AN OPTIMIZED FRAMEWORK FOR DETECTION AND TRACKING OF VIDEO OBJECTS IN CHALLENGING BACKGROUNDS

Sukanyathara J[1] and Alphonsa Kuriakose[2]

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

ABSTRACT

Segmentation and tracking are two important aspects in visual surveillance systems. Many barriers such as cluttered background, camera movements, and occlusion make the robust detection and tracking a difficult problem, especially in case of multiple moving objects. Object detection in the presence of camera noise and with variable or unfavourable luminance conditions is still an active area of research. This paper propose a framework which can effectively detect the moving objects and track them despite of occlusion and a priori knowledge of objects in the scene. The segmentation step uses a robust threshold decision algorithm which uses a multi-background model. The video object tracking is able to track multiple objects along with their trajectories based on Continuous Energy Minimization. In this work, an effective formulation of multi-target tracking as minimization of a continuous energy is combined with multi-background registration. Apart from the recent approaches, it focus on making use of an energy that corresponds to a more complete representation of the problem, rather than one that is amenable to global optimization. Besides the image evidence, the energy function considers physical constraints, such as target dynamics, mutual exclusion, and track persistence. The proposed tracking framework is able to track multiple objects despite of occlusions under dynamic background conditions.

KEYWORDS

Surveillance, segmentation, multi-background registration, threshold decision, energy minimization, tracking, computer vision.

1. INTRODUCTION

Segmentation and tracking plays an important role in Visual surveillance systems. Video tracking is the process of locating a moving object (or multiple objects) over time using a camera. Video tracking can be a time consuming process due to the amount of data that is contained in video. Adding further to the complexity is the possible need to use object recognition techniques for tracking, a challenging problem in its own right.

Video object segmentation, detection and tracking processes are the basic, starting steps for more complex processes, such as video context analysis and multimedia indexing. Object tracking in videos can be defined as the process of segmenting an object of interest from a sequence of video scenes. This process should keep track of its motion, orientation, occlusion and etc. in order to extract useful context information, which will be used on higher-level processes.

When the camera is fixed and the number of targets is small, objects can easily be tracked using simple methods. Computer vision-based methods often provide the only non-invasive solution. Their applications can be divided into three different groups: Surveillance, control and analysis.

Under various environmental assumptions, several video object segmentation algorithms have been proposed. [6] - [8] proposes several simple and efficient video object segmentation algorithms. However, the proposed algorithms cannot address dynamic backgrounds because only one background layer is employed in their background model. Some algorithms are complex and require large amount of memory. Vosters *et al*. [9] proposed a more complex algorithm, consisting of an Eigen background and statistical illumination model, which can address sudden changes of illumination, but it has very high computational requirement.

To enable the long-term tracking, there are a number of problems which need to be addressed. The key problem is the detection of the object when it reappears in the camera's field of view. This problem is aggravated by the fact that the object may change its appearance thus making the appearance from the initial frame irrelevant.

Tracking algorithms estimate the object motion. Trackers require only initialization, are fast and produce smooth trajectories. On the other hand, they accumulate error during run-time (drift) and typically fail if the object disappears from the camera view. Research in tracking aims at developing increasingly robust trackers that track "longer". The post-failure behavior is not directly addressed. Detection based algorithms estimate the object location in every frame independently. Detectors do not drift and do not fail if the object disappears from the camera view. However, they require an offline training stage and therefore cannot be applied to unknown objects.

This paper intends to:

1. Propose a new method which combines multi-background registration based object detection to detect objects under dynamic backgrounds and tracking based on continuous energy minimization.

2. To obtain better results despite of occlusions in complex backgrounds.

The rest of the paper is organized as follows: The proposed system model is explained in section 3, 4, 5 and 6. Section 7 contains the experimental results. In Section 8, conclusion of the work is given.

2. PROPOSED SYSTEM MODEL

In order to solve the problem of detection and tracking in cluttered backgrounds, a robust method which makes use of a Multi-background registration based object detection and Energy minimization based tracking is proposed in this paper. It is an enhanced method over the previous ones, and it is able to detect the area of interest in dynamic background tracks multiple moving objects along with their trajectories. Separate trajectories are assigned to the objects and those trajectories are not destroyed even if the object undergoes inter-object occlusion.

The segmentation method is memory efficient and it is able to detect objects under background clutter. The entire process consists of three major parts namely, Multi-background registration based segmentation, Threshold decision and Multiple-object tracking.

For detecting the moving objects, a background model is found out using multi-background registration and the foreground objects are detected using the built background model. The background model uses multiple background images which suits it for using in dynamic

backgrounds. Apart the other methods, this is able to track multiple objects under dynamic backgrounds along with their trajectories.

The proposed system is an enhancement over the stationary background and single object tracking systems and include three main components:

1. An efficient threshold determination for segmentation.

2. Object detection.

3. Tracking multiple objects based on Continuous Energy Minimization.

3. THRESHOLD DECISION

To better deal with dynamic background conditions, an efficient threshold decision is inevitable. This paper makes use of Gaussianity test and Noise level estimation for efficient threshold decision.
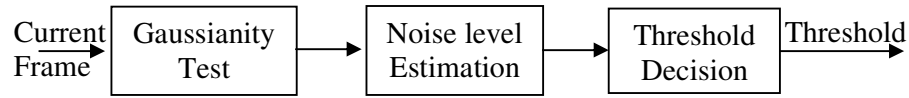

Fig. 1. Threshold decision

The Gaussianity test is applied to each block to determine if the minimal background differences in the block are Gaussian distributed or not. The camera noise is assumed to be Gaussian distributed.

3.1. Gaussianity Test

Divide the frame into a number of non-overlapping blocks of size $M_b * N_b$. Apply Gaussianity test to each block to determine if the minimal background differences in the block are Gaussian distributed or not. The Gaussianity test can be shown as the following equations:

$$I_{r=}\frac{1}{M_b*N_b}\sum_{m=1}^{M_b}\sum_{n=1}^{N_b}[BDmin(m,n)]^r \qquad (1)$$

$$H(I_1, I_2, I_3, I_4) = I_3 + I_4 - 3I_1(I_2 - I_1^2) - 3I_2^2 - I_1^3 - 2I_1^4$$

1) Gaussian: $|H(I_1, I_2, I_3, I_4)| < G_{th}$

2) Non-Gaussian: $|H(I_1, I_2, I_3, I_4) \geq G_{th}$

Where the smaller the H value, the closer the distribution of BD_min is to the Gaussian distribution, and G_th is the threshold value for binarizing the decision. If the minimal background differences in a block are Gaussian distributed, the block belongs to the background region because the (minimal) difference between the current frame and the background images is only caused by noise.

3.2. Noise Level Estimation and Threshold Decision

The optimal threshold $BDth^k$ is found out using the following equation:

$$BDth=max\{|BDmin(i,j)||(i,j) \in Background\ blocks \qquad (2)$$

The background blocks are indicated by the Gaussianity test described in the previous section. Note that $BDth$ and $BDmin(i,j)$ are all random variables.

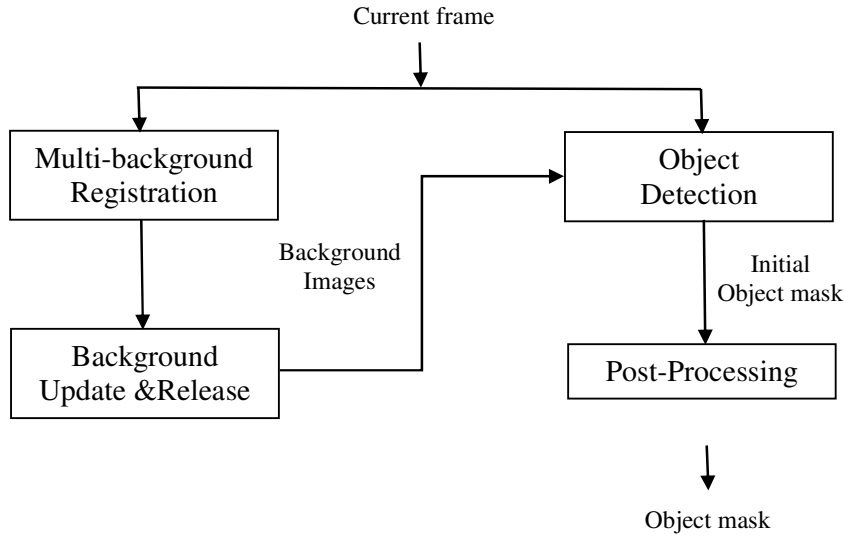## 4. VIDEO OBJECT SEGMENTATION WITH MULTI-BACKGROUND REGISTRATION

Fig. 2. Multi-background registration

The segmentation method is based on an online multilayer background modeling technique known as Multi-background registration (MBReg). The key concept in this algorithm is the fact that it models the background with N layers of background images instead of a single background layer. For each pixel position, the corresponding pixel in each layer of the background image represents one possible background pixel value.

As shown in Fig. 2. , the background model is established and maintained in the MBReg and background update and release blocks. In the MBReg block, each input pixel of the current frame $CurFrm(i,j,t)$, where (i,j) is the pixel position and t is the time index, is compared with the corresponding background pixels in the multi-background image $BImg(i,j,t-1,k)$, where $k \in [1,N]$, and a matching flag, $match(i,j,k)$, is recorded by the following equation:

$$match(i,j,t,k) = \begin{cases} 1 & ;if\ BD(i,j,t,k) \leq BDth(i,j,t,k) \\ 0 & ;\ otherwise \end{cases} \qquad (3)$$

The background difference $BD(i,j,t,k)$ can be calculated using the equation:

$$BD(i,j,t,k) = |CurFrm(i,j,t) - BImg(i,j,t-1,k)| \qquad (4)$$

In the background update and release block an 'unmatched background' counter $CntSno(i,j,k)$ and a weighting coefficient $Wgt(i,j,t,k)$ are maintained to record the duration when a background pixel is unmatched to the input pixel and the confidence of each background pixel where BDF is the background decaying factor.

$$CntSno(i,j,t,k) = \begin{cases} 0 ; & if\ match(i,j,t,k) = 1 \\ CntSno(i,t-1,k) + 1; & otherwise \end{cases} \quad (5)$$

$$Wgt(i,j,t,k) = \begin{cases} Wgt(i,j,t-1,k) + 1 ; & if\ match(i,j,t,k) = 1 \\ Wgt(i,j,t-1,k) - 1; & if\ CntSno(i,j,t,k) > BDF(i,j,t,k) \\ Wgt(i,j,t-1,k) & ; \ otherwise \end{cases} \quad (6)$$

Using the unmatched background counter and weighting coefficient, the background model can be updated or released with the following equations:

$$BImg(i,j,t,k) = \begin{cases} UpdBckgnd(i,j,t,k) & ; if\ match(i,j,t,k) = 1 \\ 0(release) & ; if\ k \in Built\ background\ layers \\ & and\ Wgt(i,j,t,k) < RELth(i,j,t,k) \\ 0(release) & ; Wgt(i,j,t,k) = 0 \end{cases} \quad (7)$$

4.1. Object Detection

Object detection is done using the background layers and the current frame. When each frame is encountered, it is compared with the background layers and the pixels which belong to the foreground are set as 1, while the pixels identified as background are set as 0. This results in the initial object mask. Using post processing methods we eliminate the remaining false detections if any.

The foreground pixels in a frame is found out using the built background model as follows:

$$Ini\_Object\_Mask(i,j) = \begin{cases} 0 \ \exists k\ , Built(i,j,k) = 1 \\ \qquad matched \\ 1 \qquad , otherwise \end{cases} \quad (8)$$

Morphological opening, closing, or removal of spur pixels can be done on the initial object mask to obtain the final object mask, which highlight the objects of interest.

5. TRACKING MULTIPLE OBJECTS

Tracking of multiple objects is seen as a function of continuous energy minimization here. Other than a number of recent approaches, it focus on designing an energy function that represents the problem as faithfully as possible. It uses a suitable optimization scheme to find strong local minima of the proposed energy. The scheme extends the conjugate gradient method with periodic trans-dimensional jumps. These moves allow the search to escape weak minima and explore a much larger portion of the variable-dimensional search space, while still always reducing the energy.

The aim of this method is to find an optimal solution for multi-target tracking over an entire video sequence. In other words, each target needs to be assigned a unique trajectory for the duration of

the video, which matches the target's motion as closely as possible. To this end, a global energy function is defined which depends on all targets at all frames within a temporal window, and thus represents the existence, motion and interaction of all objects of interest in the scene. Tracking is performed in world coordinates, i.e. the image evidence is projected onto the ground plane. Additionally, the evidence is weighted with a height prior to reduce false detections.

The state vector X consists of ground plane coordinates of all targets at all times. The (x, y) location of target i at frame t is denoted $x_i^t F$ and N indicate the total number of frames and targets respectively. In this formulation the position of each target is always defined and considered when computing the energy, even in case of occlusion.

5.1. Energy Function

The energy function is made up of five terms: an observation term based on image data; three physically motivated priors for object dynamics, collision avoidance and object persistence; and a regularizer which tries to keep the number of trajectories low:

$$E(x) = E_{obs} + \propto E_{dyn} + \beta E_{exc} + \gamma E_{per} + \delta E_{reg} \quad (8)$$

Where $\propto, \beta, \gamma$ and δ are labels used for minimizing the energy function so as to get the optimized trajectory locations.

5.1.1. Observation Model

This makes use of the object detection step. Here, pedestrians are detected and interpreted as a kind of "intelligent smoothing", which takes into account the other energy terms rather than blindly smooth the nodes of the trajectory curve.

It does however go beyond smoothing, for example it helps to prevent identity switches between crossing targets (since it favors straight paths). The energy should be minimum when the detection of an object precisely match the tracked location.

5.1.2. Dynamic Model

It uses a constant velocity model:

$$E_{dyn}(x) = \sum_{t=1}^{F-2} \sum_i^N \left\| v_i^t - v_i^{t+1} \right\|^2 \quad (9)$$

Where $v_i^t = x_i^t = x_i^{t+1} - x_i^t$ is the current velocity vector of target i.

Dynamic model can be interpreted as a kind of "intelligent smoothing", which takes into account the other energy terms rather than blindly smooth the nodes of the trajectory curve.

5.1.3. Mutual Exclusion

The most obvious physical constraint is that two objects cannot occupy the same space simultaneously. An occluded object has the possibility to lose its track when the detection becomes unavailable for a certain period. In case of occlusion, either inter-object occlusion or occlusion due to an obstacle, the property of mutual exclusion provides both the objects a unique track. This constraint is included to the energy function by defining a continuous exclusion term where s_g is the scale factor.

$$E_{exc}(x) = \sum_{t=1}^{F} \sum_{i \neq j} \frac{s_g}{\left\| x_i^t - x_j^t \right\|^2} \tag{10}$$

5.1.4. Target Persistence

Another constraint one would in most cases like to integrate into the energy function is the fact that targets cannot appear or disappear within the tracking area (but nevertheless can enter or leave the area). However, only a soft constraint is imposed, since otherwise one would have to explicitly model entry/exit locations (e.g. doors) and long term occlusion. Hence the sigmoid penalty:

$$E_{per}(x) = \sum_{t=1}^{N} \sum_{t \in \{1,F\}} \frac{1}{1 + \exp{(1 = q.b(x_i^t))}} \tag{11}$$

where $b(x_i^t)$ and $b(x_j^t)$ are distances of the start, respectively end points of trajectory i to the border of the frame.

5.1.5. Regularization

The regularization drives the minimization towards a simpler explanation of the data, i.e. a model with fewer targets and longer trajectories:

$$E_{reg}(x) = N + \sum_{t=1}^{N} \frac{1}{F(i)} \tag{12}$$

where $F(i)$ is the temporal length of trajectory i in frames. The regularization balances the model's complexity against its fitting error, and discourages over-fitting, fragmentation of trajectories, and spurious identity changes.

Similar to any non-convex optimization, the result depends on the initial value from which the iteration is started. Empirically, even a trivial initialization with no targets work reasonably well, however it will take many iterations to converge.

5.1.6. Continuous Energy Minimization

The optimum values of tracked locations is obtained when the energy term has the minimum values. After minimization, the trajectories obtained are smooth in nature, in contrast with the sharp edged trajectories obtained before optimization.

Algorithm: Energy Minimization

Input: S initial solutions, D detections

Output: Best of $<=$ S solutions

      for $s = 1 \to S$ **do**

            while $\neg\, converged$ **do**

                  for $m \in \{grow, shrink, add, remove, merge, split\}$

                  do

                        for $i \in 1, 2, \dots, N$ **do**

                        try jump move m on trajectory i

                        (greedy parameter selection)

$$\text{if } E^{new}(X_s) < E^{old}(X_s) \text{ then}$$

perform jump move m

end if

end for

end for

end while

end for

Return: $arg\ min_{X_s} E(X_s)$

The algorithm for energy minimization, which is based on [11] is able to find the most accurate trajectory locations for each target. The trajectories can grow, shrink, add, remove, merge and split to precisely match the detection. Initial solutions are given as input to the energy minimization algorithm, which in turn provides the optimum solutions which maintains the energy as minimum.

The initialization uses the output of an arbitrary simpler tracker as a more qualified initial value. For initialization, per-target extended Kalman filters (EKFs) is used, where the data association is performed in a greedy manner using a maximum overlap criterion, to quickly generate a variety of starting values. The system can keep track of all the objects along with their trajectories.

6. EXPERIMENTAL RESULTS

The system is implemented in Matlab which ease the manipulation of the video frames. The system was tested with different video sequences with different frame rates. Although the framework is not meant to detect objects in crowded scenes, the system efficiently detect and track the moving objects under complex cases of background clutter such as fountains, camera noise, swaying leaves, rotating fans etc.

The objects vary in shape while moving, but the segmentation based detection is very robust in this case. The energy minimization algorithm performs well in cases of occlusion by its properties such as mutual exclusion, target persistence, etc. The false positives are eliminated using the regularizer, which prevents the number of tracks from increasing abruptly. The framework is robust to background clutter conditions and occlusions such as inter object occlusion or the object hidden by any obstacle.

6.1. Segmentation Results

Performance of the proposed segmentation method have been tested with various datasets and object detection provided robust results towards background clutter. The various challenging sequences used for testing include video with fountain as background, video taken inside a forest with swaying leaves all over, campus, etc. with size varying targets. The results of the segmentation method are showed in Fig. 3. for different test cases. The foreground object mask of the corresponding frame of the video is shown for 6 different videos in the experimental evaluation.
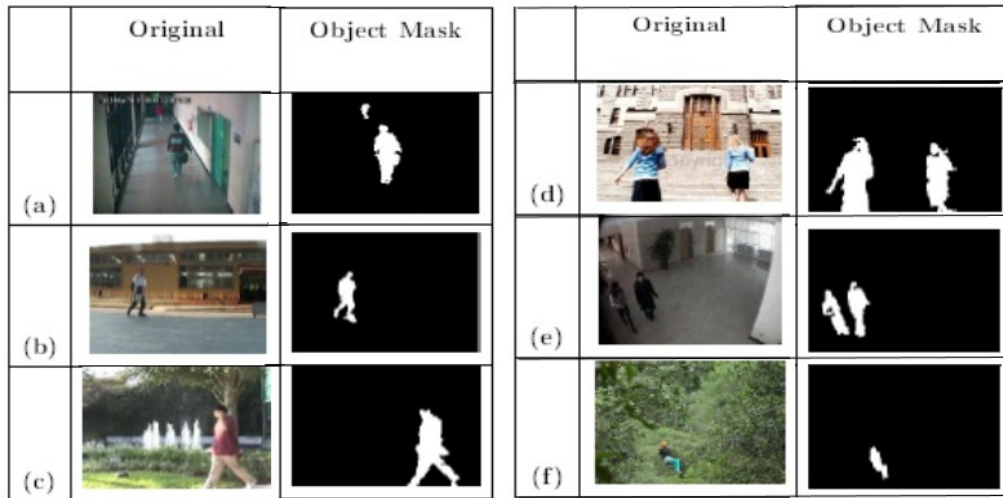
Fig. 3. Segmentation Results (a) IPPR_data2 [12] (b) fans (c) fountain (d) stairs
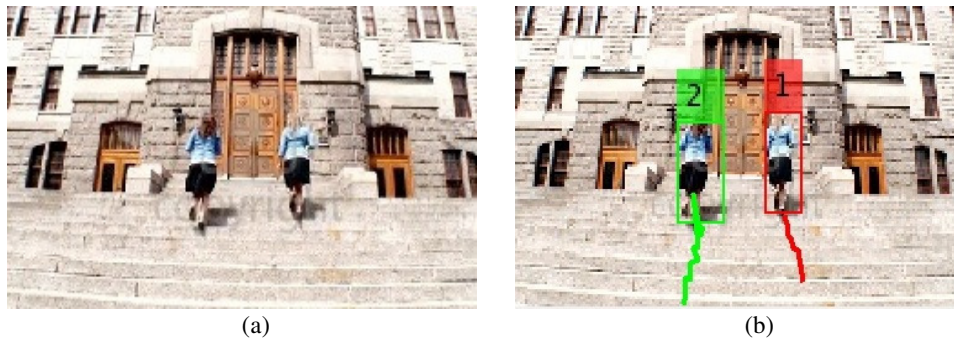
(e) IPPR_data1 [12] (f) forest

6.2. Tracking Results

(a)   (b)

Fig. 4. Tracking on *Stairs* data (a) Input frame (b) Tracking result

(a)   (b)

Fig. 5. Tracking on *IPPR_data2* [12] (a) Input frame (b) Tracking result

Fig. 6. *Campus* [13] Test case with object crossover (inter-object occlusion)

Fig. 7. Test case of object occluded by a roof (occlusion due to obstacle)

The robustness towards background clutter and occlusion were tested for various challenging video datasets and all the objects in the videos were correctly tracked by the framework. Fig.6. shows a video dataset with huge background clutter caused by swaying leaves and distortions. The case of object crossover is depicted in the results. It correctly tracks the objects without identity switching. Fig.7. illustrates a case when occlusion is handled even though the object moves and enters below a roof. The execution time per frame is less than 0.5 sec during the process.

7. CONCLUSION AND FUTURE WORK

In this paper we have proposed a novel method for detection and tracking of objects in complex backgrounds. Compared to the previous methods proposed for detection and tracking such as particle filter and extended Kalman filter, the method based on detection in complex backgrounds and multiple object tracking with continuous energy minimization robustly and efficiently detect and track multiple objects in complex environments for video surveillance. Moreover the method is robust to occlusions and it tracks without a priori knowledge of the number of targets, which is a difficult problem to tackle. Tracking is considered as a function of energy minimization which is more suitable for real world applications. The proposed detection method also overcomes the limitations of frame differencing based methods of segmentation. For crowded scenes, techniques for crowd analysis have to be integrated with the framework.

REFERENCES

[1]   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.

[2]   Shao-Yi Chien, Shyh-Yih Ma, and Liang-Gee Chen," Efficient Moving Object Segmentation Algorithm Using Background Registration Technique", IEEE Trans. Circuits And Systems For Video Technology, Vol. 12, No. 7, July 2002.

[3]   W.-K. Chan and S.-Y. Chien," Real-time memory-efficient video object segmentation in dynamic background with multi-background registration technique", IEEE Workshop Multimedia Signal Processing, 2007, pp. 219–22.

[4]   A. Cavallaro, O. Steiger, and T. Ebrahimi, "Tracking video objects in cluttered background, IEEE IEEE Trans. Circuits Syst. Video Technology, vol. 15, no. 4, Apr. 2005.

[5]   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.

[6]   C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Realtime tracking of the human body," IEEE Trans. Pattern Anal. Machine Intell., vol. 19, no. 7, pp. 780–785, Jul. 1997.

[7]   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.

[8]   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.

[9]   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.

[10] Anton Milan, Konrad Schindler, Stefan Roth, "Detection- and Trajectory-Level Exclusion in Multiple Object Tracking", in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Portland, June 2013.

[11] Anton Milan, Konrad Schindler, Stefan Roth, "Continuous Energy Minimization for Multi-Target Tracking", in Proc. of the IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, No. 1, Jan. 2014

[12] IPPR. Dataset of IPPR design contest [Online].
Available: http://archer.ee.nctu.edu.tw/contest/data.htm

[13] Campus Sequence [online] Available:http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html

[14] M. N. Gurcan, Y. Yardimci, and A. E. Cetin, "Influence function based Gaussianity tests for detection of micro calcifications in mammogram images," in Proc. Int. Conf. Image Process., Oct. 1999, pp. 407–411.

[15] B. Benfold and I. Reid. Stable multi-target tracking in real-time surveillance video. In CVPR 2011.

[16] J. Berclaz, F. Fleuret, and P. Fua. Robust people tracking with global trajectory optimization. In CVPR 2006.

[17] M. Enzweiler, A. Eigenstetter, B. Schiele, and D. M. Gavrila. Multicue pedestrian classification with partial occlusion handling. In CVPR 2010.

[18] A. Ess, B. Leibe, K. Schindler, and L. van Gool. Robust multiperson tracking from a mobile platform. IEEE T. Pattern Anal. Mach. Intell., 31(10):1831–1846, Oct. 2009.

[19] J. Ferryman and A. Ellis. PETS2010: Dataset and challenge. In Advanced Video and Signal Based Surveillance (AVSS), 2010.

[20] J. M. Ferryman and A. Shahrokni. PETS2009: Dataset and challenge. In Winter-PETS, Dec. 2009.

[21] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Multi-target tracking using joint probabilistic data association. In IEEE Conf. on Decision and Control, volume 19, pages 807–812, Dec. 1980.

[22] W. Ge and R. T. Collins. Multi-target data association by tracklets with unsupervised parameter estimation. In BMVC 2008.

[23] H. Grabner, J. Matas, L. Van Gool, and P. C. Cattin. Tracking the invisible: Learning where the object might be. In CVPR 2010.

[24] P. Green. Reversible jump markov chain monte carlo computation and bayesian model determination.Biometrika, 82(4):711–732, 1995.

[25] J.Henriques, R. Caseiro, and J.Batista. Globally optimal solution to multi-object tracking with merged measurements. In ICCV 2011.

[26] H. Jiang, S. Fels, and J. J. Little. A linear programming approach for multiple object tracking. In CVPR 2007.

[27] L. Kratz and K. Nishino. Tracking with local spatio-temporal motion patterns in extremely crowded scenes. In CVPR 2010.

[28] C.-H. Kuo, C. Huang, and R. Nevatia. Multi-target tracking by on-line learned discriminative appearance models. In CVPR 2010.

[29] B. Leibe, K. Schindler, and L. Van Gool. Coupled detection and trajectory estimation for multi-object tracking. In ICCV 2007.

[30] Y. Li, C. Huang, and R. Nevatia. Learning to associate: Hybrid boosted multi-target tracker for crowded scene. In CVPR 2009.

[31] 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.

AUTHORS

Sukanyathara J received her B Tech degree in Computer Science & Engineering from M G University, Kerala, India in 2012. She is currently working towards the MTech degree in Computer Science and Engineering from MG University, Kerala, India. Her research interests include Digital image processing, Video Object Segmentation, Multi-target Tracking and Object Detection in real-world complex background scenarios.

 Alphonsa Kuriakose received her M Tech degree in Computer Science & Engineering from M G University, Kerala, India in 2012. She is currently working as Assistant Professor in Computer Science and Engineering at Viswajyothi College of Engineering and Technology, Kerala, India. Her research interests include Computer Networks, Information security, and Digital Image Processing.