

Robust Detection and Tracking of Multiple Objects under Complex Backgrounds and Occlusions

Sukanyathara J

Department of Computer Science
Viswajyothi College of Engineering & Technology
MG University, Kerala, India

Alphonsa Kuriakose

Department of Computer Science
Viswajyothi College of Engineering & Technology
MG University, Kerala, India

Abstract—Detection and tracking are two important aspects of visual surveillance applications which is gaining importance rapidly. Certain conditions like cluttered background, camera noise, target appearance variation, and occlusion are barriers to robust detection and tracking, especially in case of multiple moving objects. Object detection under complex backgrounds is an area of active research. Contrary to recent approaches, this paper focus on developing a framework which can effectively detect multiple moving objects and track them despite of background clutter and prior knowledge of targets in the scene with two major contributions. First, a segmentation method which is robust in dynamic background conditions such as swaying leaves, fountains, and other complex backgrounds. Second, a multiple object tracking algorithm using Kalman filter which perform efficient tracking of occluded objects. By modelling the background and foreground, the system can accurately detect the real moving objects. The video object tracking method assigns each object a unique track and maintains it over time. Experimental results convey that the framework perform well for several challenging sequences, and our proposed framework is effective for the aforementioned challenges.

Keywords—computer vision; complex backgrounds; occlusion handling; background clutter; multiple object tracking.

I. INTRODUCTION

As a technology for monitoring people using computer vision, video surveillance has received increased attention. Automated systems which help security personals to detect and track people are gaining importance. Detection and tracking of moving objects is very essential in monitoring public transportation, and critical assets. The existing detection and tracking based approaches for video object tracking are unreliable in complex surveillance videos due to problems like occlusions, and target appearance variations.

Object tracking, in general, is a challenging problem. Difficulties in tracking objects can arise due to abrupt object motion, changing appearance patterns of both the object and the scene, non-rigid object structures, object-to-object occlusions, etc. Tracking is usually performed in the context of higher-level applications that require the location and/or shape of the object in every frame. Typically, assumptions are made to constrain the tracking problem in the context of a particular application.

Several methods have been proposed for detection and tracking of video objects but those methods have some

limitations. In [1] a segmentation method is introduced using multi-background registration technique but it doesn't keep track of the objects over time. In [2], the multi-background registration based segmentation is integrated with a particle filter tracker which could track only a single object. Gaussian mixture model is used for segmentation in [3] but it results in false positives and not suitable for dynamic backgrounds. [4], [5], [6] proposes several simple but efficient video object segmentation algorithms. However, these algorithms cannot address dynamic backgrounds because only one background layer is employed in their background model. Various algorithms are complex and require large amount of memory. Vosters et al. [7] proposed a more complex algorithm, using an Eigen background and statistical illumination model, which can address sudden changes of illumination, but it has very high computational requirement.

This paper proposes a framework which robustly detect multiple moving objects in cluttered background and track them despite of occlusion and a priori knowledge of objects in the scene. The segmentation algorithm makes use of background and foreground model which is used to find the initial foreground mask of the input frame. Morphological operations are then applied on the initial foreground mask to get the final foreground mask highlighting the foreground pixels only. The blobs are analyzed from the foreground mask and the objects are detected along with their current position in the frame, width, and height. This information is important in keeping track of the objects over time.

Tracking is done using Kalman filter. Each object is assigned an identifier. When a new object is detected, a new track is assigned to the object with corresponding identifier. Tracks are assigned only if the object remains visible for a particular number of frames. This makes the Kalman filter show wrong identifiers to newly tracked objects. In case of a false detection, the object size will be small, but a track is created and when it disappears before the frame threshold, the next track will be assigned the next id only. Thus new tracks will get large track ids. To overcome this drawback, a height threshold is specified before creating a new track.

The proposed tracking framework is able to track multiple objects despite of occlusions under dynamic background conditions. The framework is able to track objects correctly even if the appearance factors such as size and shape of the target varies along the lifetime of the targets. The

trajectories of the targets are plotted and somewhat smooth trajectories are obtained.

The segmentation and tracking framework is proposed with two major contributions. First, a segmentation algorithm which make use of background model and foreground likelihood to detect the foreground objects in cluttered background or low quality, noisy videos. Second, a video object tracking framework based on Kalman filter which is able to track multiple moving objects despite of occlusion, target appearance variation, non-rigid object motion, and plotting their trajectories. It is suitable in complex background conditions.

In this paper, a video object segmentation and tracking framework is proposed for application in visual surveillance networks with two major contributions. First, an algorithm which is robust to background clutter for video object segmentation with foreground and background models. Second, a video object tracking algorithm with Kalman filter which is able to track multiple moving objects and tolerate occlusion to a good extend. It is suitable in complex background conditions such as swaying leaves in a forest, rotating fans, fountains, etc., and also robust in low camera quality, low lighting conditions and shape-varying targets.

II. VIDEO OBJECT SEGMENTATION

To formulate an object detection strategy which is robust to background clutter, modelling of the foreground as well as the background are considered. A log likelihood function is then applied to find out the most likely foreground and background pixels when the current frame is analyzed. The method is meant for fixed cameras mounted for security

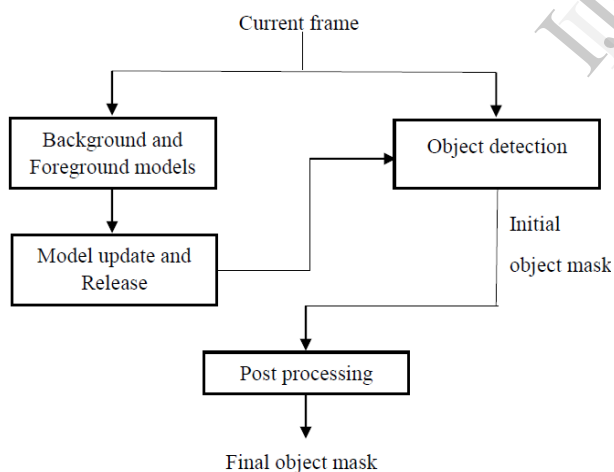


Fig.1. Segmentation Block Diagram

purposes.

The block diagram of the proposed segmentation algorithm is shown in Fig.1. The functional blocks are background and foreground modelling, background update and release, object detection and post processing. In the presence of dynamic textures, cyclic motion, and non-stationary backgrounds in general, the “correct” model of spatial uncertainty often has an arbitrary shape and may be bimodal or multimodal, but structure exists because by

definition, the motion follows a certain repetitive pattern. Such arbitrarily structured data can be best analyzed using nonparametric methods since these methods make no underlying assumptions on the shape of the density.

A. Modelling the Foreground and Background

Pixel-wise models ignore the dependencies between proximal pixels and it is asserted here that these dependencies are important. In order to build a background model, consider the situation at time t , before which all pixels, represented in 5 dimensional space, form the set $\psi_b = \{y_1, y_2, \dots, y_n\}$ of the background. Given this sample set, at the observation of the frame at time t , the probability of each pixel-vector belonging to the background can be computed using the kernel density estimator.

$$P(X/\psi_b) = n^{-1} \sum_{i=1}^n \psi H(x - y_i) \quad (1)$$

The kernel density estimator is a nonparametric estimator and under appropriate conditions the estimate it produces is a valid probability itself. Thus, to find the probability that a candidate point x belongs to the background b , an estimate can be computed.

In this paper, temporal persistence is utilized as a property of real foreground objects, i.e. the objects of interest tend to remain in the same spatial vicinity and tend to maintain consistent colours from frame to frame. The joint representation used here allows competitive classification between the foreground and background. To that end, models for both the background and the foreground are maintained. An appealing feature of this representation is that the foreground model can be constructed in a consistent fashion with the background model, a joint domain-range nonparametric density $\psi_f = \{z_1, z_2, \dots, z_n\}$.

At any time instant the probability of observing a foreground pixel at any location (i, j) of any colour is uniform. Then, once a foreground region is been detected at time t , there is an increased probability of observing a foreground region at time $t + 1$ in the same proximity with a similar colour distribution. Thus, foreground probability is expressed as a mixture of a uniform function and the kernel density function:

$$P(X/\psi_f) = \alpha\gamma + (1-\alpha)m^{-1} \sum_{i=1}^n \psi H(x - z_i) \quad (2)$$

Where α is the mixture weight, and γ is a random variable with uniform probability, where,

$$\gamma R, G, B, X, Y(r, g, b, x, y) = \frac{1}{(R * G * B * M * N)} \quad (3)$$

H is a symmetric positive definite $d \times d$ bandwidth matrix.

$$P(X/\psi_b) = n^{-1} \sum_{i=1}^n \psi H(x - y_i) \quad (4)$$

All the pixels detected as foreground are used to update the foreground model.

B. Background Update and Release

In case of dynamic backgrounds, the value of background pixels change over time. Update the foreground model and the non-active background by replacing the pixels in the foreground, model and the previous background model.

To update the model,

1. Find all the pixels detected as foreground to the foreground model ψf .
2. Remove all pixels in ψf from $k f$ frames ago.
3. Append all pixels of the image to the background model ψb
4. Remove all pixels in ψb from $k b$ frames ago.

This also releases the candidate background value belongs to noise or the foreground object.

C. Object Detection

Next step is to find the foreground objects using the background and foreground model built. For each pixel, the likelihood ratio for a pixel to become foreground is found using the foreground model. Similarly find the likelihood ratio for becoming the background.

To find out whether the detected pixel is an object of interest or not, the following equation is used:

$$\tau = -\ln \frac{P(X/\psi b)}{P(X/\psi f)} \quad (5)$$

The pixels in a frame with τ value greater than a predefined threshold are considered to be the foreground pixels. The pixels detected as foreground are set as 1 and all the remaining positions remain 0. This gives the initial foreground mask.

D. Post Processing

After obtaining the initial object mask which contains the foreground objects, morphological operations are performed on the mask, so as to remove noise to avoid false positives. The pixels closely related are joined together to get the final foreground mask.

III. MULTIPLE OBJECT TRACKING

The aim of this method is to track multiple video objects for the time in which they are present. In other words, each target needs to be assigned a unique trajectory for the duration of the video, which matches the target's motion. The trajectories of each target is maintained over time based on the spatio-temporal movement of the objects.

A. The Proposed Tracking Method

The detailed flow diagram of the proposed multiple object tracking is shown in Fig.2. The outputs obtained from the segmentation step and the foreground mask are the inputs to the tracking algorithm. The next state of each object is predicted using a Kalman Filter. Each object is assigned a trajectory and this information is maintained throughout the processing.

When a new object arrives, check whether it is already assigned with an id and track. The objects which appear for the first appearance are assigned a new track. In case of a detection which already has a track assigned, the track is updated. At each state, check whether the detections are assigned to tracks.

For all assigned detections, predict the next states. In case of inter-object occlusion or object occluded by an obstacle, the tracks must be reassigned when the detection is regained. For this purpose, it is essential to check whether the detection is regained. If the detection is regained within a certain time, the tracks are reassigned.

A detection which is not regained is a detection which is lost. The lost and expired tracks are deleted. Based on the spatio-temporal movement of the objects, the trajectories of all tracked objects are maintained and it is used for plotting the path of the object. The method is able to track multiple objects under cluttered dynamic backgrounds.

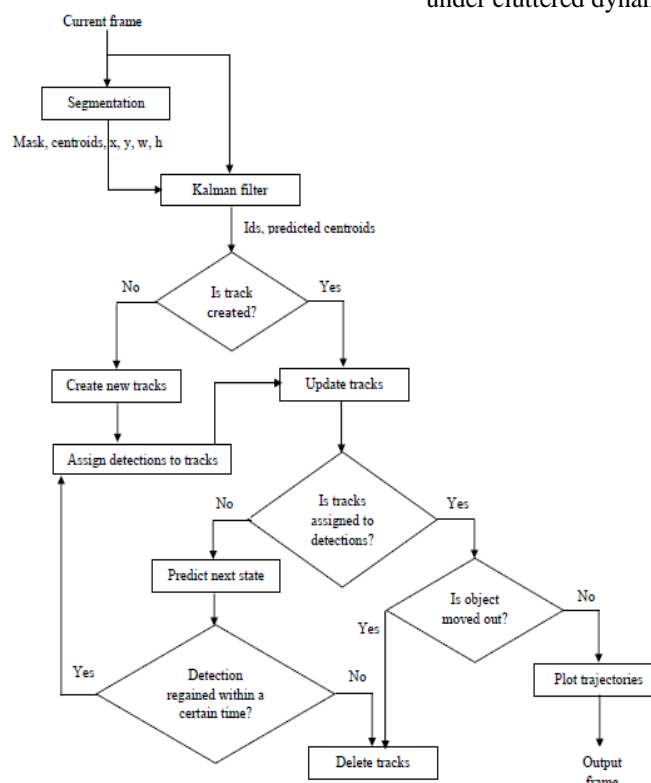


Fig.2. Flow Chart of the Proposed Tracking Method

IV. EXPERIMENTAL RESULTS

The framework is implemented with MATLAB and tested for its robustness and all the results showed that the framework is robust towards occlusions and heavy background clutter. Although this method is not meant to be used in crowded scenes, the performance in non-crowded scenes are very robust.

The test cases for cluttered background includes swaying leaves in the forest, man walking in front of fountain, and people walking in a room with low lighting conditions. For tracking evaluation, the system was tested under occlusions and multiple people walking simultaneously. The tracker showed its efficiency in all the test cases.

A. Segmentation Results

Segmentation step is the basic part for getting an accurate object detection. The results obtained for various test cases are depicted in Table I. There are six test sequences with high amount of background clutter used for testing the robustness

TABLE I. RESULTS OF SEGMENTATION

	Original	Background Likelihood	Foreground Likelihood	Object Mask
(a)				
(b)				
(c)				
(d)				
(e)				
(f)				

(a) PETS_data2 (b) fans (c) fountain (d) stairs (e) PETS_data1 (f) forest

of the segmentation algorithm. The test sequences are (a) PETS_data2 (b) fans (c) fountain (d) stairs (e) PETS_data1 (f) forest. In all the cases, there were no false positives or noise in the resultant foreground object mask, which proves the robustness of object segmentation. The objects are then detected as blobs which in turn become the input of the tracking algorithm.

TABLE II. TRACKING RESULTS FOR VARIOUS DATASETS

Dataset	Original	Tracked Objects	Target Trajectories
stairs			
fans			
leaves			
forest			
corridor			

B. Results of Multiple Object Tracking

Multiple objects are tracked throughout its lifetime in a video under heavy background clutter and occlusions. Table II shows different test cases of object tracking. The trajectories of each object is also maintained and plotted over time, thus an efficient tracking of the object's movement is done. Each object is correctly identified by its unique id wherever it goes.

C. Robustness towards Occlusion

A case of an object occluded by an obstacle is shown in Fig.3. In this case the visibility of the object is zero when it moves under the roof. The states of the objects are predicted for a certain life span so that, if the object reappears, it will be

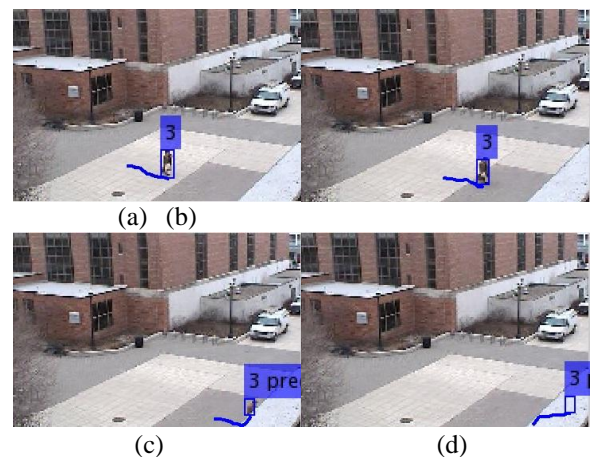


Fig.3. Occlusion by an Obstacle (person went under roof)

reassigned with its id.

Another case of occlusion is inter object occlusion which happens due to object crossover. A test case [11] of object crossover is depicted in Fig.4. The two objects which cross one another maintain their ids without identity switching. (a), (b), (c) and (d) are frames showing the object crossover sequence. The framework is robust towards inter-object occlusion due to object crossover.

A video with enormous background clutter and noise is used as test data in Fig.4. In this video, there are cases of object crossover, presence of multiple objects, target appearance variation, and occlusion due to obstacle. The

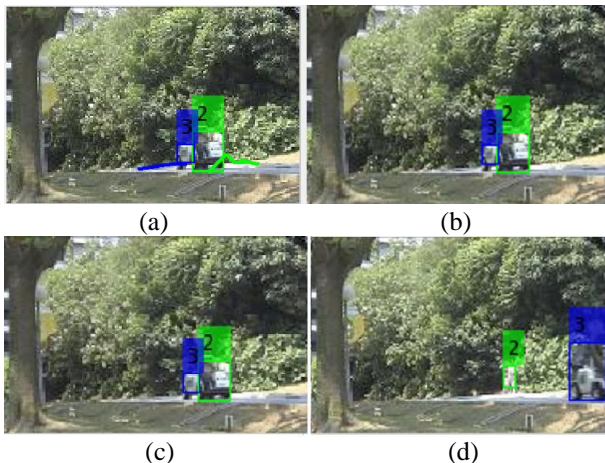


Fig.4. Inter-object occlusion (Campus) [11]

proposed segmentation algorithm provided best results without any false positive so that no background pixel was considered as foreground. The previous methods [2], [3], [4], [5] and [6] fail to correctly segment objects in this video. In contrast, the proposed detection and tracking framework provide efficient performance in all the aforementioned challenges.

V. CONCLUSION AND FUTURE WORK

This paper proposes a multiple object detection and tracking framework which is robust to background clutter and occlusions. It accurately track the objects having appearance variations since the movement of the object is tracked. Several training based techniques fail when the appearance of the target changes, or they require enormous amount of initial training data, which is surely a memory killer. But apart from such methods, the segmentation method used here efficiently tolerates background clutter caused by rotating fans, swaying leaves, fountain etc., and the tracking algorithm robustly track multiple objects of interest. Non-rigid object motion can also be tracked by using this framework. Although, this algorithm is not meant to be used in very crowded scenes, where all objects are identified together, but it is very accurate in non-crowded scenes. To use it with crowded scenes, some crowd

analysis method must also be incorporated and this can be one of the future works. The proposed framework can be installed in places such as indoor, corridor, parking area, outdoor with dynamic background, and any other non-crowded place which requires attention.

REFERENCES

- [1] W.-K. Chan and S.-Y. Chien, "Real-time memory-efficient video object segmentation in dynamic background with multibackground registration technique", IEEE Workshop Multimedia Signal Processing, 2007.
- [2] Shao-Yi Chien, Wei-Kai Chan, Yu-Hsiang Tseng, and Hong-Yuh Chen, "Video Object Segmentation and Tracking Framework with Improved Threshold Decision and Diffusion Distance", IEEE Trans. Circuits and Systems for Video Technology, vol. 23, no. 6, June 2013.
- [3] Benjamin Langmann, Seyed E. Ghobadi, Klaus Hartmann, Otmar Loffeld, "Multi-Modal Background Subtraction Using Gaussian Mixture Models", IAPRS, Vol. XXXVIII, Part 3A – Saint-Mande, France, September 1-3, 2010.
- [4] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," IEEE Trans. Pattern Anal. Machine Intell., vol. 19, no. 7, pp. 780–785, Jul. 1997.
- [5] S.-Y. Chien, S.-Y. Ma, and L.-G. Chen, "Efficient moving object segmentation algorithm using background registration technique," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 7, pp. 577–586, Jul. 2002.
- [6] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," IEEE Trans. Pattern Anal. Machine Intell., vol. 25, no. 10, pp. 1337–1342, Oct. 2003.
- [7] L. Vosters, C. Shan, and T. Gritti, "Background subtraction under sudden illumination changes," in Proc. IEEE Int. Conf. Advanced Video Signal Based Surveillance, Aug. 2010.
- [8] N. Funk, "A study of the Kalman Filter applied to Visual Tracking", Project for CMPT, December 2003.
- [9] Y. Sheikh and M. Shah, "Bayesian Modeling of Dynamic Scenes for Object Detection" in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 11, November 2005.
- [10] IPPR. Dataset of IPPR design contest [Online] Available: <http://archer.ee.nctu.edu.tw/contest/data.htm>
- [11] Campus Sequence [online] Available: http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html
- [12] J.Henriques, R. Caseiro, and J.Batista. Globally optimal solution to multi-object tracking with merged measurements. In ICCV 2011.
- [13] H. Jiang, S. Fels, and J. J. Little. A linear programming approach for multiple object tracking. In CVPR 2007.
- [14] L. Kratz and K. Nishino. Tracking with local spatio-temporal motion patterns in extremely crowded scenes. In CVPR 2010.
- [15] C.-H. Kuo, C. Huang, and R. Nevatia. Multi-target tracking by on-line learned discriminative appearance models. In CVPR 2010.
- [16] B. Leibe, K. Schindler, and L. Van Gool. Coupled detection and trajectory estimation for multi-object tracking. In ICCV 2007.
- [17] Y. Li, C. Huang, and R. Nevatia. Learning to associate: Hybrid boosted multi-target tracker for crowded scene. In CVPR 2009.
- [18] S. Oh, S. Russell, and S. Sastry. Markov chain Monte Carlo data association for multi-target tracking. IEEE Transactions on Automatic Control, 54(3):481–497, 2009.
- [19] W. Brendel, M. Amer, and S. Todorovic. Multiobject tracking as maximum weight independent set. CVPR 2011.
- [20] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. CVPR 2005.