Mathematical Statistician and Engineering Applications
ISSN:2094-0343
2326-9865

A Novel Algorithm for Foreground Moving Object Detection

V. Naga Bushanam[1], M. Rama Krishna Murthy[2]
[1]Research Scholar , Andhra University , Dept of CS&SE, Visakhapatnam& CIET,Lam,Guntur, India.
[2]Professor, Dept of CSE, ANITS,Visakhapatnam, India.
Corresponding Author email: bushaniit2@gmail.com

Article Info
Page Number: 325 – 334
Publication Issue:
Vol.71 No.3s (2022)

Article History
Article Received: 22 April 2022
Revised: 10 May 2022
Accepted: 15 June 2022
Publication: 19 July 2022

Abstract

Recognition and tracking of moving object recognition from the video sequences is an important study subject because it may be utilized in a variety of applications. Tracking tries to find and predict specific movements of observed objects throughout a time period, whereas recognition allows for the return of object forms identified in the picture. As a result, detection may have a significant influence on the tracking process as a whole. The topic of detection is the emphasis of this article. The Optical flow technique, Frame difference (FD) technique and Background subtraction (BS) technique are most used detection methods right now. We provide a detection technique depending on the BS and FD methods because it is especially well suited to quick actual operations; nevertheless, optical flow has a greater computational expense due to the high density estimations. It is possible to accomplish sparse detection quickly by combining the BS and FD methods with the Edge detectors and Laplace filters. One of the major contributions offered as a result of this, is the development of an algorithm for the moving object recognition based on a quite extended combination of fundamental real-time surveillance processes. For typical benchmark datasets, experimental findings demonstrate that the proposed technique has superior noise suppression and detection accuracy than existing methods.

Keywords: Canny edge detection, Laplacian, Erosion, Dilation, Frame difference, Background subtraction

---

Introduction

The technique that processes image by extracting the moving objects from a succession of photos using picture attributes including edges, colors, and textures is known as Moving object detection [1]. It is without a doubt an essential topic of study, for real time smart security and surveillance purposes [2], used in 2D motions and 3D settings [3], automated identification and tracking of vehicle, personnel monitoring, and many other applications. Multi-target detection's overall goal is the estimation of target number and its trajectories using noisy sensor readings at each observation period. Multiple object detection approaches may be divided into two categories, according to a recent study [4], Detection Free Tracking (DFT) and Detection Based Tracking (DBT). Prior to estimating their trajectories, the former incorporates an object detection phase. Given a predetermined initialization, the latter concentrates only on the tracking process. It's worth mentioning that DBT enables objects to emerge and vanish and has a broader use, while its behavior is mostly determined by the detection quality that offers operational observations such as calculation of trajectory. Only the detection phase will be discussed in this work. As is the case with many detection algorithms [5] we may also concentrate on the quality of shape aand contours of the detected entities in pictures. The Optical flow technique, Frame difference (FD) technique [6] and Background subtraction (BS) technique [7] are most used detection methods right now. Fully Convolutional Networks (FCN) and Convolutional Neural Networks (CNN) have grown more effective [8] by wide range of programs, but community training still requires supervised learning with a large amount of actual data..

Because the optical glide method calculates the displacement between two photos, it must not only accurately identify every pixel but also find the spots between the two input images [9]. To put it another way, the computational complexity of the optical waft approach is enormous. As a consequence, since it computes a dense optical waft topic, it takes longer and is more complex than other algorithms. The heritage subtraction method is a popular target extraction method that eliminates items using a basic set of principles. It is straightforward to set up, but it is sensitive to little changes [10]. The body difference technique is one of the most basic ways in laptop vision. The frame distinction method has the advantage of needing considerably less processing, but it is susceptible to noise [11], and it sometimes seems to suffer from the empty phenomenon, which contains small apertures and gaps, causing its findings to be erroneous. Regardless of the many challenging conditions with body difference and backdrop eradication, recent area investigations have shown that these issues are being treated by a number of higher procedures.

Weng et al. developed an entirely new inter-body difference technique for item recognition and tracking with heritage reduction in 2010. This technique is not the most efficient in the near term, but it is more dependable and adaptive. Gang et al. [13] introduced a technique that combines three-body differential approach with the Canny region detection set of criteria in 2013. Liu et al. [14] evidenced that an underwater robot could carry out submerged operations and find a moving target using submerged video using a technique combining background reduction and 3-body differentiation without bothered by changes in illumination conditions or sensitive scenarios in 2014. The use of siamese FCNs to separate the street area for avenue identification was suggested by Wang et al. [8] in 2017. This method can recognize more precise street areas than prior algorithms, and integrating a pre-mixed region may help to improve overall detection performance. Yuan et al. published a comprehensive deep learning device for recognizing traffic indications in difficult settings in 2019. A visibility collection focus module is used in addition to the densely linked frequency hopping connection and background subtraction layer to increase overall detection performance.

Despite the fact that several research on detection and tracking have been conducted, it seems that no systematic strategy has yet to be developed. The issue as a whole is still unsolvable,

ISSN:2094-0343
2326-9865

with a variety of methods having different strengths and weaknesses. With this in mind, we'll concentrate on the detection phase as well as a few of the most important real-time strategies, such as history removal and body difference-based algorithms. We suggest a unique solution that integrates a variety of commonly used equipment in this situation. The body difference, heritage removal, the Laplace filter, and the Canny edge detector are the most significant components of the method.

The latest technology is expected to clean the edge of a moving body and fills the unfilled openings using a number of mathematical morphological methods. The new mix includes data from both the BS and FD techniques, and instead of the standard, it runs the FD on three frames. This novel combination outperforms independent BS or FD systems in tests while maintaining real-time execution.

The following is the layout of the paper's relaxation: The "Presented item Detection set of rules" describes the approaches and processes employed in the main method we proposed. "Test and assessment" describes the experimental data. In "application to real-Time Video Processing," the suggested technique is used to real time video processing settings. Then, findings are offered, as well as recommendations for further study. Filters & Definitions for amateurs Pre-processing and post-processing are two different types of processing. It is impossible to overestimate the importance of image pretreatment and postprocessing in this project. Because we're only interested in the detection phase of gadgets, the detection result is represented as a binary image from a series of input images. Observed issues are made up of person-linked components. This data is utilized to build the foundation for subsequent detection tactics, so it's wonderful to have an impact on others. As a consequence, the binary output must precisely reproduce the item forms, including delimiting edges and filling object interiors. It demonstrates a variety of visual representations of detection approaches, including contour, rectangular container and silhouette. This article uses a discrete snapshot as an actual data to assess the final shape's quality using both subjective and quantitative methods. The most fundamental item identification processing processes are image binarization, shade conversion, side detection, and filtering process. The majority of these center solutions are simple to set up and use, and they are often compatible with actual software. Most of the filters have a temporal complexity of O(N), where N is pixel range. Furthermore, parallel execution on a graphics processing unit (GPU) system is already possible. The typical techniques that might be used in the suggested object identification algorithm are discussed in detail.

Preliminaries:

Color to Grayscale Conversion

The three primary colors are represented by the letters RGB, which stand for red, green, and blue. On which the mentioned color combinations have been mixed in a number of ways to produce a wide range of colors with different weights. A grayscale image is one where the value of every pixel is a sample identifying a sort of light, i.e., it carries intensity values between black to white at its most basic level. Because colour scale photos often include a great amount of data, processing them needs the computer analyzing all of the recorded information, which increases processing time and leads to their being confused with image processing and calculation. The color scale image should be converted to grayscale to improve computer efficiency.

Image Binarization

The Binary images are divided into values corresponding to black and white respectively. The pixels having grey degree greater than the edge are charged one, while the rest are charged zero. It means that a photograph might be taken with white item against the background having black based on camera settings, which is often used to distinguish a foreground photograph from a history photograph. Grayscale photographs have a grayscale value of zero to 255, with 255 signifying white and zero designating black, while black and white photographs have just zero and one values, with zero denoting black and one denoting white. Binarization of photographs aims to blend information while speeding up rational decision-making. As a consequence, binary images may be effective in improving computer reputation performance.

Filtering Process

The structure and execution of a rejector that fulfills photo processing criteria is known as filter processing. The mean clear out, median clear out, Gaussian filter out, and Laplace filter are the most often used filters.

The second order derivative of the Gaussian equation is the Laplacian. In terms of polishing influence and side localization, the second one-order differential exceeds the main one-order differential. Image sprucing, unlike a Gaussian clean out, which blurs a picture, enhances gray contrast and clarifies a fuzzy image. Because the Laplacian is a differential operator, it may be used to boost the grayscale mutation region while weakening the slowly converting grayscale area in an image. As a consequence, the polishing procedure may also choose the original image for Laplacian assessment, allowing you to create a shot that accurately shows the quick grayscale variation. Then, to get a crisper view, the Laplacian image has been placed over the source photo. The following is the fundamental Laplacian polishing technique where x and y are the coordinates of pixel:

$$\frac{\partial}{\partial x} G_\sigma(x,y) = \frac{\partial}{\partial x} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (1)$$

$$\frac{\partial^2}{\partial x^2} G_\sigma(x,y) = \frac{x^2}{\sigma^4} e^{-\frac{x^2+y^2}{2\sigma^2}} - \frac{1}{\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (2)$$

$$\nabla^2 G_\sigma(x,y) = \frac{\partial^2 G_\sigma(x,y)}{\partial x^2} - \frac{\partial^2 G_\sigma(x,y)}{\partial y^2} \qquad (3)$$

According to the research, the Laplacian filter should be included in the blended history Subtraction/body difference detection technique. The Laplacian filter, which sharpens the image while preserving historical details, may be used. The gray tone of the image may be preserved, and other capabilities can be highlighted.

Canny Edge Detector

The side detection method is used to locate the edges of objects in a photograph. Facet detection techniques come in a variety of sizes and forms. Some of the most extensively used area detection techniques include Sobel, Prewitt, Roberts, and Canny. In 1986, John F. Canny designed the Canny edge detector. The Canny technique's three main goals are low error rates, exceptional localization, and mark specialization. The astute operator persisted through a multiple stage procedure because of its superior capacity to fulfil the mentioned conditions.

(i) Using a Gaussian filter, we simplified the shot and removed the noise. To limit the impact of noise on area detection and avoid false detection, the noise should be removed:

$$H_{ij} = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{(i-(k+1))^2+(j-(k+1))^2}{2\sigma^2}\right\} \quad (4)$$

where Ws is the window length and P is the pixel's brightness price. The noise sensitivity of a detector decreases as its length increases.

(ii) For every pixel in the picture, compute the gradient strength and path. The image employment of a variety of operators is recognized by Canny algorithm since the rims of the shot may factor in any path. The gradient intensity value G as well as the gradient direction are described as

$$G_x = S_x * W_s \quad (5)$$

$$G_y = S_y * W_s \quad (6)$$

$$G = G_x^2 + G_y^2 \quad (7)$$

$$\theta = \arctan(G_y / G_x) \quad (8)$$

where S_x is the x directional edge detection operator in y direction, and S_y is the y directional edge detection operator in x direction. The x and y gradient values are denoted by G_x and G_y, respectively.

(iii) To get rid of erroneous side detection answers, use non-most suppression. All gradient readings beyond the closest highest reading reduced to null using the non-maximum suppression region sparse technique. N, NW, NE, S, SW, SE, W and E are the guidelines that make up the gradient. Pixel P's gradient direction is, and pixels P1 and P2's gradient linear interpolation GP1 and GP2 are as follows:

$$\tan\theta = G_y / G_x \quad (9)$$

$$G_{P1} = (1-\tan\theta)\times E + \tan\theta \times NE \quad (10)$$

$$G_{P2} = (1-\tan\theta)\times W + \tan\theta \times SW \quad (11)$$

(iv) Determine the genuine and probable edges using double-threshold detection.

(v) Suppress the isolated weak edges to complete the edge detection process.

(vi) **Frame Differencing Method**

The body difference technique may be used to examine a sequence of photos. To get details about moving targets, gradient vectors and gray values have been used. The approach analyzes factor-by-factor grey values from two successive snap photos to build a frame distinction image. The difference between frames is determined using the equation

$$D_k(x,y) = \left| I_k(x,y) - I_{k-1}(x,y) \right| \qquad (12)$$

where I_K is the photo gray fee of the current body, I_{K-1} is the photo gray charge of the next body, and D_K is the photo gray fee after the difference between I_K and I_{K-1}.

Background Subtraction Method

To remove the backdrop picture from the present frame, heritage subtraction employs a distinction calculation. The procedure is divided into two phases, as shown below. The video collection is initially used to gather the cutting-edge body image Kth and the historical picture. The differences between the most recent body photo and the most recent background photo are computed to generate a body distinction photograph. Zhang and Liang [6] use historical Background subtraction (BS) in combination with contour projection analysis and morphological filtering. The findings may be inaccurate due to small interferences. The goal is for developing a heritage improvement approach which can consistently identify moving objects while filtering out these inevitable disruptions. As a consequence, a more accurate Canny detector has been presented, as well as a more effective legacy reduction strategy.

Proposed Method:

The processing flowchart for the recommended technique is shown in Figure 1. After converting a shade picture to a grayscale photo, the Laplace filter's dominating feature will increase the contour and detail of the grayscale target. Second, each of the three bodies and the heritage distinction processes is completed one by one. To find and obtain side data, Canny part detection and threshold binarization is used. Then, the final shifting object forms are created utilizing a combination of the first two procedures plus a logic OR succeeded by a feature extraction,.

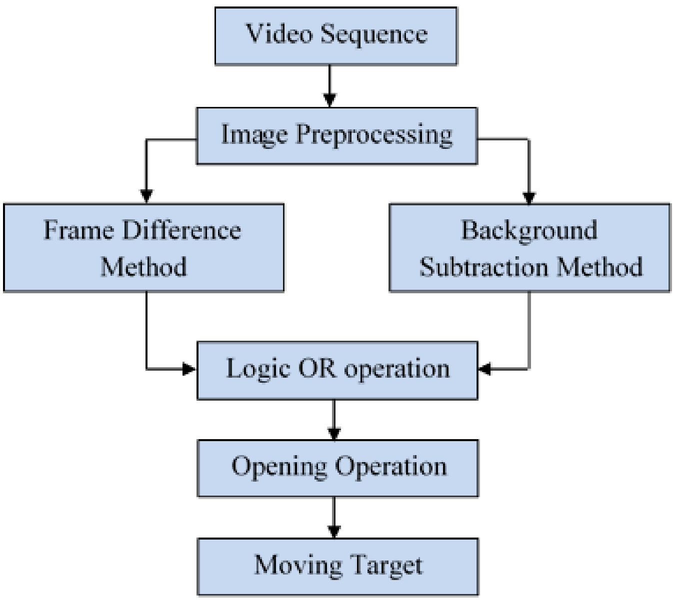

Figure 1: Proposed Algorithm Flowchart

Results and Discussion:

Datasets:

The datasets SABS, Wallflower, and Multivision are used as benchmarks in these studies. They're utilized for visual displays, comparison trials, and numerical evaluations based on a variety of factors. The SABS dataset1 [11] is a pixel-based benchmark for background model assessment. SABS is a video surveillance system that can change the exterior appearance of video sequences in nine distinct ways. It was decided to preserve the usage of Gaussian noise and global illumination. Imperfect labeling is substantially less of a problem in the SABS dataset than it is in other handcrafted floor reality datasets. SABS evaluates detection using both ground reality and additional shadow annotation.

The Wallfower dataset2 [12] has seven check cases. Each circumstance presents a distinct historical renovation situation that may be unsettling. The algorithm's output is separated into legacy and foreground images each with its own assessment image. The assessment image is separated into three tiers to manage with the unique challenges which emerge at the various scales: pixel, vicinity, and frame. For training, assessment, and assessment activities, these training photos, check images, and assessment image are important. The Multivision dataset3 [13] is a platform that enables clients to assess the hardware and software of real-time vision systems that use many cameras. The purpose of a creative and prescient technology is to turn an image into a complete record and to provide a visible solution that effectively analyzes photos from several cameras while also supplementing estimates in a reliable and durable manner. The benchmark is a dataset with floor reality segmentation that enables for goal assessment of body distinction and history discount approaches, which is necessary for our study.

Evaluation Criterion:

To examine and analyze statistics outputs from different detection structures, a number of assessment standards based on floor truth assessment have been devised. The accuracy of sample identification and fact retrieval is used to gauge the return of actual relevant findings. On this observation, the correctness, take into consideration, accuracy, and F-degree are used to rate the pleasantness of the experimental outcome. Accuracy is defined as ,

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (13)$$

Where FP is the range of history pixels wrongly classified as foreground, TP is the range of foreground pixels correctly categorized as foreground, FN is the range of foreground pixels and TN is the range of background pixels efficiently represented as background.

To calculate recall, the number of true positives (TP) is divided by the total number of genuine positives and false negatives (FN).

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \qquad (14)$$

To calculate precision, the number of true positives (TP) is divided by the total number of genuine positives and false positives (FP).

$$\mathrm{Pr}\,ecision = \frac{TP}{TP + FP} \qquad (15)$$

The harmonic suggest of remembering and precision is the F-measure.

$$F - measure = \frac{2\,\mathrm{Re}\,call * \mathrm{Pr}\,ecision}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision} \qquad (16)$$

A high F-measure cost indicates that the classifier produces accurate (high precision) results and most individuals of all great outcomes (excessive, remember).

The SABS-Bootstrap sequences of 352 288 snap pictures were utilized to demonstrate the implications. These images depict the outcomes of the classic BS method, the FD approach, and the recommended strategy (i). The recommended approach separates riding factors such as transferring autos, walking individuals, and wind-blown trees without difficulty.

Comparative Analysis

Most of the set of rules factors have been set as described in the preceding sections of the comparative analysis, and they remained consistent across all of the trials. The 10 picture sequences which includes Chair container, lcd screen, Floor Reality images, Camouflage are utilized and are based on Multivision and Wallflower datasets. Table 1 shows a numerical judgements made mostly on the basis of actual data. Under ten distinct photo sequences, three alternative detection algorithms are tested on the basis of precision, accuracy, F-measure and recall. According to these data, the recommended set of rules exceeds the traditional BS and FD procedures in terms of overall detection performance.

Table 1Performance metrics of various datasets

Datasets	Proposed Method			BS Method			FD Method		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Camouflage	0.914	0.909	0.938	0.864	0.948	0.948	0.452	0.588	0.012
F.A	0.932	0.823	0.941	0.926	0.909	0.802	0.752	0.575	0.187
GT-S	0.889	0.521	0.692	0.942	0.946	0.537	0.914	0.871	0.319
Chair Box	0.923	0.898	0.636	0.931	0.991	0.612	0.853	0.626	0.373
Hall Way	0.912	0.855	0.684	0.892	0.992	0.486	0.802	0.676	0.076
Lab Door	0.955	0.837	0.685	0.953	0.865	0.632	0.899	0.513	0.131
LCD Screen	0.959	0.848	0.686	0.957	0.941	0.582	0.914	0.684	0.141
Wall	0.961	0.691	0.631	0.964	0.813	0.495	0.941	0.479	0.092
Crossing	0.958	0.834	0.911	0.841	0.597	0.177	0.831	0.476	0.025
Suitcase	0.981	0.943	0.789	0.899	0.297	0.319	0.932	0.657	0.065

Conclusion

This study proposes a more desired object identification technique by precisely integrating crucial components of history removal and frame differentiation methodologies often used in real-time surveillance detection. The body distinction technique, Canny part detector, historical past removal method, and Laplace clear out are all used in real time in this method. The proposed approach was compared to traditional BS and FD algorithms and tested on common datasets using accuracy, recall, and precision as assessment criteria using floor fact evaluation. The whole computation time is comparable to a typical video charge on a home computer, and the findings were more accurate than previous methods. Furthermore, all the mentioned techniques are almost similar in design, software programs development in combination with GPU systems are being investigated to speed up solutions.

References

1. Yazdi M, Bouwmans T. New trends on moving object detection in video images captured by a moving camera: a survey. Comput Sci Rev. 2018;28:157–77.
2. Baek I, Davies A, Yan G, Rajkumar RR. Real-time detection, tracking, and classifcation of moving and stationary objects using multiple fsheye images. In: 2018 IEEE intelligent vehicles symposium (IV). IEEE; 2018. p. 447–52.
3. Hu HN, Cai QZ, Wang D, Lin J, Sun M, Krahenbuhl P, Yu F. Joint monocular 3D vehicle detection and tracking. In: Proceedings of the IEEE international conference on computer vision; 2019. p. 5390–9.
4. Luo W, Xing J, Milan A, Zhang X, Liu W, Zhao X, Kim TK. Multiple object tracking: a literature review. ArXiv Journal; 2014. arXiv:1409.7618.
5. Shotton J, Blake A, Cipolla R. Contour-based learning for object detection. In: Tenth IEEE international conference on computer vision (ICCV'05), vol. 1. IEEE; 2005. P. 503–10.
6. Zhang L, Liang Y. Motion human detection based on background subtraction. In: 2010 Second international workshop on education technology and computer science, vol. 1. IEEE; 2010. p. 284–7.
7. Zhang Y, Wang X, Qu B. Three-frame difference algorithm research based on mathematical morphology. Proc Eng. 2012;29:2705–9.
8. Wang Q, Gao J, Yuan Y. Embedding structured contour and location prior in siamesed fully convolutional networks for road detection. IEEE Trans IntellTransp Syst. 2017;19(1):230–41.
9. Dosovitskiy A, Fischer P, Ilg E, Hausser P, Hazirbas C, Golkov V, Van Der Smagt P, Cremers D, Brox T. Flownet: learning optical fow with convolutional networks. In: Proceedings of the IEEE international conference on computer vision; 2015. p. 2758–66.
10. Niu L, Jiang N. A moving objects detection algorithm based on improved background subtraction. In: 2008 eighth international conference on intelligent systems design and applications, vol. 3. IEEE; 2008. p. 604–7.
11. Zhan C, Duan X, Xu S, Song Z, Luo M. An improved moving object detection algorithm based on frame diference and edge detection. In: Fourth international conference on image and graphics (ICIG 2007). IEEE; 2007. p. 519–23.
12. Weng M, Huang G, Da X. A new interframe diference algorithm for moving target detection. In: 2010 3rd international congress on image and signal processing, vol. 1. IEEE; 2010. p. 285–9.
13. Gang L, Shangkun N, Yugan Y, Guanglei W, Siguo Z. An improved moving objects detection algorithm. In: 2013 International conference on wavelet analysis and pattern recognition. IEEE; 2013. p. 96–102.
14. Liu H, Dai J, Wang R, Zheng H, Zheng B. Combining background subtraction and three-frame diference to detect moving object from underwater video. In: OCEANS 2016-Shanghai. IEEE; 2016. p. 1–5.

15. Yuan Y, Xiong Z, Wang Q. VSSA-NET: vertical spatial sequence attention network for trafc sign detection. IEEE Trans Image Process. 2019;28(7):3423–34.
16. Brutzer S, Höferlin B, Heidemann G. Evaluation of background subtraction techniques for video surveillance. In: CVPR 2011. IEEE; 2011. p. 1937–44.
17. Toyama K, Krumm J, Brumitt B, Meyers B. Wallfower: principles and practice of background maintenance. In: Proceedings of the seventh IEEE international conference on computer vision, vol. 1. IEEE; 1999. p. 255–61.
18. Fernandez-Sanchez EJ, Rubio L, Diaz J, Ros E. Background subtraction model based on color and depth cues. Mach Vis Appl. 2014;25(5):1211–25.