International Journal on Emerging Technologies **11**(3): 402-411(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Video Object Segmentation and Classification with Improved Kalman Filter-based Method for People Tracking

Dr. Matheswari Rajamanickam
Assistant Professor, Department of Computer Science,
M.E.S. College of Arts, Commerce and Science, Bengaluru (Karnataka), India.

(Corresponding author: Dr. Matheswari Rajamanickam)
(Received 19 February 2020, Revised 15April 2020, Accepted 17 April 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Video surveillance system focuses on object tracking over a particular environment or boundary. Object recognition and tracking in crowded areas, identifying chromatic shadows of objects, occlusion detection and tracking of multiple objects in scene is still a promising task in video surveillance system. The recent advent in digital video processing technology also led to the development of advanced research work in the field. However, not all the existing methods are efficient and powerful in terms of the computational time and complexity involved in the detection and tracking of moving objects. To alleviate this problem, the proposed research work introduces a new integrated framework for segmentation and classifications of moving objects with an extended Kalman filter algorithm for tracking the pattern-matched moving objects. The contributions of the research work lies in three folds namely, Segmentation, Classification and Pattern matching. In the first phase for efficient segmentation of moving objects, Multi-textured Object Segmentation (MTOS) method is proposed. Further, the research work presents a Multi-Class Spatial Classifier (MCSC) method of classifying moving objects. Finally, the research work presents a quantization of signals using extended Kalman filter-based pattern matching (QKF-PM) method for improving the efficiency of pattern matching method for the process of detecting and tracking multiple moving objects. The experimental result shows that the newly proposed approach MTOS, MCSC and QKF-PM methods produces much improved results in detecting and tracking of moving objects.

Keywords: Bounding box, Kalman filter, MCSC, MTOS, QKF-PM, Top-down approach.

Abbreviations: ACMs, Active Contour Models; AFABS, Advanced Fuzzy Aggregation-based Background Subtraction; BISVM , Batch Incremental SVMs; SIFT, Scale Invariant Feature Transform; E-RBPF, Enhanced Rao-Blackwellized Particle Filter; MLPs, Multi-Layer Perceptron; RNNs, Recurrent Neural Networks; CFOL, Correlation Filters and Online Learning; MTOS, Multi-textured Object Segmentation; MCSC, Multi-Class Spatial Classifier; QKF-PM, Quantized Kalman Filter based Pattern Matching; MRF, Markov Random Fields; MAP, Maximum a Posteriori; PSE, a Posteriori State Estimation; FG, Foreground; LQE, Linear Quadratic Estimation; SA, Segmentation Accuracy; ST, Segmentation Time; PSNR, Peak Signal-to-Noise Ratio; CA, Classification Accuracy; CT, Classification Time; FPR, False Positive Rate; PMA, Pattern Matching Accuracy; PMT, Pattern Matching Time; MODR, Moving Object Detection Rate.

I. INTRODUCTION

Object recognition system identifies the presence of an object from the given input image frames. Moving regions are obtained by subtracting the current image on pixel-by-pixel basis from a reference background image. Background subtraction models are sensitive to dynamic changes caused by the instances like a parked van suddenly moving out of vehicle parking area, illumination variations etc. As a way of addressing these issues, temporal differencing of consecutive frames on a pixel-by-pixel basis is performed. For a moving camera, the background also keeps changing with respect to time. Some research approaches makes use of temporal differencing method to minimize the false detection rate.

An object detection and tracking methods reads a sequences of image frames from input video stream. Image processing methods are applied for each frames extracted from a video sequence. Fig. 1, illustrates the steps involved in the detection and tracking of moving objects.

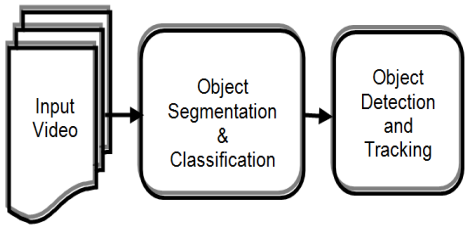


Fig. 1. Object Detection and Tracking.

II. RELATED WORKS

Video surveillance is used to monitor the behavior, activities and other dynamic information about people, vehicle or objects of any kind intended for the purpose of providing security in geographical space. There has been several researches focused on these tasks and different methods have been adopted for automation.

Active Contour Models (ACMs) developed [1]. This system is suitable for tracking of moving targets of any

colors, in varied scenarios like lighting, or change in the target object. However, the method lacks in its performance in the presence of dynamic shadows of objects. An advanced fuzzy aggregation-based background subtraction (AFABS) was designed [2] for moving object detection in dynamic background conditions. However, the performance efficiency of the system declines on heavily dynamic background conditions

An intelligent system was designed [3] using Batch incremental SVMs (BISVM) classifier for categorizing the moving objects. The methods needs to be combined with feature extraction and classification algorithms for crowded urban regions.

A multi-sensor fusion framework was introduced [4] for Moving Object Detection and Tracking for reducing false detections. However, the classification precision changes constantly to different driving scenarios compromising in its reliability. Hence, to improve the reliability of classifying objects and detecting them appropriately, a context-based learning phase methods need to be incorporated to estimate and evaluate the different parameters of moving objects.

A method for point extraction called scale invariant and transforming of features was presented [5]. The research work initially identifies the location where minimum motion occurs. The research work needs to consider color information for more strong point matching and object tracking process.

An improvised Rao-blackwellized particle filtering method called (E-RBPF) was presented in [6]. A method for detecting and tracking multiple moving objects using kinematic planar and redundant manipulators was presented [7]. However, the methods [6-7], failed to detect moving objects in dense crowded environment.

A discriminative model based on an ensemble of multilayer perceptron (MLPs) to track moving objects was introduced [8]. An online multi-target tracking system based on recurrent neural networks (RNNs) was proposed in [9]. An end-to-end learning method was used for online multi-target tracking. An efficient pattern matching algorithm viz, strip subtraction and strip division was described more precisely and the adoptability of this algorithm under different scenarios are presented in the study [10].

An appearance based modeling using MRF framework was developed [11] for segmenting objects in video sequences. The method needs to incorporate the appearance model as supplementary nodes and borders inside the MRF structure in a graph cut. An adaptive compressed sensing system was developed [12] for detecting and tracking objects based on the salience of the scene. The research work aims at reducing the computational cost and therefore the system resources are scaled to meet only the prerequisite of the scene.

A recurrent neural network for tracking static and dynamic objects was developed [13]. The model do not require any supervision and is capable of predicting the future states of objects based on the input video sequence.

A combined texture and shapes based object detection and recognition model was developed [14]. A convolutional neural network for deep metric learning was developed [15] for enhancing the tracking performance. The appearance metric was used to predict the pedestrian's trajectories.

A feature descriptors for network-flow formulation based on data association through back-propagation method was developed [16]. This method makes use of differentiable function for finding the optimum smoothed network flow problem. A system to address the issues of online tracking and categorization of multiple objects. The method adaptively learns the shape and motion of target objects and employs this method for efficient tracking and classification of objects [17]. However, the above mentioned works [16-17] needs to integrate advanced hybrid algorithms for faster object matching.

An object tracking method which performs online classification of objects based on super-pixel values was proposed [18]. The method is an integration of local and holistic models. These models are representative to discriminative and generative methods used for detection, classification and tracking process. The Correlation Filters and Online Learning (CFOL) system was developed [19] for visual tracking. A new sampling method integrated with online learning and update strategy was performed for effective target tracking. A tracking method for identifying each human using a single camera and across multiple disjoint cameras was presented in the study [20]. A method to detect and track objects captured using single camera, an adaptive multiple kernel system was used. For tracking objects captured using multiple cameras, appearance and context features of objects are combined to detect the couples automatically. However, these algorithms [18-20] does not consider threshold decision in segmenting and classifying objects.

The canny edge detection method based on fuzzy logic was proposed [20-21] for gray scale images. However, the method does not consider size, shape and texture features of objects in the image.

It is observed that under changing illumination condition, detecting shadows of objects and occluded parts identification, is still an open issue in the video surveillance system. Also handling multi-target objects is still far from reality. The consideration of size, shape and texture features of objects also will improve the accuracy of detection and tracking of objects.

Therefore the above mentioned research gaps are overcome using methods proposed in the study for detection and tracking of moving objects.

III. MATERIALS AND METHODS

From extensive literature review and the results obtained from the existing methods, it is observed that the preprocessing of videos is also one of the emergent areas that can improve the detection results. The segmentation and classification process can be made more effective by classifying pixels belonging to foreground and background using threshold methods. Also, devising a scheme for background modelling and filtering techniques play an important role in the field. Therefore, the present research work is focused on improving the performance of object detection and tracking process using the proposed methods for segmentation, classification and pattern matching of objects. The present research work would reduce the false positive identification of objects with an improved moving objects' detection rate. Therefore, a new

integrated framework for identifying moving objects in lesser time period is an essential and demanding requirement in the field of research.

The proposed method lies in three folds. A Multi-textured Object Segmentation (MTOS) method is proposed to improve the segmentation of objects presents in the image frames. A Multi-Class Spatial Classifier (MCSC) method is developed for improving the classification accuracy. An extended Kalman filter with quantization of signals is introduced for improving the performance of detection and tracking of moving objects.

A. Object Segmentation using adaptive threshold based on multiple texture features

The research works reads input video sequence as shown in Fig. 2 (a). The image frames extracted from the video sequence are mathematically expressed as follows:

$$VFi = VF1, VF2, VF3. . , VFn \quad (1)$$

(a) Input Video

(b) Extracted Image frames.

Fig. 2.

The extracted image frames are normally subjected to different noises like impulse noise, Gaussian noise, shot noise, quantization noise etc. The noisy data prevents from determining the edges of an image objects. To alleviate this problem, the proposed method performs pre-processing of images using median filtering using the Eqn. (2). The median value for a 3 × 3 square window size can be expressed as follows:

$$\text{Median}\,[m,n] = med\left\{\frac{m*n}{VF_n}\right\} \quad (2)$$

The adaptive threshold-based approach is more appropriate for efficient segmentation and is used for categorizing the pixels according to their properties say, a. spatial values, b. the gradient of their gray levels and c. the homogeneity of their textures. Sharp edges are segmented using appropriate threshold values. This effectively reduces the dynamic shadows of moving objects. Fig. 3 (b) shows the resultant image after applying adaptive threshold value.

Labels are assigned to all the pixel values present in the image. Based on the pixel value, foreground (FG) and background (BG) objects are identified. An object is marked as FG object if the pixel value is greater than threshold values; otherwise it is marked as BG image. The following equation defines the texture features, which are extracted from two images of distinctly different textural characteristics.

$$f_1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(\frac{P(i,j)}{R}\right)^2 \quad (3)$$

$$f_2 = \sum_{i=0}^{N_g-1} n^2 \left\{\sum_{|i-j|=n}\left(\frac{P(i,j)}{R}\right)\right\} \quad (4)$$

$$f_3 = \frac{\sum_{i=1}^{N_g}\sum_{i=1}^{N_g}\left[\frac{ijP(i,j)}{R}\right]-\mu_x\mu_y}{\sigma_x\sigma_y} \quad (5)$$

The Eqns. (3-5), 'f_1', 'f_2' and 'f_3' denotes texture features of person 1, texture feature of person 2 and the texture feature of hall respectively.

The method assigns values to each pixel in the image as either belonging to foreground or background. The adaptive threshold method is formulated as follows:

$$AT = AT[a,b,p(a,b),f(a,b)] \quad (6)$$

In the above equation (6), the adaptive threshold 'AT' is measured based on the pixel coordinates '(a,b)' of the threshold value point. Here, $p(a,b)$ and $f(a,b)$ represents the gray level frame pixels. Based on the threshold property, the moving object region (foreground) is extracted using texture features on pixel-by-pixel basis. The value of threshold for all pixels is evaluated as either 0 or 1. The threshold frame *g(a,b)* is expressed as follows:-,

$$g(a,b) = \begin{cases} 1 & if \quad p(a,b)f(a,b) > T_{H.} \\ 0 & otherwise \end{cases} \quad (7)$$

From Eqn. (7), it is clear that if the pixel value of the frame is greater than the threshold value, then it is considered as the foreground value and it is represented as 1. Otherwise it is considered as a background. From the threshold measurement, the moving object regions are obtained from the background for performing segmentation of objects. Fig. 3 (a) shows the texture extraction, adaptive threshold object and the result of segmentation based on multiple texture features.

Fig. 3 (a) Multiple texture extraction (b) Adaptive threshold object. (c) Segmented image.

B. A Multi-Class Spatial Classification of Objects

Once the objects are segmented, the next step is to classify objects in to different classes. The classification of objects performed in spatial domain with respect to spatial features like areas, region, roads, rivers etc. gives significant improvement in the detection and tracking process. The method uses Markov Random Fields (MRF) principle for identifying both fixed and moving objects. The two different label fields say the region plots 'R' and the spatial region plots 'S' are considered. The spatial region plot helps in identifying moving objects, object boundary and occluded objects. The labels like motion labels, boundary labels, occlusion labels etc. are assigned appropriately. The MRF principle classifies the input video frames into region plot and the spatial region plot. The resultant label field 'X' is obtained by combining these two labels. The joint pair of 'R' and 'S' are evaluated using the Hammersley-Clifford theorem. Hence, the joint probability density function of the Markov Random Field is mathematically expressed as follows:-,

$$P(R_l, X_l, \varphi_l) = \frac{1}{Z}\exp(-E_l(R, X|\varphi_l)) \quad (8)$$

In the above equation (8), the normalization factor is represented as 'Z'. Here, the expression '$E_l(R, X|\varphi_l)$',

denotes the local energy function which is used to measure, "how well 'R' and 'S' fit together around the location 'l'". The local joint neighborhood function around the location 'l' is represented as 'φ_l'. The energy function is measured as follows:-,

$$E_l(R,X|\varphi_l) = -\sum_{t\in\varphi_l} \delta(R_t,R_l)\,\delta(X_t,X_l) \quad (9)$$

In the above equation (9), '$\delta(X_t,X_l)$' and '$\delta(R_t,R_l)$', is the Kronecker delta function and is expressed as follows:-,

$$\delta(X_t,X_l) = \begin{cases} 1 & if\ \ X_t = X_l \\ 0 & otherwise \end{cases} \quad (10)$$

The above equation (10) is the fusion model. This fusion model is used to determine the spatial similarity of the joint pair of Markov Random Fields. The mathematical calculation for determining the resultant label field is expressed as follows:-,

$$X = arg\ min_x \sum_{l\in L} E_l(R,X|\varphi_l) \quad (11)$$

From the above equation (11), the resultant label field is represented as 'X'. Fig. 4 shows the output of the shape extraction using MRF principle in frame1 and frame 2 respectively.

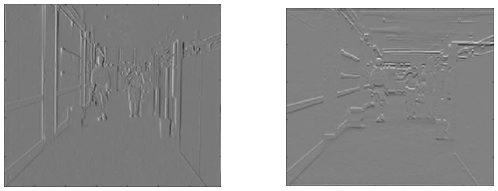


Fig. 4 (a) MRF object in frame 1. (b) MRF object in frame 2.

Next the proposed MCSC method performs the edge-strength estimation for preserving the object boundary. The edge-strength estimation is performed using fuzzy '$IF\ldots THEN\ldots$' rules. Let us consider the two gradient operators 'G_x' and 'G_y' at a point '(x,y)' in the image. The gradient of the object is mathematically defined as follows:-,

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (12)$$

In the above Eqn. (12), '∇f' denotes the gradient of the object. Here, the derivative with respect to the gradient in 'x'direction is represented as '$\frac{\partial f}{\partial x}$', and the derivative with respect to the gradient in 'y' direction is represented as '$\frac{\partial f}{\partial y}$' respectively.

The fuzzy triangular membership function is used to determine the edge-strength for preserving the object boundary. From Fig. 5, it is clear that the three resultant values low, medium and high are obtained using the fuzzy sets.

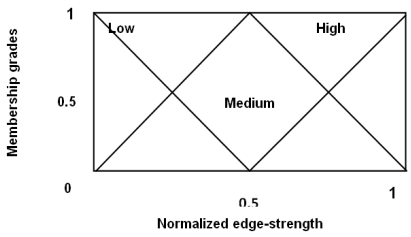


Fig. 5. Fuzzy Membership Function.

The normalized edge-strength, '$N(x,y)$', is measured based on the number of horizontal and vertical intensity gradients at pixel location *(x, y)* in the object. The input image with fuzzy edge-strength measure is shown in the following Fig. 6.

Fig. 6. Fuzzy normalized edge-strength for frame 1.

C. Spatial Classification of Objects Using MAP Estimation

Next the research work performs spatial classification of objects using maximum a posteriori (MAP) estimation.

This MAP estimation function classifies the objects present in the video frames effectively based on different class labels. Likelihood of each pixel's class label is derived from the classification model to produce the final classified results. Bayes' theorem is used for classifying the segmented objects for reducing the false positive rate. The posterior probability function is formulated as follows:

$$P(X|D) = \frac{P(D|X)\,P(X)}{P(D)} = \frac{\text{liklihood} * \text{prior information}}{\text{dxnormalizing factor}} \quad (13)$$

From the above Eqn. (13), the expression, '$P(X|D)$', represents the posterior probability function. Here, P(X) is the probability of prior information representing the true classification i.e., the ratio of the number of location and the different class labels. Therefore, Bayes inference rule produces the maximum a posteriori (MAP) estimation in equation, which is expressed as follows:-,

$$arg\ max\, P(X|D) = arg\ max\ \ P(X|D)\,P(D) \quad (14)$$

The MAP function finds the maximum value of the posterior using Eqn. (14). Here, the posterior is the probability density function. Based on the above formulation, the maximum posterior probabilities for different class labels are obtained.

Let us consider a set of segmented objects say, '$O_i = o_1, o_2, o_3.., o_n$' with the corresponding class labels as '$c_i \in C$'. The "C_{MAP}", represents the MAP estimation. The object classification using the Bayes classifier is mathematically expressed as follows:-,

$$c_{MAP} = arg\ max_{c_i\in C}\, P(c_i|O_i = o_1, o_2, o_3.., o_n) \quad (15)$$

In the above Eqn. (15), the classifier predicts the probability of the class which has most likelihood value. Hence, the MAP function returns the class value for these probabilities as high in the given frames. Objects are classified with different class labels. Fig. 7 shows the spatial classification (SC) of input image.

Fig. 7. SC at Level 1, Level 2, Level 3.

D. Quantization of signals and Extended Kalman Filter-Based methods for Object Detection cum Tracking Process

The proposed method is designed using top-down approach to identify the chromatic shadows of objects. The FG motion objects are traced with respect to objects at time 't'. The initial Kalman filter (KF) is used to track the detected objects. Fig. 8 (a, b, c) shows the input frames, shadows of object at time 't' and foreground (FG) objects at time 't' respectively.

Fig. 8. (a)The Input image frames (b) Shadows objects at time "t" (c) Moving FG objects at time "t".

The Kalman Filter, also termed as linear quadratic estimation (LQE) takes the sequence of images captured over time 't' as input. The general first order Kalman Filter is mathematically expressed as follows:

$$x_k = S_k\, x_{k-1} + b_k\, c_k + n_k. \quad (16)$$

From the above Eqn. (16), "x_k" denotes the "k^{th}" frame. The state transition model is represented using the variable "S_k". This is applied to the prior state "x_{k-1}".

The measurement of the true state at any given time 't', can be formulated as follows:

$$y_k = C_k x_k + u_k \quad (17)$$

From the above Eqn. (17), the observation model is represented as 'C_k' that maps the true state space into the observed space.

The Kalman Filter includes a two-step process. The prediction phase and the update phase. Using the prior state, the prediction performs the state estimation for obtaining the current state as follows:-,

Prediction State Estimation = $a_{k|k-1}$

$$= E_k a_{k-1|k-1} + D_k j_k \quad (18)$$

Prediction Error Covariance = $p_{k|k-1}$

$$= E_k p_{k-1|k-1} {E_k}^T + Q_k \quad (19)$$

From the above Eqns. (18) and (19), the variable "E_k" denotes state transition model to the prior state "a_{k-1}". Here, the variable "D_k" denotes the control-input model. The variable, "j_k" and "Q_k" denotes the control vector covariance. The observation information with the current a priori state prediction are combined and the updated a posteriori state estimation (PSE) are derived. The PSE can be expressed as given below.

$$a_{k|k} = a_{k|k-1} + K_k\, r_k \quad (20)$$

The updated state estimation is represented using the variable "$a_{k|k}$". The optimal Kalman gain is represented using 'k' and "K_k". The the pre-fit and post-fit residual can be expressed as given below.

$$r_k = w_k - H_k a_{k|k-1} \quad (21)$$

$$r_{k|k} = w_k - H_k a_{k|k} \quad (22)$$

From equation (21) and (22), "w_k" denotes the time '*k*' measurement. Here, "H_k" represents the different observation matrices. The optimal Kalman gain is computed as follows:-,

$$K_k = p_{k|k-1} H_k^T U_k^{-1} \quad (23)$$

From Eqn. (23), "U_k" denotes the pre-fit residual covariance. The Updated a posteriori covariance estimation can be formulated as given below.

$$p_{k|k} = (1 - K_k H_k) p_{k|k-1} \quad (24)$$

The variable "$p_{k|k}$" is the updated covariance estimation at discrete time 'k'.

The Kalman Filter method examines the association between foreground and shadows to ensure temporal reliability in matching objects. In order to perform moving FG object matching, the FG and shadow objects are associated at time 't and 't-1'.

Next, to improve the identification of chromatic shadows of objects, top-down approach has been used and shadow objects are updated at the tracking end. The following figure shows the updated shadow images.

(a) Input Image frame (b) FG and Shadow blobs associations.

Fig. 9.

After chromatic shadows of objects are identified, the FG objects are detected using kernel pattern matching method. The input mapping using kernel pattern matching can be expressed as given below.

$$K_m(x, y) = \emptyset\,(x).\emptyset(y) \quad (25)$$

From the Eqn. (25), "K_m" is the kernel pattern segment function and "$\emptyset\,(x_i).\emptyset(y_i)$", represents the product between two seed points in the consecutive image frames. The feature vector of the two seed points (x,y) can be expressed as follows:-,

$$K_m(x, y) = \exp\left(\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (26)$$

From the Eqn. (26),"$\|x - y\|$"represents the squared Euclidean distance between the two seed points. The variable "σ" represents the free parameter. The association of the seed points in two consecutive image frames are expressed as given below.

$$SP_1 = ASP_2 \quad (27)$$

The variable SP_1 and SP_2 are the two known seed points of the frames. The basic 3X3 matrix is represented using "A" and the normalization can be obtained as follows:

$$W \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_2 \\ a_3 & a_4 & a_5 \\ a_6 & a_7 & a_8 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \quad (28)$$

However, for detecting FG objects, four pairs of seed pairs are needed. Therefore, the Euclidean distance pair are measured as stated below.

$$E_{Dist} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (29)$$

After measuring E_{Dist}, the FG and BG regions are classified based on the threshold *(T_H)* values of these pixels and is expressed as follows:

$$P(x, y) = \begin{cases} FG\ , if\ E_{Dist} > T_H \\ BG\ \ , otherwise \end{cases} \quad (30)$$

The $(k-1)^{th}$ and the "Kth" frame's FG seed points are expressed as given below.

$$P_k = \left\{P_{(x,y,k)} \middle| \forall\, p_{(x,y,k)} \in FG\right\} \quad (31)$$

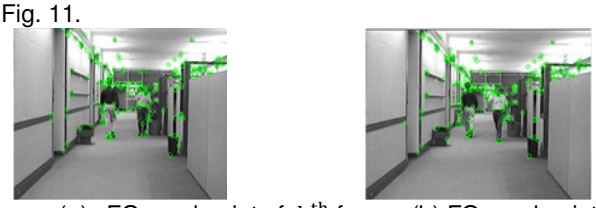
$$P_{k-1} = \left\{P_{(x,y,k-1)} \middle| \forall\, p_{(x,y,k-1)} \in FG\right\} \quad (32)$$

The FG seed points (SP) of the two frames are shown in Fig. 10.

(a) SP of 'k^{th}' Input frame (b) '$(k-1)^{th}$' frame.

Fig. 10.

The pattern matching of these input frames are performed and FG moving objects are identified. Similarly, the FG seed points are extracted using functions defined in Eqns. (31) and (32) and is shown in Fig. 11.



(a) FG seed point of k^{th} frame. (b) FG seed point of $(k-1)^{th}$ frame.

Fig. 11.

The feasible positions of the object image and pattern matching of FG objects are mathematically expressed as follows:

$$f(P_{(x,\ y,\ k-1)}) = P_{(x+\Delta x,\ y+\Delta y,\ k)} \quad (33)$$

$$F(P_{k-1}) = \{f(P_{(x,\ y,\ k-1)}) | \forall P_{(x,\ y,\ k-1)} \epsilon P_{k-1}\} \quad (34)$$

Here, the function $f(P_{(x,y,K-1)})$describes the seed points of the FG object. Also, the updated FG seed point is represented using the function $P_{(x+\Delta x,y+\Delta y,k)}$. The FG objects regions are further derived as follows:

$$P_k^{''} = F(P_{k-1}) \cup P_k \quad (35)$$

Here, "$P_k^{''}$" denotes the video frames and the updated matching function is denoted as "F". These updated seed points are used to identify the moving objects and these regions are marked as the FG moving objects' region.

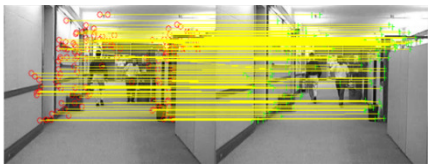


Fig. 12. Object matching with updated FG feature points.

Fig. 12 illustrates the updated FG feature points of two successive video frames. Next, one horizontal segmentation and two vertical segmentations are performed and the final object tracking with minimum bounding box approach is illustrated in the following Fig. 13.

Fig. 13. FG marked with bounding boxes.

From Fig. 13, it is clear that the red color represents the bounding box and is framed to mark objects inside the rectangular region. Finally, objects are tracked within bounding box using the QKF-PM method.

The present research work applies quantization of signals with an attempt to minimize the loss of data. This further improves the accuracy of tracking. Finally, the quantization of signals which is based on the center of seed points of two successive frames of the moving object regions are computed as follows:

$$C = \frac{1}{N}\sum_{i=1}^{n} P(x_i, y_i) \quad (36)$$

In the Eqn. (36), the center of seed point affinity features are referred by the variable "C" and "N" represents the total number of seed points across the regions of FG moving objects. .

In order to perform object tracking, the FG search regions (FG_{SR}) needs to be computed and can be derived as follows:

$$FG_{SR} = \rho W_{Box} + \tau H_{Box} \quad (37)$$

From the above Eqn. (37), the variable "W_{Box}" represents the width property of the bounding box. Similarly, the height property of the bounding box is represented using the variable "H_{Box}". The constants parameters are represented using the variables "ρ" and "τ". The resultant FG moving objects are tracked as shown in the figure below.

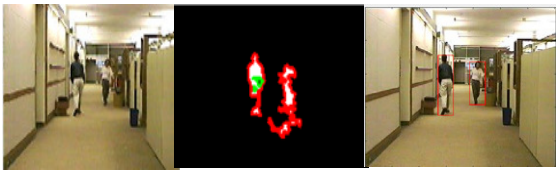


Fig. 14. Original Frame, Moving Object Detection and Tracking using Bounding box.

Fig. 14, clearly shows the original frame, object detection and tracking the moving object within the minimum bounding box.

IV. EXPERIMENTAL SETUP

The proposed system is tested on Actions as Space-Time Shapes dataset from the standard benchmark database to establish its competence. The research work is implemented in MATLAB. The proposed methods MTOS, MCSC and QKF-PMare compared with the existing methods like Active Contour Models (ACMs) in [1],Batch incremental SVMs (BISVM) in [3], Video Stabilization for MODT using SIFT developed in [5] and an Enhanced Rao-Blackwellized Particle Filter (E-RBPF) developed in [6] respectively. The performance of these above said methods are evaluated using the factors like segmentation accuracy, segmentation time, peak signal-to-noise ratio, Classification accuracy, Classification time, False positive rate, Moving Object Detection Rate.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

The segmentation accuracy (SA) is the ratio between the total number of video frames containing moving objects being correctly segmented and the total number of video frames. The Segmentation Accuracy (SA) is calculated in Equation 38.

$$SA = \frac{\text{No.of frames with moving objects correctly segmeneted}}{\text{Total number of video frames/s}} \times 100 \quad (38)$$

In the above Eqn. (38), the SA is measured in terms of percentage (%). If the SA is high, then the proposed MTOS method is said to be more efficient.The segmentation time (ST), is mathematically expressed as follows,

$$ST = \text{Total no. of video frames} * \text{time taken to segment the objects present in single frame} \quad (39)$$

In the above equation (39), ST is measured in milliseconds. Lesser the segmentation time, the method is said to be more efficient.

The Peak Signal-to-Noise Ratio is defined as the ratio between the total number of reference video frames and the power of distorting noise that affects the image. A greater value of PSNR is appropriate since it implies that the ratio of signal to noise is higher. Here, the signal is the ground truth image and the noise is the error caused due to distortion. The PSNR rate is defined in a logarithmic scale, in (decibels) and is expressed as follows:-,

$$\text{PSNR} = 20 * \log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \quad (40)$$

From the above Eqn. (40), the variable 'MAX_f' is the maximum possible intensity of the image (with size 255) and MSE is the mean square error. If the Peak Signal-to-Noise Ratio is high, then the proposed MTOS technique is said to be more efficient.

Table 1, clearly shows that the proposed MTOS method have improved by 12% and 8% segmentation accuracy than the existing ACMs [1] and Video Stabilization for MODT using SIFT methods [5] respectively. Hence the segmentation accuracy is significantly improved using the proposed MTOS method than the other existing methods. Next, the proposed method have reduced segmentation time by 23% and 18% than the existing ACMs in [1] and Video Stabilization for MODT using SIFT methods in [5] respectively. Finally, the proposed method have improved the PSNR rate by 25% and 18% than the existing ACMs [1] and Video Stabilization for MODT using SIFT methods [5] respectively. This is because the proposed method performs preprocessing of video frames and appropriate threshold values are chosen using adaptive threshold method based on multiple texture features of objects.

The classification accuracy is determined as the ratio of the number of video frames containing moving objects being correctly classified and the total number of video frames containing objects. The classification accuracy is calculated as follows:

$$CA = \frac{\text{No.of frames with moving objects correctly classified}}{\text{Total number of video frames containng objects}} \times 100 \quad (41)$$

The classification accuracy, denoted as 'CA', is measured in terms of percentage (%). The higher value of classification accuracy ensures the better performance of the proposed MCSC method.

The classification time is measured as the total computational time taken to classify objects in single frame. The classification time (CT), is mathematically expressed as follows:-,

$$CT = \text{Number of vidoe frames} \times \text{time taken to classify objects in single frame} \quad (42)$$

From the above Eqn. (42), classification time 'CT', is measured in terms of milliseconds (ms). The lesser the classification time, the method is said to be more reliable in its performance.

The false positive rate can be formulated as the ratio of the number of image frames containing moving objects being incorrectly classified by the method and the total number of video frames. The false positive rate (FPR) is mathematical expressed as follows:

$$FPR = \frac{\text{Number of frames with moving objects being incorrectly classified}}{\text{Total number of video frames}} \times 100 \quad (43)$$

In the above Eqn. (43), false positive rate, 'FPR', is measured in terms of percentage (%). The lower value of false positive rate ensures the better performance of the proposed MCSC method.

Table 2, clearly shows that the proposed MCSC method have improved by 27% and 9% classification accuracy than the existing Batch incremental SVMs (BISVM) [3] and Video Stabilization for MODT using SIFT methods [5]. The proposed method is seen to have reduced the classification time by 30% and 18% than the existing Batch incremental SVMs (BISVM) [3] and Video Stabilization for MODT using SIFT methods [5] respectively. The proposed method also have reduced the false positive rate by 19% and 12% than the existing Batch incremental SVMs (BISVM) [3] and Video Stabilization for MODT using SIFT methods [5] respectively. The proposed method achieved improved results after applying MRF principle for identifying the shape of moving objects. The application of fuzzy rules for preserving object boundary and the application of MAP estimation on spatial domain further improved the classification results.

The accuracy of pattern matching of objects can be formulated as the ratio of the number of video frames containing moving objects being correctly matched and the total number of video frames. The pattern matching accuracy is calculated as follows:

$$PMA = \frac{\text{No.of frames containing moving objects correctly matched}}{\text{Total number of video frames}} \times 100 \quad (44)$$

In the above Eqn. (44), the pattern matching accuracy is measured in terms of percentage (%). If the pattern matching accuracy is high, then the proposed QKF-PM technique is said to be more efficient. The pattern matching time is defined as the computational time taken to perform pattern matching of objects with the training patterns with respect to the total number of video frames. The pattern matching time is defined as follows:

$$\text{PMT} = \text{No. of video frames} \times \text{Computational time taken for matching objects} \quad (45)$$

In the above Eqn. (45), the pattern matching time is measured in terms of milliseconds (ms). The detection rate of foreground moving objects can be formulated as the rate at which moving objects in video frames are correctly detected and the total number of video frames. The moving object detection rate is calculated as follows:

$$MODR = \frac{\text{No. of frames with moving objects correctly detected}}{\text{Total number of video frames}} \times 100 \quad (46)$$

In the above Eqn. (46), the moving object detection rate is measured in terms of percentage (%). If the moving

object detection rate is high, then the proposed QKF-PM technique said to be more efficient.

Table 3, clearly shows that the proposed QKF-PM method have improved pattern matching accuracy rate by 21% and 11% than the existing Enhanced Rao-Blackwellized Particle Filter (E-RBPF) [6] and Video stabilization for MODT using SIFT methods [5] respectively. The proposed method have reduced the pattern matching time by 31% and 27% than the existing Enhanced Rao-Blackwellized Particle Filter (E-RBPF) [6] and Video stabilization for MODT using SIFT methods [5] respectively. Finally, the proposed QKF-PM method have improved moving object detection rate by 16% and 13% than the existing Enhanced Rao-Blackwellized Particle Filter (E-RBPF) [6] and Video stabilization for MODT using SIFT methods [5] respectively. Hence the moving object detection rate is significantly improved using the proposed QKF-PM method than the other existing methods. The research work identified chromatic shadows of objects using the top-down approach. The kernel pattern function and the application of bounding box approach improved the performance of object detection and tracking.

Table 1: Tabulation for SA, ST and PSNR.

Metrics	Segmentation Accuracy			Segmentation Time			PSNR(db)			
No. of video frames/s	ACM	Video Stabilization for MODT using SIFT	MTOS	ACM	Video Stabilization for MODT using SIFT	MTOS	Video frame size (KB)	ACM	Video Stabilization for MODT using SIFT	MTOS
20	66	69	74	30	28	25	14.5	33	37	40
40	67	70	75	32	31	26	14.6	35	38	42
60	68	71	77	37	35	27	14.7	37	37	44
80	69	72	78	40	38	29	14.8	37	38	45
100	71	73	79	41	39	30	14.9	38	42	48
120	73	74	81	43	41	32	15.0	45	45	51
140	75	76	83	45	42	34	15.8	44	44	54
160	77	78	85	46	43	36	15.9	47	47	62
180	78	79	87	47	44	37	16.0	49	56	68
200	80	82	88	51	46	40	16.1	51	59	71

Table 2: Tabulation for CA, CT and FPR.

No. of video frames/s	Classification Accuracy (%)			Classification Time (ms)			False Positive Rate (%)		
	BISVM Classifier	Video Stabilization for MODT using SIFT	MCSC	BISVM Classifier	Video Stabilization for MODT using SIFT	MCSC	BISVM Classifier	Video Stabilization for MODT using SIFT	MCSC
20	42	63	77	24	20	16	39	36	32
40	53	70	79	28	23	18	40	37	34
60	58	77	81	30	27	20	43	41	36
80	71	78	84	34	29	22	49	44	37
100	74	80	85	38	31	24	50	46	38
120	76	83	89	39	32	28	53	47	40
140	77	79	88	40	35	31	54	48	42
160	80	86	91	44	38	33	56	50	45
180	81	85	90	46	40	35	58	52	49
200	83	87	91	50	43	38	59	56	50

Table 3: Tabulation for PMA, PMT and MODR.

No. of video frames/s	Pattern Matching Accuracy (%)			Pattern Matching Time (ms)			Moving Object Detection Rate (%)		
	E-RBPF	Video Stabilization for MODT using SIFT	QKF-PM	E-RBPF	Video Stabilization for MODT using SIFT	QKF-PM	E-RBPF	Video Stabilization for MODT using SIFT	QKF-PM
20	63	68	75	20	17	12	64	66	77
40	60	66	73	26	23	15	68	68	78
60	59	64	74	28	27	20	65	67	78
80	62	67	78	33	32	23	67	71	80
100	63	66	80	36	34	26	71	73	81
120	67	71	77	40	39	31	74	76	86
140	65	73	79	47	45	33	73	74	84
160	71	78	82	53	50	37	75	78	85
180	72	80	83	55	53	39	79	81	90
200	75	81	90	57	55	42	81	84	95

From Table 1-3, it is evident that, the proposed methods performs well compared to the existing methods in [1, 3, 5, 6] respectively.

VI. CONCLUSIONS AND FUTURE WORKS

Computer Vision plays an important role in many aspects of our life. Object tracking involves two related processes like, object detection and tracking the detected objects. The research work aimed at improving the accuracy of detection and tracking of multiple moving objects and reduce the computational time taken for segmentation, classification and pattern matching of objects. The experimental results shows that the proposed methods Multi-textured Object Segmentation, Multi-Class Spatial Classifier and Quantized Kalman Filter-based Pattern Matching methods produce improved experimental results when compared with the existing techniques. Experimental results reveals that the research work achieved all the stated objectives and is suitable for human object detection and tracking system. As scope for future extension, the current research work can be extended for the development of an intelligent framework for human motion analysis with speech recognition system.

Conflict of Interest. No.

REFERENCES

[1]. Silva, A.S., Severgnini, F. M. Q., Oliveira, M. L., Mendes, V. M. S., & Peixoto, Z. M. A. (2016). Object Tracking by Color and Active Contour Models Segmentation. *IEEE Latin America Transactions, 14*(3), 1488 - 1493.

[2]. Chiranjeevi, P., & Sengupta, S. (2013). Neighborhood supported model level fuzzy aggregation for moving object segmentation. *IEEE Transactions on Image Processing*, *23*(2), 645-657.

[3]. Liang, C. W., & Juang, C. F. (2015). Moving object classification using a combination of static appearance features and spatial and temporal entropy values of optical flows. *IEEE Transactions on Intelligent Transportation Systems*, *16*(6), 3453-3464.

[4]. Chavez-Garcia, R. O., & Aycard, O. (2015). Multiple sensor fusion and classification for moving object detection and tracking. *IEEE Transactions on Intelligent Transportation Systems*, *17*(2), 525-534.

[5]. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91-110.

[6]. Bhaskar, H., Dwivedi, K., Dogra, D. P., Al-Mualla, M., & Mihaylova, L. (2015). Autonomous detection and tracking under illumination changes, occlusions and moving camera. *Signal Processing*, *117*, 343-354.

[7]. Jin, H., Zhang, H., Liu, Z., Yang, D., Bie, D., Zhang, H., & Zhao, J. (2017). A synthetic algorithm for tracking a moving object in a multiple-dynamic obstacles environment based on kinematically planar redundant manipulators. *Mathematical Problems in Engineering*, *3*(1), 1-15.

[8]. Mondal, A., Ghosh, A., & Ghosh, S. (2018). Scaled and oriented object tracking using ensemble of multilayer perceptrons. *Applied Soft Computing*, *73*, 1081-1094.

[9]. Milan, A., Rezatofighi, S. H., Dick, A., Reid, I., & Schindler, K. (2017). Online multi-target tracking using recurrent neural networks. In *Thirty-First AAAI Conference on Artificial Intelligence*, 4225-4232.

[10]. Dev, D. S., & Kisku, D. R. (2017). Improved Pattern Matching Algorithm. *Applied Mathematics & Information Sciences, 11*(4), 1163-1184.

[11]. Yang, J., Price, B., Shen, X., Lin, Z., & Yuan, J. (2016). Fast appearance modeling for automatic primary video object segmentation. *IEEE Transactions on Image Processing*, *25*(2), 503-515.
[12]. Wells, J. W., & Chatterjee, A. (2018). Content-aware low-complexity object detection for tracking using adaptive compressed sensing. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, *8*(3), 578-590.
[13]. Dequaire, J., Ondrúška, P., Rao, D., Wang, D., & Posner, I. (2018). Deep tracking in the wild: End-to-end tracking using recurrent neural networks. *The International Journal of Robotics Research*, *37*(4-5), 492-512.
[14]. Kallakunta, R. K., Prakash, V. B., Shyam, V., & Anil Kumar, M. (2016). Texture and Shape based Object Detection Strategies. *Indian Journal of Science and Technology, 9*(30), 1-4.
[15]. Thoreau, M., & Kottege, N. (2018). Improving Online Multiple Object tracking with Deep Metric Learning. *arXiv preprint arXiv:1806.07592*, 1-6.
[16]. Schulter, S., Vernaza, P., Choi, W., & Chandraker, M. (2017). Deep network flow for multi-object tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6951-6960.
[17]. Wong, S. C., Stamatescu, V., Gatt, A., Kearney, D., Lee, I., & McDonnell, M. D. (2017). Track everything: Limiting prior knowledge in online multi-object recognition. *IEEE Transactions on Image Processing*, *26*(10), 4669-4683.
[18]. Chan, S., Zhou, X., & Chen, S. (2018). Online classification for object tracking based on superpixel. *Neurocomputing*, *286*, 88-108.
[19]. Zhang, X., Xia, G. S., Lu, Q., Shen, W., & Zhang, L. (2018). Visual object tracking by correlation filters and online learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *140*, 77-89.
[20]. Lee, Y. G., Tang, Z., & Hwang, J. N. (2017). Online-learning-based human tracking across non-overlapping cameras. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(10), 2870-2883.
[21]. Patel, A. K., & Khandelwal, A. (2019). A review of techniques used for Edge Detection in Image. *International Journal of Electrical, Electronics and Computer Engineering, 8*(2), 06-09.
[22]. Singh, D., & Shrivastava, A. (2019). Analysis of Video Streaming based on Canny Edge Detection Algorithm. *International Journal of Electrical, Electronics and Computer Engineering, 8*(1), 50-54.

How to cite this article: Rajamanickam, M. (2020). Video Object Segmentation and Classification with Improved Kalman Filter-based Method for People Tracking. *International Journal on Emerging Technologies*, *11*(3): 402–411.