Dynamic Object Identification Using Background Subtraction and Its Gradual Development

[1]Mr.B.Ramesh Naik and [2]Mr.N.V.Ramanaiah

Abstract— The project proposes efficient dynamic identification and people counting based on background subtraction using dynamic threshold approach with mathematical morphology. Here these different methods are used effectively for object identification and compare these performance based on accurate detection. Here the techniques frame differences, dynamic threshold based detection will be used. After the object foreground detection, the parameters like speed, velocity motion and sensitivity will be manipulated. For this, most of previous methods depend on the assumption that the background is static over short time periods. In dynamic threshold based object detection, morphological process and filtering also used effectively for unwanted pixel removal from the background. The background frame will be updated by comparing the current frame intensities with reference frame. Along with this dynamic threshold, mathematical morphology also used which has an ability of greatly attenuating color variations generated by background motions while still highlighting dynamic objects. Finally the simulated results will be shown that used approximate median with mathematical morphology approach is effective rather than prior background subtraction methods in dynamic texture scenes and performance parameters of moving object such sensitivity, speed and velocity will be evaluated.

Index Terms—Background modeling, background subtraction, video segmentation, video surveillance.

————————————— ◆ —————————————

I. INTRODUCTION

Dynamic object identification in real time is a challenging task in visual surveillance systems. It often acts as an initial step for further processing such as classification of the identified moving object. In order to perform more sophisticated operations such as classification, we need to first develop an efficient and accurate method for identifying dynamic objects. A typical dynamic object identification algorithm has the following features: (a) estimation of the stationary part of the observed scene (background) and obtaining its statistical characteristics (b) obtaining difference images of frames taken at different times and difference images of the sequence with the image of the stationary part of the scene (c) discrimination of regions belonging to objects, identification of these objects, determining the trajectories of motion of these objects, and their classification (d) adaptation of the stationary part of the background for changing detection conditions and for changing the content of the scene (e) registration of situations and necessary messages. One of the simplest and popular method for dynamic object identification is a background subtraction method which often uses background modeling, but it takes long time to detect moving objects [1-6]. Temporal difference method is very simple and it can detect objects in real time, but it does not provide robustness against illumination change. The foreground extraction problem is dealt with the change detection techniques, which can be pixel based or region based. Simple differencing is the most intuitive by arguing that a change at a pixel location occurs when the intensity difference of the corresponding pixels in two images exceeds a certain threshold. However, it is sensitive to pixel variation resulting from noise and illumination changes, which frequently occur in complex natural environments. Texture based boundary evaluation methods are not reliable for real time dynamic object identification.

This paper proposes a new method to detect dynamic objects from a stationary background based on the improved edge localization mechanism and gradient directional masking for video surveillance systems. In the proposed method, gradient map images are generated from the input and background images using a gradient operator.

The gradient difference map is then calculated from gradient map images. Finally, moving objects are detected by using appropriate directional masking and thresholding. Simulation results indicate that the proposed method provides better results than well-known edge based methods under different illumination conditions, including indoor, outdoor, sunny, and foggy cases for detecting moving objects. Moreover, it detects objects more accurately from input images than existing methods.

The rest of this paper is organized as follows. Section 2 provides a brief description of related research. Section 3 introduces the proposed method for identifying dynamic objects. Section 4 compares the performance of the proposed method with well-known edge based methods. Section 5 concludes the paper.

II. RELATED WORK

For object identification 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 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 carried out on standard videos and comparative analysis has been performed with competitive schemes.

III. THE PROPOSED LIBS SCHEME

The proposed System architecture for image segmentation using background subtraction and EM technique shown in below figure1

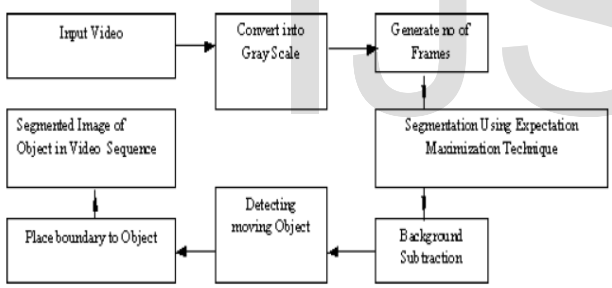


Figure 1 System Architecture

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.

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

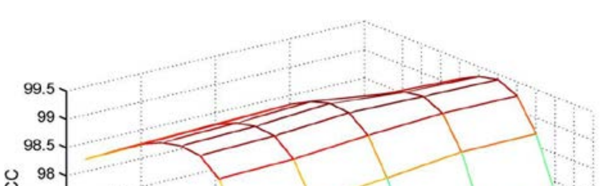


Fig. 2. Variation of percentage of correct classification(PCC) with window size(W) and constant(C) ..

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 {*f1,f2…..fn*} , where 20 ≤ n ≤ 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**
      $\vec{V} \leftarrow [f_k(i,j), f_{k+1}(i,j), \dots, f_{k+(W-1)}(i,j)]$
      $\sigma \leftarrow$ standard deviation of $\vec{V}$
      $\mathbf{D}(p) \leftarrow |V(k+(\lfloor W \div 2 \rfloor)) - V(p)|$, 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 $\vec{\mathbf{D}}$
      if $\mathbf{S} \leq \lfloor W \div 2 \rfloor \times \sigma$ **then**
        label $f_{k+(\lfloor W \div 2 \rfloor)}(i,j)$ stationary
      else
        Label $f_{k+(\lfloor W \div 2 \rfloor)}(i,j)$ 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 $s = \lceil W \div 2 \rceil \ldots n - (\lfloor W \div 2 \rfloor)$ and $f_s(i,j)$ is stationary
  end for
end for

IJSER

ISSN 2229-5518

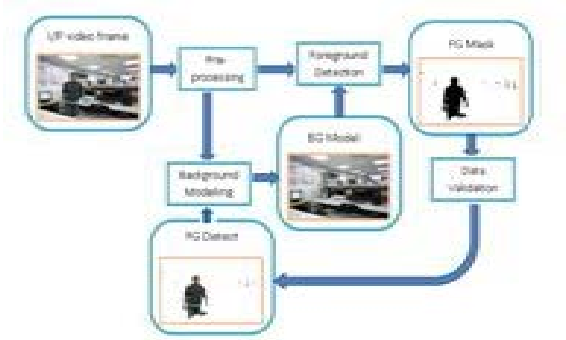


Figure 3: Three main processes background subtraction flow: Background modeling, moving object detected, and updates periodically background model

B. Extraction of Foreground Object

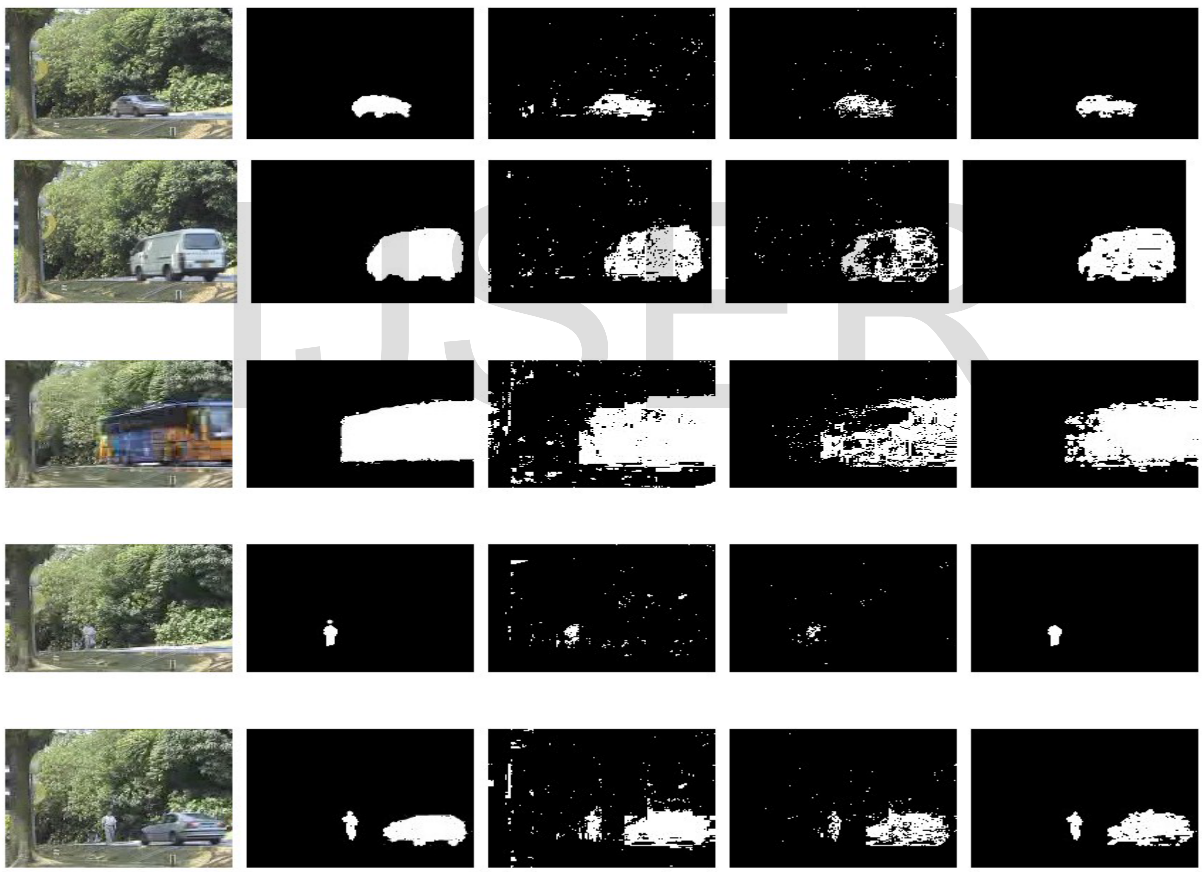
After successfully developing the background model, a local thresholding based background subtraction is used to find the foreground objects. A constant C is considered that helps in computing the local lower threshold (T_L) and the local upper threshold (T_U).

These local thresholds help in successful identification 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***



(a) (b) (c) (d) (e)

Figure 4. (a) Example frames from the video from the I2R dataset, (b) Ground-truth foreground mask and foreground mask estimation using: (c) STILP-PKDE [6], (d) WBS [8], (e) the proposed method

IV. 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 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.

TABLE I
COMPARATIVE ANALYSIS 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

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=[(TP+TN)/TPF]×100……..(1)

Where TP is true positive that represents the number of correctly detected foreground pixels and TN is true negative rep-resenting 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 I. 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 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.

V. 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 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.

Authors

[1] **Mr.B.Ramesh Naik**, Assistant Professor, ECE department of Seshachala Institute of Technology,Puttur.. He has completed M.Tech in ECE from SVCET, Chittoor. His research areas are LowPower VLSI and Communication system.

[2]**Mr.N.V.Ramanaiah**,M.tech(Student) in ECE department of Seshachala Institute of echnology,Puttur. His research areas are Digital image Processing,
Signal Processing and Communications

IJSER