

# A Survey and Comparative Analysis of Moving Object Detection and Tracking

A.Ramya<sup>1</sup> Dr. P.Raviraj<sup>2</sup>

PG Scholar, Kalaignar Karunanidhi Institute of Technology, Coimbatore, TN, India<sup>1</sup>  
Professor, Kalaignar Karunanidhi Institute of Technology, Coimbatore, TN, India<sup>2</sup>

## Abstract

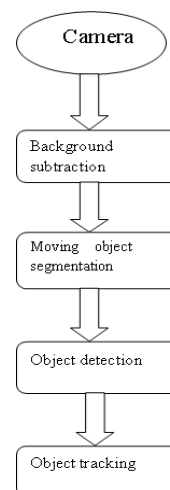
Moving object detection in a video sequence is a critical task in vision application and it is an active research topic; it can be handled by indoor and outdoor environments. Identifying moving objects is an important task for automated video analysis. This paper presents the survey on object detection and tracking in automated video analysis in vision applications. Background subtraction is the foremost approach for detecting moving objects. Background may be static or dynamic and estimation of background detection is essential to detect an object in a video sequence. Tracking an object in a video sequence means identifying the location of the object continuously when either the object or camera is moving. Object tracking is required in vision applications that require the location and shape of object in every frame. This paper highlights survey of various object detection and background subtraction techniques.

**Keywords:** Moving object detection, Low rank component, sparse component, Background subtraction, Motion segmentation

## 1. Introduction

In automated video analysis [2], there are three key steps for automated video analysis; object detection, object tracking and behaviour recognition. As a first steps, object detection locate and segment objects in a video. Then the objects can be tracked from frame to frame and the behaviour of the tracked object is analysed. In this both object detection and tracking plays an important role in many practical applications. Use of cameras and other sensors to monitor activities with the goal of understanding events happening at a site automatically. The automatic detection of events in videos would enable efficient chronicle and automatic annotation. The detection and tracking of moving objects draw great attention from the researchers in the area of computer vision. Object detection is usually performed by object detectors or background subtraction Detection of moving objects provides a classification of the pixels in the video sequence into either foreground or background. This classification of pixels to detect moving objects uses an approach

background subtraction. In background subtraction, each pixel in the video frame gets deviate from the background taken as moving objects for applications such as surveillance so, [8] there are many challenges in developing a good background subtraction algorithm. First, background subtraction must be robust against illumination changes. Second, detection of non-stationary background objects and shadows cast by moving objects should be avoided. A good background model should have a sense of reacting quickly to changes in background and adapting itself to accommodate changes occurring in the background such as moving of a stationary object from one place to another. For a real-time system, good foreground detection rate and the processing time for background subtraction is essential. Object detection is divided into different stages as shown in figure1. For applications such as surveillance, Background subtraction is a hierarchy of techniques for segmenting out objects. Motion segmentation is a part of background subtraction. In motion segmentation, object is detected by classifying pixels according to motion patterns which usually called as motion-based object detection.



**Figure 1. General block diagram of video surveillance system.**

In motion segmentation [2] object which moves continuously moves present in the scene and due to motion of camera, background also moves.

Optical flow field defines translation of each pixel in a region which is a dense field of displacement vectors. Partition of Optical flow field is a most common approach for motion segmentation. Optical flow field [2] in a scene should be smooth in each layer of motion and at layer boundaries, only sharp motion changes should occur. Optical flow field and segmentation can work in presence of large camera motion however in general not true in practice. Foreground may be complex with non-rigid shapes and background also is complex when considering under varying textures and illumination changes. Object detection and background subtraction techniques are further discussed in the following sections.

## 2. Background Subtraction Techniques

### 2.1. Eigen Background Subtraction

Eigen background subtraction proposed by Oliver, et al[3] uses an Eigen space to model the background for moving object segmentation. In this method, background model is learned from unconstrained video sequences is its ability, even when they contain moving foreground objects segmentation. PCA is used to reduce the dimensionality of the space. After PCA is performed, reduced space should represent only the unmoved parts of the image, if also moving objects presents in the space.

The main steps of the algorithm are as follows,[1]

- 1) A sample of  $n$  images with each image contains  $p$  pixels for all the images mean  $\mu_b$  is computed.
- 2) Covariance matrix is computed and when a new image  $I$  enters, it is projected on to Eigen space,  $I' = \varphi_{Mb}(I - \mu_b)$  and  $I'$  is projected back as  $I'' = \varphi_{Mb}^T I' + \mu_b$ .
- 3) Difference  $I$  and  $I''$  is computed.
- 4) Foreground points detected at location where  $|I - I''| > T$ .

An Eigen vector formed in subspace represents the static unmoved parts of the scene.

### 2.2. Kernel Density Estimation

Elgammal et. el [4] proposed kernel density function for building statistical representation of background and fore-ground. A background model and background subtraction process based on non-parametric kernel density estimation uses pixel intensities a basic feature for modelling the background. In non-parametric kernel density approach, density function is estimated directly from data without any assumptions about underlying distribution. The choice of suitable kernel bandwidth is an important issue in kernel

density estimation. A major drawback of using non-parametric kernel density estimator is the computational cost.

### 2.3. Sparse Signal Recovery

Volkan et.al [6] proposed lattice matching pursuit for stable recovery to represent sparse signals using markov random field (MRF). In sparse signal representation non-zero coefficients are clustered together. In LAMP, likelihood of signal support is evaluated iteratively and optimised under an Ising model. Data residual is calculated in matching pursuit as a first step. Sparse signal is estimated using graphical model such as support and signal model. In graphical model, sparse support decreases the ambiguity and size of search space for unknown part determines the speed of the algorithm. LaMP achieves best performance at  $M=2.5$  K when compared with other methods such as FPC and COSAMP.

Jin Zhou et.al [5] proposed a new greedy sparse recovery algorithm in which data residue in iterative process pruned according to sparsity and group clustering rather than only sparsity. Proposed algorithm requires 5 main steps in each iteration,

1. Pruning the residue estimation
2. Merging the support sets.
3. Signal estimation by least square
4. Pruning the signal estimation and
5. Updating the signal/residue estimation and support set.

Data pruning is estimated in step1 and step 4 using DGS rather than  $k$ -sparse approximation. Sparsity ranges can be set to run the AdaDGS recovery algorithm till the condition is true. As a result optimised background subtracted image and background image is got through using AdaDGS based background subtraction.

Julien et.al [7] proposed proximal operator using network flow algorithm for solving structured sparse problems including sum of  $l_\infty$ -norms over the group of variables. This proximal operator which is associated with structured norm can be computed efficiently by solving quadratic min-cost flow problem.

### 2.4. Motion Segmentation

For a robust foreground segmentation spagnolo.et.al [8] proposed an algorithm which combines background image and temporal image analysis. Updating of background image is done to all pixels including pixels covered by foreground objects. In this approach, to have effective moving points, use radiometric similarity between subsequent regions of consecutive frames and also between background image which is referenced and temporal image to segment objects of foreground.

Update of each pixel in the background image is according to the variations of pixels in image with same intensity value. For each pixel photometric gain is calculated and mean photometric gain calculated for pixels with same intensity value. This algorithm can withstand for light conditions for continuous variations, reduce noise in image, detect object accurately when foreground object moves after a long period of time. Drawback of this approach is it mistakenly detect object that are motionless in image and it cannot eliminate shadows.

### 3. Object Detection Techniques

#### 3.1. Principal Component Pursuit

Wright.et.al [9] proposed an algorithm principal component pursuit (PCP) in which recovery of both low-rank component and sparse component of a data matrix. This PCP is a convex program solved for the recovery of low-rank  $L_0$  and sparse  $S_0$  by simply minimising the weighted combination of nuclear and  $l_1$  norm. Let  $\|M\|_* = \sum_i \sigma_i(M)$  denote the nuclear norm of the matrix  $M$  (i.e) sum of the singular values of  $M$  and let  $\|M\|_1 = \sum_{ij} |M_{ij}|$  denotes  $l_1$  norm of  $M$  as seen long vector in  $R^{n1 \times n2}$ . PCP estimate solving,

$$\begin{aligned} & \text{Minimise } \|L\|_* + \lambda \|S\|_1 \\ & \text{Subject to } L+S=M \end{aligned}$$

can exactly recover low-rank  $L_0$  and sparse  $S_0$ . Principal components of a data matrix can be recovered even though the entries are randomly corrupted. To get a robust principal component analysis (RPCA),

$$M=L_0 + S_0 + N_0 \quad N_0 \text{ is a dense noise.}$$

$L_0$  can be approximated as low-rank and small errors obtained can then be added to all data entries (i.e) to get robust PCA combine sparse gross errors and dense small noise.



Figure 2. a) Original frames

b) Low-rank L



c) Sparse S

Figure . Shows that Background modelling from video.

#### (a) Frames of original video M.

#### (b)-(c) Low-rank L and sparse component S obtained by PCP[9]

#### 3.2. Mixture of Gaussian (MoG)

Stauffer[12] proposed Gaussian mixture model in which each pixel in the background is modelled as a mixture of Gaussian. Each and every pixel value is compared with the existing set of models to find the match. Parameters that are considered in this method are  $\alpha$  and  $T$  .parameters for the matched model gets updated based on the learning rate. If there finds no match, the least model got is discarded and it is replaced by new Gaussian with initialization by current pixel value. (i.e.) pixel values that don't fit in the background are considered to be background. MoG method is robust when dealing with different cameras and scene and also with slow changes in lighting. Due to automatic pixel-wise threshold, method can recover quickly when background reappears

#### 3.3. Detecting Contiguous Outliers in the Low-rank Representation (DECOLOR)

Xiaowei et.al [2] proposed an algorithm for moving object detection. In DECOLOR both object detection and estimation of background is performed simultaneously without training sequence. Assume background images in videos are linearly connected, so a matrix can be formed consisting of vectorized video frames can be estimated by low-rank. In the low-rank matrix, moving objects can be detected as outliers. This outlier detection makes easier estimation of foreground from background.  $D=\{I_1, \dots, I_n\} \in IR^{m \times n}$  is a matrix which represents  $n$  frames.  $B \in IR^{m \times n}$  is a matrix which denotes underlying background images.  $S \in \{0,1\}^{m \times n}$  is a matrix which denotes foreground support

$$S = \begin{cases} 0, & \text{if } ij \text{ is background} \\ 1, & \text{if } ij \text{ is foreground} \end{cases}$$

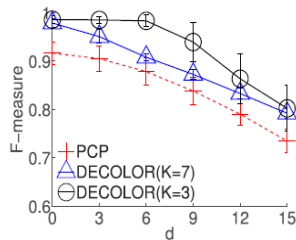
To estimate foreground and background, background model, foreground model and signal model combined to minimize the energy function can be formulated as,

$$\begin{aligned} & \min_{B, S_{ij} \in \{0,1\}} \frac{1}{2} \sum_{ij: S_{ij}=0} (D_{ij} - B_{ij})^2 \\ & + \beta \sum_{ij} S + \gamma \sum_{(ij,kl) \in \epsilon} |S_{ij} - S_{kl}|, \end{aligned}$$

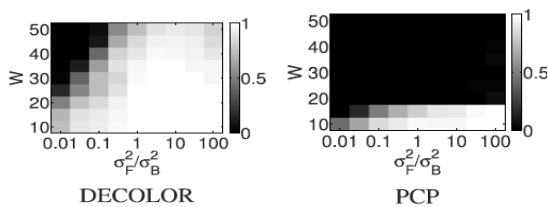
$$\text{s.t rank}(B) \leq K.$$

This formulates that background images should form a low-rank matrix. Fig. 3(a) shows the averaged F-measure as a function of  $d$  with foreground not move for  $d$ -frames, here  $\text{rank}(B_0)=3$ . the accuracy of DECOLOR decrease as long as  $d > 0$  when  $K=7$  which is a default

parameter. this is due to case DECOLOR overfits the static foreground. on decreasing K to 3 DECOLOR performs stable until  $d > 6$  which demonstrates that DECOLOR can tolerate a temporary stopping of foreground. Result of PCP is also shown for comparison in figure 3b. DECOLOR performs high accuracy when compared to PCP. If the object is large or moving slowly its interior region remains unchanged which makes foreground to fit into the low-rank model so, DECOLOR fails.



**Figure 3. (a) F-measure as a function of d where d is the number of frames within with foreground stops moving.**



**Figure 3. (b) Fraction of accurate foreground detection as a function of  $\sigma_f$  and  $W[2]$**

In example of smoke detection, smoke is detected as foreground which makes background behind the smoke cannot be recovered which is occluded.

## 4. Proposed Work

### 4.1. Graph Cut

The objective of graph-cut is to segment objects in a video. The mathematical description of a graph is as follows, [10]

$$G = (V, E)$$

is a graph with V denote vertices and E denote edges. A directed graph G consists of set of vertices and set of ordered pairs of edges. An s-t graph is a weighted directed graph with 2 nodes source s and sinks t nodes. An s-t cut in a graph G is set of edges  $E_{cut}$  such that from source to sink there is no path. The cost of a cut is the sum of the edges weights in  $E_{cut}$ . The value of the flow is defined as,

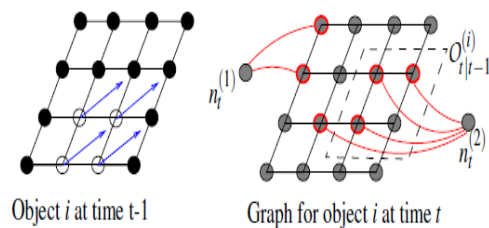
$$|f| = \sum_{v \in V} f(s, v)$$

Represent the amount of flow passing from source to sink. The max-flow/min-cut problem is to maximize the value of flow and minimize the cost.

Aureli. et. al proposed [11] a method to track and segment multiple objects in a video using

min-cut/max-flow algorithm. Object segmentation with use of graph-cut gives a modest computational cost. A binary labelling objective function is introduced for each object which combines low-level pixel-wise features, high-level observations obtained through an independent detection module. P denotes the set of N pixels from an input image sequence. t is associated as a feature vector

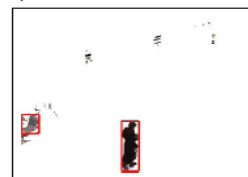
$Z_{s,t} = (Z_{s,t}^{(C)}, Z_{s,t}^{(M)})$  where  $Z_{s,t}^{(C)}$  is a 3-dimensional vector in RGB color space and  $Z_{s,t}^{(M)}$  is a 2-dimensional vector of optical flow values to each pixel. Assume at time t  $k_i$  objects are tracked and  $i^{th}$  object at time t is denoted as  $O_t^{(i)}$ . An energy function is introduced which is minimized using graph-cut.



The above graph describes about the energy minimization at time t-1 in left figure, optical flow vectors for the object is shown in blue, white nodes represented as objects and black nodes as background. The right figure describes the graph at time t, red circled pixel nodes correspond to mask of the two observations and dashed box indicates predicted mask.



**Figure 4. a) Reference frame b) Current frame**



**c) Result of background subtraction (pixels in black are labelled as foreground) and derived object detections (indicated with red bounding boxes) [11].**

The algorithm is robust to partial and complete occlusion, illumination changes, missing observations. The use of secondary level multi-level energy function in the method allows individual tracking and segmentation of objects and the above observations of the figure are obtained by

simple background subtraction based on a single reference frame.

## 5. Conclusion

Availability, efficiency of usage and application automation of videos, along with the increasing popularity of video on internet and versatility of video application heavily rely on object detection and tracking in videos. Object detection gives accuracy at different scenarios based on the application. This paper discussed

some techniques to detect moving object in video frames and their advantage and limitations. Graph-cut algorithm is robust to occlusion and complex background. DECOLOR gives high accuracy when background is static but performs poorly when camera is large and foreground moves slowly (or) occluded. This shows that considering graph-cut algorithm can perform better when background is complex and occluded.

### *Comparative Analysis of Object Detection Techniques*

Techniques	Advantages	Disadvantages
Eigen Background Subtraction	<ul style="list-style-type: none"> <li>Robust to unstable background</li> </ul>	<ul style="list-style-type: none"> <li>Difficulty in keeping up to date video streams</li> </ul>
Kernel density Estimation	<ul style="list-style-type: none"> <li>Handle situations when background scene is not completely static</li> <li>Adaptive to changes in illumination</li> </ul>	<ul style="list-style-type: none"> <li>High computational cost</li> </ul>
Mixture of Gaussian (MoG)	<ul style="list-style-type: none"> <li>Deals with slow lighting changes</li> <li>Deals with multi model distribution</li> </ul>	<ul style="list-style-type: none"> <li>Sensitive to sudden changes in global illumination</li> <li>Computationally intensive</li> <li>Scene remains stationary for a long period of time; variance of background becomes small</li> </ul>
Background and temporal analysis	<ul style="list-style-type: none"> <li>Robust to illumination changes</li> <li>Effective when applied in different outdoor and indoor environments</li> </ul>	<ul style="list-style-type: none"> <li>Mistakenly detect object that are motion less</li> <li>Not able to eliminate shadows, especially when objects are highly contrasted on background</li> </ul>
Principal Component Pursuit(PCP)	<ul style="list-style-type: none"> <li>Robust to sparse gross errors and dense noise when combined</li> <li>Scalable</li> </ul>	<ul style="list-style-type: none"> <li>High computational and memory cost</li> </ul>
Detecting Contiguous Outliers in the Low-Rank Representation (DECOLOR)	<ul style="list-style-type: none"> <li>Robust to static background</li> </ul>	<ul style="list-style-type: none"> <li>Poor performance when camera is large and foreground moves slowly or occluded</li> <li>Complex background.</li> </ul>
Graph Cut	<ul style="list-style-type: none"> <li>Robust to complex background, Sudden illumination changes, occlusion</li> </ul>	-

## References

- [1]. M.Piccardi, "background subtraction techniques: a review," proc. IEEE int'l conf. systems, man, and cybernetics, 2004.
- [2]. Xiaowei Zhou, student member, IEEE, Can Yang, and Weichuan Yu, member, IEEE "moving object detection by detecting contiguous outliers in the low-rank representation", IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 3, march 2013
- [3]. Kinjal a Joshi, Darshak g. Thakore "a survey on moving object detection and tracking in video surveillance system", INTERNATIONAL JOURNAL OF SOFT COMPUTING AND ENGINEERING (IJSCE) issn: 2231-2307, volume-2, issue-3, july 2012
- [4]. Ahmed Elgammal, Ramani Duraiswami, member, IEEE, David Harwood, and Larry s. Davis, fellow, IEEE "background and foreground modeling using nonparametric kernel density estimation for visual surveillance".
- [5]. Junzhou Huang, Xiaolei Huang, Dimitris Metaxas "learning with dynamic group sparsity"
- [6]. Volkan Cevher, Marco f. duarte, Chinmay hegde, richard g. Baraniuk, "sparse signal recovery using markov random fields".
- [7]. Julien Mairal, Rodolphe jenatton, guillaume obozinski, francis bach, "network flow algorithms for structured sparsity"
- [8]. Sen-Ching S.Cheung and Chandrika Kamath "Robust techniques for background subtraction in urban traffic video"
- [9]. P. Spagnolo, t.d' orazio \*, m. leo, a. Distante "moving object segmentation by background and temporal analysis" ELSEVIER image and vision computing 24 (2006) 411-423.
- [10]. E.candes, X. Li, Y. Ma, and J. Wright, "robust principal component analysis?" J. ACM, vol. 58, article 11, 2011.
- [11]. Zhayida Simayijiang, Stefanie Grimm, "Segmentation with Graph Cuts".
- [12]. Chris Stauffer, W.E.L Grimson, "Adaptive background mixture models for real-time tracking", Proc.IEEE Conf. Computer Vision and Pattern Recognition 1999.