separability property still presents in aerial images. As shown in Fig. 6, we can observe that vehicle colors and nonvehicle colors have less overlapping regions under the (u,v)color model. Therefore ,The color transform to obtain (u,v)components first and then use a support vector machine (SVM) to classify vehicle colors and non vehicle colors. When performing SVM training and classification, we take a block n X m of pixels as a sample. More specifically, each feature vector is defined as [u_1,v_1…..u_{nxm},v_{nxm}] . Notice that we do not perform vehicle color classification via SVM for blocks that do not contain any local features. Those blocks are taken as non vehicle color areas.

As we mentioned in Section I, the features are extracted in a neighborhood region of each pixel in our framework. Considering an N X N neighborhood Ap of of pixel p, as shown in Fig. 7, we extract five types of features, i.e. S,C,E,A ,and Z,for the pixel. These features serve as the observations to infer the unknown state of a DBN, which will be elaborated in the next subsection. The first feature S denotes the percentage of pixels in that are classified as vehicle colors by SVM, as defined in (5). Note that $N_{vehiclecolor}$ denotes to the number of pixels in Ap that are classified as vehicle colors by SVM, i.e.

$$S = \frac{N_{\mathrm{Vehicle\ color}}}{N^2}. \tag{5}$$

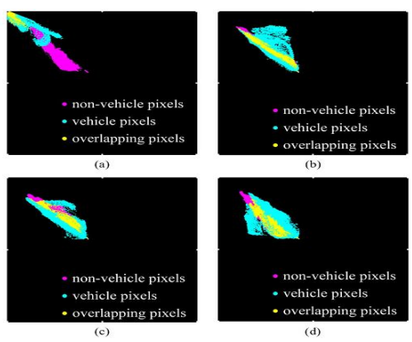


Fig 6: Vehicle colors and non vehicle colors (a)u-v (b)R-G ,(c)G-B ,and(d)B-R planes.

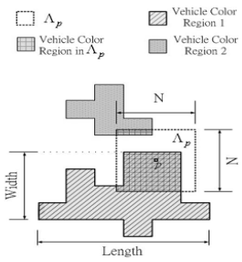


Fig 7: Neighborhood region for feature extraction.

Features C and E are defined, respectively, as

$$C = \frac{N_{Corner}}{N^2} \quad (6)$$

$$E = \frac{N_{Edge}}{N^2}. \quad (7)$$

Similarly, N_{corner} denotes to the number of pixels in Ap that are detected as corners by the Harris corner detector, and N_{edge} denotes the number of pixels in that are detected as edges by the enhanced Canny edge detector. The pixels that are classified as vehicle colors are labeled as connected vehicle-color regions. The last two features A and Z are defined as the aspect ratio and the size of the connected vehicle-color region where the pixel p resides, as illustrated in Fig. 7. More specifically, a=Length/Width, and feature Z is the pixel count of "vehicle- color region 1" in Fig. 7.

2) DBNs for Classification:

Dynamic Bayesian Networks (DBNs) are used for vehicle classification in video. Bayesian networks offer a very effective way to represent and factor joint probability distributions in a graphical manner which makes them suitable for classification purposes. A Bayesian network is defined as a directed acyclic graph G =(V;E) where the nodes (vertices) represent random variables from the domain of interest and the arcs (edges) symbolize the direct dependencies between the random variables. For a Bayesian network with n nodes$X_1, X_2, \ldots X_n$ the full joint distribution is defined as

$$\begin{aligned} p(x_1, x_2, \ldots, x_n) &= p(x_1) \times p(x_2|x_1) \times \ldots \\ &\quad \times p(x_n|x_1, x_2, \ldots, x_{n-1}) \\ &= \prod_{i=1}^{n} p(x_i|x_1, \ldots, x_{i-1}) \end{aligned} \quad (8)$$

but a node in a Bayesian network is only conditional on its parent's values so

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i|parents(X_i)) \quad (9)$$

Where p(x1; x2; : : : ; xn) is an abbreviation for p(X1=x1^x2……. xn).In other words, a Bayesian network models a probability distribution if each variable is conditionally independent of all its non-descendants in the graph given the value of its parents.

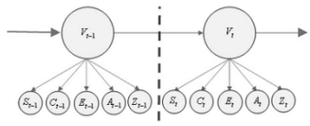


Fig 7: DBN model for pixel wise classification.

We perform pixel wise classification for vehicle detection using DBNs [13]. The design of the DBN model is illustrated in Fig. 7.NodeV_t indicates if a pixel belongs to a vehicle at time slice t .The state of V_t is dependent on the state of. Moreover, at each time slice ,state has influences on the observation nodes S_t , C_t, F_t,A_t, and Z_t . The observations are assumed to be independent of one another .The definitions of these observations are explained in the previous subsection.

Discrete observation symbols are used in our system. We use K-means to cluster each observation into three clusters, i.e., we use three discrete symbols for each observation node. In the training stage, we obtain the conditional probability tables of the DBN model via expectation-maximization [13] algorithm by providing the ground-truth labeling of each pixel and its corresponding observed features from several training videos. In the detection phase, the Bayesian rule is used to obtain the probability that a pixel belongs to a vehicle, by considering 8 and 9we can deduce i.e.

$$\begin{aligned} P(V_l|S_l, C_l, E_l, A_l, Z_l, V_{l\ 1}) &= P(V_l|S_l)P(V_l|C_l) \\ &\times P(V_t|E_t)P(V_t|A_t)P(V_t|Z_t)P(V_t|V_{t-1})P(V_{t-1}). \end{aligned} \quad (10)$$

The joint probability $P(V_t|S_t,C_t,F_t,A_t,Z_t,V_{t-1})$is the probability that a pixel belongs to a vehicle pixel at time slice l given all the observations and the state of the previous time instance. According to the naive Bayesian rule of conditional probability, the desired joint probability can be factorized since all the observations are assumed to be independent. Term P $(V_t|S_t)$is defined as the probability that a pixel belongs to a vehicle pixel at time slice t given observation S_t at time instance[is definedin(5)].Terms P $(V_t|C_t)$,$P(V_t|F_t)$,$P(V_t|A_t)$ $P(V_t|Z_t)$,and $P(V_t|V_{l-1})$are similarly defined.The proposed vehicle detection

framework can also utilize a Bayesian network (BN) to classify a pixel as a vehicle or non- vehicle pixel.When performing vehicle detection using BN, the structure of the BN is set as one time slice of the DBN model in Fig. 7.

C. Post Processing

We use morphological operations to enhance the detection mask and perform connected component labeling to get the vehicle objects. The size and the aspect ratio constraints are applied again after morphological operations in the post processing
stage to eliminate objects that are impossible to be vehicles. However, the constraints used here are very loose.

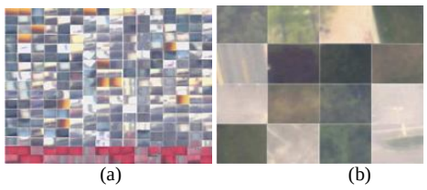
3. EXPERIMENTAL RESULTS

Experimental results are demonstrated here. To analyze the performance of the proposed system, various video sequences with different scenes and different filming altitudes are used. It is infeasible to assume prior information of camera heights and target object sizes for this challenging data set. When performing background color removal, we quantize the color histogram bins as 16X16X16. Colors corresponding to the first eight highest bins are regarded as background colors and removed from the scene. Fig. 8(a) displays an original image frame, and Fig. 8(b) displays the corresponding image after
background removal.

(a) (b)

Fig 8 : Background color removal results

When employing SVM, we need to select the block size to form a sample and perform vehicle color classification (see Fig. 9)



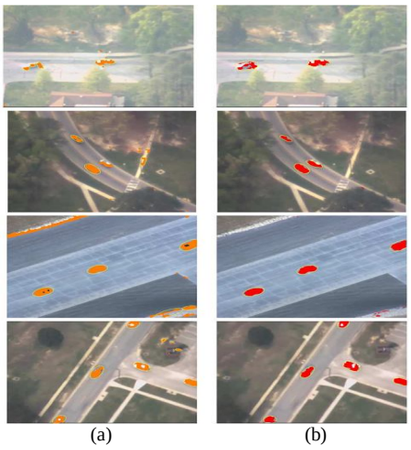
(a) (b)

Fig 9.Training images for vehicle color classification: (a) Positive samples and (b) negative samples.

Fig.10. shows the results of color classification by SVM after background color removal and local feature analysis.

Fig 10: Results of vehicle color classification.

In Fig. 11, we display the detection results using BN and DBN [see Fig. 10(a) and (b), respectively].The colored pixels are the ones that are classified as vehicle pixels by BN or DBN. The ellipses are the final vehicle detection results after performing post processing. DBN outperforms BN because it includes information along time.When observing detection results of consecutive frames, we also notice that the detection results via DBN are more stable. The reason is that, in aerial surveillance, the aircraft carrying the camera usually follows the vehicles on the ground, and therefore, the positions of the vehicles would not have dramatic changes in the scene even when the vehicles are moving in high speeds. Therefore, the information along the time contributed by helps stabilize the detection results in the DBN.



(a) (b)

Fig 11.Vehicle detection results: (a) BNs and (b) DBNs.

4. CONCLUSION

An automatic vehicle detection system for aerial surveillance does not assume any prior information of camera heights, vehicle sizes, and aspect ratios. This system, performs region-based classification , which would highly depend on computational intensive color segmentation algorithms such as mean shift. We have not generated multiscale sliding windows that are not suitable for detecting rotated vehicles either.Instead, we have proposed a pixel wise classification method for the vehicle detection using DBNs.In spite of performing pixel wise

classification, relations among neighboring pixels in a region are preserved in the feature extraction process. Therefore, the extracted features comprise not only pixel-level information but also region-level information. Since the colors of the vehicles would not dramatically change due to the influence of the camera angles and heights, we use only a small number of positive and negative samples to train the SVM for vehicle color classification. Moreover, the number of frames required to train the DBN is very small. Overall, the entire framework does not require a large amount of training samples. We have also applied moment preserving to enhance the Canny edge detector, which increases the adaptability and the accuracy for detection in various aerial images.

5. REFERENCES

[1] Hsu-Yung Cheng, Member, IEEE, Chih-Chia Weng, and Yi-Ying Chen, "Vehicle Detection in Aerial Surveillance Using Dynamic Bayesian Networks " March 21, 2012. *Lecture Notes in Computer Science* Project 99-2628-E-008-098.

[2] S. Hinz and A. Baumgartner, "Vehicle detection in aerial "images using generic features, grouping, and context," in Proc. DAGM-Symp., Sep.2001, vol. 2191, Lecture Notes in Computer Science, pp. 45–52.

[3] J. Y. Choi and Y. K. Yang, "Vehicle detection from aerial images using local shape information," Adv. Image Video Technol., vol. 5414, Lectures Notes in Computer Science, pp. 227–236, Jan. 2009.

[4] R. Lin, X. Cao, Y. Xu, C.Wu, and H. Qiao, "Airborne moving vehicle detection for urban traffic surveillance," in Proc. 11th Int. IEEE Conf. Intell. Transp. Syst., Oct. 2008, pp. 163–167

[5] R. Lin, X. Cao, Y. Xu, C.Wu, and H. Qiao, "Airborne moving vehicledetection for video surveillance of urban traffic," in Proc. IEEE Intell.Veh. Symp., 2009, pp. 203–208.

[6] J. Y. Choi and Y. K. Yang, "Vehicle detection from aerial images using local shape information," Adv. Image Video Technol., vol. 5414, Lecture Notes in Computer Science, pp. 227–236, Jan. 2009

[7] H. Cheng and D. Butler, "Segmentation of aerial surveillance video using a mixture of experts," in Proc. IEEE Digit. Imaging Comput. — Tech. Appl., 2005, p. 66.

[8] W. H. Tsai, "Moment-preserving thresholding: A new approach, "Comput. Vis.Graph., Image Process., vol. 29, no. 3, pp. 377–393, 1985.

[9] C. G. Harris and M. J. Stephens, "A combined corner and edge detector," in Proc. 4th Alvey Vis. Conf., 1988, pp. 147–151

[10] J. F. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[11] W. H. Tsai, "Moment-preserving thresholding: A new approach,"Comput. Vis.Graph., Image Process., vol. 29, no. 3, pp. 377–393, 1985.

[12] L. W. Tsai, J. W. Hsieh, and K. C. Fan, "Vehicle detection using normalized color and edge map," IEEE Trans. Image Process., vol. 16, no.3, pp. 850–864, Mar. 2007.

[13] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach,2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2003