Target Tracking with Background Modeled Mean Shift Technique for UAV Surveillance videos

Athilingam R [#1] Mohamed Rasheed A[#2] Senthil Kumar K [#3] Kaviyarasu A[#4] Thillainayagi R[#5]

[# 1,2,3,4,5] Division of Avionics, Department of Aerospace Engineering, Madras Institute of Technology, Anna University, Chromepet, Chennai – 600044, India

[1] aathikannan@gmail.com [4]isrokavi@gmail.com [5]thillaimit@gmail.com

ABSTRACT - Detection of the motion pattern of the object under study is the initial and major task in aerial surveillance. Object detection and tracking is the primary approach to find the interactions between the objects. Here the background modeled mean shift tracking algorithm is proposed for target detection in videos taken by Unmanned Aerial Vehicle. Traditional Mean shift tracking technique works well for dynamic objects with static background but UAV videos are of dynamic nature. The target may increase or decrease in size invariantly in consequent frames. Thus the proposed technique involves updation of background model adaptively for objects with random size variation and movement. The MATLAB based Simulation for the sample frames of UAV Videos is done and the performance evaluation is done to justify the efficiency of the proposed technique

Key words - Unmanned Aerial Vehicle, Video Surveillance, Object recognition, Object detection, Video signal processing

I. INTRODUCTION

In Computer vision, tracking an object is the primary and challenging task for surveillance applications. It Is the base for vehicle navigation, human computer interaction, l surveillance, traffic monitoring and autonomous navigation. In the field of aerial surveillance, the tracking methods involve finding the location and the displacement along with trajectory of the target under surveillance. The Unmanned Aerial Vehicle (UAV) (Fig 1) based surveillance videos have dynamic properties since each information in it is non stationary. Also the targets under surveillance do not move in fixed or predetermined pattern. Their movement will be random and undeterministic[1]. Thus an iterative and periodically updating technique is required to track targets so as to eliminate the target loss problem.

Fig 1: Unmanned Aerial Vehicle

Due to robustness, less computations, efficiency and ease of implementation Mean shift based tracking methods play a vital role in object extraction and tracking But, object occlusion, complex and unpredictable motion and varying shapes of objects that exist in mean shift techniques result in false and inaccurate tracking and cause tracking a complex task. Unlike usual videos, due to dynamic nature of UAV based Surveillance videos, an adaptive technique is required to track object effectively. Thus a background modeled Mean Shift tracking technique is proposed with background updated frame by frame. Thereby the problem of occlusion gets eliminated. The performance evaluation to detect tracking efficiency and error rate is done.

The organization of the paper is as follows. Section 2 describes the related works in the field of tracking, Section 3 describes the traditional and proposed mean shift tracking methods , Section 4 describes the results and relevant discussions, Section 5 is conclusion and the references,

II. RELATED WORKS

Abundant object tracking methods have been implemented based on different object representations and features. Target representation, detection and tracking for a single model is based on points, shapes contours silhouettes and articulated models[2]. Also the models are based on probability density function of target with Gaussian, Parzen window[3], and histogram model based on geometric shapes and color, texture, gradient magnitude. The color feature of the target and background is one of the most popular features. Yilmaz.A *et.al,* [2] classified target tracking into three categories – point, kernel and silhouette. Targets are considered as points in point tracking technique and point matching is done for the position of point in current frame and old frame. Kernel tracking is based on object shape and appearance. Silhouette tracking uses the region of object in each frame. Particle filter based appearance model and background subtraction is proposed by Bose *et.al,* [4]. Mean shift algorithm proposed by Comaniciu *et.al,* [5] determines the relation between frames by blobs. Mean shift tracking is robust since it is not adaptable to change in shape of size of object. The performance is dependent to the kernel size defined at the initialization stage. Presence of shadows and occlusion may result in errors. In the CAMSHIFT algorithm proposed by Chen J J *et.al,* [6] the target model is predicted in regular intervals and updated based on zeroth order moment of each pixel considered. Additional spatial information was fed to mean shift tracker by Birchfield *et.al,*[7] and Cai.N *et.al,* [8] where spatiogram is used to extract the spatial features. The technique prorposed by Fukunaga *et.al.*[9], uses fixed size tracking window with non parametric statistical approach. A study is done by Collins RT *et.al,* [10] to eliminate incorrect tracking which use difference of Gaussian kernel to improve the performance of tracking. Scale and orientation properties are considered to make modifications in kernel is proposed by Qifeng Q *et.al,* [11] and K.M.Yi *et.al,* [12] but the adaptation in size of target is not considered. Along with scale and orientation, other features like number, mean covariance of color pixels are also defined by C.W.Juan *et.al,* [13]. The technique proposed by Peng NS *et.al,* [14] uses updation using the estimation of color histogram. Two kinds of multidimensional histograms are proposed by J.Ning *et.al,* [15], N.E.O Connor *et.al,* [16] and H.R.Tacakoli *et.al,* [17] combines color and texture information of the object in each frame. Contourlet transform is employed in the techniques proposed by X.Tian *et.al,* [18] and J.Wu *et.al,* [19] for better performance towards spatial features. A robust feature called Adaptive contour feature is proposed by W.Gao *et.al,* [20] for human detection.

III PROPOSED APPROACH

A. Traditional mean shift tracking

The general tracking model of Mean shift algorithm involves initialization, creating model and similarity measure stages. The tracking window is defined on initial frame Fi and symmetric kernel is used to determine color histogram. The initialization stage involves definition of color system, size of the bin of histogram, the kernel function used and defining the area of the target. The models in the Mean Shift Tracking are of two categories - candidate model and target model. Target model represents the description of the target. The candidate model is used to compare the target model based on the similarity function. If candidate model is not equal to target model the current window is shifted. In tracking, the target is defined by ellipsoidal or rectangular region of an image. Due to robustness and being independent to partial occlusions Color histogram based target representation is usually preferred. On the initial frame the color feature of the window is classified into 'u' colors. Let the normalized pixels of target region be $\{X_i^*\}_{i=1..n}$ centered at the origin of the target area comprising n pixels. Let *u (u=1,2,…m) be* the feature of target model and the probability is computed as

$$\begin{cases} q^* = \{q_u^*\} \\ q_u^* = C\sum_{i=1}^{n} k(\|x_i^*\|)^2 \ \delta[\mathrm{b}(\mathrm{x_i^*}) - \mathrm{u}] \end{cases} \quad (1)$$

Where q* is target model, q_u* is the probability of u^{th} element of q*. The term b $\{x_i^*\}$ associates the histogram b_{in}, the pixel x_i* and isotropic kernel profile K(x). δ is the Kronecker function and it is 1 if the variables are equal.

$$\delta_{i,j} = \begin{cases} 0, if\ i \neq j \\ 1, if\ i = j \end{cases} \quad (2)$$

The normalization constant C is defined as

$$C = 1 \Big/ \sum_{i=1}^{n} k(\|x_i^*\|^2) \quad (3)$$

The probability of feature *u* with center position as y of the target model is

$$\begin{cases} p^*(y) = \{pq_u^*(y)\}_{u=1\ldots m} \\ q_u^* = C_h \sum_{i=1}^{nh} k(\|(y - x_i)/\ \mathrm{h}\|)^2 \ \delta[\mathrm{b}(\mathrm{x_i}) - \mathrm{u}] \end{cases} \quad (4)$$

$$C_h = {}^{1}/_{\sum_{i=1}^{nh} k(\|(y - x_i)/\ \mathrm{h}\|^2)} \quad (5)$$

Here, p*(y) is the target model, p_u*(y) is the probability of u^{th} element, h is the bandwidth, C_h is the independent normalization function. To calculate the similarity of the target and candidate model Bhattacharyya coefficient for two normalized histograms is defined as

$$\rho[p^*(y), q^*] = \sum_{u=1}^{m} \frac{1}{2}[p_u^*(y) q_u^*] \quad (6)$$

The distance between P*(y) and q* is calculated by

$$d\{\rho[p^*(y), q^*]\} = \sum_{u=1}^{m} \frac{1}{2}[1 - \rho[p_u^*(y), q^*]] \quad (7)$$

By Taylor expansion, the linear approximation of the coefficient is

$$\rho[p_u^*(y), q^*] \approx \frac{1}{2}\sum_{u=1}^{m} 1/2(p_u^*(y_0) q_u^*) + \frac{C_h}{2}\sum_{u=1}^{n_h} w_i k\ (\|y - x_i\|^2) \quad (8)$$

$$w_i = \sum_{u=1}^{m} \sqrt{\frac{q_u^*}{p_{u\,(y_o)}^*}}\, \delta[\mathrm{b}(\mathrm{x_i}) - \mathrm{u}] \quad (9)$$

The first term in eqn (8) id independent of y, the second term denotes the estimate of kernel density and it is calculated by kernel profile 'k' at the point 'y' on the current frame.

The estimated new position of target from y to y_1 is

$$y_1 = \frac{\sum_{i=1}^{nh} x_i w_i g[\left(\left\|\frac{y - x_i}{h}\right\|\right)^2]}{\sum_{i=1}^{nh} w_i g(\left\|\frac{y - x_i}{h}\right\|)^2} \quad (10)$$

Here the kernel chosen is Epanechnikov kernel, thus g(x) = -k(x) = 1.

The optimal location of 'y' can be obtained on each frame. Thus the new position is reduced to

$$y_1 = \frac{\sum_{i=1}^{nh} x_i w_i}{\sum_{i=1}^{nh} w_i} \quad (11)$$

Traditional Mean shift tracking method is robust to appearance changes and low computational complexity. But it assumes the size of target as fixed. But, practically the target varies in size frame to frame. The task becomes worse in case both the background and the target are moving. Tracking being dynamic process needs update of its feature dynamically.

B. Background modeled mean shift tracking

Object Tracking is the process of estimating the location/trajectory of a target with respect to time. Based on the current target position (state) gathered from the Target Detection step, and the previous states, the new location of the target are predicted. The objective is to discover the relation between the features of the current frame and the corresponding features of the previous frame. The proposed algorithm for tracking mainly comprises of three basic steps: flow estimation, background subtraction and mean shift tracking.

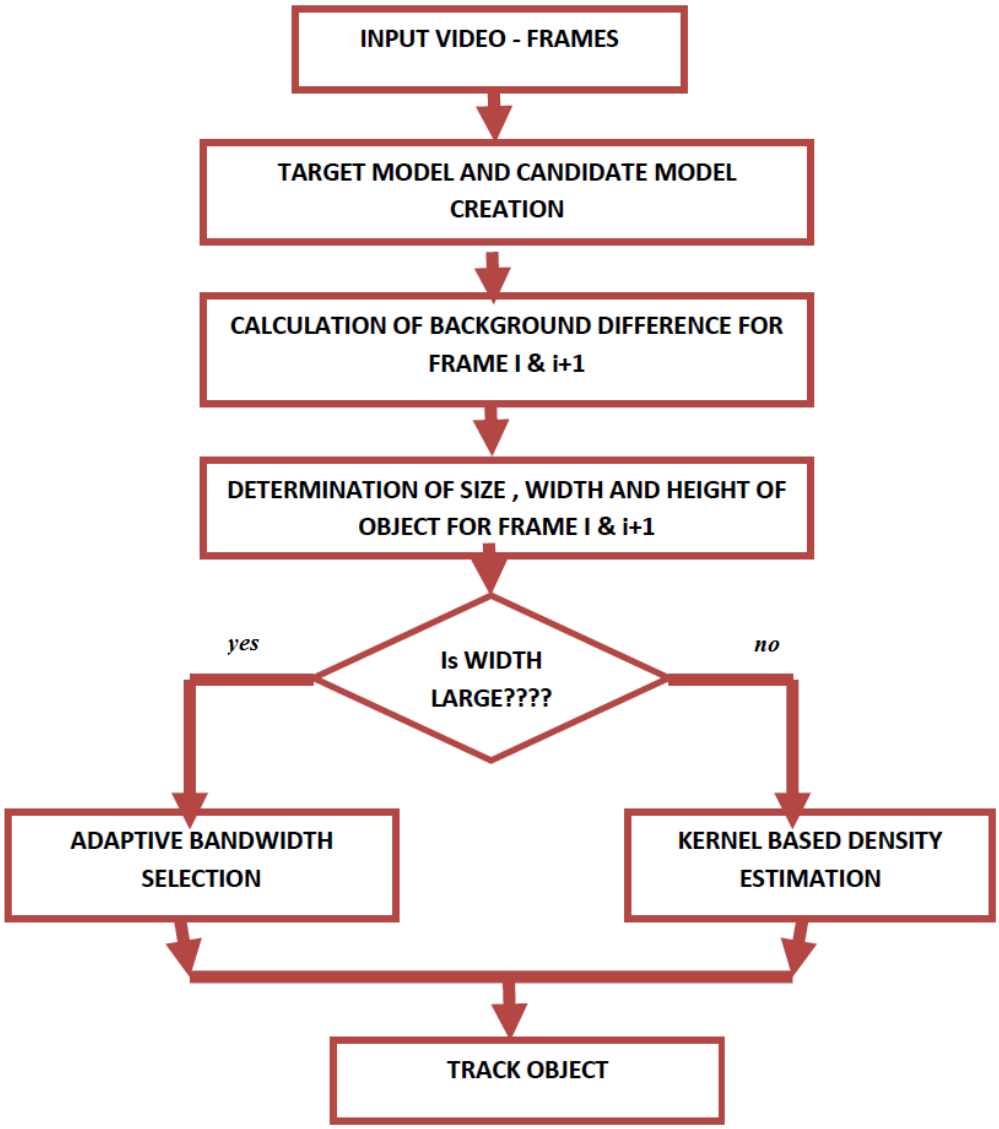

Fig 2 : Flow Diagram

C. Flow estimation

An affine transformation based flow estimation is proposed which is a linear 2-D geometric transformation and geometric model which maps variables at position (x,y) in an source image into new variables in an output image (x',y'). The motion between two successive frames can be modelled as an affine mapping. An affine flow field has six parameters. If the optic flow vector at an image location (x, y) is (v_x ,v_y), the first-order model is:

$$[v_x \; v_y] = [x \; y \; 1] \begin{bmatrix} d+s_1 & s_2+r \\ S_2-r & d-s_1 \\ v_{x0} & v_{y0} \end{bmatrix} \quad (12)$$

Where, s1 is the shear along the main image axes; s2 is the shear along the diagonal axes, (v_{x0}, v_{y0}) is the optic flow at the origin; d is the rate of dilation; r is the rate of rotation.

D. Background subtraction

Running average based background subtraction method is used due to its computational efficiency and low memory requirements. It involves subtraction of consequent frames based on the background model with a learning rate α

$$\mu_t = \alpha I_t + (1-\alpha)\mu_{t-1} \quad (13)$$

Where α is an empirical weight often chosen as a trade-off between stability and quick update, I is the pixel's current value and μ the previous average; Consider the input video with n frames $F_1, F_2 \ldots, F_n$. The background model Bi is the first frame of the input video is initialized. The next background model B_{i+1} is obtained by the difference between subsequent frames for i=1, 2…, n , α ranges from 0 to 1.

$$B_i = F_i, \; i=1 \tag{14}$$

$$B_{i+1} = \alpha F_i + (1-\alpha) B_i \tag{15}$$

The background update Gi is obtained by subtracting the subsequent frames Fi with the background model Bi.

$$F_i - B_i = G_i \tag{16}$$

The resultant binary image R_i depends on the value of threshold T. Here the threshold is made adaptive by calculating the mean of all the pixels in each frame.

$$T = mean\,(G_i) = \frac{1}{n^2}\sum_{i,j} G(i,j) \tag{17}$$

E. Adaptive Bandwidth Mean shift tracking

The correctness of mean shift tracking relies on exact candidate model assignment. The proposed method(Fig 2) involves detecting the target and defining the target model from the first frame. The target model is defined as a rectangle with Epanechnikov kernel. The features are centered radially symmetric and normalized distance based on Epanechnikov kernel

$$k(x) = \begin{cases} 1 - x^2, \; if\; 0 \le x \le 1 \\ 0, \quad otherwise \end{cases} \tag{18}$$

This initialization stage involves determination of area of the target model with center *y*. The color system and the number of bins for the histogram are defined. Each bin of an m-bin histogram depicts the number of same color pixels.

$$q^* = \{q_u^*\}_{u=1,\ldots m} \tag{19}$$

Here q_u^* is the number of same color pixels divided by the total pixels in the target area. The current incoming frame is assumed as y_o^*. The sum of probability should be equal to 1. Let $b(x_i)$ be the index of index of the bin which the pixel x_i belongs to and *u* be the bin number in the histogram. Assuming h as bandwidth of frame, m be the total number of bins and n as the number of pixels in the current window, C as the normalization constant , target model q_u^* is defined as

$$q_u^* = C \sum_{i=1}^{n} k\left(\left\|\frac{y-x_i}{h}\right\|\right) \delta_u(b(x_i)) \tag{20}$$

$$C = 1 \Big/ \sum_{i=1}^{n} k\left(\left\|\frac{y-x_i}{h}\right\|\right) \tag{21}$$

$$\delta_u(x) = \begin{cases} 1, if\; x-u=0 \\ 0, otherwise \end{cases} \tag{22}$$

$$p_u^*(y_0^*) = \begin{cases} k\left(\left\|\frac{y-x_i}{h}\right\|\right)\delta_u\big(b(x_i,J)\big) - \\ \sum_{i=1}^{n} \; g(x_i)\delta_u(b(x_i\,J-1))), \\ if\; b\,(x_i,J) = b(x_i,J-1) \\ \sum_{i=1}^{n} k\left(\left\|\frac{y-x_i}{h}\right\|\right)\delta_u\big(b(x_i,J)\big), \; otherwise \end{cases} \tag{23}$$

Also, g(x) = -k(x) = 1.The similarity of the target and candidate model is calculated by Bhattacharyya coefficient for two normalized histograms is defined as

$$\rho[p^*(y_0^*), q^*] = \sum_{u=1}^{m} \frac{1}{2}[p_u^*(y_0^*) q_u^*] \tag{24}$$

The distance between P*(y) and q* is

$$d\{\rho[p^*(y_0^*), q^*]\} = \sum_{u=1}^{m} \frac{1}{2}[1 - \rho[p_u^*(y_0^*), q^*]] \tag{25}$$

With *n* be the number of pixels in the current window, the new position y_i^* is

$$y_i^* = \frac{\sum_{i=1}^{n}\sum_{u=1}^{m} x_i \sqrt{\frac{q_u^*}{p_u^*(y_0^*)}}\, \delta_u(b(x_i))}{\sum_{i=1}^{n}\sum_{u=1}^{m} \sqrt{\frac{q_u^*}{p_u^*(y_0^*)}}\, \delta_u(b(x_i))} \tag{26}$$

Consider area (*A*) of the background subtracted object in current frame and previous frame to determine the difference in change in background. . Let *l* be length of vector that defines the distance between

position of last frame and current frame. Let *F=1* if there is background. Let Area A(x,y,u) where x,y are the pixel locations and u is the frame number, then the vector and the respective slope *m* is calculated by

$$F = \begin{cases} 1, & if\ A(x_1, y_1^*, J) - A(x_0, y_0^*, J-1) \neq \emptyset \\ 0, & otherwise \end{cases} \quad (27)$$

$$\Delta x = \frac{y^*_{(1,x)} - y^*_{(0,x)}}{2} \quad (28)$$

$$\Delta y = \frac{y^*_{(1,y)} - y^*_{(0,y)}}{2} \quad (29)$$

$$l = \sqrt{\Delta x^2 + \Delta y^2} \quad (30)$$

$$m = \frac{y^*_{(1,y)} - y^*_{(0,y)}}{x^*_{(1,x)} - x^*_{(0,x)}} = \frac{\Delta x}{\Delta y}, if\ \Delta x \neq 0 \quad (31)$$

$$l^2 = (y^*_{(1,y)} - y^*_{(0,y)})^2 + (y^*_{(1,x)} - y^*_{(0,x)})^2 \quad (32)$$

After determination of new position of the window and the candidate model , the similarity function is calculated.

$$\rho[p^*(y_1^*), q^*] = \sum_{u=1}^{m} \frac{1}{2} [p_u^*(y_1^*) q_u^*] \quad (33)$$

And if $\rho[p^*(y_1^*), q^*] < \rho[p^*(y_0^*), q^*]$, $y^*_{1 <-- 1/2\,(y_1^* + y_0^2)}$ (34)

The resultant frame is stored as the previous frame for the incoming next frame.

FOR all frames 1 to n

If frame =1

Define bin size , object , kernel function, color system, and Create Target model

For frames 2 to n

Create candidate model Yo.

Store as previous frame and compare with next frame to obtain background subtraction.

If previous object = next object

Assign next frame as previous frame and shift to new position and use KDE based tracking.

Else

Define area of new background and Calculate new background model and shift to new position

End If

End IF

IV. IMPLEMENTATION

Tracking is done and quantitative analysis is done by calculating Pixel Wise Performance Metrics. Videos obtained from Unmanned Aerial are considered for analysis. The videos are transmitted with the transmission frequency of 1.2 GHz from UAV DHAKSHA to Ground Control Station, with altitude ranging from 10 to 50 meters. The properties of the videos are listed in table 1

TABLE 1: INPUT VIDEO PROPERTIES

Property	Video
No of frames	76
Frame rate	25 FPS
Size	240 x 432

V. RESULTS AND DISCUSSION

Videos obtained from Dhaksha-UAV are first converted into frames. The resultant frames are used for further processing. The sample input frame is shown in fig 3. The result of traditional mean shift algorithm is shown in.4.and 5. The affine flow computation is shown in fig 6. The background subtracted image is shown in fig 7. The samples frames from result of Adaptive MST are shown in fig 8 and 9.

Fig 3: Input Frame

Fig 4 MST Output Video frame 1

Fig 5 MST Output Video frame 2

Fig 6: Flow Field

vx0: -31.8589 vy0: -20.9576
d: -0.0167 r: -0.0057
s1: 0.0199 s2: 0.0360

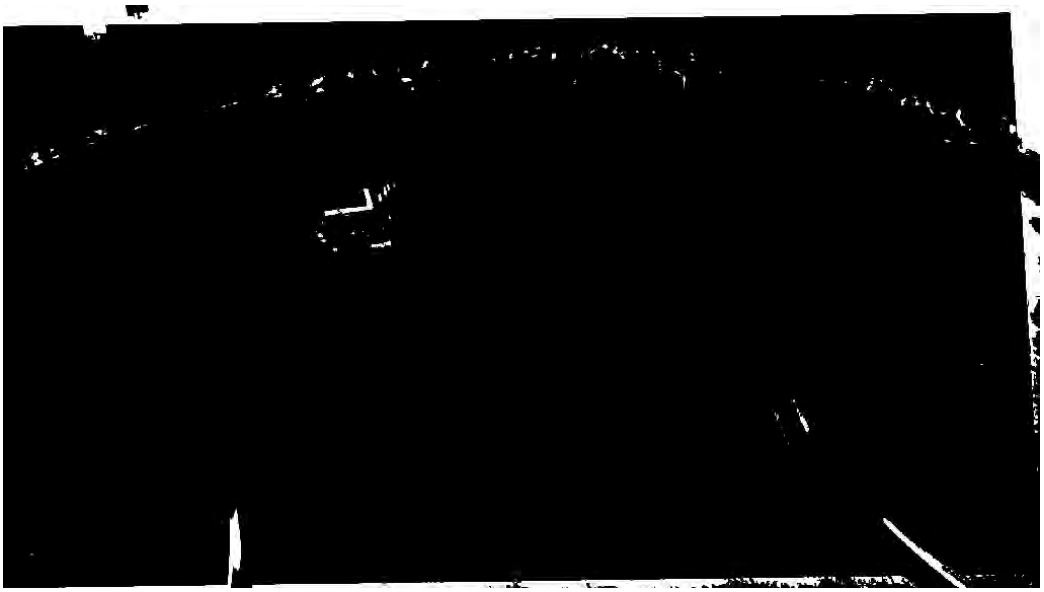

Fig 7: Background subtraction

Fig 8 ABMST Output Video 1 frame 1

Fig 9 ABMST Output Video 1 frame 1

To evaluate the performance of the proposed technique and traditional technique tracking rate, Duration of tracking and object tracking error are considered. Duration of Successful tracking (DST) calculates the ratio of total seconds of successful tracking and total seconds in video. Tracking rate (TR) calculates the percentage of successful tracking of target by the ratio of number of successful detection to the total frames. Object tracking error (OTE) calculates the tracking error by comparing expected position of center in incoming frame and the position actually obtained. The results are listed in table 2. The comparative results show that the proposed technique is efficient.

$$Tracking\ Rate = \frac{Number\ of\ successfully\ tracked\ frames}{Total\ number\ of\ frames\ in\ video} \tag{35}$$

$$Duration\ of\ Successful\ tracking = \frac{Number\ of\ seconds\ of\ correct\ tracking}{Total\ seconds\ of\ video} \tag{36}$$

$$Object\ tracking\ error\ = \frac{1}{F_{overlap}} \sum_{i=1}^{n\ frames} \sqrt{(x_i^{exp} - x_i^{obt})^2 + (y_i^{exp} - y_i^{obt})^2} \tag{37}$$

TABLE 2 : PERFORMANCE METRICS

METRIC (rounded in %)	MEAN SHIFT	ABMST
TR	77	86
DST	79	90
OTE	27	16

VI. CONCLUSION

In this paper, a new approach towards mean shift tracking is proposed for effective tracking of targets with change in size. The background model is updated for each frame and the target model representation is updated accordingly. The experimental result shows that Adaptive background mean shift tracking method is efficient for videos of dynamic nature. The proposed method provides a good degree of correctness in tracking targets with irregular changing nature.

REFERENCES

[1] Lin F, Lum K Y and Chen B M., "Development of a vision based ground target detection and tracking system for a small unmanned helicopter", Sci. China Ser F Inf. Sci., Vol 52 (11), 2009, pp 2201–15.

[2] Yilmaz A, Javed O and Mubarak S, "Object Tracking: A Survey", ACM J Comput. Surv. Vol 38(4), 2006, pp 1-45.

[3] Chen J J, An G C and Zhang S F, "A Mean shift algorithm based on modified Parzen window for small target tracking", In Proceedings of IEEE Int. Conf Acoust. Speech Signal Process. Dallas, 2010, pp 1166–69.

[4] Bose, B, Wang X and Grimson E, "Detecting and Tracking Multiple Interacting Objects without Class Specifc Models", Technical report, Massachusetts Institute of Technology, MIT-CSAIL-TR-2006-027, 2006.

[5] Comaniciu D, Ramesh V and Meer P, "Kernel-based object tracking", IEEE Trans. Pattern Analysis and Mach. Intell. Vol 25(5), 2003, pp 564–77.

[6] Chen J J, Zhang S F and An G C, "A generalized mean shift tracking algorithm", Sci China Inf Sci, Vol 54 (11), 2011, pp 2373 – 85.

[7] N.Cai and N.Zhang, "Infrared target tracking in sea clutter background based on spatiograms", Laser & Infrared, Vol 40(8), 2010, pp 910–16.

[8] Birchfield S.T and Rangarajan S. "Spatiograms versus histograms for region-based tracking". In Proceedings of IEEE Conf. Comput. Vis. Pattern Recogni., San Diego, CA USA, 1158-63, 2005.

[9] Fukanaga k and Hosteler LD, "The estimation of the gradient density function with application in pattern recognition", IEEE Trans.Info Theory, 21(1), 1975, Pp 32 -40.

[10] Collins R. T. Mean-shift blob tracking through scale space. In Proceedings of IEEE Comput. Society Conf. Comput. Vis. Pattern Recognit., Wisconsin, USA, 2003, pp 234– 40.

[11] Qifeng Q, Zhang D, and Peng Y, "An adaptive selection of the scale and orientation in Kernel based tracking", In Proceedings of IEEE Conf. Signal - Image Technol. Internet Syst., Shanghai, China, 2007, pp 659–64.

[12] K.M.Yi, H.S.Ahn and J.Y.Choi, "Orientation and scale invariant mean shift using object mask based kernel", In Proceedings of 19th Int. Conf. Pattern Recognit. Tampa, FL, USA, 2008, pp 1- 4.

[13] C.W.Juan and J.S.Hu, "A new spatial color mean shift object tracking algorithm with scale and orientation estimation", In proceedings of IEEE Int. Conf. robot. Automat. Pasadena, Ca, USA, 2008.

[14] Peng NS , Yang J, Liu Z and Zhang FC, "Automatic selection of kernel bandwidth for mean shift object tracking", J of Software, 16, 2005, pp 1542 -1550.

[15] J. Ning, L. Zhang, D. Zhang and C. Wu, "Robust object tracking using joint color texture histogram", Int.J. Pattern Recognit. Artif. Intell. Vol 23(7), 2009, pp 1245–63.

[16] N.E. O'Connor, P. Kehoe, C.O'Conaire and A.F. Smeaton, "Vehicle tracking in UAV video using multi-spectral spatiogram models", In Proceedings SPIE, Vol 6974, 2008.

[17] H.R. Tavakoli and M.S. Moin, "Mean shift video tracking using color LSN histogram", In Int. Symp. Telecommun, Tehran, 2010, pp 812–16.

[18] X. Tian, D. Yang and C. Du, "Image retrieval method based on color and Contourlet histogram", Comput. Eng. Vol 36(1), 2010, pp 224–27.

[19] J. Wu and Z. Cui, "Research on vehicle tracking algorithm using Contourlet transform", In Int. IEEE Conf. Intelligent Transp. Sys., Washington DC, 2011, pp 1267–72.

[20] W Gao, H. Ai and S. Lao, "Adaptive contour features in oriented granular space for human detection and segmentation", In Proceedings of Comput. Vis. Pattern Recognit., Miami, FL,2009, pp 1786 – 93.