Intensity Range based Background Subtraction for Effective Object Detection *

Kalyan Kumar Hati[†], Pankaj Kumar Sa[‡], and Banshidhar Majhi[§¶]

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 defines an intensity range for each pixel location in the background to accommodate illumination variation as well as motion in the background. 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: Video surveillance, video segmentation, background modeling, background subtraction.

---

*This paper is accepted for publication in the *IEEE Signal Processing Letters*. BibTeX provided in the last page can be used to cite this paper. The final version is available at `http://dx.doi.org/10.1109/LSP.2013.2263800`

[†]`KalyanKumarHati@nitrkl.ac.in`

[‡]`PankajKSa@nitrkl.ac.in`, Corresponding Author

[§]`BMajhi@nitrkl.ac.in`

[¶]The authors are with the Department of Computer Science and Engineering, National Institute of Technology Rourkela, Rourkela – 769 008, Odisha, India.

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 interaction, 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 segmentation, background subtraction, temporal differencing, active contour models, and generalized Hough transforms. 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 2 describes some of the related works. In Section 3 the proposed algorithms are presented. Simulation results are discussed in Section 4. Finally, Section 5 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 determine 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 foreground. 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 an hierarchical manner. At the pixel level, Gaussian mixture model (GMM) is used, whereas at the global level statistical features of the shadow is 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 carried 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 W (an odd length window) and a constant C for its computation. The optimal values are selected experimentally. Both stages of LIBS scheme are described as follows.

3.1 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.

3.2 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 ($\mathbf{T_L}$) and the local upper threshold ($\mathbf{T_U}$). These local thresholds help in successful detection of objects suppressing shadows if any. The steps of the algorithm are outlined in *Algorithm* 2.

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, “Time of Day”, “PETS2001”, “Intelligent Room”, “Campus”, “Fountain”, and “Lobby”. The first sequence is from wallflower dataset. It describes an indoor scenario where brightness changes during the entire span of the movie. Along with the change in brightness, a person enters the room, sits, reads book, and leaves out of the room. He performs same

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**
      $\overrightarrow{V} \leftarrow \left[f_k(i,j), f_{k+1}(i,j), \ldots, f_{k+(W-1)}(i,j)\right]$
      $\sigma \leftarrow$ standard deviation of $\overrightarrow{V}$
      $\mathbf{D}(p) \leftarrow |V\left(k + (\lfloor W \div 2 \rfloor)\right) - V\left(p\right)|$,
      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**
    $M(i,j) = \min\left[f_s(i,j)\right]$ and
    $N(i,j) = \max\left[f_s(i,j)\right]$,
    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

Algorithm 2 Background Subtraction for a frame f

for $i \leftarrow 1$ to height of frame **do**
  for $j \leftarrow 1$ to width of frame **do**
    Threshold $\mathbf{T}(i,j) = [M(i,j) + N(i,j)] \div C$
    $\mathbf{T_L}(i,j) = M(i,j) - \mathbf{T}(i,j)$
    $\mathbf{T_U}(i,j) = N(i,j) + \mathbf{T}(i,j)$
    if $\mathbf{T_L}(i,j) \leq f(i,j) \leq \mathbf{T_U}(i,j)$ **then**
      $\mathbf{S}_f(i,j) = 0$ //Background pixel
    else
      $\mathbf{S}_f(i,j) = 1$ //Foreground pixel
    end if
  end for
end for

activities twice. Second sequence is chosen from the PETS2001 data set, which has been recorded in a changing background and illumination conditions. The third sequence is from computer vision and robotics research laboratory of University of California, San Diego. It is recorded inside a room where a person enters the room, gives few poses and walks away. The last three sequences are from I2R dataset. The "Campus" sequence depicts an outdoor scenario with moving vehicle and human beings on a road. It is also observed that the leaves of the tree on the roadside are found to be waving. "Fountain" sequence illustrates a scenario with a water fountain in the background. "Lobby" sequence is recorded inside a room with changing illumination. Considering the characteristics of selected video sequences, they are the most suitable representatives for validation of generalized behavior of the proposed scheme.

For comparative 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]. Percentage of correct classification (PCC) is used as the metric for comparison, and is defined as,

$$PCC = \frac{TP + TN}{TPF} \times 100 \tag{1}$$

where TP is true positive that represents the number of correctly detected foreground pixels and TN is true negative representing the number of correctly detected background pixels. TPF represents the total number of pixels in the frame. TP and TN are measured from a predefined ground truth.

Further, the window size (W) used during classification of a pixel as stationary or non-stationary is chosen experimentally by varying $W = 5, 7, 9, 11, 13$. Similarly, for each window the constant C used for calculating the local threshold, is varied between 3 and 13 in a step of 1. For each combination of W and C, the PCC is computed. A graphical variation among these three parameters is shown in Fig. 1 for the "Lobby" video sequence. It may be observed that for $W = 9$ and $C = 7$, the PCC achieved maximum of 99.47%. Similar observations are also found for other video sequences. The objects detected in different sequences are depicted in Fig. 2. It may be observed that, LIBS accurately detects objects in almost all cases with least misclassified objects. Moreover, shadows in "Intelligent Room" sequence are also removed by the proposed algorithm. Furthermore, 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 PCC obtained in each case is listed in Table 1. The higher accuracy of PCC 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. Such 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 falls on any surface, decreases its intensity by some factor. Therefore, 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.

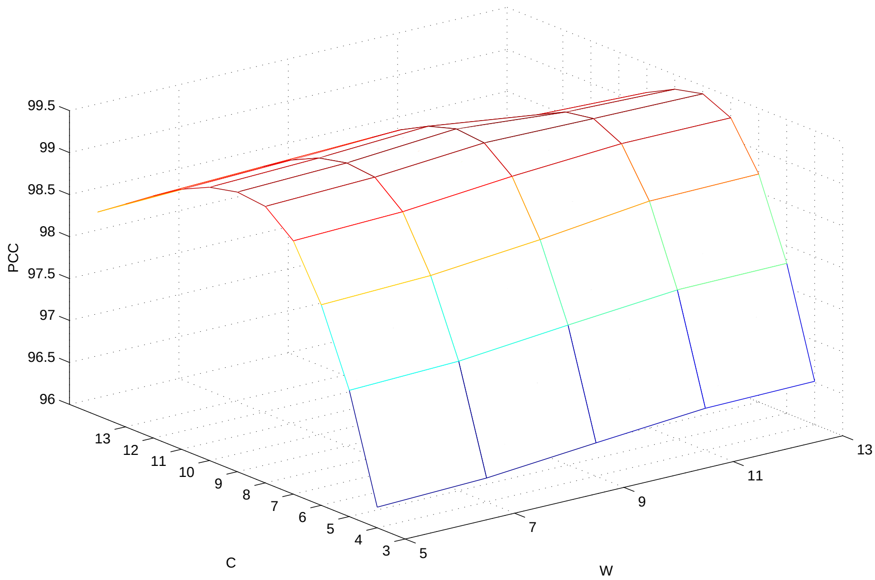

Figure 1: Variation of percentage of correct classification (PCC) with window size (W) and constant (C)

Table 1: Comparative analysis of PCC

Scheme	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 2: Left to right. Time of Day, PETS2001, Intelligent Room, Campus, Fountain, and Lobby frame sequence. Top to bottom. Original frame, Ground truth, and results of GMM, EGMM, Readdy *et. al.*, and LIBS.

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 outperforms others and detects objects in all possible scenarios considered.

References

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computing Survey*, vol. 38, Dec 2006.

[2] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 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 Computer Society Conference on CVPR*, 1999, pp. 246 – 252.

[4] I. Haritaoglu, D. Harwood, and L. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 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 Symposium on Computer Graphics and Image Processing*, Oct 2005, pp. 189 – 196.

[6] J. McHugh, J. Konrad, V. Saligrama, and P. Jodoin, "Foreground-adaptive background subtraction," *IEEE Signal Processing Letters*, vol. 16, no. 5, pp. 390 – 393, May 2009.

[7] O. Barnich and M. Van Droogenbroeck, "ViBe: A Universal Background Subtraction Algorithm for Video Sequences," *IEEE Transactions on Image Processing*, 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 Processing Letters*, vol. 19, no. 3, pp. 127 – 130, Mar 2012.

[9] V. Reddy, C. Sanderson, and B. Lovell, "Improved foreground detection via block-based classifier cascade with probabilistic decision integration," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 1, pp. 83 – 93, Jan 2013.

[10] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," *Comupter Vision and Image Understanding*, 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," *Comupter Vision and Image Understanding*, 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 Transactions on Circuits and Systems for Video Technology*, 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 *Proceedings of the eleventh ACM international conference on Multimedia*, Nov 2003, pp. 2 – 10.

[14] Z. Zivkovic, "Improved adaptive gausian mixture model for background subtraction," in *Proceedings IEEE International Conference on Pattern Recognition*, Aug 2004, pp. 28 – 31.

```
------
BibTeX
------
@article{LIBS,
   author={Kalyan Kumar Hati and Pankaj Kumar Sa and Banshidhar Majhi},
   title={Intensity {R}ange based {B}ackground {S}ubtraction
          for {E}ffective {O}bject {D}etection},
   journal={IEEE Signal Processing Letters},
   volume={20},
   number={8},
   pages={759~--~762},
   month={August},
   year={2013},
   DOI={10.1109/LSP.2013.2263800}.
}
```