International Journal of Innovative Computing, Information and Control
Volume 11, Number 4, August 2015

ICIC International ©2015 ISSN 1349-4198
pp. 1139–1153

AN EFFICIENT HUMAN OBJECT DETECTION AND TRACKING WITH THE AID OF MORPHOLOGICAL OPERATION AND OPTIMIZATION ALGORITHM

Paramasivam Mukilan[1] and Amitabh Wahi[2]

[1]Department of Electronics and Communication Engineering
C.M.S. College of Engineering and Technology
Coimbatore-641032, Tamilnadu, India
mukilan0677@gmail.com

[2]Department of Information Technology
Bannari Amman Institute of Technology
Sathyamangalam-638401, Tamilnadu, India

Received July 2014; revised December 2014

Abstract. *Rapid advancements in the field of multimedia and other fields urge the processing of a huge database of video clips. The detection and tracking of different objects in the video clips has become an interesting area of research in the computer vision and it has wide applications in many fields like service robots, surveillance systems, public security systems, and virtual reality interfaces. Target tracking mainly concerns with the evaluation of object's speed and position over time using one or multiple sensors. In this paper, we proposed an efficient human object detection and tracking from a moving sequence. Our proposed method can be divided into three different phases. In the first phase the image segmentation from the video is being carried out using shot segmentation. The second phase of our proposed method is concerned with selection of threshold values. In order to identify the effective threshold value optimization technique is being carried out using ABC algorithm. Once the threshold value is selected we go for morphological operation and difference image formation in order to detect the object. The final phase of our work is to track the human object from a video sequence. We have used parallel Kalman filtering for the tracking process. The experimental results show that our proposed method delivers efficient results which proves that our proposed method is better suited for human object detection from any video sequence.*

1. **Introduction.** With the increasing threat of terrorist, the advanced video surveillance system has to be put into use. The advanced video surveillance system needs to analyze the behaviors of people in order to prevent the occurrence of the potential dangerous case. The analysis of behaviors of people requires the detection and tracking system. In recent years, the development of detection and tracking system has been going forwards for several years; many real time systems have been developed. However, there are still some challenging technologies which need more researches: foreground segmentation and false alarm elimination [1]. Tracking is usually performed in the context of higher-level applications that require the location and/or shape of the object in every frame. Assumptions are made to constrain the tracking problem in the context of a particular application [13]. The tracking methods are categorized on the basis of the object and motion representations used, provide detailed descriptions of representative methods in each category, and examine their pros and cons. The use of object tracking is pertinent in the tasks of: motion-based recognition, automated surveillance, video indexing, human-computer interaction and traffic monitoring and vehicle navigation.

Visual analysis of motion has recently persuaded more studies in the computer vision area. It attempts to detect, track, and identify people, and more generally, to understand object behaviors from image sequences involving human. Moving object detection is the first step processes for nearly every system of vision-based analysis. The aim on moving object detection is to segment the regions corresponding to people from the rest of an image sequence.

It is known to be a significant and difficult research problem. Background subtraction is a particularly popular method for motion segmentation uses dynamic appearance models to track people. A recursive convex hull algorithm is used to find body part locations for single person [3]. The background subtraction also deals with removal of noise in images, complex object motion, non-rigid or articulated nature of objects, partial and full object occlusions, complex object shapes, scene illumination changes and real-time processing requirements [11]. One can simplify tracking by imposing constraints on the motion and/or appearance of objects. For example, almost all tracking algorithms assume that the object motion is smooth with no abrupt changes. One can further constrain the object motion to be of constant velocity or constant acceleration based on a priori information. Prior knowledge about the number and the size of objects, or the object appearance and shape, can also be used to simplify the problem.

The aim of an object tracker is to generate the trajectory of an object over time by locating its position in every frame of the video. Object tracker [9] may also provide the complete region in the image that is occupied by the object at every time instant. The task of detecting the object and establishing correspondence between the object instances across frames can either be performed separately or jointly. In the first case, possible object regions in every frame are obtained by means of an object detection algorithm, and then the tracker corresponds to the objects across frames [15]. In the latter case, the object region and correspondence are jointly estimated by iteratively updating object location and region information obtained from previous frames.

Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video [4]. A common approach for object detection is to use information in a single frame. Some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections [10]. This temporal information is usually in the form of frame differencing, which highlights changing regions in consecutive frames. Given the object regions in the image, it is then the tracker's task to perform object correspondence from one frame to the next to generate the tracks.

Selecting the right features plays a critical role in tracking. In general, the most desirable property of a visual feature is its uniqueness so that the objects can be easily distinguished in the feature space [19]. Feature selection is closely related to the object representation. Mostly all the tracking algorithms use a combination of these features: color, edges, optical flow and texture. Most features are chosen manually by the user depending on the application domain. However, the problem of automatic feature selection has received significant attention in the pattern recognition community.

The rest of the paper is organized as follows. Section 2 reviews some of the recent researches related to our method. Section 3 describes our proposed methodology that involves shot segmentation, feature extraction, and finally object detection and tracking. Section 4 discusses the results of the proposed method. Section 5 is the concluding remarks of the method.

2. **Review of Recent Researches.** Background subtraction was very important part of surveillance applications for successful segmentation of moving objects from video sequences. Hussain and Kharat [14] have presented a robust algorithm, for motion detection and tracking in dynamic scenes based on background modeling technique to analyze the illumination change for detection & tracking of moving objects. Successive frame difference was taken and compared for the required set threshold for the changing pixel detection. Experimental result shows the high performance of the proposed method for tracking in noisy backgrounds.

In the sports video analysis, the most important part is the movement of player (object) detection and tracking. Manikandan and Ramakrishnan [5] have proposed a method to detect the movement of player from the background image in video sequence and for the player tracking. They proposed new method to detect players based on background subtraction. A reliable background updating model was established. A dynamic optimization threshold method was used to obtain a more complete behavior of moving player and tracking. Motion of a moving player and tracking in a video stream are studied and its velocity was detected. The centroid of object was computed to use in the analyses of the position of the moving object. The experimental results show that the proposed method runs quickly, accurately and fits for the real time detection.

Hsia and Guo [6] proposed an efficient modified directional lifting-based 9/7 discrete wavelet transform structure to further reduce the computational cost and preserve the fine shape information in low resolution image. The experimental results documented that the proposed low-complexity MDLDWT scheme could not provide more precise detection rate for multiple moving objects, and the fine shape information could not be effectively preserved for the real time video surveillance applications in both indoor and outdoor environments.

Cuevas and García [7] proposed a novel background modeling that is applicable to any spatio-temporal non-parametric moving object detection strategy. Through an efficient and robust method to dynamically estimate the bandwidth of the kernels used in the modeling, both the usability and the quality of previous approaches are improved. Empirical studies on a wide variety of video sequences demonstrated that the proposed novel background modeling reduces the quality of previous strategies while maintaining the computational requirements of the detection process.

Karasulu and Korukoglu [8] proposed Simulated Annealing-Background Subtraction (SA-BS) to determine the optimal threshold for the foreground-background segmentation and to learn background model for object detection. The obtained performance results and statistical analysis showed that the proposed method is more preferable than regular BS method. During the restarted SA-BS processes, all the sampled points were stored in a single sampling map. After considering all samples from all intermediate states of the SA-BS process, not just the last sample points to which it has converged, we lose the global convergence of SA-BS, and false peaks were detected in proposed SA-BS method.

Wang and Thorpe [2] developed the first outdoor real-time system solving both Simultaneous Localization and Mapping (SLAM) and Detecting and Tracking Moving Objects (DATMO) simultaneously for urban environments from a ground vehicle. To correct the vehicle odometry they used an ICP-based matching scan method and moving objects were detected based on a simple geometric analysis. He also presented a mathematical framework integrating both SLAM and DATMO and showed that they can be mutually beneficial from each other.

Kirkpatrick *et al.* [16] studied the Simulated Annealing, which springs from a property of solids where the heating was followed by the gradual cooling and the solid attains an almost perfect crystalline state yielding the minimum free energy. The SA was formed as the basis of an optimization technique for combinatorial problems. When a solid was heated beyond its melting point, its particles were arranged randomly in the liquid phase. As it was cooled down, the particles rearranged themselves into lower energy configurations associated with decreasing free energy. If the cooling was carried out sufficiently slowly, at the zero temperature the solid would attain the global minimum energy configuration.

According to the result of moving object detection research on video sequences, Madhavi and Ganeswara Rao [18] have proposed a method to detect moving object based on background subtraction. First of all, establish a reliable background updating model based on statistical and use a dynamic optimization threshold method to obtain a more complete moving object. And then, morphological filtering was introduced to eliminate the noise and solve the background disturbance problem. At last, contour projection analysis was combined with the shape analysis to remove the effect of shadow, and the moving object was accurately and reliably detected. The occlusion was one of the most common events in object tracking and object centroid of each object was used for detecting the occlusion and identifying each object separately. Video sequences have been captured in the laboratory and tested with the proposed algorithm. The algorithm works efficiently in the event of occlusion in the video sequences.

Schindler *et al.* [20] have presented an approach for multi-object tracking which considers object detection and space time trajectory estimation as a coupled optimization problem. It was formulated in a hypothesis selection framework and builds upon a state-of-the-art pedestrian detector. At each time instant, it searches for the globally optimal set of space time trajectories which provided the best explanation for the current image and for all evidence collected so far, while satisfying the constraints that no two objects may occupy the same physical space, nor explain the same image pixels at any point in time. Successful trajectories hypotheses were fed back to guide object detection in future frames.

Fuzzy inference systems have been used to solve a lot of real-world problems. Adaptive network fuzzy inference system (ANFIS) was one of the most important fuzzy inference systems. ANFIS was originally proposed for prediction and regression problems. When ANFIS have been used for time series forecasting, the inputs of ANFIS have been generally other simultaneous time series in the literature. However, it was well-known that lagged variables should be used to obtain better forecasts in a time series forecasting process. Also, some advantages can be obtained if lagged variables are used as inputs for ANFIS. Erol *et al.* [26] have proposed an ANFIS by redesigning ANFIS approach for time series forecasting problem. Fuzzy c-means method was used for fuzzification in the suggested ANFIS. In addition, in the proposed approach, it was not necessary to utilize fuzzy numbers for input memberships. And, the parameters of output membership function are determined by using particle swarm optimization method. The proposed method was applied to two real-world time series. In order to compare the forecasting performance of the proposed approach, some other forecasting methods available in the literature are also applied to the time series.

The proposed approach is designed in such a way to improve all these drawbacks of the existing system. We have used the technique for tracking the human object through effective segmentation and then parallel Kalman filter which makes the exact tracking of human object more efficient. Our techniques prove to deliver better precision and F-measure.

3. Proposed Methodology for Tracking of Object.

3.1. Object detection and tracking. The detection and tracking of moving object has become an attractive field for researchers in the recent years as there is for developing different methods of tracking and detection. The moving object detection from a video sequence is a monotonous job as there will be distinction in background in each frame as well as the camera resolution. Object detection and tracking is gaining more importance recently as there is vast need for improvement in such sort of detection process. In this paper we have developed an efficient system for the human object detection. Human object detection from a video sequence can be done based on the movement of the object in each frame of the video sequence. In order to make the detection process more efficient we have also utilized an optimization algorithm.

3.2. Proposed human object detection steps. The human object detection and tracking is an interesting topic in the research field and hence it is a basic requirement to develop more efficient technique for this detection purpose. The proposed method of object detection and tracking system consists of the processes such as Segmentation, Feature Extraction and Tracking. The first step in our approach is to segment the database video clips into different frames or shots. For the segmentation we have employed shot segmentation method. Once the shot segmentation process is completed next we use thresholding method in order to find the values that suit the detection process. After thresholding operation, we optimize these values in order to select the suitable threshold value. For optimization purpose we have employed Artificial Bee Colony algorithm. After optimization the difference images are found out and then morphological operation is carried out in order to extract features and in the final process we use Kalman filtering to track and detect the human object. The flow diagram for our proposed method is shown as Figure 1.

3.2.1. *Shot segmentation.* A large amount of digital videos have been generated due to the rapid development of computing and network infrastructures. In general, videos can be represented by a hierarchical structure, while shots are the basis units for constructing high level semantic scenes. Thus, shot boundary detection is an important preprocessing step for efficient browsing and further content analysis. A shot consists of consecutive frames which are usually captured by a single camera action. Typically, there are no significant content changes between successive frames in the shot [25].

In our proposed method we use clustering approach to video shot segmentation with the assumption that the frames in one cluster constitute a shot. The dominant-set clustering approach, along with the similarity matrix, is usually easy to cluster the frames which may not be consecutive in frame index into the same shot due to their high similarity. Smoothing and elimination process is done after the clustering process in order to make segmentation process more effective. This algorithm provides better clustering of the image. Here a set of samples were considered, for which an undirected edge-weighted graph with no self-loops is built for each vertex represents a sample and two vertices are linked by an edge whose weight represents the similarity of the two vertices. A dominant set of the weighted graph is iteratively found for clustering the samples into coherent groups and then removed from the graph until the graph is empty. When comparing with other clustering algorithm, the dominant-set clustering determines the number of the clusters automatically and possesses low computational cost.

The algorithm is as shown below.

Consider the input matrix M

1. Initialize M^i, $i = 1$

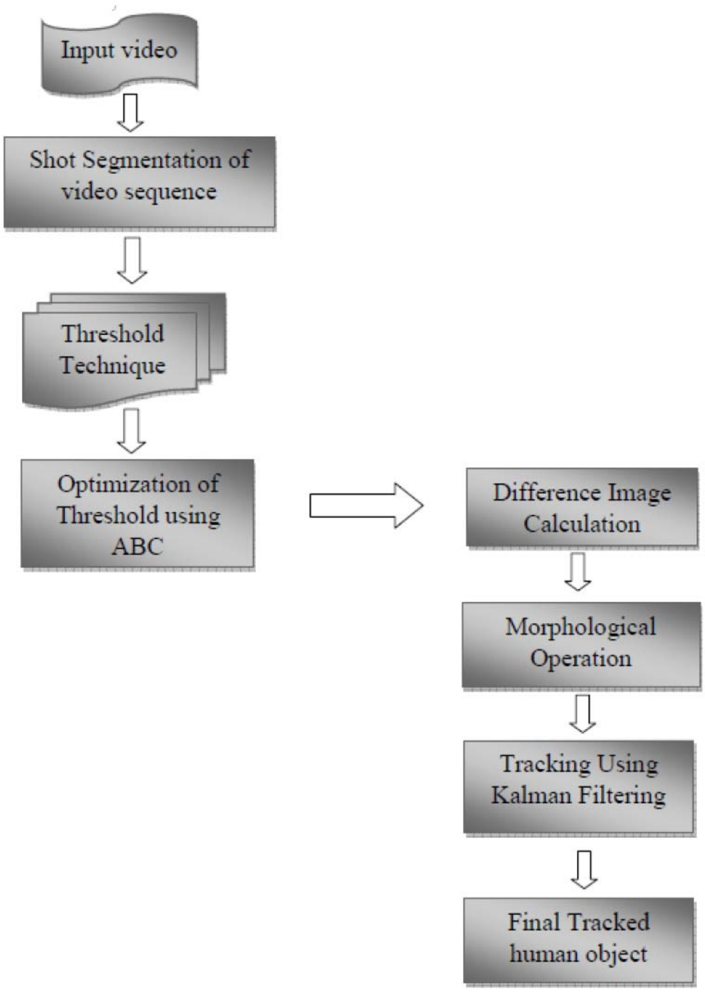


FIGURE 1. Proposed work flow

2. Calculate the local solution of g^i and $F(g^i)$
3. Get the dominant set: $S^i = \rho\left(g^i\right)$
4. Split out M^i and get a new similarity affinity matrix M^{i+1}
5. If M^{i+1} is empty, break, else $M^i = M^{i+1}$ and $i = i + 1$ then go to step 2;

Output: $G_{n=1}^{i}\left\{S^n, g^n, F(g^n)\right\}$

3.2.2. *Selection of object points using gray threshold method.* Threshold is one of the methods used for image segmentation process. Threshold is used for refining the foreground image from the background. In the thresholding process, based on the particular threshold value being selected, the pixels in the image are marked as the object pixel if the value of the pixel is greater than the selected threshold value. If the pixel value is lesser than threshold value then those pixels are marked as background pixels. In our proposed method the threshold value is selected based on the position of the object in the image. Thresholding is an effective method for image segmentation when the image consists of a large number of similar objects for selecting the required object. By choosing a suitable threshold between the dominant intensities of the background and object, the original image can be partitioned into regions where the human objects alone are separated from the other image parts. The image is segmented into the object and background pixels as per the expression given below.

For each pixel (x, y),

$$P(x,y) = \begin{cases} 1 & \text{for } Q(x,y) \geq t \\ 0 & \text{for } Q(x,y) < t \end{cases} \tag{1}$$

where '$Q(x, y)$' is the input image for segmentation, '$P(x, y)$' is the output image after segmentation and 't' is the selected threshold value. The object which possesses the value above the threshold value will be chosen as the pixel point for further processing. After the process the original image and the segmented image are compared. The gray threshold helps to level the images which further help in separating the human object from the other similar objects in the images. Once the human object is separated from the image the next process is to refine the corresponding object in the sequence. For this purpose we carried out the morphological operations.

3.2.3. *Proposed artificial bee colony for optimization of weights in neural network.* Artificial bee colony is one of the most efficient optimization techniques available at present. The goal of bees in the ABC model is to find the best solution, the position of a food source represents a possible solution to the optimization problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution [24]. After sharing their information with onlookers, every employed bee goes to the food source area visited by herself at the previous cycle since that food source exists in her memory, and then chooses a new food source by means of visual information in the neighborhood of the one in her memory and evaluates its nectar amount [23].

3.2.3.1. Employee bee phase.

The colony of artificial bees contains three groups of bees: employed bees, onlookers and scouts. A bee waiting on the dance area for making decision to choose a food source is called an onlooker and a bee going to the food source visited by itself previously is named an employed bee. A bee carrying out random search is called a scout. First half of the colony consists of employed artificial bees and the second half constitutes the onlookers. For every food source, there is only one employed bee. In other words, the number of employed bees is equal to the number of food sources around the hive.

At the initialization stage, a set of food source positions are randomly selected by the employed bees and their nectar amounts are determined. Then, these bees come into the hive and share the nectar information of the sources with the onlooker bees waiting on the dance area within the hive. Initially, ABC generates a randomly distributed initial population represented by p_i having n solutions where each solution is the food source position and S_p is the population size. Each solution is represented by h_i, where $1 \leq i \leq n$ is an N-dimensional vector, where N is the number of optimization parameters taken into consideration. After initialization, the population of the positions is subjected to repeated cycles of the search processes of the employed bees, the onlooker bees, and scout bees.

3.2.3.2. Onlooker bee phase.

In this phase, selection of the food sources by the onlookers after receiving the information of employed bees and generation of new solution is carried out. The onlooker bee prefers a food source area depending on the nectar information distributed by the employed bees on the dance area. As the nectar amount of a food source increases, the probability with which that food source is chosen by an onlooker increases, too. Hence, the dance of employed bees carrying higher nectar recruits the onlookers for the food source areas with higher nectar amount.

An onlooker bee chooses a food source depending on the probability value associated with that food source (P_i) given by the expression:

$$P_i = \frac{f_i}{\sum\limits_{a=1}^{n} f_a} \tag{2}$$

where f_i is the fitness value of the solution, and n is the number of food sources which is equal to the number of employed bees.

After arriving at the selected area, onlooker chooses a new food source in the neighborhood of the one in the memory depending on visual information. Visual information is based on the comparison of food source positions. When the nectar of a food source is abandoned by the bees, a new food source is randomly determined by a scout bee and replaced with the abandoned one. An artificial onlooker bee probabilistically produces a modification on the position (solution) in her memory for finding a new food source and tests the nectar amount (fitness value) of the new source (new solution).

Let the old position be represented by $x_{i,a}$ and the new position is represented by $q_{i,a}$, which is defined by the equation,

$$x_{i,a} = q_{i,a} + \sigma_{i,a}(q_{i,a} - q_{j,a}), \quad i \neq j \tag{3}$$

where, $j = \{1, 2, \ldots, n\}$, $a = \{1, 2, \ldots, N\}$, and $\sigma_{i,a}$ is a random number in the range $[-1, 1]$.

The position update equation shows that as the difference between the parameters of the $q_{i,a}$ and $q_{j,a}$ decreases, the perturbation on the position $q_{i,a}$ also decreases, too. Thus, as the search approaches to the optimum solution in the search space, the step length is adaptively reduced.

Rearranging the position updating step, we have:

$$x_{i,a} - q_{i,a} = \sigma_{i,a}(q_{i,a} - q_{j,a}) \tag{4}$$

As $x_{i,a}$ is the position update from $q_{i,a}$ in the previous step, representing in the time domain, we can write $q_{i,a}$ as z_T when $x_{i,a}$ is taken as z_{T+1}. Hence we have:

$$z_{T+1} - z_T = \sigma_{i,a}(q_{i,a} - q_{j,a}) \tag{5}$$

The left side $z_{T+1} - z_T$ is the discrete version of the derivative of order $\alpha = 1$. Hence we have:

$$W^{\alpha}[z_{T+1}] = \sigma_{i,a}(q_{i,a} - q_{j,a}) \tag{6}$$

3.2.3.3. Scout bee phase.

The employed bee whose food source is exhausted by the employed and onlooker bees becomes a scout and it carries out random search. The food source whose nectar is abandoned by the bees is replaced with a new food source by the scouts. This is simulated by randomly producing a position and replacing it with the abandoned one. Here, if a position cannot be improved further through a predetermined number of cycles called limit then that food source is assumed to be abandoned. In the classic ABC algorithm a scout searches the vicinity of the hive in a random way. This exploration feature of scout can be beneficial in the initial iterations; however, performing a completely random movement in the final iterations may not be effective. Therefore, in this approach, a scout explores the search space globally in the initial iterations and locally in the final iterations. Since in the last iterations improvement of the best food source may not happen, therefore, it may be selected as a scout and removed from the population. The threshold value is chosen with the help of optimization.

3.3. **Morphological operation.** After segmentation of the image I by adjusting the contrast and intensity of the image using threshold segmentation and after optimization, morphological operation is performed on the image. Morphological operations are affecting the form, structure or shape of an object. They are used in pre or post processing or for getting a representation or description of the shape of objects/regions. A set of functions that are valuable for processing and decomposition of shapes in arbitrary dimensions is provided by the mathematical morphology. Set-theoretic operations like union and intersection are used to define morphological operations. The two inputs given to morphological operation are binary image and structuring element. After adjusting the contrast and intensity of the image I_m, the image is converted to the binary form I_{bm}. Then, by applying Equation (3) an enhanced image is obtained through the morphological operations like imdilate, which utilizes the structuring element

$$[I_{bm} \oplus E](x, y) = \max\{I_{bm}(x - m, y - n)\}, \quad \text{where } m, n \in E \tag{7}$$

The above expressions are used for calculating the morphological function. By using these morphological operation maximum intensity pixels of the image alone is selected. Thus, the operation employed contrast and intensity adjusted image is further enhanced by utilizing the morphological operation. Once the morphological operation in completed the various image properties are found. The image properties are calculated using the Regionprobs method where the area of the particular human object is found out along with the mean. The area calculation helps in detecting the object based on its location and these can further help in tracking the particular object from the entire objects present in the particular video sequence.

3.4. **Motion estimation using parallel Kalman filter for human object detection and tracking.** Generally motion estimation is the process of detecting the motion vectors which forms the transition from one frame to other in a video sequence. The human object detection and tracking in any video can be detected based on the motion of the object at consecutive sequences and this can help us in detecting the object. In our proposed method we utilize Kalman filter for motion estimation process. The Kalman filtering is performed in 8×8 block of the image. The Kalman filter used here is the parallel Kalman filter which is a modified form of Kalman filter and it can be used to estimate the motion of the object more accurately than normal filter. In this motion estimation process, consecutive motion vectors of the block are taken through the Kalman filter. Based on these values the motion of the object is estimated. The Kalman filter model for the system and a particular time instant t is given as shown below,

$$K_{fi} = P_i K_{fi} + R_i \tag{8}$$

where K_{fi} is the n-dimensional vector and P_i is the $n \times n$ matrix and R_i is the random sequence vector in the system.

Consider that at time t, there is an m dimensional vector which is available and is corrupted by a noise, and then the expression is given as shown below,

$$S_i = L_i K_{fi} + W_i \tag{9}$$

where L_i is called the $m \times n$ observation matrix and the vector W_i is the additive noise in the processing.

Assume that the vectors, R_i and W_i are mutually correlated to one another, which results in the expression,

$$G\left[R_i W_i^J\right] = \sigma \quad i = 0, 1, \dots \tag{10}$$

Here the value σ represents the null matrix.

Based on these considerations we derive the equation for the Kalman filtering which is given in the below expression,

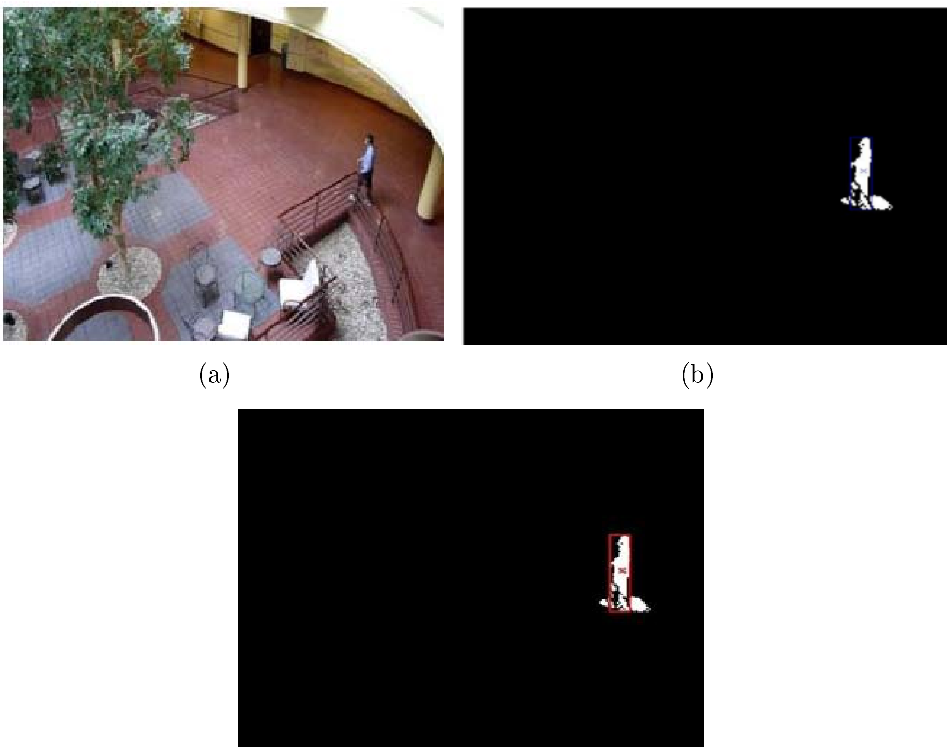
$$\bar{K}_{fi} = P_i \bar{K}_{fi-1} + F_i[A_i - L_i \bar{K}_{fi-1}] \quad (11)$$

Equation (9) gives the parallel Kalman filter model equation which is utilized to estimate the motion estimation in order to track the object in the video.

The detected object from the image is tracked by utilizing the motion estimation process as mentioned above. The parallel Kalman filter tracks the human object that is being detected by identifying the motion estimated by the human object in the various frames of the video. The Kalman filter utilizes the motion estimation of the human objects in the frame for effectively tracking the specified objects in the frames which has been taken from the input video clips. By estimating the motion, the tracking can be done accurately and this proposed method provides an efficient tracking system.

4. **Results and Discussion.** The proposed object detection and tracking system using the level features was implemented in the working platform of MATLAB. The detection and tracking process is tested with different frames of video and the upcoming result of the proposed work has been shown below. Initially, the videos are segmented to different shots or frames and then features are extracted followed by the detection and tracking process. We have used some videos based on motion of human object recorded from surveillance cameras. The results are shown as Figures 2-5.

Figure 2(a) is the original image of the human object obtained from the shot segmentation. Figure 2(b) is the output after the difference image. Figure 2(c) shows the tracked



(a) (b) (c)

FIGURE 2. Results of object tracking in 40th frame: (a) Input frame; (b) Difference image; (c) Tracked output

image of the object in the first frame. Similarly for different frames the process is repeated and finally the object is tracked.

Figure 3(a) is the original image of the human object obtained from the shot segmentation. Figure 3(b) is the output after the difference image. Figure 3(c) shows the tracked image of the object in the 50th frame.

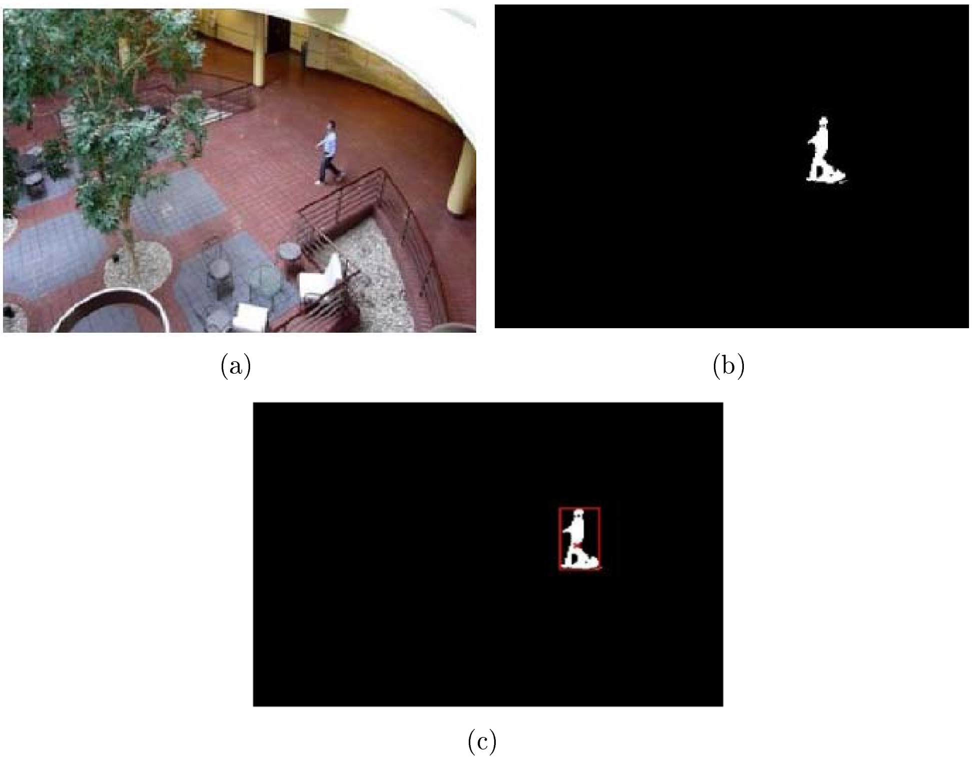


FIGURE 3. Results of object tracking in 50th frame: (a) Input frame; (b) motion extracted output; (c) Tracked output

Figure 4(a) is the original image of the human object obtained from the shot segmentation. Figure 4(b) is the output after the difference image. Figure 4(c) shows the tracked image of the object in the 80th frame.

Figure 5(a) is the original image of the human object obtained from the shot segmentation. Figure 5(b) is the output after the difference image. Figure 5(c) shows the tracked image of the object in the first frame. Similarly for different frames the process is repeated and finally the object is tracked. The proposed methodology proved to be more effective and accurate in object detection and tracking.

Performance Analysis:

The precision and recall value for the proposed method are calculated for analyzing the performance. Leting the object to be tracked be denoted by O_h and the tracked output is denoted as T_h, then precision and recall are expressed as,

$$precision = \frac{\{O_h \cap T_h\}}{\{T_h\}} \tag{12}$$

$$recall = \frac{\{O_h \cap T_h\}}{\{O_h\}} \tag{13}$$

Precision measures how much of T_h covers the O_h and recall measures how much of O_h is covered by the T_h. Using Equation (12) and Equation (13), the precision and recall

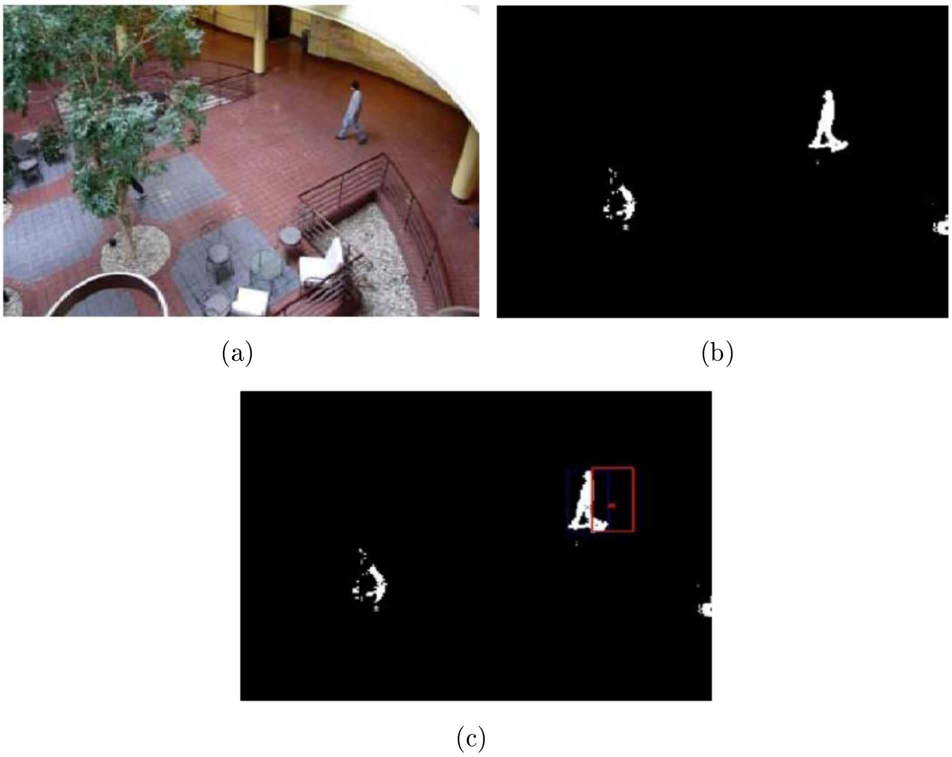


FIGURE 4. Results of object tracking in 80th frame: (a) Input frame; (b) Motion extracted output; (c) Tracked output

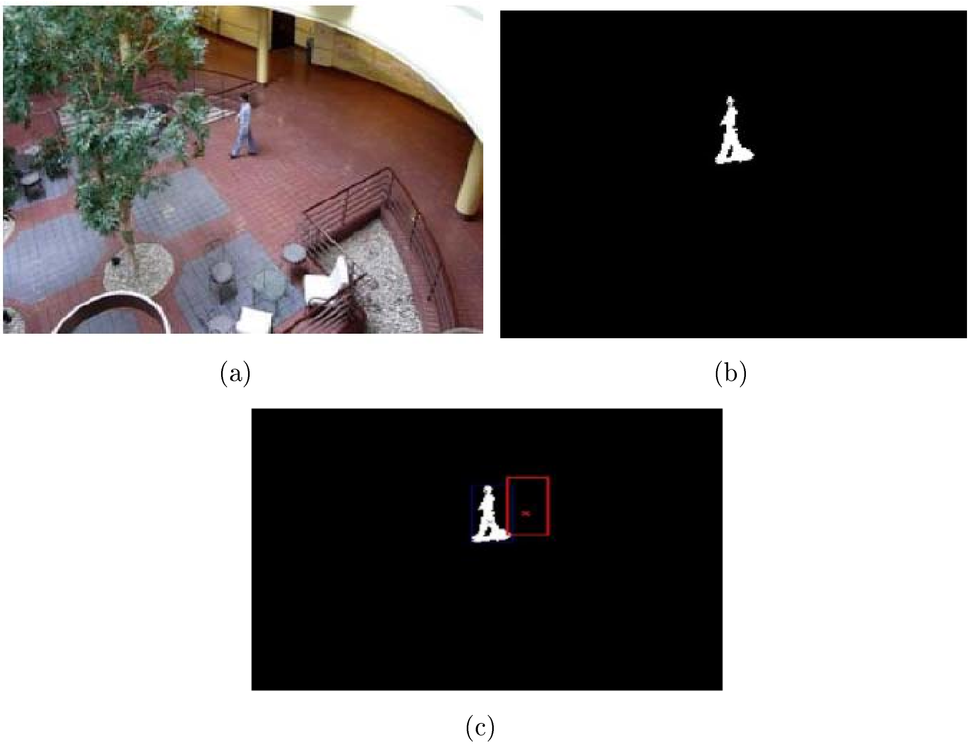


FIGURE 5. Results of object tracking in 100th frame: (a) Input frame; (b) Motion extracted output; (c) Tracked output

values for the query image are calculated for the proposed method and also for the existing method. The F-measure for the proposed method is then calculated using the expression,

$$F = 2\left(\frac{precision.recall}{precision + recall}\right) \tag{14}$$

The values obtained from the calculation are given in Table 1. These values are used for the analysis of performance between the proposed and existing method. Here the existing method is the previous paper where object detection and tracking using low level features is performed. Each values relating to the methods are entered in the table for comparison and from the table it is clear that our proposed method delivers better precision and recall than the existing method. Here the existing method is the vision based object detection and tracking [22].

TABLE 1. Precision and recall for the proposed method

S. No	Performance Analysis				F-Measure	
	Precision		Recall			
	Proposed Method	Existing Method	Proposed Method	Existing Method	Proposed Method	Existing Method
1	0.95	0.71	0.45	0.15	0.6107	0.2477
2	0.87	0.65	0.54	0.24	0.6664	0.3506
3	0.79	0.59	0.64	0.35	0.7071	0.4394
4	0.74	0.52	0.78	0.46	0.7595	0.4882

TABLE 2. Average F-measure for proposed and existing methods

F-measure	Methods	
	Proposed	Existing
Average F-measure	0.6859	0.3815

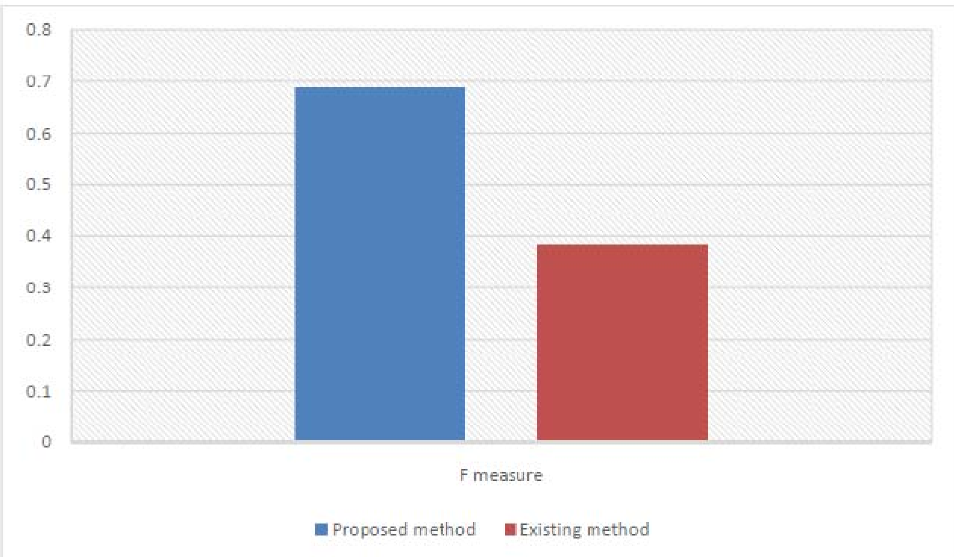


FIGURE 6. Graphical representation of average F-measure for proposed and existing method

The average F-measure value for the proposed and existing method is found out and the corresponding graph is shown in Figure 6.

The proposed method proves to be effective in terms of detecting the human object. Proper detection of human object based on motion can be thus applied in the field of human identification systems, user authentication and in military surveillance as it can be used to track the intruders towards the protected areas.

5. **Conclusion.** Human object detection from the video has been a tedious process. Various researches were performed in the field of human object detection from video sequences. In this paper we have proposed an efficient human object detection and tracking system. We developed a unique method where optimization algorithm is used in order to select the threshold value required for the object to be detected from the video. As the results show the proposed methodology proved to be more efficient and accurate in human object detection and tracking than the existing methods. To prove the effectiveness of our proposed method we have compared the precision and recall value along with F-measure of the proposed method with existing method for the human object detection and tracking process. As per the performance analysis, it is clear that our proposed method provides better F-measure value when comparing with other method. As a result it can be concluded that our proposed method is efficient in the field of human object detection and tracking.

REFERENCES

[1] J. Zhou and J. Hoang, Real time robust detection and tracking system, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[2] C.-C. Wang and C. Thorpe, Simultaneous localization and mapping with detection and tracking of moving objects, *Proc. of the IEEE International Conference on Robotics and Automation*, 2002.

[3] E. Gedikli and M. Ekinci, Motion detection, tracking and analysis for automated surveillance, *Proc. of the 3rd International Conference on Information and Communication Technologies*, 2008.

[4] C. Kim and J. N. Hwang, Fast and automatic video object segmentation and tracking for content-based applications, *Circuits and Systems for Video Technology*, vol.12, no.2, pp.122-129, 2002.

[5] R. Manikandan and R. Ramakrishnan, Object detection and tracking using background subtraction for sports applications, *International Journal of Advanced Research in Computer and Communication Engineering*, vol.2, no.10, 2013.

[6] C. H. Hsia and J. M. Guo, Efficient modified directional lifting-based discrete wavelet transform for moving object detection, *Signal Processing*, vol.96, pp.138-152, 2014.

[7] C. Cuevas and N. García, Improved background modeling for real-time spatio-temporal non-parametric moving object detection strategies, *Image and Vision Computing*, vol.31, no.9, pp.616-630, 2013.

[8] B. Karasulu and S. Korukoglu, Moving object detection and tracking by using annealed background subtraction method in videos: Performance optimization, *Expert Systems with Applications*, vol.1, pp.33-43, 2012.

[9] C. Rasmussen and G. Hager, Probabilistic data association methods for tracking complex visual objects, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.23, no.6, pp.560-576, 2001.

[10] A. Mittal and L. S. Davis, M2Tracker: A multi-view approach to segmenting and tracking people in a cluttered scene, *International Journal of Computer Vision*, vol.51, no.3, pp.189-203, 2003.

[11] J. Friedman, T. Hastie and R. Tibshirani, Additive logistic regression: A statistical view of boosting, *Annals of Statistics*, vol.2, pp.337-374, 2000.

[12] C. Veenman, M. Reinders and E. Backer, Resolving motion correspondence for densely moving points, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.23, no.1, pp.54-72, 2001.

[13] J. Xing, H. Ai and S. Lao, Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses, *Computer Vision and Pattern Recognition*, pp.1200-1207, 2009.

[14] M. A. Hussain and G. U. Kharat, Robust human motion detection and tracking in dynamic background, *International Journal of Soft Computing and Engineering*, vol.2, no.6, 2013.

[15] C. Wren, A. Azarbayejani and A. Pentland, Real-time tracking of the body, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.19, no.7, pp.780-785, 1997.
[16] S. Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi, Optimization by simulated annealing, *Science*, vol.220, no.4598, pp.671-680, 1983.
[17] L. Matthies, R. Szeliski and T. Kanade, Kalman filter-based algorithms for estimating depth from image sequences, *International Journal of Computer Vision*, vol.3, no.3, pp.209-236, 1989.
[18] B. S. M. Madhavi and M. V. Ganeswara Rao, A fast and reliable motion detection and tracking based on background subtraction, *IOSR Journal of Electronics and Communication Engineering*, vol.1, no.1, 2012.
[19] A. Jepson, D. Fleet and T. Elmaraghi, Robust online appearance models for visual tracking, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.25, no.10, pp.1296-1311, 2003.
[20] B. Leibe, K. Schindler and L. Van Gool, Coupled detection and trajectory estimation for multi-object tracking, *IEEE the 11th International Conference on Computer Vision*, pp.1-8, 2007.
[21] J. Fernandeza, R. Guerreroa, N. Mirandaa and F. Piccolia, Multi-level paralelism in image identification, *Mecanica Computational*, vol.28, pp.227-240, 2009.
[22] K. R Jadav, M. A. Lokhandwala and A. P. Gharge, Vision based moving object detection and tracking, *National Conference on Recent Trends in Engineering & Technology*, 2011.
[23] D. Karaboga and B. Akay, A comparative study of artificial bee colony algorithm, *Journal of Applied Mathematics and Computation*, vol.214, pp.108-132, 2009.
[24] D. Karaboga and C. Ozturk, Fuzzy clustering with artificial bee colony algorithm, *Journal of Scientific Research and Essays*, vol.5, no.14, pp.1899-1902, 2010.
[25] L. Li, X. Zeng, X. Li, W. Hu and P. Zhu, Video shot segmentation using graph-based dominant-set clustering, *Proc. of the 1st International Conference on Internet Multimedia Computing and Service*, New York, USA, 2009.
[26] E. Erol, H. A. Cagdas, Y. Ufuk and B. Eren, A new adaptive network based fuzzy inference system for time series forecasting, *Aloy Journal of Soft Computing and Applications*, vol.2, no.1, 2014.