Indonesian Journal of Electrical Engineering and Computer Science
Vol. 2, No. 1, April 2016, pp. 151 ~ 160
DOI: 10.11591/ijeecs.v2.i1.pp151-160

Background Modeling to Detect Foreground Objects Based on ANN and Spatio-Temporal Analysis

N Satish Kumar*[1], Shobha G[2]
CSE Department, R V College of Engineering, Bangalore, Karnataka, India
*Corresponding author, e-mail: satish.rmgm@gmail.com[1], shobhag@rvce.edu.in[2]

Abstract

This paper presented an approach to building background model for moving object detection using unsupervised artificial neural network (ANN) without any prior knowledge about foreground objects. First, using local binary pattern (LBP) which is texture feature, builds a statistical Background Model using ANN, then, comparing the behavior of next incoming frame with model and decide each pixel whether is deviating from a model or not. And based on if method detects foreground objects then background model is updated to make this model adaptive. Also, spatial-temporal information has been exploited in this method to suppress sudden illumination variation and to suppress false foreground pixels. It was demonstrated and proved, by qualitative and quantitative metrics that the newly presented approach is adaptive, generic and can address all issues and challenges for background subtraction. To evaluate the performance of the presented approach this paper compared with recent approaches by using standard metrics and proved that presented method outperforms many existing recent approaches.

Keywords: *Local Binary Pattern, Illumination variation, spatial-temporal, ANN, HCI, Background Subtraction*

1. Introduction

The ability to extract moving foreground objects from a complex video sequence is first and foremost step of many computer vision problems [1, 2], traffic monitoring [3], human detection and tracking or human-machine interface (HCI) [4, 5, 6] video summarization, among other applications. Background Subtraction is nothing but discriminating moving objects from static in a given video sequence.

There may be many different algorithms has been used for many years in many computer vision applications, example object detection and tracking, the target recognition, human tracking etc., [7, 8]. Even though results of the existing background subtraction algorithms are fairly good, however, many of these algorithms are vulnerable to both global and local illumination changes such as shadows and highlights. These problems cause many computer vision applications to fail.

As an instance if optical flow based Background Subtraction [9, 10] is analyzed, this is computationally expensive and not suitable for real-time scenarios. Therefore, there is the need for an algorithm which should be computationally affordable for real-time scenarios and generic. Apart from these requirements there exist important problems in background subtraction those are: sensitivity to dynamic background changes, as countermeasure model, has to adapt via background maintenance or updating. Following are some of the familiar issues and challenges in background maintenance:

- ***Light changes:*** The model should adapt to gradual & sudden illumination changes.
- ***Moving background:*** The model should not detect changing background which is not *of interest for visual surveillance, such as waving trees, rippling the water.*
- ***Cast shadows:*** The model should detect and suppress moving cast shadows.
- ***Bootstrapping:*** The model should be accurately initialized even in the absence of static (free of moving objects) training set at the time of building model.
- ***Camouflage:*** Moving objects should be properly detected even if there is a chromatic similarity to those of the background model.

It is very much desirable to have accurate and efficient results of non-stationary and foreground objects detection in video sequence without shadows and illumination effects. These

Received January 21, 2016; Revised March 21, 2016; Accepted March 31, 2016

problems are the underlying motivation of this work described in this paper. A lot of research has been done in background subtraction field, but still there is a requirement to develop a robust, efficient and generic background subtraction algorithm which is able to also detect and suppress shadows from all kinds of complex videos. Even shadow detection is very useful to many applications such as Shape from Shadow problems [11]. The described method in this paper must also address problems such as sensitivity, reliability, robustness, shadow detection and speed of foreground object detection. In this paper, we present a generic adaptive background subtraction algorithm with shadow suppression for detecting moving objects from all types of complex background videos.

Here is the organization of the paper. Section II describes overview of existing approaches of background subtraction. Section III explains background model methodologies and shadow suppression. In Section IV, we reported results achieved with experimentation of the proposed approach in terms of accuracy and efficiency, comparing them with several other existing popular methods. Section V concludes and sheds lights on further research directions in background subtraction and shadow detection.

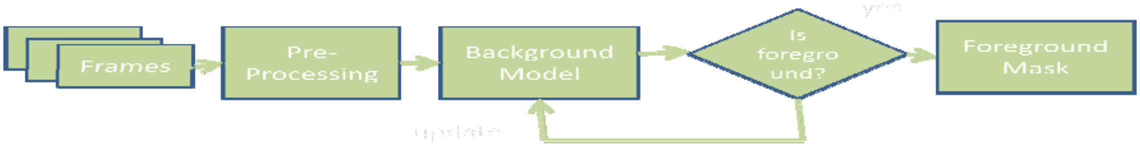


Figure. 1. Background subtraction flow diagram

2. Literature Review

There are many traditional approaches to moving object detection which includes optical flow [10], temporal differencing [12], and background subtraction [13]. Temporal differencing works by taking differences in consecutive video frames, which is easy to distinguish static objects from moving foreground objects. This approach will incorporate adaptive nature to dynamic environments and can solve sudden illumination variation. But it is subject to the foreground aperture problem. Optical flow techniques aim at computing an approximation of the 2D motion field from the spatio-temporal information of image pixel values. Even though they can detect moving objects in the presence of camera motion, most optical flow computation methods are computationally expensive, and cannot be applied in real-time videos.

Obviously Background subtraction is the common and efficient method of detecting moving foreground objects from the stationary camera (e.g. [13]). It works based on the differencing of current frame sequence with reference background model without any prior knowledge about how many objects, velocities of moving objects and should not have foreground aperture problem. But optical flow is very sensitive to illuminations variations due to various reasons. Even though these are detected, they leave behind holes, where newly entered objects differ from background model. There will be a false alarm rate for a short period of time.

Apart from above said state-of-art methods, further sections will give insight into many recent algorithms in order to succeed in detecting and extracting moving foreground objects. There are many pixel-based algorithms such as ViBe [14] and PBAS [15] but they differ from traditional approaches as they consider on random pixel sampling and label diffusion. Even though pixel-based methods are simple, lightweight, and effective they are not considered the spatial relationship of pixels and are not object-based. Spatial-based methods on the other hand, attempt to harness this information using features or block descriptors [16, 17] and local color histograms [18] in order to achieve better and efficient results.

There are typical cases where this concept is useful, for example, foreground occlusions with pixel intensities which are equal to that of background and global illumination variations. In order to overcome these issues temporal and spatiotemporal-based methods have been proposed by many researchers which take into account temporally recurring behavior between the background model and the previous, current and upcoming frames of the video. Such useful information can play a very significant role in analyzing and detecting moving foreground objects without any noise and accuracy. This information is also useful in estimation of short-term intensity changes (dynamic backgrounds e.g. rippling water, swaying trees, etc.), sudden illumination changes, and moving cast shadow detection. Such a solution is presented using bidirectional temporal analysis [19].

There have been proposing many approaches to merge the concepts of different models that rely on multiple algorithm techniques simultaneously including post-processing, advanced morphological operations. However with these combinations are very successful at improving the performance of their respective algorithms, but often they suffer from increasing computational expenses, time, and some required prior training phase which is practically infeasible for real-time applications. Heikkila and M Pietikainen [20] proposed background subtraction methodology by exploiting local binary pattern (LBP) as a feature, which is proved to be reliable and efficient texture based method. Since then, many researchers have proposed alternatives: Yoshinaga [21] presented a methodology by integration of both spatial-based and pixel-based approaches using Mixture of Gaussians (MoGs). On the other hand, Zhang et al. proposed an approach by combining spatial texture and temporal motion analysis using weighted LBP histograms. Object Tracking using Camshift and MoG [22, 23] was proposed to track moving object at real-time.

However there are many background subtraction algorithms proved to be very efficient and reliable, there is no single generic algorithm which can solve all the issues and challenges mentioned in the previous section.

This paper proposed an approach to building background model for moving object detection and shadow suppression which is based on unsupervised simple Artificial Neural Network (ANN) without any prior knowledge about foreground and shadow. The idea consisting of adopting biologically inspired ANN to model the background, comparing the behavior of next incoming frame with model and deciding per pixel whether is deviating from a model or not. And also the proposed method makes use of spatial information of the background model to detect foreground objects. It was demonstrated and proved, by qualitative and quantitative metrics, that the new proposed method is adaptive, generic can address all above said issues and challenges for background subtraction.

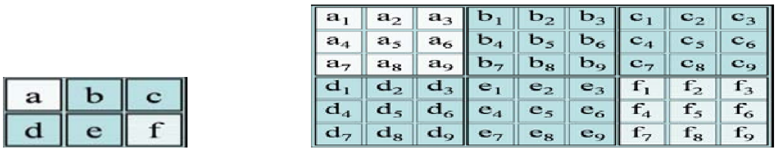
3. Proposed Approach

This presented a new approach of building a background model based on LBP (texture) feature and ANN, inspired by Kohonen [24]. This paper presented a method which employed a simple 2-D flat grid of nodes to build a background model. Each node j (output neuron) has weight vector Wj. It builds a neuronal map for each pixel which consists of nine weight vectors. Features at each pixel have clustered into the set of weight vectors based on Euclidian distance. The LBP feature vectors are presented to all the neurons as inputs. Then, for each input vector, the neuron c has selected with minimum Euclidian distance. Foreground moving object detection carried out by checking the difference between current frame and background model by Euclidian distance. If incoming pixel exhibits same behavior that of the model, then it is termed as background, otherwise as a foreign foreground pixel. Background Model is updated if any pixel is classified as background. Background Model building, foreground detection using Euclidian distance and updating the Model is given in following sections.

3.1 Background Model

The background model is built using the first frame sequence from the video; that is, each of the nine weight vector is assigned with corresponding LBP operator of a pixel of the first frame sequence. In this approach, to represents weight vector, LBP feature of a pixel has been chosen, which is very robust and invariant to illumination and color.

The set of weight vectors of an image I with N rows and M columns is represented as 2-D flat grid of neurons A with 3xN and 3xM dimensions, and weight vectors for pixel (x, y) are at position (i, j), i=3x,…, 3x+2 and j=3y,…, 3y+2. Example 2-D flat grid of neuronal map is demonstrated for an image I with 2 rows and 3 columns in figure 1.



a	b	c
d	e	f

a_1	a_2	a_3	b_1	b_2	b_3	c_1	c_2	c_3
a_4	a_5	a_6	b_4	b_5	b_6	c_4	c_5	c_6
a_7	a_8	a_9	b_7	b_8	b_9	c_7	c_8	c_9
d_1	d_2	d_3	e_1	e_2	e_3	f_1	f_2	f_3
d_4	d_5	d_6	e_4	e_5	e_6	f_4	f_5	f_6
d_7	d_8	d_9	e_7	e_8	e_9	f_7	f_8	f_9

Figure 2. (a) A simple image (b) the neuronal map structure

Figure (a), is image with 2 rows and 3 columns, (b) represents weight vectors (a1, …, a9) store in 3x3 matrix of neuronal map A, and similarly, (f1, …, f9) weight vectors with respect to pixel f in image (a) and so on.

3.2 Feature Extraction

This paper employed LBP texture features to build background model, which is very efficient and considers the neighborhood of each pixel and converted into a binary number. Due to its simplicity, LBP operator has become a very popular feature for many computer vision applications. The most important property that made LBP feature to select for building background model is its computational simplicity and its robustness to illumination variation. This section describes how to calculate LBP features from an image. This is demonstrated in figure 2. To extract LBP, a circular neighborhood denoted by (S, R) is considered, where S represents sampling points and R is the radius of the neighborhood.

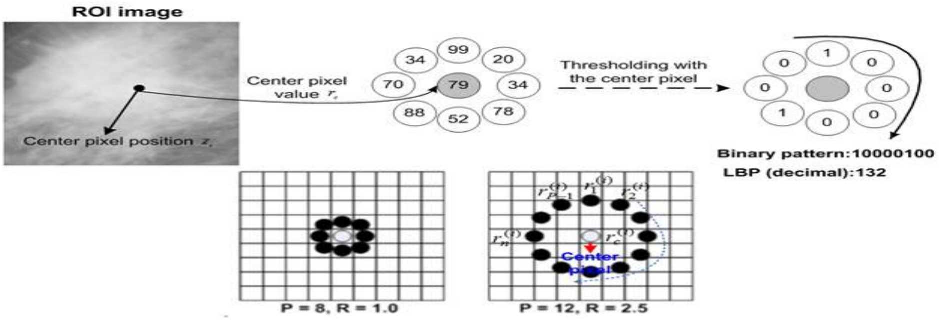


Figure 3. An example of LBP computation

The points around the pixel (x, y) located at coordinate points (x_p, y_p) is given by:

$$(x_p, y_p) = (x + R\cos(2\pi p/P), y - R\sin(2\pi p/P)) \tag{1}$$

With the equation (1) if a sampling point does not yield at integer coordinates, the value is interpolated. The LBP label for the center pixel (x, y) of image f(x, y) is obtained through the equation (2).

$$LBP_{P,R}(x,y) = \sum_{p=0}^{P-1} s\left(f(x,y) - f(x_{p,}y_p)\right) 2^p \tag{2}$$

Where s(z) is the threshold function and is given by,

$$\mathrm{s(z)} = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \tag{3}$$

3.3 Finding the Best Match

Given the current pixel p at time t, the value I_t(p) is compared to the current pixel model, given by M_{t-1}(p), to determine the weight vector BM(p) that best matches it:

$$d(BM(p), I_t(p) = \min_{i,j=0,\dots,n-1} d(m_{t-1}^{i,j}(p), I_t)) \tag{4}$$

Where d(:,:) is the Euclidian distance between two vectors between background model and incoming frame pixel.

3.4 Foreground Detection and Updating the Model

By subtracting the current image from the background model, each pixel p_t of the t^{th} sequence frame, I_t is compared to the current pixel weight vectors to determine if there exists a weight vector that best matches it. The best matching weight vector is used as the pixel's encoding approximation, and therefore p_t is detected as foreground if no acceptable matching weight vector exists; otherwise it is classified as background.

$$B_t(q) = (1 - \alpha_t B_{t-1}(q) + \alpha_t I_t(p) \qquad \forall q \in N_{\overline{p}} \tag{5}$$

Where $N_{\overline{p}}$ represents the 2-D neighborhood with width 2k+1 ∈N

The spatial analysis is introduced using K-Nearest Neighbor (KNN) Search algorithm in Background Model, and the background subtraction mask Dt(p) is computed as:

$$D_t(p) = \begin{cases} 1 & if\ KNN(BG, I_t) \leq n/2 \\ 0 & otherwise \end{cases} \tag{6}$$

Where KNN is result of KNN search on finding best match in the background model, n is the number of best matches found in Background Model.

4. Experimentation and Result Analysis

Experimental results for Background Model and foreground detection have been performed for all types of challenging video sequences. 5 different types of videos with (moving background, illumination variation, water surface, camera jitter) with frame rate 15 fps and 320x240 resolutions are considered. The selected parameters for experimentation is as follows: The number of model chosen is 3x3, that is each pixel has to be repeated 9 times, distance threshold is 1.0 for training and 0.008 for testing phase, learning rate fixed to 1 for training and 0.5 for testing.

4.1 Performance Measure Method

To evaluate the performance of the proposed method with state-of-the art methods three measures were used: Recall, Precision, and F-measure. Those metrics definition are given as following.

Table 1. A contingency table

	Correct Foreground	Correct Background
Classified as Foreground	True Positives (TP)	False Positives (FP)
Classified as Background	False Negatives (FN)	True Negatives (TN)

$$Recall = \frac{TP}{TP + FN} \quad if\ TP + FN > 0\ , otherwise\ undefined \tag{9}$$

$$Precision = \frac{TP}{TP + TN} \quad if\ TP + FN > 0, otherwise\ undefined \tag{10}$$

In order to obtain high Recall, actually Precision has to be scarified and vice versa, so there is a trade-off between Recall and Precision. To avoid these misleading, the paper used F-measure [25] as another very important performance metric which considers both Recall and Precision results simultaneously. The F-measure expression is given below as:

$$F_1(r, p) = \frac{2pr}{p + r} \tag{11}$$

where r: Recall; p: Precision

4.2 Experimental Results

In order to evaluate the performance of the proposed methodology, five video sequences from the Li dataset have to be used. The results from the proposed method was compared with those from MoGv2 [25], GMG [26], and Texture BGS [20]. The background subtraction results obtained from proposed method and other state-of-art methods were demonstrated in further sections.

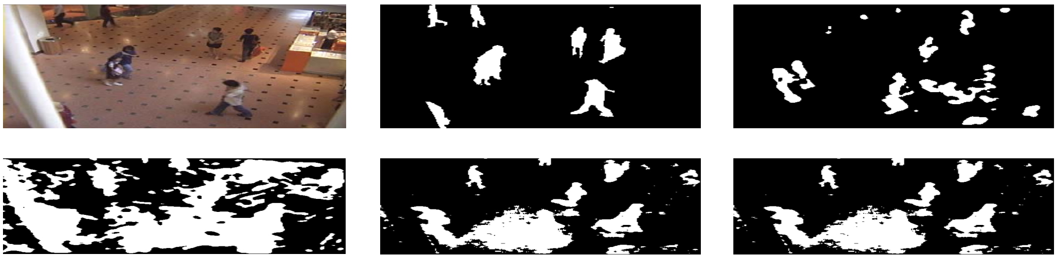


Figure 5. Shopping Mall Video: (a) current frame; (b) Ground Truth foreground Mask; (c) MoGv2; (d) GMG; (e) Texture BGS; (f) Proposed method

Here moving Escalator (SS) video sequence has been considered. This sequence is an example for complex background and same is demonstrated in figure 6.

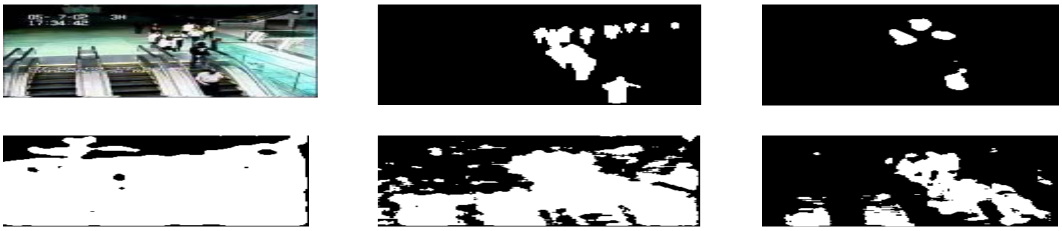


Figure 6. Escalator Video: (a) current frame; (b) Ground Truth foreground Mask; (c) MoGv2; (d) GMG; (e) Texture BGS; (f) Proposed method

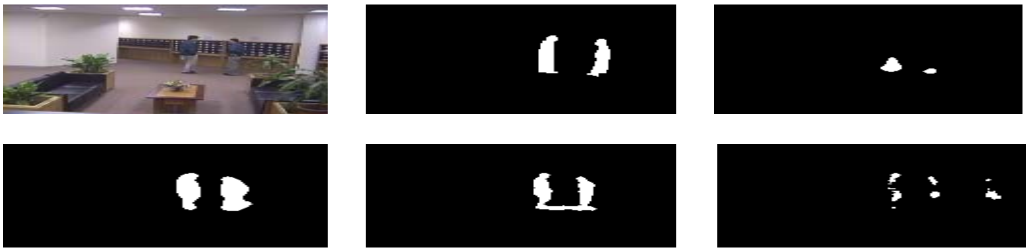


Figure 7. Indoor Light Switch Video: (a) current frame; (b) Ground Truth foreground Mask; (c) MoGv2; (d) GMG; (e) Texture BGS; (f) Proposed method

Lobby (LB) video sequence has been considered, which is an indoor environment; contains 1,545 frames. This demonstrated sudden illumination variation in an indoor environment as shown in figure 7. To compare the proposed method with the recent popular background subtraction algorithms, paper used parameters given in those reference papers or by repeating the experiments. All the Background Subtraction methods including proposed, MoGv2, GMG, and Texture BGS algorithms are implemented in MATLAB. In this paper ground truth foreground mask frames are employed provided by Li dataset. The recall values acquired by applying all the methods including proposed method are demonstrated in figure 8. The recall values obtained by applying proposed method outperforms all the existing methods and same is shown in figure 8.

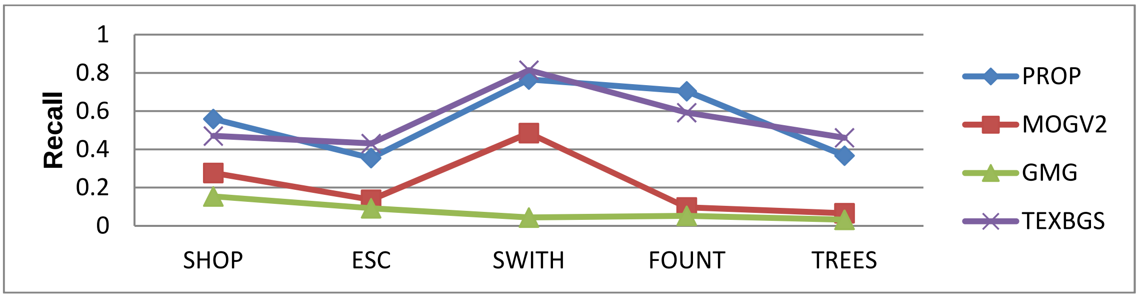


Figure 8. Recall results for proposed and other methods for all video sequences of Li dataset

The precision values obtained by proposed and those comparisons with other methods are demonstrated in figure 9. The precision values of proposed method are better than all the existing methods and same is shown in figure 9.

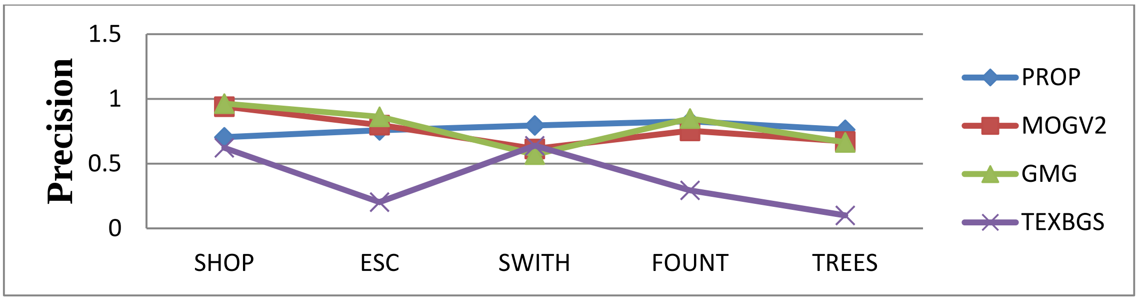


Figure 9. Precision results for proposed and other methods for all video sequences of Li dataset

The F1 measure values for all video sequences of Li dataset have been shown in figure 11. The F1 result is a very good performance metric which takes both recall and precision simultaneously. F1 values are more for the proposed when compared with other methods.

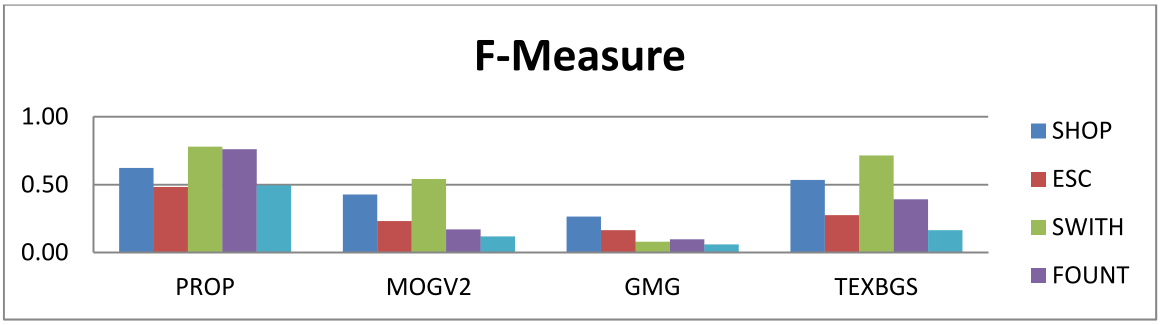


Figure 10. F-Measure results for proposed and other methods for all video sequences of Li dataset

Figure 11 shows the average precision, recall and F-measure for all video sequence of Li dataset. The proposed method shows good average values of all performance metrics for all videos.

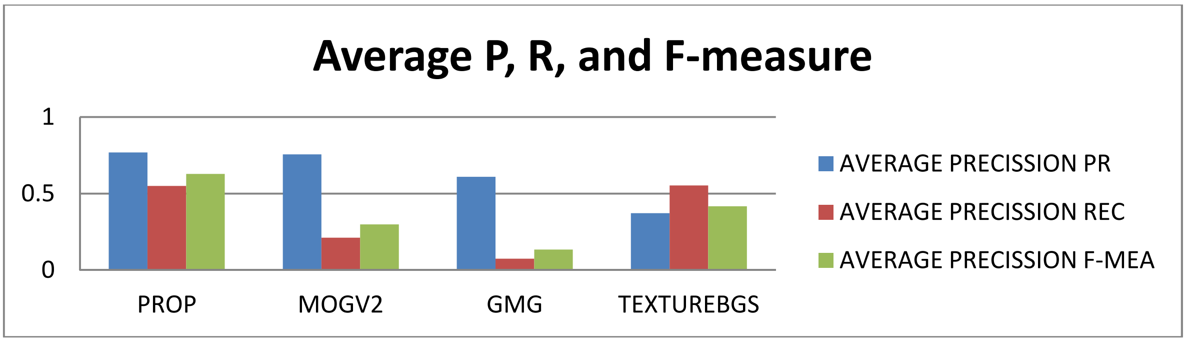


Figure 11. Average Precision, Recall and F-Measure for all video sequences of Li dataset

There is another performance metric called, percentage of wrong classification (PWC) [27]. This gives the percentage of pixels classified as foreground but those are not actual foreground pixels. This is also called percentage of false alarms.

PWC (% of wrong classification): 100(FN+FP)/ (TP+FN+FP+TN)* (12)

Figure 12 shows PWC of the proposed and all other methods, also figure demonstrates that PWC values for all video sequences are less compared with other methods. That means proposed method results in fewer false alarms when compared with other methods.

5. Conclusion

Very less research has been done for proposing generic background subtraction algorithm which can solve almost all the issues and challenges of foreground detection problem. The proposed work is a contribution to the new background subtraction method using ANN, spatio-temporal information. This algorithm also makes use of gradient information whenever necessary in order to compensate sudden illumination variation for indoor environment. Performance of the proposed method with all existing is reported. Analysis shows that proposed method is very robust and outperforms many existing algorithms for all types of video sequences. Proposed method can be applied for any type of video sequences and method is generic and achieves good results for many challenging video sequences.

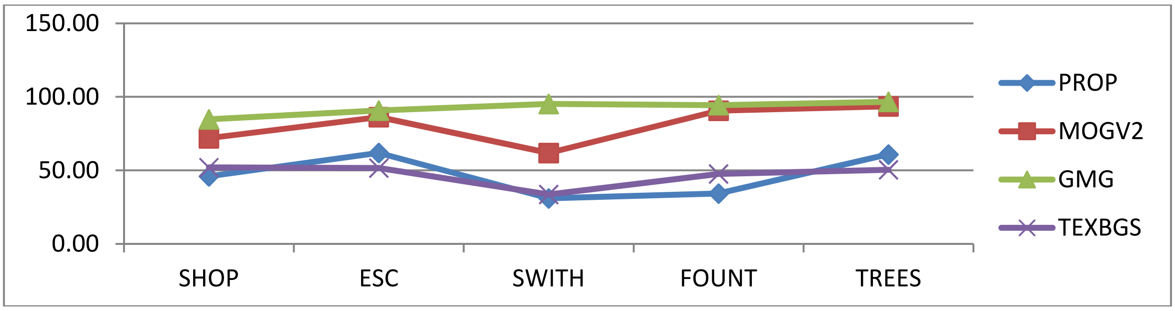


Figure 12. Percentage of Wrong Classification for all video sequences of Li dataset

References

[1] Ismail Haritaoglu, David Harwood, and Larry S Davis. "*Real-time System for Detecting and Tracking People*". Proc. the third IEEE International Conference on Automatic Face and Gesture Recognition (Nara, Japan), IEEE Computer Society Press, Los Alamitos, Calif. 1998: 222-227.

[2] D Butler, V Bove, and S Shridharan. "Real time adaptive background/foreground segmentation". *EURASIP Journal on Applied Signal Processing*. 2005; 14: 2292–2304.

[3] N Friedman and S Russell. "*Image Segmentation in Video Sequences: A Probabilistic Approach",* Proceedings of the 13th Conference on Uncertainty in Artificial intelligence, Morgan Kaufmann. 1997.

[4] CR Wren, A Azarbayejani, T Darrell, and A Pentland. "Pfinder: Real-time Tracking of the Human Body". *IEEE Trans. on Pattern Analysis and Machine Intelligence.* IEEE Computer Society Press, Los Alamitos, Calif. 1997; 19(7): 780-785.

[5] J Ohya, et al. "Virtual Metamorphosis". *IEEE Multimedia. DOI: 10.1109/93.771371*. 1999; 6(2): 29 – 39.

[6] J Davis, and A Bobick. "*The representation and Recognition of Action using Temporal Templates*". Proceedings of Conference on Computer Vision and Pattern Recognition. 1997.

[7] A Utsumi, H Mori, J Ohya, and M Yachida. "*Multiple-human tracking using multiple cameras*". In Proc. the thrid IEEE International Conf. Automatic Face and Gesture Recognition (Nara, Japan). IEEE Computer Society Press, Los Alamitos, Calif. 1998.

[8] M Yamada, K Ebihara, and J Ohya. "*A new robust real-time method for extracting human silhouettes from color images*". In Proc. the third IEEE International Conf. Automatic Face and Gesture Recognition (Nara, Japan), IEEE Computer Society Press, Los Alamitos, Calif. 1998: 528–533.

[9] T Horprasert, I Haritaoglu, C Wren, D Harwood, LS Davis, and A Pentland. "*Real-time 3d motion capture*". In Proc. 1998 Workshop on Perceptual User Interface (PUI'98), San Francisco. 1998

[10] Daniel D Doyle, Alan L Jennings, Jonathan T Black "Optical flow background estimation for real-time pan/tilt camera object tracking". *Journal of the International Measurement Confederation*. 2014; 48: 195–207.

[11] Xue Yuan, Xiaoli Hao, Houjin Chen, Xueye Wei. Background Modeling Method based on 3D Shape Reconstruction Technology. *TELKOMNIKA e-ISSN: 2087-278X*. 2013; 11(4): 2079~2083

[12] Álvaro Bayona, Juan C San Miguel, José M. Martínez. "*Stationary Foreground Detection Using Background Subtraction and Temporal Difference In Video Surveillance". Proceedings of 2010 IEEE 17th International Conference on Image Processing.* Hong Kong. 2010.

[13] C Stauffer and E Grimson. Adaptive background mixture models for real-time tracking. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. 1999: 246–252.

[14] O Barnich and M Van Droogenbroeck. "ViBe: a powerful random technique to estimate the background in video sequences." *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009*. 2009: 945–948.

[15] M Hofmann, PTiefenbacher, G Rigoll. "Background Segmentation with Feedback: The Pixel-Based Adaptive Segmenter", in proc of IEEE Workshop on Change Detection. 2012.

[16] Haritaoglu I, Harwood D, David LS. W4: "Real-time Surveillance of People and their Activities." *IEEE Trans. on PAMI*. 2000; 22(8): 809–830.

[17] Pierre-Marc Jodoin, Max Mignotte, and Janusz Konrad. "Statistical Background Subtraction Using Spatial Cues". *IEEE Transactions on Circuits and Systems for Video Technology*. 2007; 17(12).

[18] Shengping Zhang, Hongxun Yao, Shaohui Liu. "Dynamic Background Subtraction Based on Local Dependency Histogram". *The Eighth International Workshop on Visual Surveillance - VS2008*, Marseille, France. 2008.

[19] Atsushi Shimada, Hajime Nagahara, Rin-ichiro Taniguchi, "Background Modeling based on Bidirectional Analysis". *CVPR*. 2013.

[20] Heikkila M, Pietikainen M. "A texture-based method for modeling the background and detecting moving objects". *IEEE Transactionson Pattern Analysis and Machine Intelligence*. 2004.

[21] S Yoshinaga, A Shimada, H Nagahara and R Taniguchi. "Background model based on intensity change similarity among pixels". *In Frontiers of Computer Vision, (FCV), 2013 19th Korea-Japan Joint Workshop*, 2013: 276–280.

[22] Li Zhu, Tao Hu. Research of CamShift Algorithm to Track Motion Objects. *TELKOMNIKA, e-ISSN: 2087-278X*. 2013; 11(8): 4372~4378.

[23] Hong-xun, Zhang and De Xu. "Fusing color and gradient features for background model". *In 8th International Conference on Signal Processing*. 2006.

[24] Davide Ballabio, Viviana Consonni, Roberto Todeschini. "The Kohonen and CP-ANN toolbox: A collection of MATLAB modules for Self Organizing Maps and Counterpropagation Artificial Neural Networks. *Chemometrics and Intelligent Laboratory Systems*. 2009; 98(2): 115–122.

[25] Andrew B Godbehere, Akihiro Matsukawa, Ken Goldberg. “Visual Tracking of Human Visitors under Variable-Lighting Conditions for a Responsive Audio Art Installation”. *IEEE American Control Conference Fairmont Queen Elizabeth, Montréal*, Canada. 2012.

[26] F El Baf, T Bouwmans, and B Vachon. “Type-2 fuzzy mixture of Gaussians model: Application to background modeling”. *International Symposium on Visual Computing, ISVC*. 2008: 772–781.

[27] Agung Nugroho Jati, Ledya Novamizanti, Mirsa Bayu Prasetyo, Andy Ruhendy Putra. Evaluation of Moving Object Detection Methods based on General Purpose Single Board Computer. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2015; 14(1): 123 ~ 129.