

A robust single and multiple moving object detection, tracking and classification

T. Mahalingam and M. Subramoniam
Sathyabama University, Chennai, India

Received 12 July 2017
Revised 4 January 2018
Accepted 4 January 2018

Abstract

Surveillance is the emerging concept in the current technology, as it plays a vital role in monitoring keen activities at the nooks and corner of the world. Among which moving object identifying and tracking by means of computer vision techniques is the major part in surveillance. If we consider moving object detection in video analysis is the initial step among the various computer applications. The main drawbacks of the existing object tracking method is a time-consuming approach if the video contains a high volume of information. There arise certain issues in choosing the optimum tracking technique for this huge volume of data. Further, the situation becomes worse when the tracked object varies orientation over time and also it is difficult to predict multiple objects at the same time. In order to overcome these issues here, we have intended to propose an effective method for object detection and movement tracking. In this paper, we proposed robust video object detection and tracking technique. The proposed technique is divided into three phases namely detection phase, tracking phase and evaluation phase in which detection phase contains Foreground segmentation and Noise reduction. Mixture of Adaptive Gaussian (MoAG) model is proposed to achieve the efficient foreground segmentation. In addition to it the fuzzy morphological filter model is implemented for removing the noise present in the foreground segmented frames. Moving object tracking is achieved by the blob detection which comes under tracking phase. Finally, the evaluation phase has feature extraction and classification. Texture based and quality based features are extracted from the processed frames which is given for classification. For classification we are using J48 ie, decision tree based classifier. The performance of the proposed technique is analyzed with existing techniques k-NN and MLP in terms of precision, recall, f-measure and ROC.

Keywords Surveillance, Moving object detection and tracking, Mixture of Adaptive Gaussian (MoAG), Fuzzy morphological filter and blob analysis

Paper type Original Article

1. Introduction

Now-a-days, due to various security reasons these surveillance systems became more popular and necessary. This enormous requirement increased the growth and technology improvement in tracking the moving the objects. It is globally widely used by the military, intelligence monitoring, human machine interface, virtual reality, motion analysis in tracking

© T. Mahalingam and M. Subramoniam. Published in *Applied Computing and Informatics*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Publishers note: The publisher wishes to inform readers that the article “A robust single and multiple moving object detection, tracking and classification” was originally published by the previous publisher of *Applied Computing and Informatics* and the pagination of this article has been subsequently changed. There has been no change to the content of the article. This change was necessary for the journal to transition from the previous publisher to the new one. The publisher sincerely apologises for any inconvenience caused. To access and cite this article, please use Mahalingam, T., Subramoniam, M. (2021), “A robust single and multiple moving object detection, tracking and classification”, *Applied Computing and Informatics*. Vol. 17 No. 1, pp. 2-18. The original publication date for this paper was 05/01/2018.



and detecting the object accurately. This increased requirement makes the research people to pay more attention to developing an advanced methodology. Generally, detection and tracking systems are based on two methods, in which the first one uses radar technology for tracking. Another one is image processing technology by which it will capture the target tracking [1]. In this paper, we are intended to achieve perfect results in detection and tracking. According to the image processing method, various movements of the object are tracked from each frame along with its position information to get the various objectives Trajectory.

In order to build a best video surveillance system, the basic requirement is preparing an algorithm, with the motto of obtaining a result on the basis of fast, reliable and robust moving object detection and tracking system. The existing works in most of the computer-vision applications prove identifying moving objects from a video is a complex task. In the proposed video object detection and tracking technique, three phases such as detection phase, tracking phase and evaluation phase are involved. The detection phase has foreground segmentation and noise reduction, and the object tracking comes under the tracking phase. The research work on segmentation [2] is applied for separating a video into various frames and the moving object is tracked from the static background. In recent work on foreground segmentation, adaptive Gaussian mixture models are applied for foreground/background pixel distribution [2,3]. This model can be also used in multiple works on computer vision and image processing [2,4] for getting a proper validity in a given model [4].

A video sequence is captured by a static camera and we intend to capture the foreground segmentation from the video, the major problem is enabling a mixer model for capturing each pixel. Based on that, deciding a new input frame from the foreground or background and a better result is gained in [2,3]. But still, the real challenges are still ahead that is noise based on illuminations changes, slow moving objects, shadows and other phenomena that produce non-stationary frames. The other part of the problem is mitigating that is less rigid distributions from the robust mixture model. Most of the research people use generalized Gaussian density (GGD) model for their work because of its flexibility in signal processing. As it works excellently in data with different shapes and containing outlying data, it was described in [5–8]. The proposed work carries Mixture of Adaptive Gaussian (MoAG) model for foreground segmentation by which it obtains a better fitting shape of data than the mixture of Gaussian distributions model (MoG).

Secondly, we need to reduce the noise because these noises rather than corrupting the information will damage the visual effects also. As these noise reduction techniques are carrying as one of the major parts in image processing and computer vision analysis. According to the frequency domain, high frequency components in the image detail can easily collapse the high-frequency noise. Next, a perfect filter performance and good flexibility in image processing is achieved by morphological filtering. As per the pre-study of [6,9], a morphological filtering with fuzzy theory is applied and those images were fuzzed by the fuzzy operations. As a result, those fuzzed images were filtered by morphological filters which are combined with square shape structural element (SE) and the line shape SE in order to reduce the noise.

The tracking of moving objects for this blob analysis algorithm is used, which counts and measures their characteristics [10]. The motto of applying blob algorithm is to ensure that the obtained result is accurate, logical and true. The pixel values in the images are comprised by the complex algorithm in the image processing system. The blob algorithm is defined as a region of connected pixels. The blob analysis algorithm is mainly used to identify the frames and study its regions [11,12]. The algorithm enables discerning pixels by their value and placing them as one among these two categories, which is either foreground or background. Mostly, blob considers the foreground pixels to be part which are easily identifiable by the human eyes. The remaining part such as background analysis, which consists of background pixel state, those are caused because of lighting [13].

The successful frame segmentation is the main result of blob analysis and the remaining works like eliminating everything else in the background image were consider as not the part of interest. The entire objects trajectory information was gathered by our proposed techniques. The evaluation phase comprises of feature extraction and classification. The features of all processed frames are extracted and are given to classification which uses J48, KNN and MLP classifiers. Our proposed technique is implemented step by step process and its performance is measured.

This paper develops an object detection and tracking system which identified multiple objects from the crowded scenes. The organization of the proposed system is as follows.

- (1) A two-phase Background Estimation Module (BE) is used to select the optimal background candidates for the generation of the updated background models.
- (2) An object segmentation method is done from each video frame through Object Segmentation (OS) Module.
- (3) Useful features are extracted in the activity recognition process from every tracked object through an FE module.
- (4) The objects are classified using decision tree based classifier

The paper is organized as follows: In [Section 2](#), related works are discussed. The proposed technique is given in [Section 3](#) which is sub divided into [3.1](#) detection phase, [3.2](#) tracking phase and [3.3](#) evaluation phase. Experimental results are given in [Section 4](#). Finally the [Section 5](#) concludes the paper.

2. Related works

According to current technology trend the algorithm which able to detect and track the moving objects from videos which are capture from camera, still having the drawbacks in separating background and foreground information. They had done by fitting image information with some geometric models or with static sensor data or by means of probabilistic motion models.

In past various scholars were involved in improving the moving object detection and tracking system. In that advanced driver assistance systems detect the moving object with sensors like camera, radar and Light Detection and Ranging by data fusion. According to [14], on grid map lidar (Light Detection And Ranging) uses Maximum Likelihood approach for occupying the Bayesian occupancy in the grip map to locate the vehicles. In the camera images Histograms of Oriented Gradients descriptor (S-HOG) were applied for generating visual descriptors of objects. The required target can be tracked by fetching the requiring information to the radar sensor using Interactive Multiple Model (IMM). As per [15], the objects in the color image (2D) were detected using off-the-shelf algorithms and lidar data (3D) using extracting local and global histograms. Linear SVM is applied for classifying the objects and tracked accordingly using segment matching based model. The applications of [14,15] were made simple by applying additional sensors limits along with it. In [16], here the object is configured and featured by means of two methods such as configure using set of deformable kernels. The features are collected by HOGs that is histograms of color, texture and oriented gradients. In the mean shift algorithm is used, which computes kernel motion by increasing the color and textural similarities is captured between the candidates with target model. The mean-shift algorithm on HOG feature uses the optimized deformation cost for configuring the objects. In [17], the color and optical flow magnitude from the map were taken for calculating the appearance of the objects and scoring the motion proposals. From the motion proposal the objects which are similar has high score and those are clustered. The

Support Vector Machine (SVM) detector is applied for calculating temporal consistency of each cluster frames. The spatio-temporal tube for an object is generated by adding the frames with maximum detection score from the clusters. According to [18], state-of-the-art object detector is implemented for detecting moving objects. In a graph the detection hypotheses among the several graphs were connected. Deep Matching is an estimation technique among two consecutive frames which detect the pairs with affinity measures among them. This deep matching concept is applied through multi-layer deep convolutional architecture. The graph is segmented into two tracks of objects by solving the minimum cost sub-graph and multi-cutting problem.

In [19], Optical flow (OF) measurements were used for detecting moving objects as outliers by means of linear and angular velocity of the aerial camera the ego-motions is estimated. In [20], the researcher uses a systematical approach for advanced vehicles. The author represents frames dynamic environment by 2.5D map of sensor measurements with data localization and cell value with low variance as well as height. In each frame by means of spatial reasoning the moving objects were extracted from 2.5D maps. In the work [21], the frames which are spatially weighted color histogram moved to new temporal locations by mean-shifted model using the kernels of Deformable Part Model (DPM). The mean-shifted model inferred by applying deformation costs statistically for gathering frame features using histogram of oriented gradient (HOG) model. As per [22,23], off-the-shelf algorithms is used for extracting frames feature points and further classified into foreground or background by comparing its multiple-view geometry. The differencing images implies foreground region which are integrated using the feature points. Then the foreground motion history and refinement schemes are used for detecting the moving objects. In the work [22] the Kalman filter is applied for tracking moving objects according to the center of gravity on the moving object regions. In [23], the optical flow boundaries were computed using multiple figure-ground segmentation approach and then get ranked by Moving Objectness Detector (MOD). From the resultant frames which top ranked are extended to spatio-temporal tubes by random walkers on motion affinities of dense point trajectories.

In [25], the works carried out with implementing adaptive neural self-organizing background model in order to adjust the background automatically which are captured as frames from video sequences by pan-tilt-zoom (PTZ) camera. The specialty of pan-tilt-zoom (PTZ) camera is it gives maximum 360 degree view on a particular region. So we can create a particular region as background model as per the choice. In this stage the challenging task is capturing the background images from a camera mounted on a mobile platform. This is discussed by Zamalieva et al at his work by estimating the geometric transformations among two frames applying Geometric Robust Information Criterion (GRIC) [26]. By appearance model the frames were selected with their geometric transformation estimated by series of homography transforms. As per the maximum-a-posteri Markov Random Fields (MAP-MRF) optimization framework Background/ foreground labels are addressed using the motion, appearance, spatial and temporal cues. At the work [27], from the frames locations were targeted manually or with the help of some object detection algorithms. Spatial correlations are applied between two targeted frames or its neighbourhood using spatio-temporal context model. In a robotic system [28], each frame is characterized by its Optical Flow (OF) features as dynamic or static points according to its distance from the terminal. Then per unit sphere directional statistics distribution is applied for tracking and detecting the dynamic flow vectors of the objects.

The researcher in his work [29] applied Helmholtz Tradeoff Estimator along with two motion models for compensating the global motions in the frames. By fusing the compensated frames error maps are generated bi-directionally and applied in the motion vector fields. Among the high error values connections were done using hysteresis thresholding with optimal weight according to its mean weighed. From the work [30] it's

observed the cues on applying three-dimensional (3D) coordinate system. Here as per the Markov chain Monte Carlo (RJ-MCMC) particle filtering method and poster (MAP) solution of a posterior probability tracking is resolved. In [31], the work by using ego-motion is discussed as per it set the voting decision on motion vectors by estimating and compensating the vectors to get the features of foreground or background frames. Then moving edges were corrected and enhanced according for building moving objects using morphological operations. In his work Zhou et al parametric motion model and non-convex penalty is implemented by compensating camera motions using Markov Random Fields (MRFs) model. It will detect the moving objects from the video frames [32]. The drawback is it will not applicable for detecting real-time object as it works at batch mode. From the work [33], backgrounds of the moving objects were divided according to its pixel as per the spatio-temporal distribution of Gaussian on non-panoramic adaptive background model. Lucas Kanade Tracker (LKT) is applied for estimating the camera motion in finding the frames with the background model. From H.264/AVC compressed video sequences the moving objects were tracked using the combination of adaptive motion vectors and spatio-temporal Markov random field (STMRF) model in [34]. Initially the targeted frames as first frame is selected manually and then several subsequent frames with motion vectors (MVs) were considered using intra-coded block motion approximation and global motion (GM) compensation model. On the frames the rough position were estimated by GM parameters and then MVs are using STMRF. On the work [35] two models such as temporal propagation spatial model composition were combined to generate the foreground and background models. Kratz et al. proposed crowd motion using trained Markov models hidden collections with spatio-temporal motion patterns in [36].

3. Proposed technique

In our proposed work, we intended to build a robust methodology for detecting and tracking the video objects perfectly. For that the proposed methodology is segmented into three phases such as detection phase, tracking phase and evaluation phase. It has three stages such as;

1. Foreground Segmentation
2. Noise Reduction
3. Moving object tracking

3.1 Detection phase

In the detection phase the initial stage is tracking the non-stationary object as; object detection is the approach of getting the non-stationary object from a video sequence.

3.1.1 Foreground segmentation. 3.1.1.1 Problem statement. As per in [2], the foreground segmentation of the video is done by the mixture components that frequently occur with high a priori probability and small variance. A pixel in the mixture model is analyzed evaluating the data color density at particular pixel on time. It can be expressed as mixture with components modeling the distribution of a random variable \vec{H} having D dimensions, and the probability of the vector \vec{H} can be stated as;

$$p(\vec{H}) = \sum_{y=1}^{N_m} p_y p(\vec{H} / \varphi_y)$$

By means of the above expression the mixing parameters in the mixture component are $p_y, y = 1, 2, \dots, N_m$. The term $p(\vec{H} / \phi_y)$ assigned the possibility of a multivariate Gaussian

distribution with the parameters ϕ_y . The vectors of various dimensions are described as \vec{H} in which every components in the mixture were provided with mean and a standard deviation vector indicated respectively by: $\vec{\psi}_y = (\psi_{y1}, \psi_{y2}, \dots, \psi_{yD})$ and $\vec{\phi}_y = (\phi_{y1}, \phi_{y2}, \dots, \phi_{yD})$, $y = 1, 2, \dots, N_m$. In order to do foreground segmentation of a video, as stated in [9] the mixture component were to be ordered first according to its value such as $p_y / \|\vec{\phi}_y\|$, $y = 1, 2, \dots, N_m$, where the first G components are chosen as a model for the background such that:

$$G = \arg \min_g \left(\sum_{y=1}^g p_y > t \right)$$

Then t is a threshold and $\|\cdot\|$ which assign the vector norms.

In our work for modeling the mixture components we applied formalism of General Gaussians instead of Gaussian distributions. According to this the new frames which are sequentially fetched with new mixture parameters. For a background model GGDs is used to obtain a better outlier which are occurred due to the sudden illumination change in indoor scenes or non-stationary backgrounds caused due to the swaying tree branches or shadows. Even though the GGDs have the ability to adopt as per the data shape than Gaussian; Gaussian can reduce over fitting more easily. The GGD formalization is applied for online estimating new mixer model.

3.1.1.2 Mixture of Adaptive Gaussian (MoAG) model. The 1-dimensional GGD for a variable $H \in \Lambda$ is defined as follows [16]:

$$p(H/\psi, \phi, \chi) = I(\chi) \exp\left(-Y(\chi) \left| \frac{H - \psi}{\phi} \right|^\chi\right)$$

In which $I(\chi) = \frac{\chi \sqrt{\tau(3/\chi)}}{2\phi\tau(1/\chi)}$, $Y(\chi) = \left| \frac{\tau(3/\chi)}{\tau(1/\chi)} \right|^{\frac{1}{\chi}}$ and $\tau(\cdot)$ denotes the gamma function given by $\tau(z) = \int_0^\infty p^{z-1} e^{-p} dp$, where z and p are real variables. The parameters ψ and ϕ are the pdf mean and standard deviation. Then, the parameter $\chi \geq 1$ controls the tails of the pdf and determines whether the latter is peaked or flat: The bigger value of χ is, the compiled with the pdf; the smaller is χ , the more peaked is the pdf. This gives flexibility to the pdf to fit the shape of heavy-tailed data produced by the presence of noise or outliers. Note that the Laplacian and the Gaussian distributions are particular cases for the GGD where, $\chi = 1$ and 2 respectively.

As explained in (3) it is not important to do multi-dimensional generalization of the functions. Earlier work states that non-linear regression models performance is has various powers as per the input variables which produce best result in data perfection [13]. To maximize the flexibility of GGD probabilistic model, multi-dimensional GGD with various shape parameters at every dimension were applied. If the data were correlated then it cannot be track-able. To keep the property of the shape, maintaining its dimension to be independent as common and reasonable for high-dimensional data [21]. Having a d -dimensional vector $\vec{H} = (H_1, H_2, \dots, H_D)$, the probability of the vector \vec{H} with a GGD is, then, given by:

$$p(\vec{H}/\vec{\psi}, \vec{\phi}, \vec{\chi}) = \prod_{k=1}^d I(\chi_k) \exp\left(-Y(\chi_k) \left| \frac{H_k - \psi_k}{\phi_k} \right|^{\chi_k}\right)$$

in which: $\vec{\psi} = (\psi_1, \psi_2, \dots, \psi_D)$ and $\vec{\phi} = (\phi_1, \phi_2, \dots, \phi_D)$. The parameter $\chi_k \geq 1$ controls the tails of the pdf and determines whether it is peaked or flat in the k^{th} dimension. A generalized Gaussian mixture with M components is expressed as:

$$p(\vec{H}/\Theta) = \sum_{y=1}^{N_m} p_y p(\vec{H}/\psi_y, \phi_y, \chi_y)$$

With $p_y, y = 1, 2, \dots, N_m$, are the mixing parameters where $0 < p_y \leq 1$ and $\sum_{y=1}^{N_m} p_y = 1$, and $p(\vec{H}/\psi_y, \phi_y, \chi_y), y = 1, 2, \dots, N_m$, are the conditional probabilities. The set of S parameters of the mixture with N_m classes is defined by $\Theta = \bigcup_{x=1}^4 \zeta_x$, where $\zeta_1 = \{\vec{\psi}_1, \vec{\psi}_2, \dots, \vec{\psi}_{N_m}\}$, $\zeta_2 = \{\vec{\phi}_1, \vec{\phi}_2, \dots, \vec{\phi}_{N_m}\}$, $\zeta_3 = \{\vec{\chi}_1, \vec{\chi}_2, \dots, \vec{\chi}_{N_m}\}$, $\zeta_4 = \{p_1, p_1, \dots, p_{N_m}\}$ is the set of mixing parameters. The major drawbacks of the finite mixture models are estimating the vector parameters Θ and the determination of the number of classes N_m .

3.1.2 *Noise reduction using fuzzy morphological filter.* The main idea of fuzzy morphological filter is selecting fuzzy SEs and performing fuzzy morphological operations.

3.1.2.1 *Fuzzy structural elements.* Image de-noising effect is done by different structural elements. A fixed value such as 0 or 1 is assign for the traditional binary structure elements, so it can be either involved or not in the operations. For an enhanced result such as smoothen and soften fuzzy data these fuzzy structural elements were applied. In a fuzzy set (B), every subordinate value belongs to the intervals [0, 1]. A fuzzy structural elements example is shown in the Figure 1 with shadow points as the fuzzy structural element's original point [38]. The interval values [0, 1] state its subordinate element values as indefinite (see Figure 2).

3.1.2.2 *Fuzzy morphological filtering.* Morphology is a mathematical framework for the analysis of spatial structures and is based on set theory. It is a strong tool for performing many image processing tasks. Mathematical morphology is completely based on set theory. Morphological sets represent important value. By using set operations many useful operators can be defined. The important morphological operations are basically dilation, erosion, open and close operations. Morphological operations make use of a structuring element M; which can be either a set or a function that corresponds to a neighborhood-function related to the image function. In our work, we applied cascade fuzzy opening-closing operation and expressed as;

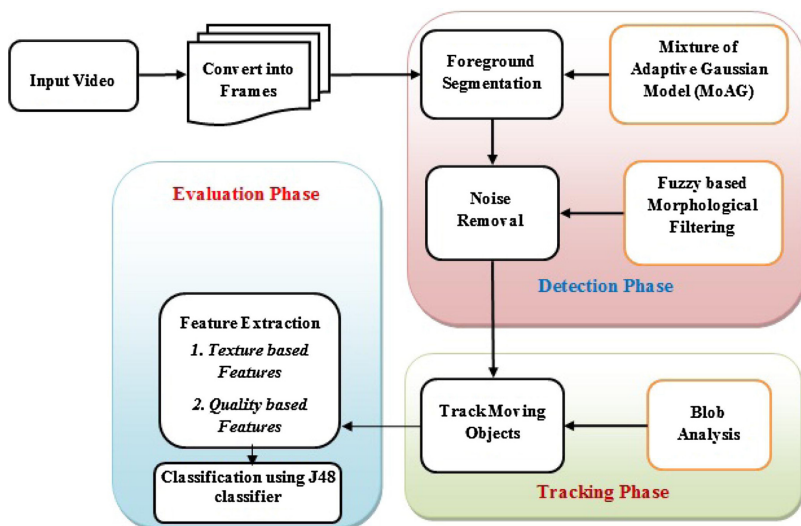


Figure 1. Architecture of the proposed system.

$$MF = \phi * X \circ Y \oplus Y + (1 - \phi) * X \bullet Y \ominus Y$$

Here fuzzy opening weight is shown as ϕ and it manipulates the final filter effects. Experiment has two operation result such as one with fuzzy opening and cascade fuzzy dilation, the next one is fuzzy closing and cascade fuzzy erosion for a better performance. Then we let ϕ equal to 0.5.

The fuzzified image is then operated through different mathematical morphological operators like “DILATION”, “EROSION”, “OPEN” and “CLOSE” with the structuring element of 3×3 mask filter. Brighter images are obtained by using the contrast intensification operator INT, on the morphologically operated image.

3.1.2.3 Morphological operations. Fuzzy morphology is nothing but a concept of applying fuzzy theory in morphology, according to which fuzzy set has input and output images. Instead of hard binarization set fuzzy set has the images. At this moment the fuzzy set operations such as images intersection and union known as fuzzy erosion and fuzzy dilate were done. The structural element is enclosed with fuzzy degree definitions of every pixel’s as per its original images. Various fuzzy operator has multiple definitions corresponding to its fuzzy morphological operations are also different as stated in [39–42]. As per the Ref. [41], the original image’s structural elements were represented on fuzziness. Then the fuzzy image erosion and dilation operations were carried out by fuzzy structural elements. The subordinating degree functions were represented as;

a. Fuzzy Dilation

$$\begin{aligned} \psi_{X \oplus Y}(a) &= \max_{b \in Y} [\max[0, \psi_X(a - b) + \psi_Y(b) - 1]] \\ &= \max[0, \max_{b \in Y} [\psi_X(a - b) + \psi_Y(b) - 1]] \end{aligned}$$

b. Fuzzy Erosion:

$$\begin{aligned} \psi_{X \ominus Y}(a) &= \min_{b \in Y} [\min[1, 1 + \psi_X(a + b) - \psi_Y(b)]] \\ &= \min[1, \min_{b \in Y} [1 + \psi_X(a + b) - \psi_Y(b)]] \end{aligned}$$

c. Fuzzy Opening:

$$X \circ Y = (X \ominus Y) \oplus Y$$

d. Fuzzy Closing:

$$X \bullet Y = (X \oplus Y) \ominus Y$$

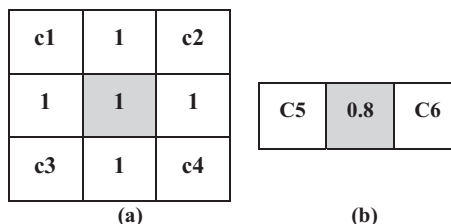


Figure 2.
(a) Fuzzy square structural elements and (b) fuzzy linear structural elements.

From the above expression x and y are the coordinate plane of the respective subordinate degree function of image and structural elements $A \mu$ and $B \mu$. According to the formula from (11) and (12), the both intervals $[0, 1]$ as the dilation's subordinate function value and fuzzy erosion's on either one basis or both. The fuzzy morphology formalizes traditional mathematical morphology with two-valued logic to fuzzy logic. According to these pixels of original images which has minimum subordinate degree will be submerged with large subordinate degrees. On fuzzy erosion the image pixel with minimum subordinate degree gets increased and maximum subordinate degree are relatively decreased. Because these corresponding subordinate values will reflect in the future possibility of a pixel to be set. The fuzzy morphological operations numerical values were $[0, 1]$ at intervals, by which it overcomes the earlier drawbacks of Fuzzy morphological operations like making the image too dark or too bright.

3.2 Tracking phase

3.2.1 Blob detection. MOG is used for background subtraction by which objects foreground is detected as blob. As per the mixture of Gaussian background subtraction a bi-level image is presented on each module to perform some basic filtering operations. The major part of the foreground is occluded and considered as tracked bob and it's matched with any of the blob centroids. As two parts come closer an object passes which is detected as single blob. That is one object is getting occluded with another. The main challenge in this section is maintaining the object labeling correctly after it splits again.

3.2.2 Blob analysis. On computer vision blob detection is explained as visual modules were expected at detecting points and the image properties are differ in their brightness or color among the surrounding. The blob detectors have two main classes with various methods according to its derivative expressions and landscape intensities. As per the modern approach these operators are referred to interest point operators, or alternatively interest region operators. There are lot of motivations in analyzing and improvising the blob detector. The main reason is it will provide the complementary information of the regions which are not gained during edge detectors or corner detectors. At an easy region the blob detection extract the region of interest easily for further process. The presence of the object is signalized in the region and object part with application image domain is tracked or recognized. On other domains like histogram analysis, blob descriptors are applied for peak detection which is useful for segmentation. The most common advantage of blob descriptors is texture analysis and texture recognition. Recent times these blob descriptors has became more popular mainly for its wide baseline stereo matching and signaling the informative image features which are more important for appearance-based object recognition based on local image statistics. It is also effective in detecting ridge detection to represent the presence of elongated objects.

3.2.3 Feature extraction. 3.2.3.1 Modified Local Self-Similarity Descriptor for texture extraction. Local self-similarity descriptor captures internal geometric layouts of local self-similarities within images/videos while accounting for small local affine deformations. It captures self-similarity of color, edges, repetitive patterns and complex textures in a single unified way. A textured region in one image can be matched with a uniformly colored region in the other image as long as they have a similar spatial layout. These self-similarity descriptors are estimated on a dense grid of points in image/video data, at multiple scales. A good match between a pair of images corresponds to finding a matching ensemble of such descriptors with similar descriptor values at similar relative geometric positions, up to small non-rigid deformations. Here the traditional local self-similarity descriptor is modified with the help of correlation value. The step by step explanation of local self-similarity descriptor is as follows,

Step 1: Initially from the input image similarity within the image is detected using the sum of squared differences $SSD_q(a,b)$.

Step 2: The resulting distance surface $SSD_q(a,b)$ is normalized and transformed into a "correlation surface" $S_q(a,b)$:

$$S_q(a,b) = \exp(-SSD_q(a,b) / \max(\text{var}_{noise}, \text{var}_{auto}(q))) \quad (1)$$

Where,

var_{noise} is a constant that corresponds to acceptable photometric variations;

$\text{var}_{auto}(q)$ takes into account the patch contrast and its pattern structure. In our implementation $\text{var}_{auto}(q)$ is the maximal variance of the difference of all patches within a very small neighbourhood of q .

Step 3: The correlation surface $S_q(a,b)$ is then transformed into log-polar coordinates centered at q , and partitioned into 80 bins (20 angles, 4 radial intervals).

Step 4: We select the maximal correlation value in each bin. The maximal values in those bins form the 80 entries of our "modified local self-similarity" descriptor vector \underline{dq} associated with the pixel q .

Step 5: Finally, this descriptor vector is normalized by linearly stretching its values to the range [0, 1] in order to be invariant to the differences in pattern and color distribution of different patches and their surrounding image regions.

Algorithm J48:

INPUT:

D //Training data

OUTPUT

T //Decision tree

DTBUILD (*D)

```
{
  T=φ;
  T= Create root node and label with splitting attribute;
  T= Add arc to root node for each split predicate and label;
  For each arc do
  D= Database created by applying splitting predicate to D;
  If stopping point reached for this path, then
    T'= create leaf node and label with appropriate class;
  Else
    T'= DTBUILD(D);
  T= add T' to arc;
}
```

Based on the above procedure we are finding the similar objects from the input video sequence.

3.3 Classification

Third phase is evaluation phase which includes feature extraction and classification. The feature extraction is discussed in experimental results in a detailed manner. Classification means labeling the images as per its features. Among that the best feature is identified by three classifiers such as KNN, J48 and MLP for result comparisons.

3.3.1 Decision Tree J48. As the name itself implies J48 is the best known decision tree based classification technique. Initially it classifies the images as per the attributes and forms tree structure respectively. The tree hierarchy is explained in an understandable way. The Decision Tree J48 is extended from ID3 and it is performed mainly for its simple methodology in identifying the hidden pixels in the images. Under classification the images were arranged in a leaf structure and get pruned. By labeling these pixels were grouped and on each pixel the information's were extracted then tested. From resultant pixel the perfect one is selected and these classifiers are appreciated for handling both discrete and continuous values.

While building a tree, J48 ignores the missing values i.e. the value for that item can be predicted based on what is known about the attribute values for the other records. The basic idea is to divide the data into range based on the attribute values for that item that are found in the training sample. J48 allows classification via either decision trees or rules generated from them.

4. Results and discussions

In this section, the efficiency and effectiveness of the proposed technique is evaluated. Our proposed detection and tracing technique is implemented in MATLAB (2013a). It has three phases namely foreground segmentation, denoising and tracking. The proposed technique correctly detects the moving objects and tracks them continuously.

4.1 Dataset description

In order to prove the performance of the proposed technique we have taken 8 video sequences which contain some challenging video characteristics like drastic scale variation, pose, and fast motion. The object detection video are taken from different datasets repositories such as MOT17, PETS2009, Football video, etc. These videos mainly consider as crowd image investigation and contain crowd count and density assessment, tracking of individual(s) surrounded by a crowd, and detection of separate flows and definite crowd occasions. The football video shows player's goal shot and goal keeper's defense. These video clips have more than 300 frames.

4.2 Performance analysis

The experimental results are given in [Figure 3](#) which has 8 rows and 4 columns. The first column contains a frame of the video and in second column foreground extracted images are given. Then the third column has clean foreground image i.e. the foreground extracted image may contain some noises which is denoised using fuzzy morphological filter. The fourth column comprises of detected and tracked moving objects. [Figure 3](#) presents the foreground segmentation, background separation object detection and tracking results which are given in each column.

In Tables 1 and 2 (as shown in annexure), V1, V2, . . . , V8 are eight videos in which Foreground (F) image, Clean Foreground (CF) image and Detected (D) image are used. In order to evaluate the performance of the proposed work we extract some features from the processed image sequence which is then given to classification. There are two types of features are extracted such as texture based features and quality based features. The statistical features has 10 metrics such as Mean, Variance, Standard deviation, Entropy, Kurtosis, Skewness, Contrast, Correlation, Energy and Homogeneity. The quality based features comprises 10 features namely Peak signal-to-noise ratio (PSNR), Mean squared error (MSE), Root Mean squared error (RMSE), Universal Image Quality Index (UIQI), Enhancement Measurement Error for original image (EMEO), Enhancement Measurement Error for processed image (EMEP), Pearson Correlation Coefficient (PCC), Signal-to-Noise

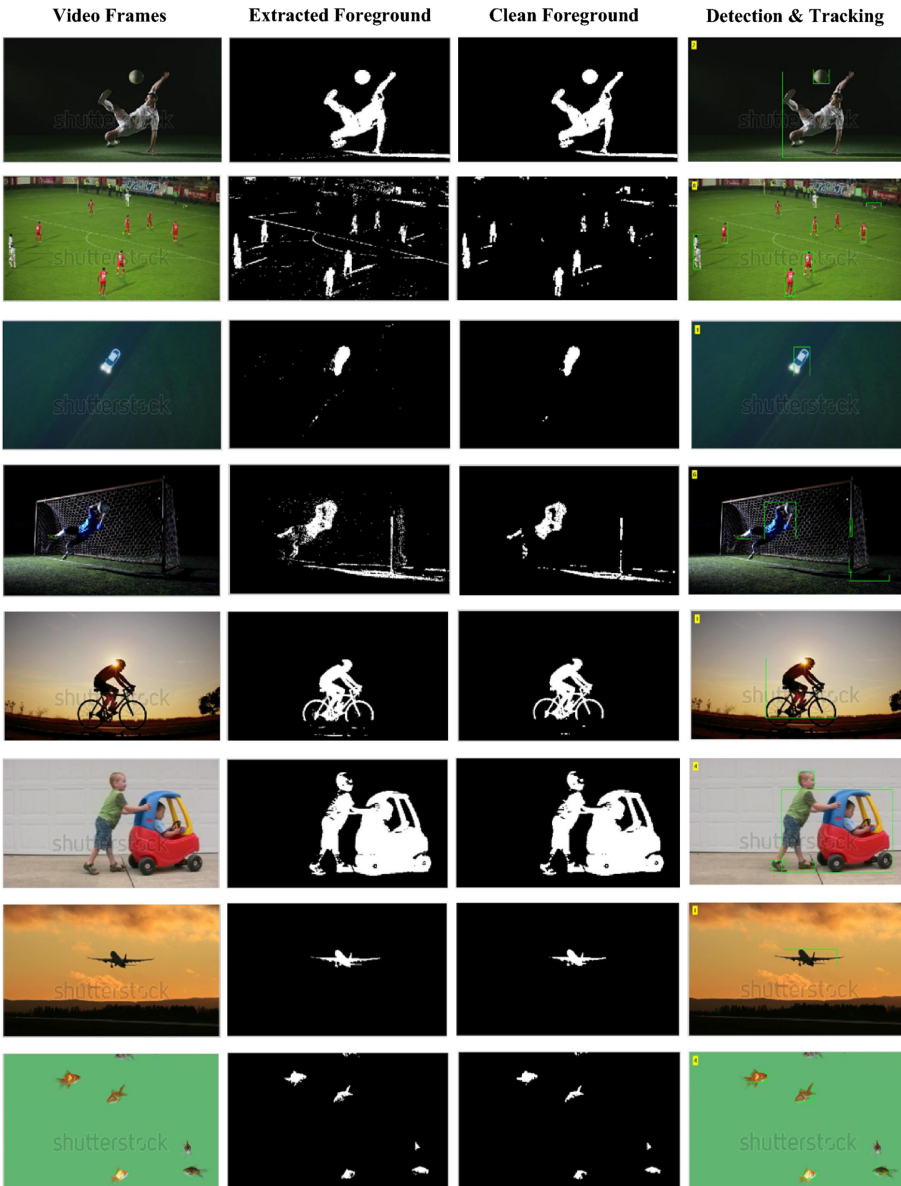


Figure 3. Detection and tracking of moving object using proposed technique.

ratio (SNR), Mean absolute error (MAE) and Root Mean square (RMS). The extracted features are tabulated. Table 1 contains quality based features and the Table 2 has texture based features (as shown in annexure).

These extracted features are converted into Attribute-Relation File Format (ARFF) and it is given to weka for classification. There are three classifiers including J48, k-nearest neighbors (k-nn) and Multilayer perceptron are used for classification. Then all three classifications are performed and its results are evaluated and tabulated.

In Table 3, the common parameters of all three classifiers are compared. Table 4, Table 5 and Table 6 presents the results of J48, KNN and MLP classifiers respectively. The performance evaluation of all three classifiers is given in Table 7 in which some of the

14

Table 3.
Comparison of all three classifiers.

Classifiers	Correctly Classified (%)	Incorrectly Classified (%)	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error (%)	Root relative squared error (%)
J48	13	7	0.3	0.3321	0.4956	66.4286	99.1289
KNN	13	7	0.3	0.365	0.5635	73	112.6943
Multilayer Perceptron	9	11	-0.1	0.5076	0.5095	101.5169	101.9078

Table 4.
Results of J48.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
0.8	0.5	0.615	0.8	0.696	0.72
0.5	0.2	0.714	0.5	0.588	0.72
0.65	0.35	0.665	0.65	0.642	0.72

Table 5.
Results of k-nearest neighbors (KNN).

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
0.6	0.3	0.667	0.6	0.632	0.65
0.7	0.4	0.636	0.7	0.667	0.65
0.65	0.35	0.652	0.65	0.649	0.65

Table 6.
Results of multilayer perceptron (MLP).

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
0	0.1	0	0	0	0.41
0.9	1	0.474	0.9	0.621	0.41
0.45	0.55	0.237	0.45	0.31	0.41

Table 7.
Performance evaluation of all three classifiers.

Classifiers	PWC	Specificity	FAR	Accuracy	PP	NP	FPR	FNR
J48	35	0.8	0.29	0.65	0.71	0.62	0.2	0.5
KNN	35	0.6	0.36	0.65	0.64	0.66	0.4	0.3
MLP	55	0	0.53	0.45	0.47	0	1	0.1

Table 8.
Comparison with existing method.

Methods	Precision	F-Measure
Proposed method	0.714	0.65
Existing method [22]	0.53	0.63

parameters like Percentage of Wrong Classification (PWC), Specificity, False Alarm Rate (FAR), Detection Rate, Accuracy, Positive Prediction (PP), Negative Prediction (NP), False Prediction Rate (FPR) and False Negative Rate (FNR) are measured and tabulated. By comparing all the parameters J48 provides better performance with reduced PWC, FAR and increased specificity, accuracy. Here in [Table 8](#) we have compared our proposed j48 based object detection method with existing method [22] in terms of precision and f-measure. Our proposed method outperforms existing method, because of its quality based and texture features.

5. Conclusion

Current trend in research field is analyzing a best approach for moving object detection and tracking than the earliest. This intention grabbed the attention among the researchers towards this field. In object tracking, a single method will not give perfect result as its accuracy are lies in the different parameters like poor resolution, change in weather condition etc. So our, proposed work is divided into three phases such as detection phase, tracking phase and evaluation phase. Here a novel technique Mixture of Adaptive Gaussian (MoAG) model is used for foreground detection. The best thing in our proposed work is applying Fuzzy based Morphological Filtering for reducing noise and achieving exact result by retaining the features of morphology operation and fuzzy theory. These things are considered as the drawbacks of traditional morphological operations will result in too dark or too bright on the images. In this order for a better de-noising effect fuzzy double structural element is implemented. Moving objects are tracked by blob detection which continuously tracks the moving objects. Features are extracted from the processed frames and are tabulated in [Tables 2](#) and [3](#). The extracted features are applied for classification which uses J48, KNN and MLP classifiers. The classification results of all three classifiers are measured and tabulated from [Tables 4](#) to [6](#). The performance evaluation is results are tabulated in [Table 7](#) which shows the J48 classifier provides better performance in terms of increased detection accuracy and reduced false alarm rate. Object detection and tracking are important and challenging tasks in many computer vision applications such as surveillance, vehicle navigation and autonomous robot navigation. In future, we would like to extend our work to detect and track object in a very crowded scene or in presence of extreme illumination variation and occlusion.

References

- [1] M.S. Allili, N. Bouguila, D. Ziou, A robust video foreground segmentation by using generalized Gaussian mixture modeling, in: Fourth Canadian Conference on Computer and Robot Vision (CRV'07), 2007, 503–509.
- [2] J. Cheng, J. Yang, Y. Zhou, Y. Cui, Flexible background mixture models for foreground segmentation, *Image Vision Comput.* 24 (5) (2006) 473–482.
- [3] C. Stauffer, W.E.L. Grimson, Learning patterns of activity using real-time tracking, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (8) (2000) 747–757.
- [4] G. McLachlan, D. Peel, *Finite Mixture Models*, Wiley Series in Probability and Statistics, 2000.
- [5] M.N. Do, M. Vetterli, Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance, *IEEE Trans. Image Process.* 11 (2) (2002) 146–158.
- [6] M. Bacchar, L.A. Gee, M.A. Abidi, Reliable location and regression estimates with application to range image segmentation, *J. Math. Imaging Vision* 11 (1999) 195–205.
- [7] R.L. Joshi, T.R. Fischer, Comparison of generalized Gaussian and Laplacian modelling in DCT image coding, *IEEE Signal Process. Lett.* 2 (5) (1995) 81–82.
- [8] K. Sharifi, A. Leon-Garcia, Estimation of shape parameter for generalized Gaussian distribution in subband decomposition of video, *IEEE Trans. Circuits Syst. Video Technol.* 5 (1) (1995) 52–56.

-
- [9] M.S. Allili, N. Bouguila, D. Ziou, Finite Generalized Gaussian Mixture Modeling and Applications to Segmentation and Tracking, Technical Report, September 2006.
- [10] R.A. Baxter, J.J. Olivier, Finding overlapping components with MML, *Statistics Comput.* 10 (1) (2000) 516.
- [11] N. Bouguila, D. Ziou, Unsupervised selection of a finite Dirichlet mixture model: an MML-based approach, *IEEE Trans. Knowledge Data Eng.* 18 (8) (2006) 993–1009.
- [12] N. Bouguila, D. Ziou, Online clustering via finite mixtures of Dirichlet and minimum message length, *Eng. Appl. Artificial Intell.* 19 (4) (2006) 371–379.
- [13] G.E.P. Box, P.W. Tidwell, Transformation of independent variables, *Technometrics* 4 (4) (1962) 531–550.
- [14] R.O. Chavez-Garcia, O. Aycard, Multiple sensor fusion and classification for moving object detection and tracking, *IEEE Trans. Intell. Transp. Syst.* 17 (2) (2016) 525–534.
- [15] S. Hwang, N. Kim, Y. Choi, S. Lee, I.S. Kweon, Fast multiple objects detection and tracking fusing color camera and 3d lidar for intelligent vehicles, in: *Proc. 13th Int. Conf. Ubiquitous Robots and Ambient Intelligence (URAI)*, 2016, pp. 234–239.
- [16] M.-C. Chuang, J.-N. Hwang, J.-H. Ye, S.-C. Huang, K. Williams, Underwater Fish tracking for moving cameras based on deformable multiple kernels, *IEEE Trans. Syst., Man, Cybern., Syst.* PP (99) (2016) 1–11.
- [17] F. Xiao, Y.J. Lee, Track and segment: an iterative unsupervised approach for video object proposals, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 933–942.
- [18] S. Tang, B. Andres, M. Andriluka, B. Schiele, Multi-person tracking by multicut and deep matching, in: *Proc. Euro. Conf. Computer Vision*, 2016, pp. 100–111.
- [19] D. Meier, R. Brockers, L. Matthies, R. Siegwart, S. Weiss, Detection and characterization of moving objects with aerial vehicles using inertial-optical Flow, in: *Proc. Int. Conf. Intelligent Robots and Systems (IROS)*, 2015, pp. 2473–2480.
- [20] A. Asvadi, P. Peixoto, U. Nunes, Detection and tracking of moving objects using 2.5 d motion grids, in: *Proc. 18th Int. Conf. on Intelligent Transportation Systems*, 2015, pp. 788–793.
- [21] L. Hou, W. Wan, K.-H. Lee, J.-N. Hwang, G. Okopal, J. Pitton, Deformable multiple-kernel based human tracking using a moving camera, in: *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 2249–2253.
- [22] W.-C. Hu, C.-H. Chen, T.-Y. Chen, D.-Y. Huang, Z.-C. Wu, Moving object detection and tracking from video captured by moving camera, *J. Visual Commun. Image Represent.* 30 (July) (2015) 164–180.
- [23] K. Fragkiadaki, P. Arbeláez, P. Felsen, J. Malik, Learning to segment moving objects in videos, in: *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4083–4090.
- [25] A. Ferone, L. Maddalena, Neural background subtraction for pantilt- zoom cameras, *IEEE Trans. Syst., Man, Cybern., Syst.* 44 (5) (2014) 571–579.
- [26] D. Zamalieva, A. Yilmaz, J.W. Davis, A multi-transformational model for background subtraction with moving cameras, in: *Proc. European Conference on Computer Vision*, 2014, pp. 803–817.
- [27] K. Zhang, L. Zhang, Q. Liu, D. Zhang, M.-H. Yang, Fast visual tracking via dense spatio-temporal context learning, in: *Proc. European Conference on Computer Vision*, 2014, pp. 127–141.
- [28] I. Marković, F. Chaumette, I. Petrović, Moving object detection, tracking and following using an omnidirectional camera on a mobile robot, in: *Proc. Int. Conf. Robotics and Automation (ICRA)*, 2014, pp. 5630–5635.
- [29] M.G. Arvanitidou, M. Tok, A. Glantz, A. Krutz, T. Sikora, Motionbased object segmentation using hysteresis and bidirectional inter-frame change detection in sequences with moving camera, *Signal Process.: Image Commun.* 28 (10) (2013) 1420–1434.

-
- [30] W. Choi, C. Pantofaru, S. Savarese, A general framework for tracking multiple people from a moving camera, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (7) (2013) 1577–1591.
- [31] F.-L. Lian, Y.-C. Lin, C.-T. Kuo, J.-H. Jean, Voting-based motion estimation for real-time video transmission in networked mobile camera systems, *IEEE Trans. Ind. Informat.* 9 (1) (2013) 172–180.
- [32] X. Zhou, C. Yang, W. Yu, Moving object detection by detecting contiguous outliers in the low-rank representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (3) (2013) 597–610.
- [33] S.W. Kim, K. Yun, K.M. Yi, S.J. Kim, J.Y. Choi, Detection of moving objects with a moving camera using non-panoramic background model, *Mach. Vision Appl.* 24 (5) (2013) 1015–1028.
- [34] S.H. Khatoonabadi, I.V. Bajic, Video object tracking in the compressed domain using spatio-temporal markov random fields, *IEEE Trans. Image Process.* 22 (1) (2013) 300–313.
- [35] T. Lim, B. Han, J.H. Han, Modeling and segmentation of floating foreground and background in videos, *Pattern Recog.* 45 (4) (2012) 1696–1706.
- [36] L. Kratz, K. Nishino, Tracking pedestrians using local spatiotemporal motion patterns in extremely crowded scenes, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (5) (2012) 987–1002.
- [38] Z. Youlian, H. Cheng, Z. Lifang, P. Lingjiao, Mixed noise reduction method based on fuzzy morphological filtering, in: 26th Chinese Control and Decision Conference (CCDC), 2014, pp. 2970–2973.
- [39] Zhou Xutong, Shi Pengfei, Fuzzy mathematical morphology based on triangle-norm logic, *J. Shanghai Jiaotong Univ.* 9 (1998) 73–77.
- [40] Xu Fengsheng, Wu Minjin, C.Y. Suen, Theoretical aspects of fuzzy morphology, *J. East China Normal Univ. (Nat. Sci.)* (4) (1996) 38–46.
- [41] D. Sinha, E.R. Dougherty, Fuzzy mathematical morphology, *Vision, Commun. Imagine Represent.* 3 (1992) 286–302.
- [42] Zhang Chengbin, Research of fuzzy morphological operator, *Software Guide* 9 (10) (2010) 23–25.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.aci.2018.01.001>.

Corresponding author

T. Mahalingam is the corresponding author and can be contacted at: lingamdivi2@gmail.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com