Motion Object Detection in a Video Sequence Using Background Subtraction

Penchala Pranay[1], Adigoppula Sreeramulu[2]

[1]Jyothishmathi Institute Technology & Science, Nustulapur, Karimnagar, A.P, India.

[2]Jyothishmathi Institute Technology & Science, Nustulapur, Karimnagar, A.P, India

Abstract: ***In this letter, we propose an intensity range based object detection scheme for videos with fixed background and static cameras. The scheme suggests two different algorithms; the first one models the background from initial few frames and the second algorithm extracts the objects based on local thresholding. The strength of the scheme lies in its simplicity and the fact that, it de- fines an intensity range for each pixel location in the background to accommodate illumination variation as well as motion in the back- ground. The efficacy of the scheme is shown through comparative analysis with competitive methods. Both visual as well as quantitative measures show an improved performance and the scheme has a strong potential for applications in real time surveillance.***

Keywords: Background modeling, background subtraction, video segmentation, video surveillance.

1. Introduction

Object detection and tracking in video is a challenging problem and has been extensively investigated in the past two decades. It has applications in numerous fields, such as video compression, video surveillance, human-computer inter- action, video indexing and retrieval etc. Object detection and object tracking are two closely related processes. The former involves locating object in the frames of a video sequence, while the latter represents the process of monitoring the object's spatial and temporal changes in each frame. Object detection can be performed through various approaches, such as region-based image segmentation, background subtraction, temporal differencing, active contour models, and generalized Hough trans- forms. In surveillance system video sequences are generally obtained through static cameras and fixed background. A popular approach called background subtraction is used in this scenario, where moving objects in a scene can be obtained by comparing each frame of the video with a background [1]. In most of the suggested schemes, the object detected is accompanied with misclassified foreground objects due to illumination variation or motion in the background. Moreover, shadows are falsely detected as foreground objects during object extraction. Presently,An additional step is carried out to remove these misclassified objects and shadows for effective object detection. To alleviate this problem, we propose a simple but efficient object detection technique, which is invariant to change in illumination and motion in the background.

The proposed approach also neutralizes the presence of shadows in detected objects.The suggested background model initially determines the nature of each pixel as stationary or non-stationary and considers only the stationary pixels for background model formation. In the background model, for each pixel location a range of values are defined. Subsequently, in object extraction phase our scheme employs a local threshold, unlike the use of global threshold in conventional schemes. The rest of the letter is organized as follows: Section II de- scribes some of the related works. In Section III the proposed algorithms are presented. Simulation results are discussed in Section IV. Finally, Section V deals with the concluding remarks.

2. Related Work

For object detection in surveillance system, background modeling plays a vital role. Wren *et al.* have proposed to model the background independently at each pixel location which is based on computation of Gaussian probability density function (pdf) on the previous pixel values [2]. Stauffer and Grimson developed a complex procedure to accommodate permanent changes in the background scene [3]. Here each pixel is modeled separately by a mixture of three to five Gaussians. The W4 model presented by Haritaoglu *et al.* is a simple and effective method [4]. It uses three values to represent each pixel in the background image namely, the minimum intensity, the maximum intensity, and the maximum intensity difference between consecutive frames of the training sequence. Jacques *et al.* brought a small improvement to the W4 model together with the incorporation of a technique for shadow detection and removal [5]. McHugh et al. proposed an adaptive thresholding technique by means of two statistical models [6]. One of them is nonparametric background model and the other one is foreground model based on spatial information.

In ViBe, each pixel in the background can take values from its preceding frames in same location or its neighbor [7]. Then it compares this set to the current pixel value in order to deter- mine whether that pixel belongs to the background, and adapts the model by choosing randomly which value to substitute from the background model. Kim and Kim introduced a novel background subtraction algorithm for dynamic texture scenes [8]. The scheme adopts a clustering-based feature, called fuzzy color histogram (FCH), which has an ability of greatly attenuating color variations generated by background motions while highlighting moving objects. Instead of segmenting a frame pixel-by-pixel, Reddy *et al.* used an overlapping block-by-block approach for detection of foreground objects [9]. The scheme passes the texture information of each block through three cascading classifiers to classify them as background or fore- ground. The results are then integrated with a probabilistic voting scheme at pixel level for the final segmentation.

Generally, shadow removal algorithms are employed after object detection. Salvador et al. developed a three step hypothesis based procedure to segment the shadows [10]. It assumes that shadow reduces the intensities followed by a complex hypothesis using the geometrical properties of shadows. Finally it confirms the validity of the previous assumption. Choi et al. in their work of [11] have distinguished shadows from moving objects by cascading three estimators, which use the properties of chromaticity, brightness, and local intensity ratio. A novel method for shadow removal using Markov random fields (MRF) is proposed by Liu et al. in [12], where shadow model is constructed in a hierarchical manner. At the pixel level, Gaussian mixture model (GMM) is used, whereas at the global level statistical features of the shadow are utilized.

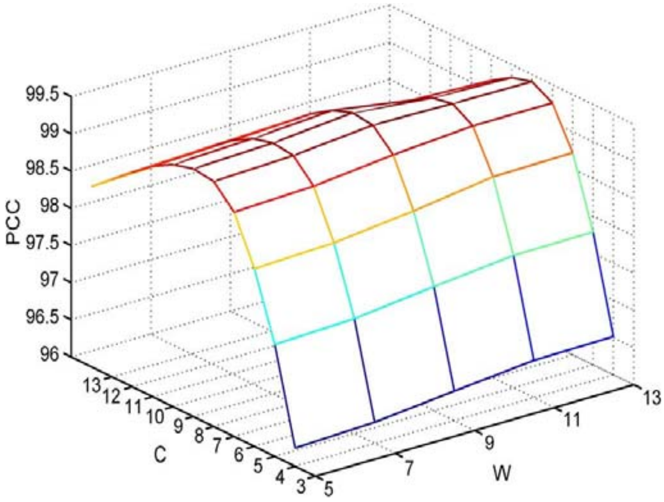
From the existing literature, it is observed that most of the simple schemes are ineffective on videos with illumination variations, motion in background, and dynamically textured indoor and outdoor environment etc. On the other hand, such videos are well handled by complex schemes with higher computational cost. Furthermore, to remove misclassified foreground objects and shadows, additional computation is also performed. Keeping this in view, we suggest here a simple scheme called Local Illumination based Background Subtraction (LIBS) that models the background by defining an intensity range for each pixel location in the scene. Subsequently, a local thresholding approach for object extraction is used. Simulation has been car- ried out on standard videos and comparative analysis has been performed with competitive schemes.

3. The Proposed Libs Scheme

The LIBS scheme consists of two stages. The first stage deals with finding the stationary pixels in the frames required for background modeling, followed by defining the intensity range from those pixels. In the second stage a local threshold based background subtraction method tries to find the objects by comparing the frames with the established background. LIBS uses two parameters namely, window size (an odd length window) and a constant for its computation. The optimal values are selected experimentally. Both stages of LIBS scheme are described as follows.

A. Development of Background Model

Conventionally, the first frame or a combination of first few frames is considered as the background model. However, this model is susceptible to illumination variation, dynamic objects in the background, and also to small changes in the background like waving of leaves etc. A number of solutions to such problems are reported, where the background model is frequently updated at higher computational cost and thereby making them unsuitable for real time deployment. Further, these solutions do not distinguish between object and shadow. To alleviate these limitations we propose an intensity range based background.



model. Here the RGB frame sequences of a video are converted to gray level frames. Initially, few frames are considered for background modeling and pixels in these frames are classified as stationary or non-stationary by analyzing their deviations from the mean. The background is then modeled taking all the stationary pixels into account. Background model thus developed, defines a range of values for each background pixel location. The steps of the proposed background modeling are presented in *Algorithm* 1.

Algorithm 1 Development of Background Model

Consider n initial frames as $\{f_1, f_2, \cdots, f_n\}$, where $20 \leq n \leq 30$.
for $k \leftarrow 1$ to $n - (W - 1)$ **do**
  for $i \leftarrow 1$ to height of frame **do**
    for $j \leftarrow 1$ to width of frame **do**
      $\overrightarrow{V} \leftarrow [f_k(i,j), f_{k+1}(i,j), \dots, f_{k+(W-1)}(i,j)]$
      $\sigma \leftarrow$ standard deviation of $\overrightarrow{V}$
      $\mathbf{D}(p) \leftarrow |V(k + (\lfloor W \div 2 \rfloor)) - V(p)|$, for each value of $p = k + l$, where $l = 0, \cdots, (W - 1)$ and $l \neq \lfloor W \div 2 \rfloor$
      $\mathbf{S} \leftarrow$ sum of lowest $\lfloor W \div 2 \rfloor$ values in $\overrightarrow{\mathbf{D}}$
      if $\mathbf{S} \leq \lfloor W \div 2 \rfloor \times \sigma$ **then**
        Label $f_{k+(\lfloor W \div 2 \rfloor)}(i,j)$ as stationary
      else
        Label $f_{k+(\lfloor W \div 2 \rfloor)}(i,j)$ as non-stationary
      end if
    end for
  end for
end for
for $i \leftarrow 1$ to height of frame **do**
  for $j \leftarrow 1$ to width of frame **do**
    $M(i,j) = \min[f_s(i,j)]$ and $N(i,j) = \max[f_s(i,j)]$, where $s = \lceil W \div 2 \rceil, \cdots, n - (\lfloor W \div 2 \rfloor)$ and $f_s(i,j)$ is stationary
  end for
end for

B. Extraction of Foreground Object

After successfully developing the background model, a local thresholding based background subtraction is used to find the foreground objects. A constant C is considered that helps in computing the local lower threshold T_L and the local upper threshold T_U These local thresholds help in successful de- tection of objects suppressing shadows if any. The steps of the algorithm are outlined in *Algorithm* 2.

Algorithm 2 Background Subtraction for a frame f

```
for i ← 1 to height of frame do
  for j ← 1 to width of frame do
    Threshold T(i, j) = (1/C)(M(i, j) + N(i, j))
    T_L(i, j) = M(i, j) − T(i, j)
    T_U(i, j) = N(i, j) + T(i, j)
    if T_L(i, j) ≤ f(i, j) ≤ T_U(i, j) then
      S_f(i, j) = 0 //Background pixel
    else
      S_f(i, j) = 1 //Foreground pixel
    end if
  end for
end for
```

4. Simulation Results and Discussions

To show the efficacy of the proposed LIBS scheme, simulation has been carried out on different recorded video sequences namely, In the video the Persons who are entering into their respective rooms with suitcases in this the video completely visible to us that coming from one room and going to another room only two persons are moving objects in this scenario. In this we have to take first frame considered as background image and remaining all frames are foreground images by subtracting Foreground image from background image will get resultant image assigning threshold values to that resultant image like (<0 means white, 1<50 means white) if the resultant image between these values by extracting the image will get the resultant image in the form of black and white image. If there is no moving object then it will show black image if there is moving object then it will show us white image by this we can detect the moving object by applying background subtraction method. Considering the characteristics of selected video sequences, they are the most suitable representatives for validation of generalized behavior of the proposed scheme.

Tabel 1: Comparative Analisys of PCC

Method	Time of Day	PETS2001	Intelligent Room	Campus	Fountain	Lobby
GMM	97.72	96.89	95.88	97.13	96.93	97.42
EGMM	98.19	98.56	96.25	97.96	98.12	98.36
Reddy *et al.*	98.93	99.33	99.08	99.34	98.83	99.39
LIBS	99.26	99.20	99.38	99.13	99.46	99.47

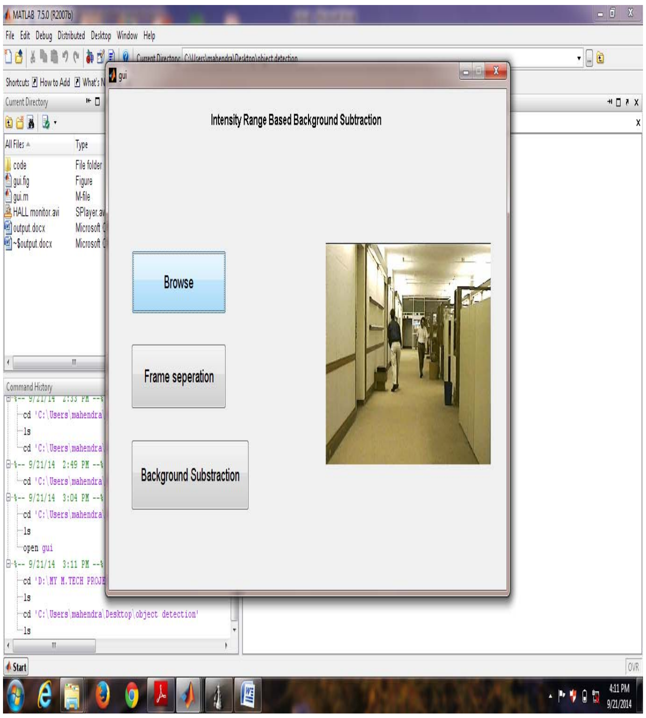


Figure: 1 Input & play the video stream

For comparative PCC analysis, the above video sequences are simulated with the proposed LIBS scheme and three other existing schemes namely, Gaussian mixture model (GMM) [13], expected Gaussian mixture model (EGMM) [14], and model of Reddy *et al.* [9]. It may be observed that for and , the achieved maximum of 99.47%. Similar observations arealso found for other video sequences. The objects detected in different sequences are depicted. It may be observed that, LIBS accurately detects objects in almost all cases with least misclassified objects. Moreover, shadows in object detection performance of LIBS scheme is superior to GMM and EGMM schemes; however it has similar performance with Reddy *et al.*'s scheme. But, LIBS scheme is computationally efficient compared to Reddy *et al.*'s scheme as the latter uses three cascading classifiers followed by a probabilistic voting scheme. The obtained in each case is listed in Table I. The Higher accuracy of is achieved due to the intensity range defined for each background pixel around its true intensity The increase and decrease in the intensity level of the background pixels due to illumination variation is handled by upper and lower part of the predefined intensity range respectively.

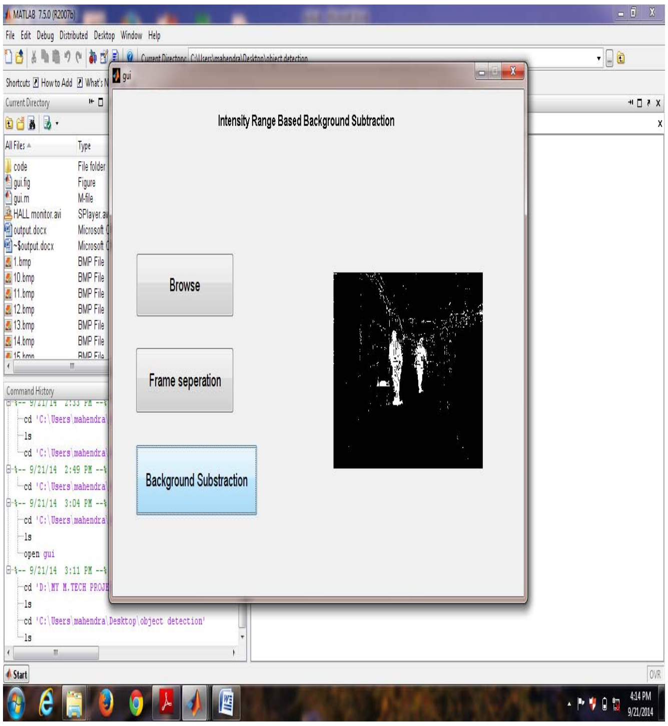


Figure: 2 Output Background Subtraction video

Such PCC increase or decrease in intensity may be caused by switching on or off of additional light sources, movement of clouds in the sky etc. Moreover, shadow having low intensity value when its intensity falls on any surface, decreases by some factor. There- fore, LIBS has an advantage of removing the shadows if any, at the time of detecting the objects. It may be noted that LIBS scheme is devoid of any assumptions regarding the frame rate, color space, and scene content.

5. Conclusion

In this work we have proposed a simple but robust scheme of background modeling and local threshold based object detection. Videos with variant illumination background, textured background, and low motion background are considered for simulation to test the generalized behavior of the scheme. Recent schemes are compared with the proposed scheme, both qualitatively and quantitatively. In general, it is observed that the suggested scheme outperfo rms others and detects objects in all possible scenarios considered. Background subtraction (BGS) is a widely used real time method for identifying foreground objects in a video stream. It is the first significant step in many computer vision applications, including human computer interaction, traffic monitoring, and video surveillance. This has prompted the development of a wide range of different BGS algorithms, along with a number of post processing techniques that aim to improve their performance. The most common paradigm for performing BGS is to build an

explicit model of the background. Foreground objects are then detected by calculating the difference between the current frame and this background model. Numerous computer vision applications depend on BGS to identify foreground objects. We have performed a comparative evaluation of BGS algorithms. This evaluation indicates that how the background is modeled does Influence performance. However, it also reveals that simple Modeling techniques.

6. Future Scope

As computers improve and parallel architectures are investigated, this algorithm can be run faster, on larger images, and using a larger number of Gaussians in the mixture model. All of these factors will increase performance. A full covariance matrix would further improve performance. Adding prediction to each Gaussian(e.g. the Kalman filter approach), may also lead to more robust tracking of lighting changes.Feature-based tracking algorithms perform recognition and tracking of objects by extracting elements, clustering them into higher level features and then matching the features between images

References

[1] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” ACM Comput. Surv., vol. 38, Dec. 2006.

[2] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, “Pfinder: Real- time tracking of the human body,” IEEE Trans. Patt. Anal. Mach. In- tell., vol. 19, no. 7, pp. 780–785, Jul. 1997.

[3] C. Stauffer and W. Grimson, “Adaptive background mixture models for real-time tracking,” in IEEE Comput. Soc. Conf. CVPR, 1999, pp.246–252.

[4] I. Haritaoglu, D. Harwood, and L. Davis, “W4: Real-time surveillance of people and their activities,” IEEE Trans. Patt. Anal. Mach. Intell., vol. 22, no. 8, pp. 809–830, Aug. 2000.

[5] J. Jacques, C. Jung, and S. Musse, “Background subtraction and shadow detection in grayscale video sequences,” in Eighteenth Brazilian Symp. Computer Graphics and Image Processing, Oct.2005, pp. 189–196.

[6] J. McHugh, J. Konrad, V. Saligrama, and P. Jodoin, “Foreground-adap- tive background subtraction,” IEEE Signal Process.Letters, vol. 16, no.5, pp. 390–393, May 2009.

[7] O. Barnich and M. Van Droogenbroeck, “ViBe: A universal back- ground subtraction algorithm for video sequences,” IEEE Trans. Image Process., vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[8] W. Kim and C. Kim, “Background subtraction for dynamic texture scenes using fuzzy color histograms,” IEEE Signal Process. Lett., vol.19, no. 3, pp. 127–130, Mar. 2012.

[9] V. Reddy, C. Sanderson, and B. Lovell, “Improved foreground detec- tion via block-based classifier cascade with probabilistic decision inte- gration,” IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 1, pp.83–93, Jan. 2013.

[10] E. Salvador, A. Cavallaro, and T. Ebrahimi, “Cast shadow segmenta- tion using invariant color features,” Comput. Vis. Image Understand., vol. 95, no. 2, pp. 238–259, 2004.

[11] J. Choi, Y. Yoo, and J. Choi, “Adaptive shadow estimator for removing shadow of moving object,”

Comput. Vis. Image Understand., vol. 114, no. 9, pp. 1017–1029, 2010.

[12] Z. Liu, K. Huang, and T. Tan, "Cast shadow removal in a hierarchical manner using MRF," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 1, pp. 56–66, Jan. 2012.

[13] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," in Proc. Eleventh ACM Int. Conf. Multimedia, Nov. 2003, pp. 2–10.

[14] Z. Zivkovic, "Improved adaptive Gausian mixture model for back- ground subtraction," in Proc. IEEE Int. Conf Pattern Recognition, Aug.2004, pp. 28–31.

Author Profile

P. Pranay has received his B.Tech Degree in Electronics & Communication Engineering in 2012 from Jyothishmathi Institute of Technological And Sciences, JNTUH. Presently he is Pursuing his M.Tech in Digital Systems & Computer Electronics from, Jyothishmathi Institute Of Technology And Science, Karimnagar, Telanagana, India. His research area includes Digital Image Processing.

A. Sreeramulu has completed his B.Tech in Electronics & Communication Engineering in 2006 from Mother Theressa College Of Engineering , JNTUH, M.TECH in Embedded Systems in 2011 from Vivekananda Institute Of Technology And Science, JNTUH and presently working as Associate Professor at Jyothishmathi Institute Of Technology And Science , Karimnagar, Telanagana, India. His area of interest includes Digital Image Processing.