

## Automatic vehicle Detection and Tracking for Event Analysis in Intelligent Aerial Surveillance Using Dynamic Bayesian Networks

K. Srinivasulu Reddy  
*M.Tech Software Engineering,  
Vardhaman College of Engineering,  
Hyderabad, India.*

A. Bhanu Prasad  
*Associate Professor, Department of IT,  
Vardhaman College of Engineering,  
Hyderabad, India.*

### Abstract

*Prediction of vehicles is a challenging domain in the traffic control department. Many of the researches, academicians, scientists and intellectual minds so far applied many techniques but didn't provide the accurate identifications of vehicle. So, in this paper we are highlighting Dynamic Back ground subtraction using colour histograms method based on Pixel wise classification. When applied this algorithm in the identification of vehicles, it is providing better accuracy when compared with previous approaches like sliding window method and region based method. In this paper we provided better identification of vehicles with the usage of pixel wise classification method.*

**Keywords:** *Aerial Surveillance, Dynamic Bayesian Networks, Soft Computing, Vehicle Detection.*

### 1. Introduction

Traffic control departments are facing problems with the identification of vehicles in proper manner during the signal time which leads to over accidents and causes serious loss to living and non living things [3]. So far many of the researches applied approaches like sliding window, region based method, hierarchical method, multiple clues and etc [4],[5], but not satisfactory with their identifications. This motivates and a research has been carried out towards the identification of vehicles. In order to identify the vehicles with exact position a model is proposed with this paper referred as Back ground subtraction using color histograms model based on pixel wise classification. With the use of this model the identification rate has increased to a maximum extent.

Identification of vehicles depends on the factors like size, shape, color, type, model and etc... As these factors changes from vehicle to vehicle identification becomes quiet complex and a challenging domain. This motivates the present result and a method is proposed referred as

back ground subtraction model based on pixelwise classification.

Videos are to be captured in the traffic. With the usage of digital cameras it is not possible and complex because streaming videos are take lots of time. So to overcome the problems use digital cameras aerial surveillance video cameras are placed. The main functionality of this cameras are to record the video in the traffic without mankind help. To identify the vehicle a part of video is needed which can be taken from aerial video by providing parameters like time, area and etc.

Vehicles are identified for security purposes. At this end clarity plays a vital role. As the vehicles are mostly in movable positions their identification becomes complex for every second the pixel value changes. Sometimes the videos are large and at this point a compression algorithm is needed. With the usage of compression algorithms due to loss phenomenon existing with the compression technique there may be a chance of video loss which parallel reduces the identification of correct vehicle. Clarity of the camera, compression technique which is using during video compression plays a key role in the accurate identification of vehicle.

Previously background subtraction algorithm window sliding method, region based method, hierarchical model, and multiple clues method, systematic property of car shape methods [3],[4],[5] are implemented, but not satisfied with their results. The results produced in the identification of vehicles by making use of these methods are not appropriate. As the traffic rate is increasing in higher rate identification of vehicles is an important, challenging task to the researchers, scientists, intellectual minds. So to identify the vehicle a new method is proposed with this paper referred to as background subtraction using colour histogram method based on pixel wise classification.

Many of the researchers applied back ground color removal after the division of pixels for image, not providing better clarity. So, in this paper a model is proposed where back ground color is removed and Edge detections 1<sup>st</sup> and then image is divided in to pixels based

on pixel wise classification method and is as shown in Fig: 1.

The rest of the paper is organized as follows. Section 2 gausses the proposed system frame work. Section 3 illustrates identification of desired vehicle. Section 4 provides conclusion and further recommendations.

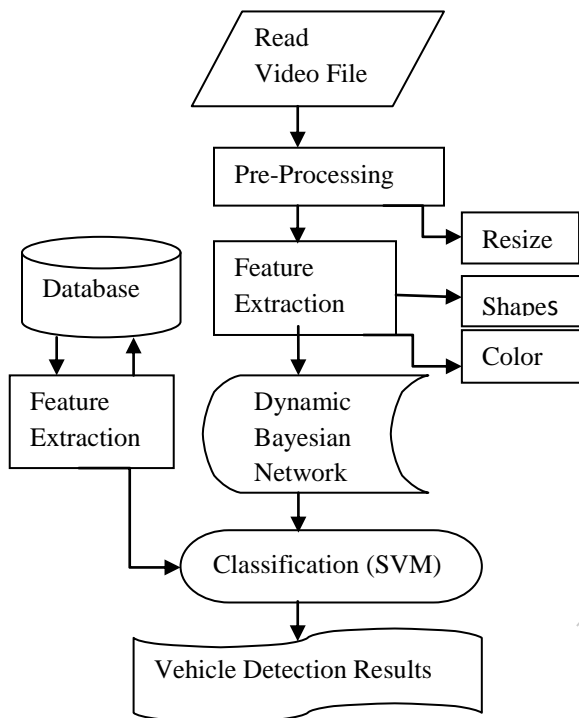


Fig 1: The proposed System Framework.

## 2. Proposed System Frame Work

### 2.1 Pre- Processing Method

The Pre-processing technique commonly involves images remaining low-frequency background noise, normalizing the intensity of individual particles images, removing, reflections, and masking portions of images. Image pre-processing is technique of enhancing data images prior to computational processing. The main intension of pre-processing technique is improvement of the data that suppresses unwanted distortions or enhance some image features improvement for feature extraction process [6][10]. Pre-processing methods are four types based on the size of pixel it will be applied. Neighborhood that is used for the calculation of new pixel brightness:

- a) Pixel brightness transformations

- b) Geometric transformations

- c) Pre-Processing methods that use a local neighbourhood of the processed pixel

- d) Image restoration that requires knowledge about the entire images

The majority of video surveillance is typically performed with one or two cameras that are mounted on a platform. Image processing performed can be divided into two stages front-end processing and scene analysis [10]. Front-end processing is applied to the source video to enhance its quality and to isolate signal components of interest such as electronic stabilization, noise and clutter reduction, mosaic construction, image fusion, motion estimation, and the computation of “attribute” images such as change and texture the scene. Scene analysis includes operations that interpret the source video in terms of objects and activities in the scene. Moving objects are detected and tracked over the cluttered scene [10].

### 2.2 Dynamic Background Subtraction Using Color Histograms

Some parts of the scenery may contain movement but should be regard as background, according to their relevance. Such movement can be periodical or irregular (ex: traffic lights, waving trees). In movement based vehicles detection is not sufficient in this situation. Because some of non vehicles also moving. In this case regions (ex: roads) are covered by non vehicles (ex: trees, buildings...etc). So we construct the color histograms of each frame and remove the color that appears most frequently in the aerial scene. It will be automatically detected the vehicles in regions.

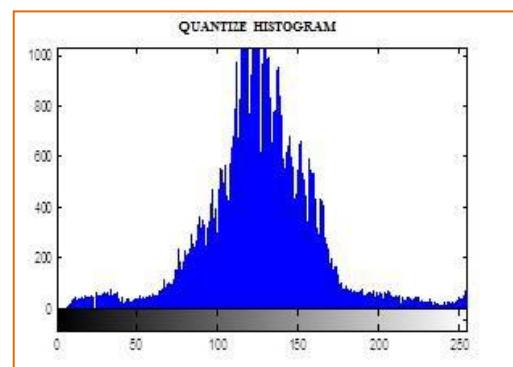


Fig 2: Color Histograms of a Frame.

The fig: 2 shows of the color are quantized into 48 histogram bins. Among all histogram bins, the 12<sup>th</sup>, 21<sup>st</sup>, and 6<sup>th</sup> bins are the highest and are the regarded

background colors are removed. These removed pixels do not need to be considered in subsequent detection processes. Performing dynamic background color removal can't only reduce false alarms but also speed up the detection process.

## 2.3 Feature Extraction Process

Feature extraction is performed in both training and detection phases. We consider local features and color features in my proposed system.

### 2.3.1 Local Feature Analysis

The image contains more information in pixels but each pixel contains corners and edges, so we use the Harris corner detector [8] is used to detect corners in vehicles. Next we use canny edge detector [9] is used to detect the edges in vehicles. In the edge detection based on the moment-preserving thresholding method will be calculating different scenes in aerial images of vehicles. In the canny edge detector, there are two importance thresholding. I.e. the lower threshold  $T_{low}$  and the higher threshold is  $T_{high}$ . As illumination in every aerial image differs to the desired threshold vary and adaptive thresholds are required in edge detection stages. The computational of Tsai's moment preserving method [8] is deterministic without iteration for L-level with  $L < 5$ . Its derivation of threshold is described as follows.



(a) Input image (b) Canny edge output

Fig 3: Canny Edge detector Results.

Let  $f$  be an image with  $n$  pixels and  $f(x, y)$  denotes the gray value  $(x, y)$ . The  $i^{\text{th}}$  moment  $m_i$  of  $f$  is defined as

$$m_i = (1/n) \sum_j n_j (z_j)^i = \sum_j p_j (z_j)^i, \quad i = 1, 2, 3, \dots \quad (1)$$

Where  $n_j$  is the total of pixels in image  $f$  with gray value  $z_j$  and  $p_j = n_j/n$ . For bi-level threshold, we would like to select threshold  $T$  such that the first three moments of image  $f$  are preserved in the resulting bi-level image  $g$ . Let all the above 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  $p_0$  and  $p_1$  based on the

moment-preserving principle. After obtaining  $p_0$  and  $p_1$ , the desired threshold  $T$  is computed using

$$P_0 = (1/n) \sum_1^T n_j \quad (2)$$

In order to detect edges, we use the gradient magnitude  $G(x, y)$  of each pixel to replace the gray scale values  $f(x, y)$  in Tsai's method. Then the adaptive threshold found by (2) is used as the higher threshold  $T_{high}$  in the canny edge detector, we set the lower threshold as  $T_{low} = 0.1 \times (G_{max} - G_{min}) + G_{min}$ , where  $G_{max}$  and  $G_{min}$  represents the maximum and minimum gradient magnitude in the images. Thresholds automatically and dynamically selected by our method give performance on the edge detections [9].

### 2.3.2 Color Transform and Color Classification

In this paper proposed new color transformation model is to separate vehicle colors and non-vehicle colors from effectively. This color transforms  $(R, G, B)$  color components into the color domains  $(u, v)$ , i.e.

$$U_p = (2Z_p - G_p - B_p) / Z_p \quad (3)$$

$$V_p = \max \{((B_p - G_p) / Z_p), ((R_p - B_p) / Z_p)\} \quad (4)$$

Where  $(R, G, B)$  is the R, G, and B color components of pixel  $p$  and  $Z_p = (R_p + G_p + B_p)/3$ . It has been shown in that all vehicle colors are concentrated in a much smaller areas on the  $u-v$  plane than in other color spaces and are therefore easier to be separated from non vehicle colors. SVM classification is used to non-vehicle color areas identified. The extraction process is five types is S, C, E, A, Z for a pixels. These features serve as observations to infer the unknown state of a DBN, which will be elaborated in the next sections.

S denotes the percentage of pixels in  $\Lambda_p$  that are classification as vehicle colors by SVM, as details in below

$$S = N_{\text{vehicle color}} / N^2 \quad (5)$$

Feature C and E are defined, respectively as

$$C = N_{\text{corner}} / N^2 \quad (6)$$

$$E = N_{\text{edge}} / N^2 \quad (7)$$

Similarly  $N_{\text{corner}}$  denotes to the number of pixels in  $\Lambda_p$  that are detected as corners by the Harris corner detector [8], and  $N_{\text{edge}}$  denotes the number of pixels in  $\Lambda_p$  that are detected as edge by the enhancement canny edge detector. The pixels that are classified as vehicle colors are labelled as connected vehicle color regions. A, Z are defined as the aspect ratio and the size of the connected vehicle color region where the pixel P resides.

$$A = \text{Length}/\text{Width} \tag{8}$$

$$Z = \text{count of pixels in "vehicle color region1"}. \tag{9}$$

Color classification is using SVM machine.

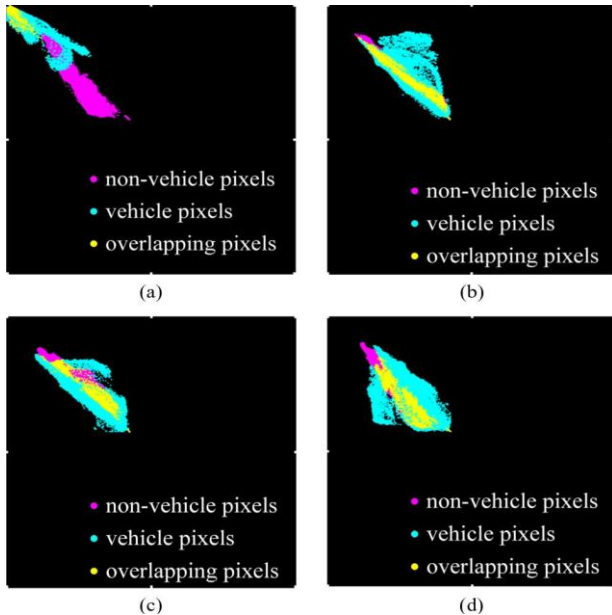


Fig 4: Vehicle color and non vehicle colors in Different Color Spaces (a) U-V, (b) R-G, (c) G-B, (d) B-R Planes.

### 2.3.3 Dynamic Bayesian Network

The Dynamic Bayesian Network is a Bayesian Network which relates variables to each other over adjacent time steps in the pixelwise classification method for vehicle detection using DBNs. The design of the DBN model is given fig: 5, a node  $V_t$  indicates if a pixel belongs to a vehicle at time slice  $t$ . In the state of  $V_t$  is dependent on the state of  $V_{t-1}$ . At each time slice  $t$ , state  $V_t$  has influences on the observation nodes  $S_t, C_t, E_t, A_t$ , and  $Z_t$ . The observations are not dependent in any others. Discrete observations symbols are used in our system. We use k-means to cluster each observation into three cluster types that are the training stage we obtains the conditional probability tables of the DBN model via exception maximization algorithm by providing the ground truth labelling of each pixel and its corresponding observed features from several training videos. In detection stage the Bayesian rule is used to obtain the probability that a pixel belongs to a vehicles. i.e.

$$P(V_t | S_t, C_t, E_t, A_t, Z_t, V_{t-1}) = P(V_t | S_t) P(V_t | C_t) \times P(V_t | E_t) P(V_t | A_t) P(V_t | Z_t) P(V_t | V_{t-1}) P(V_{t-1}) \tag{10}$$

Joint probability  $P(V_t | S_t, C_t, E_t, A_t, Z_t, V_{t-1})$  is the probability that a pixel belong to a vehicle pixel at time slice  $t$  given all the observations and the state of the previous time instance. Naive Bayesian rule of conditional probability the desired joint probability can be factorized since all the observations are assumed to be independent.  $P(V_t | S_t)$  is defined as the probability that a pixel belong to vehicle at time slice given observation  $S_t$  as instance  $t$  [S is defined in eq (5)]. Terms  $P(V_t | S_t)$ ,  $P(V_t | C_t)$ ,  $P(V_t | A_t)P(V_t | Z_t)$ , and  $P(V_t | V_{t-1})$  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: 6 we will compare the detection results using BN and DBN in the next section.

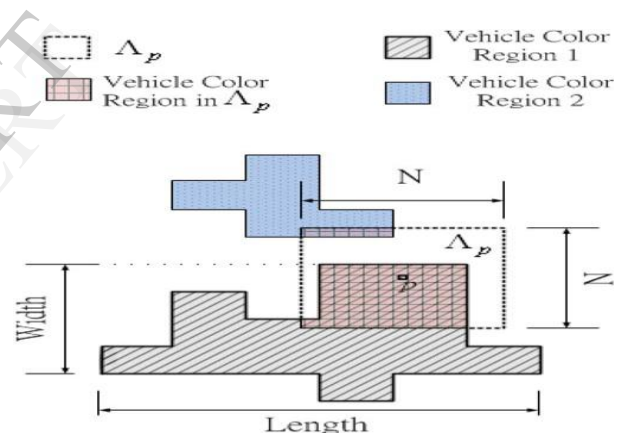


Fig 5: Neighborhood Regions for Feature Extractions.

### 2.2.4 Post Processing

We use the morphological operation to enhance the detection mask and performing connected component labeling to get the vehicle objects. The size and aspect ratio constructions 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. By using pre-processing technique reduced of the detection objects compare existing systems. If any vehicle is missing on the detection in starting stages will be detected in this stage.

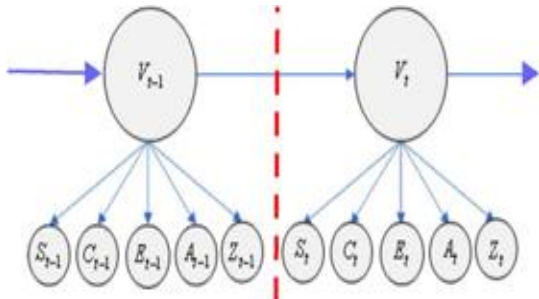


Fig 6: DBN model for Pixelwise Classification.



Fig 7: Snapshots of the experimental videos.

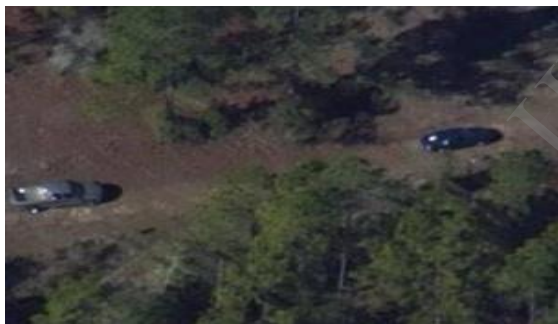


Fig 8(a): original image.

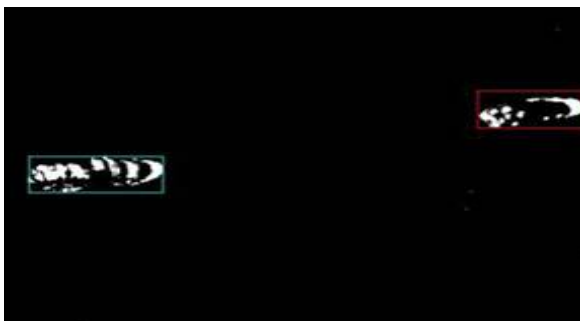


Fig 8(b): Background Substraction Using Color Histogram.

### 3. DESIRED VEHICLE 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. The experimental videos are display fig: 7; note that it is infeasible to assume prior information of camera heights and target objects sizes for this challenging data set. When performing dynamic background color, we quantize the color histograms bins as  $16 \times 16 \times 16$ . Color corresponding to the first eight heights bins are regarded as background colors and removed from the scenes. Fig: 8(a) displays an original image frame, and fig: 8(b) displays the corresponding image after background removal.

To obtain conditional probability tables of the DBN, we select the training clips from the first six experimental videos displayed. The remaining four videos are not involved in the training process. Each training clips contains 30 frames, whose ground-truth vehicle positions are manually marked. The select size of the neighborhood area for feature extraction, we list the detection accuracy using is measured by the hit rate and the number of false positives per frame. There are a total of 224025 frames in the data set. When evolving the detection accuracy, we perform evolution every 100 frames. We can observe that the neighborhood  $\Lambda_p$  with the size of  $7 \times 7$  yields the best detection accuracy.

The impacts of the enhancement Canny edge detector on vehicle detection results can be observed in fig: 3. shows the results obtained results can be observed using the traditional canny edge detector with detector with moment preserving threshold selection. Non adaptive threshold can't adjust to different scenes and would therefore results in more majorities of the dynamic background removal process and the enhanced edge detector; we list different scenes in table. We can observe that the background removal process is improved for reducing false positives and the enhanced edge detector is essential for increasing hit rates.

We compare different vehicle detection methods in fig. The moving vehicle detection with road detection method in requiring setting a lot of parameters to enforce the size constraints in order to reduce false alarms. However for the experimental data set, it is very difficult to select one dataset of parameters that suits all videos. The shape description used to verify the shape of the candidate is obtained from a fixed vehicle model and is therefore not flexible.

Table 1: Detection Accuracy Using Different Neighborhood Sizes

Size of $\Delta_p$	Hit rate	Number of positives per frame
5×5	70.91	0.523
7×7	92.31	0.278
9×9	87.06	0.281
11×11	82.35	0.401
13×13	75.58	0.415

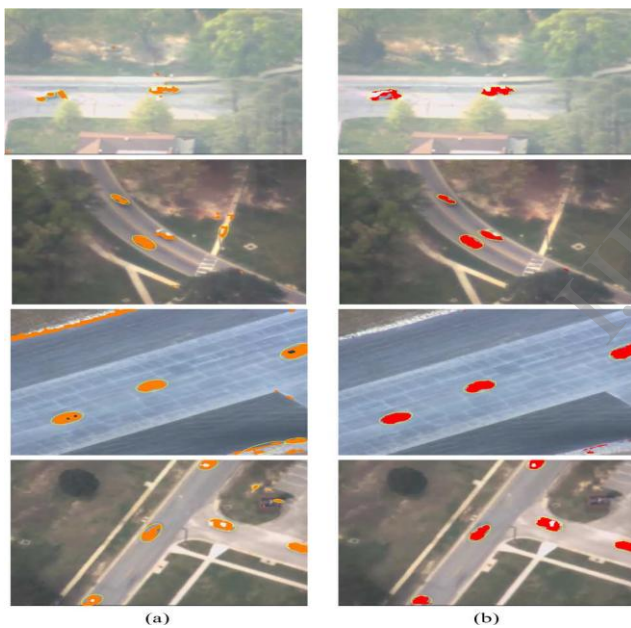


Fig 9: Vehicle Detection results (a) BN and (b) DBNs.

Compared with these methods, the proposed vehicle detection framework does not depend on strict vehicle size or aspect ratio constraints. Instead these constraints are observed that can be learned by BN or DBN. The training process does not require a large amount of training samples. The results demonstrate flexibility and good generalization ability on a wide variety of aerial surveillance scenes under different heights and cameras angles. It can be expected that the performance of DBN is better than that of the BN. In fig: 9, can be displays the detection results using BN and DBN.



Fig 10: Vehicle Detection Error Rate.

Table 2: Detection Accuracy of Four Different Scenarios

Scenarios	Hit rate	Number of false positives per frame
Without background removal and without enhanced edge detector	75.08	0.399
Without background removal and with enhanced edge detector	92.35	0.459
With background removal and without enhanced edge detector	74.96	0.297
With background removal and with enhanced edge detector	92.31	0.278

Fig: 10 show some detection error cases In fig: 10(a) displays the original image frames, and fig: 10(b) simply the detection results. The back arrow in fig: 10(b) indicates the misdetection of false positive cases. IN the first row fig: 10 (a) the rectangle structures on the building are very similar to the vehicle. Sometime this rectangle structure would be detected as vehicle incorrectly. In the second row of fig: 10(b) the miss detection is caused by the low constraints and the small

size of the vehicle .however, others vehicle are successfully detected in this challenging setting.

In fig: 11 show the average processing speeds of different vehicle detection methods. The proposed framework using BN and DBN cannot reach the frame rate of the surveillance videos, it is sufficient to perform vehicle detection every 50-100 frames. Tracking algorithm can be applied on the intermediate frames between two detection frames to track each individual vehicle. Therefore, high detection rate and low false alarm rate should be the primary considerations of designing detection methods given the condition that the execution time is reasonable.

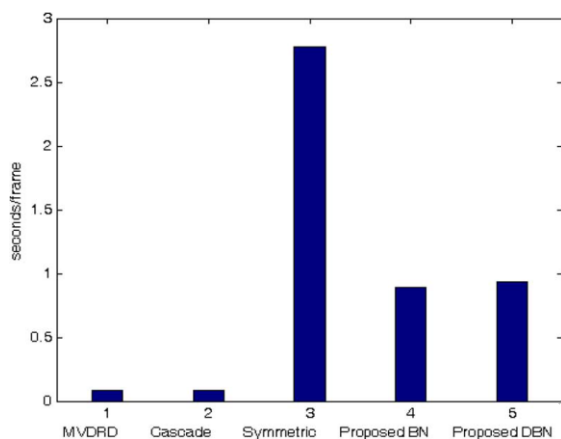


Fig 11: Preprocessing Speeds in Different methods.

#### 4. CONCLUSION

In this paper, we proposed an automatic vehicle detection system for aerial surveillance that does not assume any prior information of camera heights and vehicle sizes and aspect ratios. Instead of the region based classification, we have proposed a pixelwise classification method for the vehicle detection using DBNs. The extraction processes comprise not only pixel level information but also region level information. Vehicle color and non vehicle color identification to use SVM Classification method. More ever the number of frames required to train the DBN is very small. We have to apply moment preserving to enhance canny edge detector, which increases the adaptability and accuracy for detection in various aerial images. In the proposed method any prior information of camera in different angles and different heights to taken. In this approach is controlling of the traffic monitoring system easily compare with existing systems.

For future work, will be performing vehicle tracking on the detected vehicle can further stabilize the detection results. Automatic vehicle detection and tracking is very important aspect of the intelligent aerial surveillance system.

#### 5. References

- [1] Hsu-yung cheng, Chih-chia weng and yi-ying Chen, " *vehicle detection in aerial surveillance using dynamic Bayesian Networks*", IEEE April-2012, vol.21, no.4.
- [2] R. Kumar, H. Sawhney, S. Samarasekara, S. Hsu, T. Hai, G. Yanlin, K. Hanna, A. Pope, R. Wildes, D. Hirvonen, M. Hansen, and P. Burt, " *Aerial Video Surveillance and Exploitation*," Proc.IEEE, vol.89,n0.10,pp. 1518-1539,2001.
- [3] S. Srinivasn, H. Latchman, J. Shea, t. Wong, and J. McNair, " *Airbone traffic Surveillance systems: video surveillance of highways traffic*," in proc. ACM 2<sup>nd</sup> Int. Workshop video surveillance Scenes. Netw. 2004, pp. 131-135.
- [4] S. Hinz and 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.
- [5] H. Cheng and D. Butlr, " *Segmentation of aerial surveillance video using a mixture of experts*," in proc. IEEE Digt. Image Comput.-Tech.App., 2005, P.66.
- [6] L. Lin, X. Cao, y.Xu, c.Wu, and H.Qiao, " *Airbone moving vehicle detection for urban traffic surveillance*," in Proc. IEEE Intell. Veh.Symp. 2009, pp. 203-208.
- [7] J.Y. Choi and Y.K. Yang, " *Vehicle detection from aerial images using local shape information*" Adv. Image video Technol., vol. 5414, lecturer Notes in Computer Science, pp. 227-236, Jan. 2009.
- [8] C.G. Harris and M.J. Stephanie's, " *A Combined Corner and edge detector*," in proc.4<sup>th</sup> Alvey Vis. Conf., 198. PAMI-8, no.6, pp.679-698, Nov.1986,pp.147-151.
- [9] J.F. Canny, " *A computacional approach to edge detection*," IEEE Transs. Pattern Anal. Mach. Intel., vol.PAMI-8, no.6, pp.679-698, Nov.1986.
- [10] S.russel and P. Norving, *Artificial Intelligent: A Modern Approach*, 2nd ed. Englewood Cliffs, NJ; Prentice-Hall, 2003.
- [11] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machine and Other Kernel-Based Learning Methods*. Cambridge, U.K., Cambridge Univ.Prtses, 2000.