SURVEILLANCE OF OUTDOOR MOVING TARGETS
Matching Targets using Five Features

Nalin Pradeep S.
Vision Technologies Group,Sarnoff Innovative Technologies Private Ltd,Bangalore,India

Mayur D. Jain
Microsoft(R&D) India Private Ltd,Hyderabad,India

Keywords: Object Segmentation, Tracker, Five Features, Centroid, Shape and Color Histogram.

Abstract: The proposed video surveillance method comprises segmentation of moving targets and tracking the detected objects through five features of the target object. We introduce motion object segmentation based on mean and variance background learning model, and subtraction using both color and edge information. The cognitive fusion of color and edge information helps identifying foreground object. The combination of the five features spatial positions, LBW, Compactness, Orientation and color histogram through particle filter approach tracks the segmented objects. These five features help in matching the target tracks during occlusions, merging of targets, stop and go motion in vary challenging environmental (rainy and snowy) conditions shown in the results. Our proposed method provides solution to common problems related to matching of target tracks. We provide encouraging experimental results calculated on synthetic and real world sequences to demonstrate the algorithm performance.

1 INTRODUCTION

The main purpose of video surveillance is to allow for a secure monitoring. The primary research issue of Video Surveillance is automated detection and tracking objects, events and pattern. Nowadays video surveillance is a mature discipline aiming to define techniques and systems for processing videos from cameras placed in a specific environment to extract the knowledge of meaningful moving entities. The high level description of a video stream relies on the detection and accurate tracking of moving objects, and on the relationship of their trajectories to the scene.

Recently, a significant number of trackers have been proposed. Some deal with low-level feature tracking while others deal with high-level description such as event detection, recognition, classification as human/vehicle and even trajectory descriptions. For the success of high-level description it relies very much on accurate detection and tracking of moving objects in varying environmental conditions. For a successful tracker to exist under testing conditions it needs to overcome situations of merge, occlusions, start-stop motion of targets. Utsumi and Ohya, 1998 proposed a method of extracting a moving object region from each frame in a series of images using statistical knowledge about the target. Haritaoglu, Harwood and Davis, 2000 classified the feature trackers into several categories with their functionality (tracking single-multiple objects, handling occlusion)

In this paper, we address the problems of detection, tracking and matching the tracks of moving objects in the context of video surveillance. With the help of condensation algorithm through particle filter (Isard and Blake, 1998) we could experiment with multiple objects merging, occluding under varying conditions (like rainy, snowy). Section 2 describes the algorithm for the moving object segmentation. The approach used builds a background model using both color and gradient information and then performs background subtraction using these models (Javed, Shafique and Shah, 2002) The cognitive fusion of color and edge information gives the accurate object contour and helps to remove noise and small regions, further updated with median filter. Section 3 focuses on the

Pradeep S. N. and D. Jain M. (2006).
SURVEILLANCE OF OUTDOOR MOVING TARGETS - Matching Targets using Five Features.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 395-401
DOI: 10.5220/0001367003950401

tracker structure. The five features are extracted and target tracks matched. The state vector of a target includes spatial position (centroid), Length by Width (LBW) ratio, Orientation, compactness, color histogram. These features are updated through sample-based representation of recursive Bayesian filter applied iteratively (Esther, koller-Meier, Frank Ade, 2000). Factors such as unexpected intruders, occlusion, merging of targets with noise may affect the efficiency of tracking. To overcome this, most reliable cues motion, color and shape features are combined. Section 4 describes the matching of target tracks using the five features. The features which form the state vector of a track are updated with newly appearing or disappearing targets. In section 5, we demonstrate the result of our system on data sets. Section 6 concludes the paper. Figure 1 shows the flowchart of our system.

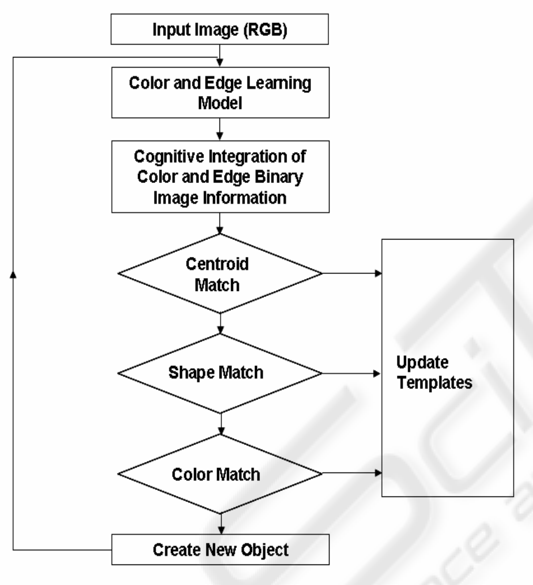

Input Image (RGB)
Color and Edge Learning Model
Cognitive Integration of Color and Edge Binary Image Information
Centroid Match
Shape Match
Color Match
Update Templates
Create New Object

Figure 1: Flowchart of the System.

2 OBJECT SEGMENTATION

The first step in the surveillance of object is separating the moving targets from the background (BGND). We have built the background model in two parts, color model and gradient model. Color model is built for each color channel. It consists of two images representing the mean and standard deviation for each color channel. For each color channel the mean image is computed as

$$u_t = \alpha x_t + (1-\alpha)u_{t-1} \qquad (1)$$

where, u_t is the mean computed up to frame *t*, α is the learning rate of model, and x_t is the intensity of the color component in frame *t*. The standard deviation image σ_t is used to normalize the confidence map during background subtraction and is computed using

$$\sigma_t^2 = \alpha(x_t - u_t)^2 + (1-\alpha)\sigma_{t-1}^2 \qquad (2)$$

The edge model is also composed of two mean images and two standard deviation images. It is computed by a horizontal and vertical Sobel edge filter. This results in horizontal (H) and vertical gradient image (V). The mean images are computed as

$$\begin{aligned} H_t &= \beta H + (1-\beta)H_{t-1} \\ V_t &= \beta V + (1-\beta)V_{t-1} \end{aligned} \qquad (3)$$

where, β is the learning rate of the model. The standard deviation images $\sigma_{H,t}$ and $\sigma_{V,t}$ are computed similar to the color model. The edge model is used to identify changes in the structure of an image. Mean image is updated continuously using the learning parameter in the color and edge model which allows the background model to adjust to gradual changes in illumination. Background subtraction is done by performing the color-based subtraction and the edge-based subtraction separately and then cognitive integration of the two results.

Color-based subtraction is performed by subtracting the current image from the mean image in each color channel. This results in three difference images which are used to create three normalized confidence maps. This is done by comparing the difference to two thresholds, $m_c\sigma$ and $M_c\sigma$, derived from the standard deviation images. For each pixel, the confidence is computed as

$$C_M = \begin{cases} 0 & D < m_c\sigma \\ \dfrac{D - m_c\sigma}{M_c\sigma - m_c\sigma} \times 100 & m_c\sigma \le D \le M_c\sigma \\ 100 & D > M_c\sigma \end{cases} \qquad (4)$$

A significant change in any color channel indicates a foreground region. A single confidence map C_C can be created by taking the maximum confidence at each pixel.

Edge-based subtraction is performed by subtracting the current horizontal difference image from the mean image H_t and vertical difference image from the mean image V_t

$$\Delta H = |H - H_t|, \quad \Delta V = |V - V_t| \qquad (5)$$

The edge gradient image is obtained as

$$\Delta G = \Delta H + \Delta V \tag{6}$$

The confidence map is computed by multiplying the ΔG by a reliability factor R and comparing the results to two thresholds $m_e\sigma$ and $M_e\sigma$. Here, σ is the sum of the horizontal standard deviation and the vertical standard deviation. For each pixel, let

$$G=|H|+|V|,\quad G_t=|H_t|+|V_t| \tag{7}$$

$$G_t^* = \max\{G, G_t\} \tag{8}$$

and reliability factor R computed as

$$R = \frac{\Delta G}{G_t^*} \tag{9}$$

The confidence for each pixel is computed as

$$C_E = \begin{cases} 0 & R\Delta G < m_e\sigma \\ \dfrac{R\Delta G - m_e\sigma}{M_e\sigma - m_e\sigma}\times 100 & m_e\sigma \le R\Delta G \le M_e\sigma \\ 100 & R\Delta G > M_e\sigma \end{cases} \tag{10}$$

The results from the color subtraction and the edge subtraction are combined by taking the maximum between the two confidence maps obtained in (4) and (10) at each pixel. The binary image thus obtained by the fusion of color and edge map is processed with median filter to remove salt and pepper noise. Figures 3(a), 3(b), 3(c) and 5(a), 5(b), 5(c) shows the moving objects (foreground) in white and background in black.

3 TRACKER STRUCTURE

We extract five types of feature in each moving object (target), which are used for tracking. The moving targets shown in Fig. 3, 5 are tracked. These five features help in accurate tracking and matching of objects in various difficult scenarios. The features LBW, Compactness, Orientation are classified as shape features of target. The objects have more complex shapes and are likely to change in time under the assumption of same segmentation method.

3.1 Centroid

Centroid (X, Y) of a target tells us the spatial position of the moving target. Centroid is calculated

$$\mathrm{X} = \frac{1}{\mathrm{A}}\sum \mathrm{x}\ ,\ \mathrm{Y} = \frac{1}{\mathrm{A}}\sum \mathrm{y} \tag{11}$$

where $(x,y) \in R$ and A is the number of pixels in a target R. $(\overline{X},\overline{Y})$ are the relative velocity of the moving target. These relative velocities are updated at each time instance by propagating the sample set maintained for each target track.

3.2 LBW (Length by Width)

At any time instance*'t'* LBW ratio is calculated for the moving target in their bounding region R. The length and width of the bounding box for a target is calculated. This ratio is maintained for individual tracks to match them

$$LBW = \left(\frac{\max x(t) - \min x(t)}{\max y(t) - \min y(t)}\right) \tag{12}$$

3.3 Compactness

Objects being tracked have complex shapes which are maintained till an object dies.

$$Compactness, C = \left\{\frac{4\pi \times Area}{Perimeter}\right\} \in R \tag{13}$$

where *Perimeter* is the sum of all boundary pixels in region *R*.

3.4 Orientation

The Orientation θ of moving target is measured across major axis and horizontal axis. The θ is updated considering all pixel positions in both x and y coordinates w.r.t to centroid within R

$$Orientation, \theta = \frac{1}{2} \times \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \tag{14}$$

where $\mu_{p,q} = \sum_x \sum_y (x - X)^p (y - Y)^q$ for p,q = 0,1,2.

3.5 Color

Color histogram is constructed for each moving target by the distribution of RGB channel values. All pixel values of an object are collected within the bounding box R. For each pixel, the RGB values of the object within this box R is normalized and distributed as histogram in 4368 bins. The probability intersection of these bins (histogram distribution) is used for object matching. For each pixel in region R, the bin for the distribution of its color channel is identified using the formula

$$\underset{\mathrm{i=0}}{\overset{\mathrm{i=4368}}{\mathrm{color}}} = (\frac{\mathrm{R}}{16}) + (\frac{\mathrm{B}\times 256}{16\times 16}) + (\frac{\mathrm{G}\times 256\times 256}{16\times 16\times 16}) \tag{15}$$

Our tracking process involves the state vector multi-dimensional and large; we cannot just sample the probability density at regular intervals. Hence, we use stochastic sampling method in the art of condensation algorithm (Esther, koller-Meier, Frank Ade, 2000). This factored sampling method helps in finding an approximation to the probability density. The state vector for any object at time‘t’ is expressed as

$$x(t) = [\ \mathrm{X},\ \mathrm{Y},\ \overline{\mathrm{X}},\overline{\mathrm{Y}}, LBW, C, \theta, \mathrm{Color}]\ .$$

4 OBJECT MATCHING

In the tracking process, if a match is found then the state vector of an object is updated by the system model

$$x(t) = Ax(t-1) + Bw(t-1) \qquad (16)$$

where, *w(t-1)* is noise, A,B are constant matrix. The state vector which denotes the identity of each individual object is updated based on measurements only if a match is found. The matching of a target at time *‘t-1’* with a target at time *‘t’* needs to be accurate and have to overcome situations of merge, occlusions, failure of detection etc. So matching is carried out at three levels using the five features of state vector. Let the observation from a detected target at time‘t’ be denoted as

$$z(t) = [\ \mathrm{X}',\ \mathrm{Y}',\ \overline{\mathrm{X}}',\overline{\mathrm{Y}}', LBW', C', \theta', \mathrm{Color}'\,]$$

4.1 Centroid Matching

An objects position from the observation and its update state vector position should be within the gating region for matching to occur. The Euclidean distance is measured and if it is within the gating region then matching occurs.

$$\Delta\mathrm{D}=\sqrt{(\mathrm{X}\text{-}\mathrm{X}')^2+(\mathrm{Y}\text{-}\mathrm{Y}')^2+(\overline{\mathrm{X}}\text{-}\overline{\mathrm{X}}')^2+(\overline{\mathrm{Y}}\text{-}\overline{\mathrm{Y}}')^2} \quad (17)$$

4.2 Shape Matching

We compute the distance from shape feature vector. The features LBW, C, θ form shape vector. The object (target) is said to have matched if the difference of sum squares is less than predefined threshold.

$$\Delta\mathrm{S}=\sqrt{(\mathrm{LBW}\text{-}\mathrm{LBW}')^2+(\mathrm{C}\text{-}\mathrm{C}')^2+(\theta\text{ - }\theta')^2} \quad (18)$$

4.3 Color Matching

We compute the match through minimum probability intersection of the histogram bins.

$$\Delta\mathrm{C}= \sum_{\mathrm{i=1}}^{\mathrm{i=4368}} \mathrm{max(color,color}')\qquad (19)$$

The sum of minimum intersection on each bin should be greater than a predefined value.

If an object does not match even after going through the above matching procedures, a new dummy target is generated for the object. This dummy target turns into true target after lasting for several frames. During this period, the dummy target is supposed to be in occlusion watch state. If during this state, the dummy target matches another measured object then we eliminate the dummy target. In Fig. 2(a) a car being identified as an object approaches a tree. In Fig. 2(b) only very small part of the car is visible and very major part of the car is hidden behind the tree, so matching fails. The object which is the car goes into occluded watch state and a dummy target is created for it.

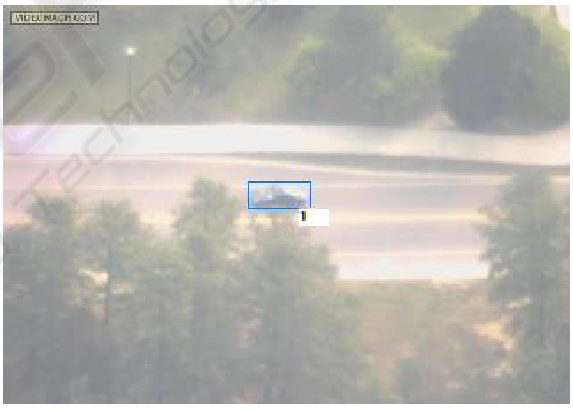


Figure 2(a): Car approaching the tree.

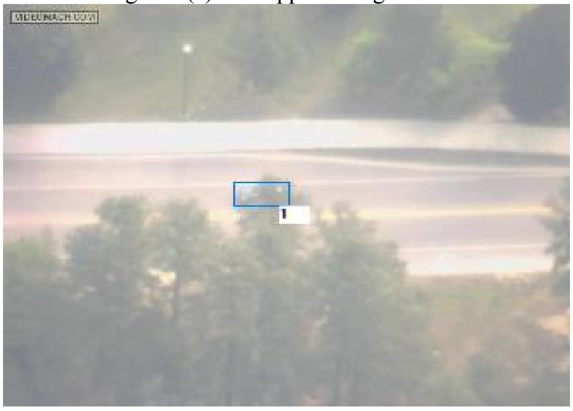


Figure 2(b): Car being occluded by the tree.

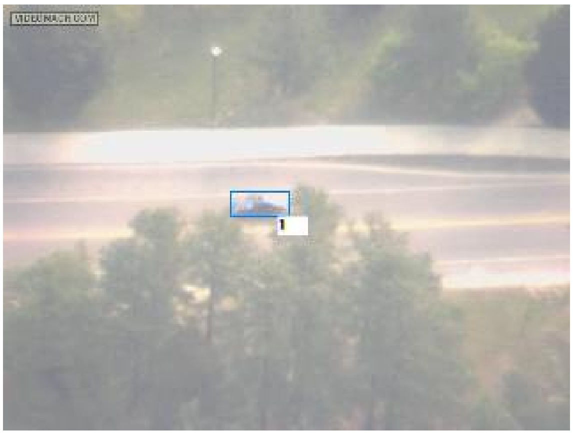


Figure 2(c): Car coming from behind the tree.

Fig. 2(c) shows the car emerging from behind the tree. The dummy target associated with the car matches the true target. During this interim occlusion period the dummy target gets updated only through the centroid component by linear prediction, using the average velocity from last three frames.

Based on the object size, camera positions, surveillance carried out in indoor/outdoor situations heuristics are used regarding the weight assigned to the above three matching. For instance if the object is tracked at an indoor situation then color, shape are given more weightage as objects are predominantly humans with complex shapes and color.

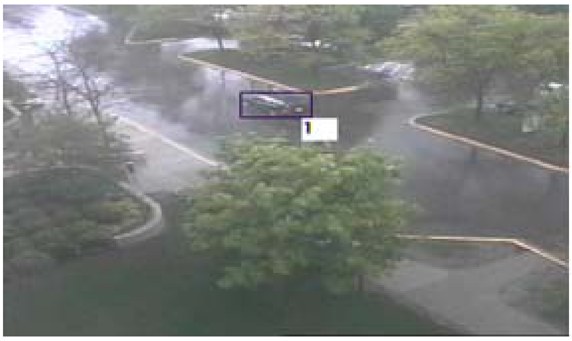
5 EXPERIMENTAL RESULTS

The proposed system is used to analyze videos with promising results. The video provided is taken in an environment where many detection and tracking failed because of the extreme rainy condition, low color illumination, tree shadow effect, and tree movements, light reflection caused on roads, vehicle color matching with BGND, and varying object size.

Fig.3 (a), 3(b), 3(c) shows segmented output of moving target and Fig 4(a), 4(b), 4(c) shows the moving car tracked and assigned label 1. The car is tracked at bizarre rainy condition with quite a lot reflection from the road. These video frames were taken at 111,123 and 148 respectively from the video clip. The detection output is not missed even for a frame under this adverse environment as explained above. The fusion of color and edge-based model helps in attaining correct object contour. The noise due to tree movement is very less and it’s further removed by using the heuristic that tree movements are subjected to random motion. The car is as well tracked without losing the object label in each frame. It can also be seen that the object is tracked till the end of the frame with the spatial information even when the shape information is insufficient.

In the other snowy video data frame shown, the merging and split case among the segmented objects is handled splendidly. Fig.5(a),5(b),5(c) shows the segmented objects of three humans. Fig. 6(a), 6(b), 6(c) shows the multiple objects tracked with label assigned to them. The labels do not get interchanged even after merge. During the merge sown in Fig.5(b), the spatial positions of one of the two objects being merged are not available so matching is carried out based on shape and color. These video frames were taken at 450,454 and 458. Track labels are maintained for multiple objects for object appearance/disappearance situations and even in occlusion cases.

Although we find the results of our tracker to be encouraging, there are still some unresolved problems. For example incase of an object being in occluded state for long period of time then the predicted spatial position can lose its way and thus enabling the failure of color and shape feature.



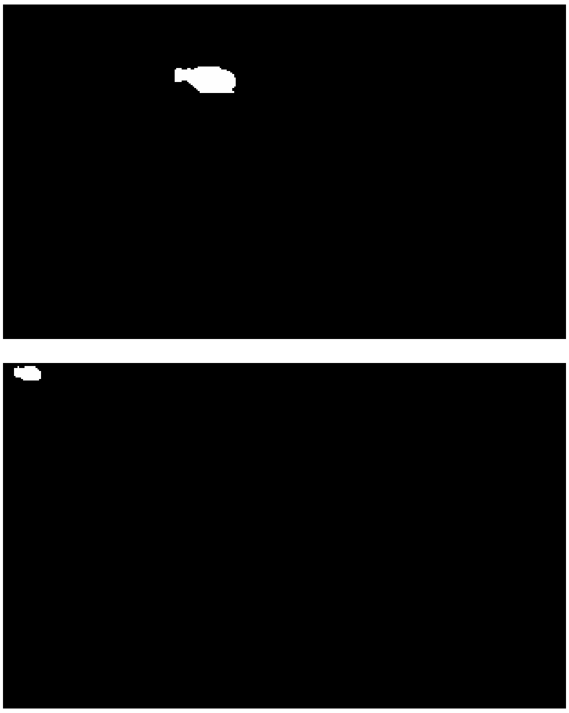
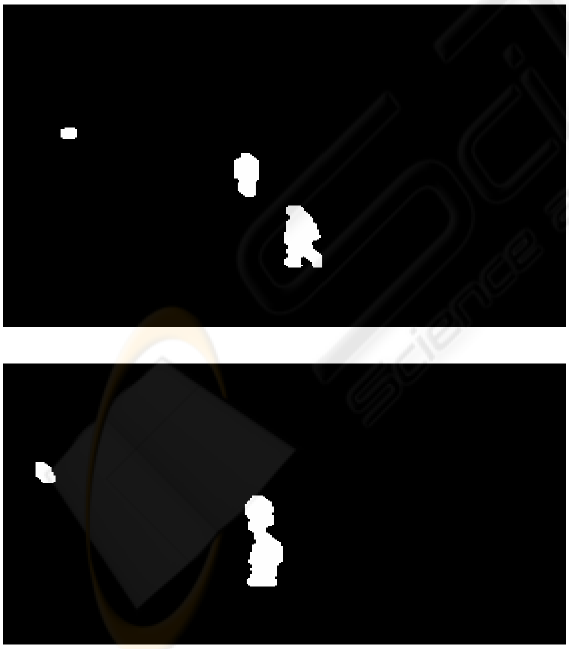


Figure 3(a), 3(b), 3(c) : Moving car target by fusion of color and edge.



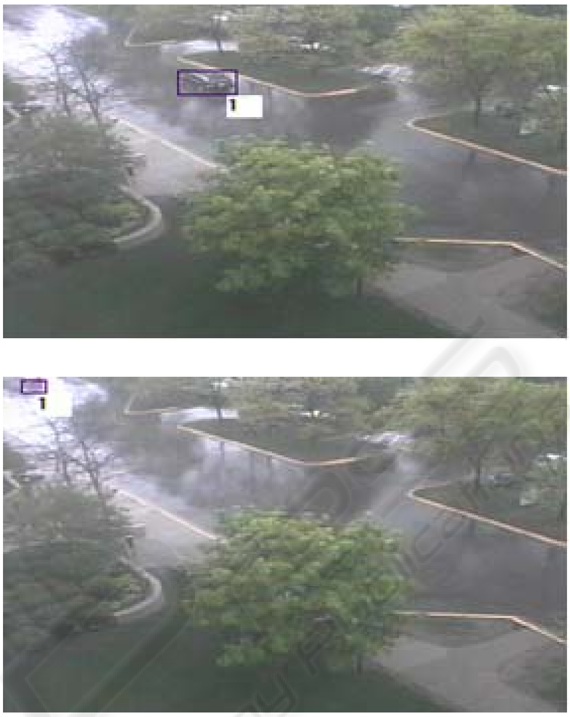
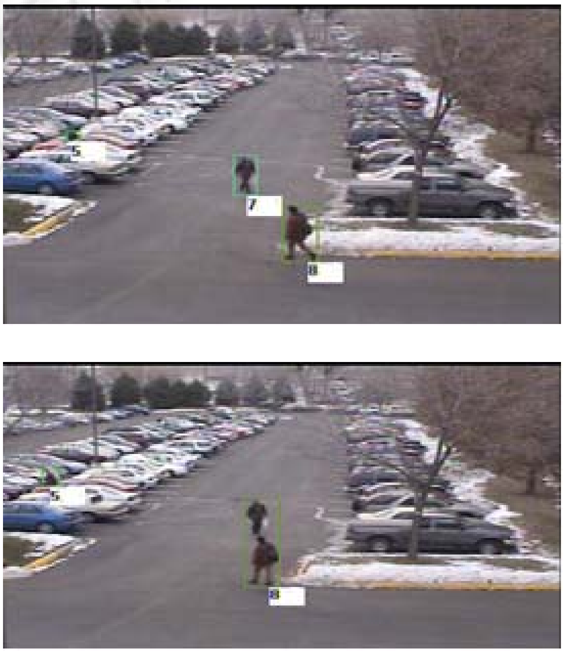


Figure 4(a), 4(b), 4(c): Car target tracked till end of the image in rainy data.



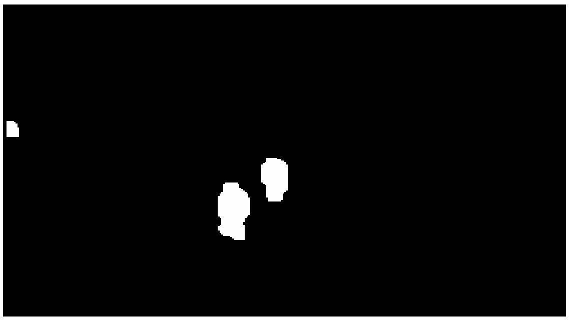


Figure 5(a), 5(b), 5(c): Multiple moving targets by fusion of color and edge.

6 CONCLUSION

In this paper, we have presented a system for detecting, and tracking moving objects in a surveillance area under varying environmental conditions. We combine edge-based and color based background model subtraction to get complete object contour, which enhances the tracking performance. Spatial position, shape and color combined together to increase the performance of tracking even in any case of split targets or non-availability of observed samples. Our future work is to resolve the cases when object is totally occluded and extreme split/merge cases of an object. Also we would work on classifying the target as humans, vehicles or group of people.

REFERENCES

M.Isard and A.Blake., volume29, n01,pp 5-28, 1998. *Condensation –conditional density propagation for visual tracking,* International Journal of Computer Vision.

A.Utsumi and J.Ohya., pp 911-916, 1998. *Image Segmentation for human tracking using sequential-image-based hierarchical adaptation*, Proceedings of IEEE computer Society Conf. on CVPR.

Tao Zhao and Ram Nevatia., pp 9–14, 2002. *Stochastic human Segmentation from a Static Camera,* Motion and Video Computing Proceedings**.**

Esther B. koller-Meier, Frank Ade., pp 93-105,2001. *Tracking Multiple Objects Using condensation Algorithm*, Journal of robotics and Autonomous systems.

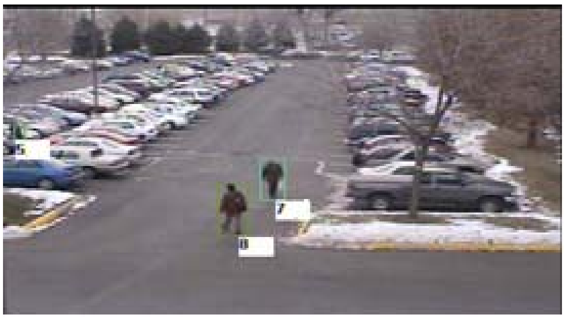


Figure 6(a), 6(b), 6(c) : Humans tracked . Merging of humans handled.

Q. Zhou and J. Aggarwal., Dec 9, 2001. *Tracking and Classifying Moving Objects from Video,* in Proc. 2nd IEEE Int'l Workshop on Performance Evaluation of Tracking in Surveillance.

I.Haritaoglu, D.Harwood and L.S Davis., vol. 22, pp 809-830,2000. *W/sup 4/:real time surveillance of people and their activities*, IEEE Trans. PAMI press.

Konstantinova Pavlina, Alexander Udvarev, Tzvetan Semerdjiev., pp. III7-1 - III7-3, 19-20 June2003. *A Study Of A Target Tracking Algorithm Using Global Nearest Neighbour Approach*, International Conference on Computer Systems and Technologies - CompSysTech'2003 , Sofia, Bulgaria,

I. Cohen, G. Medioni., June 1999. *Detecting and Tracking Moving Objects in Video Surveillance,*Proc. of the IEEE CVPR 99, Fort Collins.

O. Javed, K. Shafique, and M. Shah., Dec. 2002. *A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information*, IEEE Workshop on Motion and Video Computing, Orlando, FL.

S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld., 2000. *Detection and Location of People using Adaptive Fusion of Color and Edge Information*, In Proceedings of International Conference on Pattern Recognition.

Gonzalez and Woods., 2002. *Digital Image Processing,* 2nd edition ,Prentice Hall.

Senior A., pp 48-55,2002. *Tracking People with Probabilistic Appearance Models*, PETS.

Carnegie Mellon University, visited on 06/22/2005 http://www.cs.cmu.edu/~vsam/OldVsamWeb/research .html.