



Evaluation of an Efficient Method for Object Detection in Video

Manoj Attri¹, Rohit Tanwar^{2(✉)}, Narender¹, and Neha Nandal³

¹ Manav Rachna University, Faridabad, India
{manoj, narender}@mru.edu.in

² University of Petroleum and Energy Studies, Dehradun, India
r.tanwar@ddn.upes.ac.in

³ St. Martin's Engineering College, Hyderabad, India

Abstract. In this real world, many public or open areas are facilitated with cameras at multiple angles to monitor the activities of human for safety of people or infrastructure. The object detection is a fundamental concept of computer vision that focused on the detection of instances of objects of a certain class (such as person, animal, ghost, buildings, or vehicles) in videos. This manuscript presents a method for object detection using background subtraction and morphology. The core of the proposed work is the simple background subtraction method. In the first step we developed a background model based on some video frames that only consists of static background without any moving object. A suitable scheme is applied for updating the background model so that the challenges like camera shake, dust and fog particles in air should be resolved. The scheme uses the “learning rate” for the entire frame. In the second step we extract the foreground pixels which are in motion. After the initial foreground extraction, the morphological operators are applied for noise cleaning.

Keywords: Object detection · Object tracking · Basic background subtraction · Mathematical morphology · Learning rate

1 Introduction

In today's technical world, the motion oriented object detection is considered as an hot area of research currently in trend that attracts lots of attention of research community like computer vision, video processing, automation analysis, object representation etc. The moving object detection focused on finding out the instances of semantic objects of a various classes like as animal, human, satellites, building, crops, vehicle, geographical region or anything which can be identified in video frames, while object tracking is considered as the task where the trajectory of either single or more than one objects is followed in frames [1, 2]. The detected object can be represented in the form of point, contour, skeleton or boundary of object using segmentation etc. [3]. In the current era, the video cameras such as IP camera, CCTV camera, or some other kind of surveillance cameras are installed everywhere for surveillance. At some places, flying

cameras like drone are available for the surveillance of mob, busy places, hilly areas, deep zones etc. It enhances the development of the intelligent video surveillance system which is used for the estimation of the moving person or object.

Since, the motion oriented object detection technique involved in various real-time applications such as image analysis, image automation, auto-annotation, scene recognition and scene understanding [4–7]. But in real-time, it is still an open challenge due to the complexity of the system. It is one of the key problems in image processing and computer vision, which has received continuous attention since the birth of the field. We, as a human being, easily “detect” various objects such as car, flower, people, buildings, and automobiles etc.

In the literature, we found some challenges that needed to be faced during the detection of moving object in video frames [4–6]:

- There may be some illumination changes like sudden or gradual change due to lighting effect or due to sunlight in water.
- There may also some distracting motions like camera-shake, moving elevator, motion in water, weaving tree leaves etc.
- Facing environmental effects like dust particles in air, rain, fog in the environment etc.

The image processing provides various techniques for object detection but one that is used in this paper is basic background subtraction (BBS). This technique mainly used to model the background from initial few frames and then it identifies the moving pixels from each frame by computing the difference from current image with an effective threshold. This technique extracts the meaningful pixel more accurately. The objects not in background are also get detected by this technique however they are not in motion. The main drawback of existing BBS methods enlists that they are sensitive to the environmental change such as sudden or gradual illumination variation. Most of the existing methods applied the static background for modelling but there is requirement to update the background model along with current frames. Here, the update of the background model is again a serious challenge for background subtraction technique. Even after updating the model, experimental results were not up to the mark. So for noise cleaning morphology was used. Mathematical morphology (MM) is a branch of science based on set theory. In such situation, the post processing plays a crucial role where the morphology is used for extracting image components or geometrical features which is used for description and representation of region shape, such as skeletons, boundaries etc. [7–11]. The morphological operators are used to remove the noisy pixels or outliers from the detected pixels [9, 12, 13]. Our goal in this research is to develop a method for automatic object detection. Two morphological operators are used i.e. erosion and dilation.

2 Literature Review

Background subtraction is a straightforward way to deal with identify moving articles in video sequences. The essential thought is to subtract the present edge from a foundation picture and to order every pixel as closer view or foundation by contrasting

the distinction and an edge [18]. Morphological tasks pursued by an associated part examination are utilized to register every single dynamic district in the picture. By and by, a few troubles emerge: the foundation picture is undermined by clamor because of camera developments and vacillating articles (e.g., trees waving), light changes, mists, shadows. To manage these challenges a few techniques have been proposed [19].

A few works utilize a deterministic foundation model, e.g., by portraying the allowable interim for every pixel of the foundation picture just as the most extreme pace of progress in sequential pictures or the middle of biggest interframes supreme distinction [20, 21]. Most works anyway depend on measurable models of the foundation, accepting that every pixel is an irregular variable with a likelihood circulation evaluated from the video stream. For instance, the Pfinder framework (“Person Finder”) utilizes a Gaussian model to portray every pixel of the foundation picture [22]. An increasingly broad methodology comprises of utilizing a blend of Gaussians to speak to every pixel. This permits the portrayal of multi modular appropriations which happen in common scene (e.g., on account of shuddering trees) [23].

Another arrangement of calculations depends on spatio–fleeting division of the video signal. These strategies attempt to distinguish moving districts considering not just the worldly development of the pixel forces and shading yet additionally their spatial properties. Division is performed in a three-dimensional (3-D) district of picture time space, thinking about the fleeting development of neighbor pixels. This should be possible in a few different ways, e.g., by utilizing spatio–fleeting entropy, joined with morphological activities [24]. This methodology prompts an improvement of the frameworks execution, contrasted and customary edge distinction techniques. Different methodologies depend on the 3-D structure tensor characterized from the pixels spatial and transient subordinates, in a given time interim [25]. For this situation, identification depends on the Mahalanobis separation, expecting a Gaussian appropriation for the subsidiaries. This methodology has been executed progressively and tried with PETS 2005 informational index. Different options have likewise been considered, e.g., the utilization of a district developing strategy in 3-D space–time [26].

A critical research exertion has been done to adapt to shadows and with nonstationary foundations. Two sorts of changes must be considered: show changes (e.g., because of the sun movement) and quick changes (e.g., because of mists, downpour or unexpected changes in static articles). Versatile models and limits have been utilized to manage moderate foundation changes [27]. These systems recursively update the foundation parameters and edges so as to follow the advancement of the parameters in nonstationary working conditions. To adapt to sudden changes, various model procedures have been proposed [27] just as prescient stochastic models (e.g., AR, ARMA [28, 29]).

Another trouble is the nearness of phantoms [30], i.e., bogus dynamic areas because of statics objects having a place with the foundation picture (e.g., vehicles) which all of a sudden begin to move. This issue has been tended to by joining foundation subtraction with outline differencing or by elevated level activities [31, 32].

3 Background Subtraction

Background subtraction is a process that extracts foreground objects from the image frames more accurately as compared to others. As mentioned in [8] many other traditional object detection methods follow same value of “learning rate” for complete frame but the proposed method make use of different “learning rate” for every pixel as per given parameters. As found in the literature many object detection methods face environmental changes they can be sudden or gradual change. Because of these changes one requires to update the background model using the given “learning rate”. According to [8], background subtraction process is further classified in three parts:

Background Model Initialization

Initially the requirement is to estimate the background model. There is an assumption that few initial frames are used for modeling of the background and also the sequence of frame begins with the non-presence of an object. In [8, 10], the selective averaging technique is used to construct the model.

$$BM_N(x, y) = \frac{\sum_{m=1}^N I_m(x, y)}{N},$$

$BM_N(x, y)$: It is specified as the intensity value of the pixel (x, y) of the technique, $I_m(x, y)$ represents the intensity value of pixel (x, y) of the m^{th} frame. The total count of frames are taken to be N .

Background Subtraction

After constructing the background model, the difference is being find out among the existing frame and the background frame to detect the object [2, 10, 11, 14],

$$D_t(x, y) = |I_t(x, y) - BM_{t-1}(x, y)|,$$

$BM_t(x, y)$: represents the intensity value of pixel (x, y) at time some time t , and $I_t(x, y)$ represents the intensity value of pixel (x, y) in the existing frame at an instant t . The variation is then compared to the threshold Th_{ad} for background and foreground pixel classification.

$$(x, y) = \begin{cases} foreground & \text{if } D_t(x, y) \geq Th_{ad} \\ background & \text{if } D_t(x, y) < Th_{ad} \end{cases}$$

Background Model Update

Due to environmental changes the model can be updated iteratively for each frame. Hence in [8], updation is done in pixel-by-pixel fashion with learning rate, α_{ad} , $t(x, y)$ for each pixel. The rate of change of background dynamics is directly proportional to the rate of updating the model. The rate of updating the model is used through the learning rate. An increase in value of background dynamics will subsequently increase then the learning rate in context of the model in order to reduce the false alarms. The model is updated according to the following equation:

$$BM_t(x, y) = \alpha_{ad,t}(x, y)I_t(x, y) + (1 - \alpha_{ad,t}(x, y))BM_{t-1}(x, y),$$

Such that $0 \leq \alpha_{ad}$ and $t(x, y) \leq 1$. Here, the value of learning rate i.e. α_{ad} , $t(x, y)$ is highly dependent on two parameters and their weights, α_1 and α_2 where w_1 and w_2 are weights and $w_1 + w_2 \leq 1$.

$$\alpha_{ad,t}(x, y) = w_1 \alpha_1 + w_2 \alpha_2$$

The learning rate is assigned as per the values of following two parameters [9, 10].

1. The first parameter is greatly affected by the difference i.e. $D_t(x, y)$. The large value for α_1 is assigned for a smaller $D_t(x, y)$ and σ_1 is $Th_{ad}/5$ where σ_1 is a function of Th_{ad} .

$$\alpha_1 = \begin{cases} e^{-\frac{1 * D_t(x,y)^2}{2 * \sigma_1^2}} & \text{if } D_t(x, y) < Th_{ad} \\ 0 & \text{otherwise} \end{cases}$$

2. The another parameter α_2 depends on temporal duration of a pixel in the background [10]. The reliability and stability are calculated by finding the temporal count C_{bg} , using the equation:

$$\alpha_2 = \begin{cases} e^{-\frac{1 * (\zeta_{max} - C'_{bg})^2}{2 * \sigma_2^2}} & \text{if } C_{bg} \geq \zeta_{min} \\ 0 & \text{otherwise} \end{cases}$$

where $\sigma_2 = 15$, $\zeta_{max} = 150$ and $\zeta_{min} = 30$ are assumed and $C'_{bg} = \min(\zeta_{max}, C_{bg})$. In case a pixel is supposed to remain as background pixel and counted for greater than ζ_{min} frames then α_2 attains a non-zero value. The α_2 parameter increases with an increase in C_{bg} .

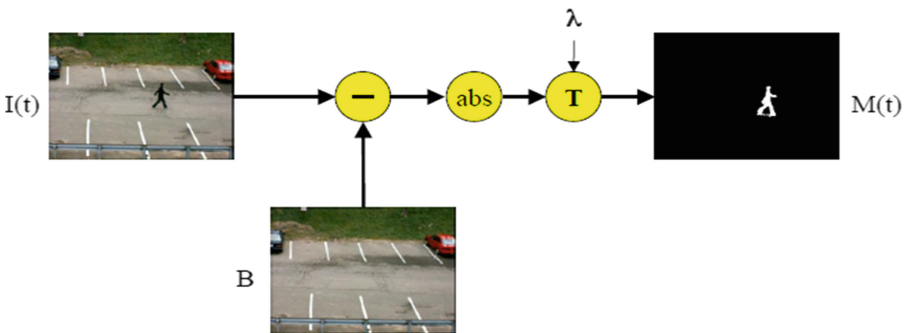


Fig. 1. Basic Steps for Background Subtraction [9] method

4 Morphological Operators

The morphological operators focused on the shape of features and used to reduce the outliers or noisy pixels. The morphological operations are applied to remove unwanted pixels after classification of pixels [10–14] (Fig. 2).



Fig. 2. Result before and after morphology [10]

- Erosion- The erosion is reduction of the objects size and removal of the small anomalies especially by doing subtraction of objects having their radius smaller as compared to the structural element [12, 14–16].
- Dilation- Dilation is a technique which expands the area of objects along the boundary, by filtering the broken areas, holes or connected regions that are separated through the structure element [13–16].

5 Methodology Used

In this work, the background subtraction based scheme is developed for dynamic background scenes that also have shadow issues.

The proposed work is completed in two steps: (i) Model the background and then compute a suitable threshold for each pixel classification, (ii) Pixel classification. The working steps of the suggested work is shown in Fig. 1. This paper mainly focuses on moving object to detect the object. As shown in Fig. 3 the colored video frame is taken as input and then result is being provided with background subtraction. But the results was not clear then after morphological operators are applied for noise cleaning.

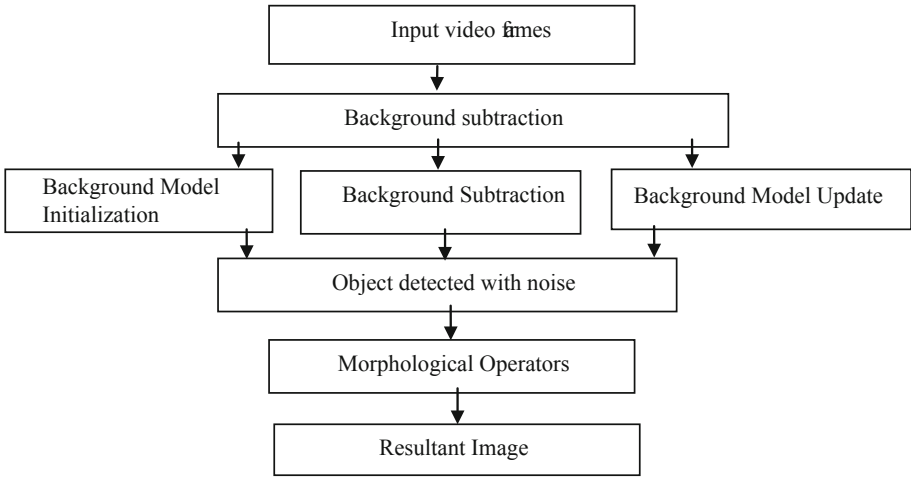


Fig. 3. Showing Block Diagram of Proposed Object Detection Algorithm

6 Experimental Results

This work is carried out on Windows XP OS with Intel Pentium (R) 4-processor having 2.46 GHz of speed and 2 GB RAM. Implementation of this research work is done with Matlab-2011 tool. The proposed work suggests significant improvement in terms of qualitative results which clearly depicts the strength in reducing the moving shadow, and other changes occurred in the background. The visual results of this experimental work depicts that the proposed results are better as shown in Fig. 4.

In the given Fig. 4, first column shows the original video frames. Second column represents the proposed results without post processing. Third column depicts the proposed results which are computed using post processing. The classified results are having some outliers which are considered as noisy pixel. So, this work has applied morphological operator to reduce the outlier or noisy pixel. Hence, the result shown in the third column are much better than the second column. So, the overall qualitative results of the proposed method using post processing is much better than results without post processing.

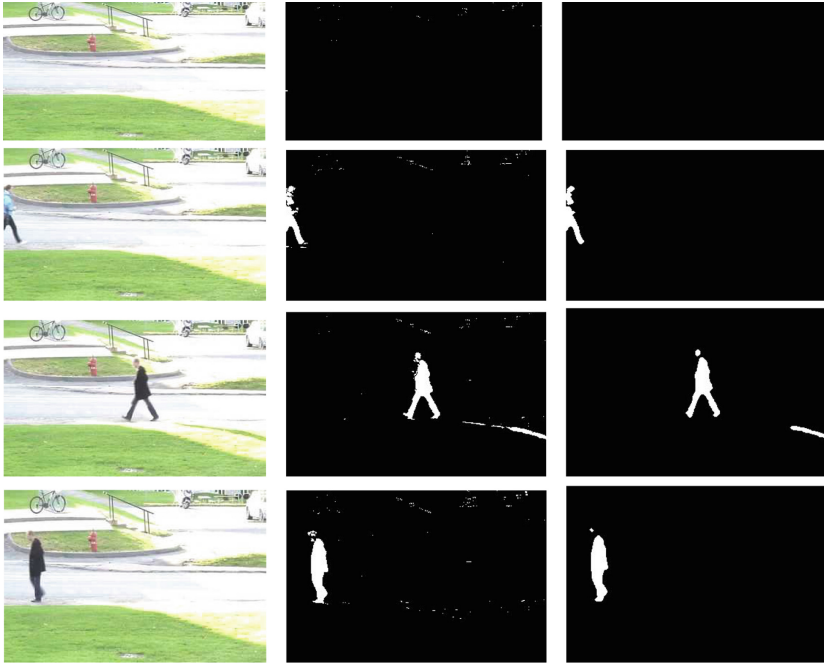


Fig. 4. Object detection results using background subtraction then applying morphology

7 Conclusion

In this paper, the object is detected using the background subtraction method and detected results have been improved using morphological operators. These operators reduced the misclassification rate. After morphological operation as exhibited in experimental results, the proposed method performs better in terms of qualitative results. This work deals with grayscale images. To resolve the wrongly classified pixel's, the purpose morphological operators are applied for noise cleaning. The proposed method produces good results as compared to other as shown in Fig. 4. The overall quality of proposed results is much better. In future, this work may be extended using GUI based application and cloud. It may be compared with [12, 17] on the basis of various parameters.

References

1. Kuralkar, P., Gaikwad, V.: Human object tracking using background subtraction and shadow removal techniques. *IJARCSSE* 2(3), 79–84 (2012)
2. Yadav, D., Singh, K.: A combined approach of Kullback-Leibler divergence method and background subtraction for moving object detection in thermal video. *Infrared Phys. Technol.* 76, 21–31 (2015)

3. Yilmaz, A., et al.: Object tracking: a survey. *ACM Comput. Surv.* **38**(4), 1–45 (2006). Article no. 13
4. Sharma, S., et al.: Cloud based emerging services systems. *Int. J. Inf. Manag.* 1–12 (2016)
5. Zhuang, L, Tang, K.: Fast salient object detection based on segments. In: *IEEE International Conference on Advanced Video and Signal Based Surveillance* (2009)
6. Davies, E.: *Computer and Machine Vision: Theory, Algorithms, Practicalities*, 4th edn. Elsevier, Amsterdam (2012). ISBN: 978-0-12-386908-1
7. Mahmoud, A., Walaa, M.: Fast and accurate approaches for image and moving object segmentation, pp. 252–259. *IEEE* (2011)
8. Jia, Z., et al.: Visual information fusion for object-based video image segmentation using unsupervised Bayesian online learning. *IET Image Process* **1**, 168–181 (2007)
9. NagaRaju, C., NagaMani, S.: Morphological edge detection algorithm based on multi-structure elements of different directions. *IJICTR* **01** (2011)
10. Ng, K.K., Delp, E.J.: Background subtraction using a pixel-wise adaptive learning rate for object tracking initialization. *VIPER*, School of Electrical and Computer Engineering, Purdue University, Indiana USA (2011)
11. Liu, Y., Pados, D.: Compressed-sensed-domain L_1 -PCA Video Surveillance. *IEEE Trans. Multimed.* **18**(3), 351–363 (2016)
12. Sharma, L., Yadav, D.K.: Histogram based adaptive learning rate for background modelling and moving object detection in video surveillance. *Int. J. Telemed. Clin. Pract.* **2**(1), 74–92 (2017)
13. Yadav, D., Singh, K.: Adaptive background modeling technique for moving object detection in video under dynamic environment. *Int. J. Spat. Temporal Data Sci.*, 1–13 (2017)
14. Yadav, D.K., Bharti, S.K.: *Edge Detection in Image Using Rough Set Theory*. L. Lambert Academic Publishing, Germany (2015)
15. Jain, A.K.: *Fundamentals of Digital Image Processing*, 1st edn. Pearson-Education, New Jersey (2011)
16. Rafael, C., Richard, E.: *Digital Image Processing*. Tata McGraw-Hill Publication, New Delhi (2008)
17. Sharma, L., et al.: Fisher’s linear discriminant ratio based threshold for moving human detection in thermal video. *Infrared Phys. Technol.* **78**, 118–128 (2016)
18. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Prentice-Hall, Englewood Cliffs (2002)
19. Cucchiara, R., Grana, C., Piccardi, M., Prati, A.: Detecting moving objects ghosts and shadows in video streams. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(10), 1337–1342 (2003)
20. Haritaoglu, I., Harwood, D., Davis, L.S.: W4: who? when? where? what? a real time system for detecting and tracking people. In: *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 222–227, April 1998
21. Haritaoglu, I., Harwood, D., Davis, L.S.: W4: real-time surveillance of people and their activities. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 809–830 (2000)
22. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfnder: real-time tracking of the human body. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 780–785 (1997)
23. Stauffer, C., Eric, W., Grimson, L.: Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 747–757 (2000)
24. Ma, Y.-F., Zhang, H.-J.: Detecting motion object by spatio-temporal entropy. In: *IEEE International Conference on Multimedia and Expo*, August 2001
25. Souvenir, R., Wright, J., Pless, R.: Spatio-temporal detection and isolation: results on the PETS2005 datasets. In: *Proceedings of the IEEE Workshop on Performance Evaluation in Tracking and Surveillance* (2005)

26. Sun, H., Feng, T., Tan, T.: Spatio-temporal segmentation for video surveillance. In: IEEE International Conference on Pattern Recognition, vol. 1, pp. 843–846, September 2000
27. Boulton, T., Micheals, R., Gao, X., Eckmann, M.: Into the woods: visual surveillance of non-cooperative camouflaged targets in complex outdoor settings. *Proc. IEEE* **89**, 1382–1402 (2001)
28. Monnet, A., Mittal, A., Paragios, N., Ramesh, V.: Background modeling and subtraction of dynamic scenes. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, pp. 1305–1312 (2003)
29. Zhong, J., Sclaroff, S.: Segmenting foreground objects from a dynamic textured background via a robust Kalman filter. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, pp. 44–50 (2003)
30. Siebel, N.T., Maybank, S.J.: Real-time tracking of pedestrians and vehicles. In: Proceedings of the IEEE Workshop on Performance Evaluation of Tracking and Surveillance (2001)
31. Cucchiara, R., Grana, C., Prati, A.: Detecting moving objects and their shadows: an evaluation with the PETS2002 dataset. In: Proceedings of the Third IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2002) in conj. with ECCV 2002, pp. 18–25, May 2002
32. Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O.: A System for Video Surveillance and Monitoring: Vsam Final Report Robotics Institute, May 2000