Brief Analysis of Methods for Detecting Moving Objects Using Computer Vision

Gaurav Sharma
Department of CSE, HSST Swami Rama Himalayan University, Dehradun,248140, Uttarakhand, India.

Abstract- In many computer vision applications, moving object detection has drawn notable interest. The scientific community has made numerous contributions to addressthe significant difficulties of moving object detection in practical settings. The research thoroughly analyzes several moving object recognition methods, whichare divided into four groups: methods based on background modeling, Approaches rooted in frame differences, methods based on visual motion estimation, and methodologies based on deep learning. Additionally, thorough explanations of numerous techniques ineach category are offered.

Keywords: Scrutiny, Moving Object, ComprehensiveSurvey,

1. INTRODUCTION

Computerized video scrutiny has become a prominent field of study in machine vision due to its wide-ranging applications in intelligent video systems over the past few decades. Video sequence analysis typically comprises three core steps: identifying individual in-motion items, tracking this item enclosure by frame, and analyzing the tracks to forecast their activity or behavior. Additionally, surveillance systems are vital in the public and private sectors' defense against criminal activity and terrorist threats. It depends on the capability to recognize moving objects in indoor and outdoor environments, this is a critical phase in the information extraction process for Visual perception software. Generally, the term object denotes its universal format, which includes people walking on the street and man-made items with distinct borders that exist independently of their surroundings [1].

However, detecting moving objects and foreground objects has proven to be difficult due to the following factors [2]: These challenges include the potential existence of multiple moving objects within the scene, the presence of small, inadequately textured moving objects, rapid and adverse changes in lighting conditions, and the occurrence of multiple occlusions. In recent decades, significant attempts have been made to create various techniques for identifying various moving objects, including automobiles and humans, in indoor or outdoor scenes. Despite the available techniques, there remains a requirement for an exhaustive examination and experimental scrutiny of the relevant literature to generic object detection. In this current survey, we classified the object detection methods into groups according to the methodologies they utilized to identify prominent moving objects. In particular, this survey will help the researchers better understand this study area and choose the best method to meet their requirements.

2. LITERATURE REVIEW BASED ON BACKGROUND MODELLING

One of the simplest methods for spotting moving items in scenes is background modeling and removal. ln all video sequences, there are both background and foreground elements. The foreground retains solely pertinent data, encompassing the object of interest. if we can eliminate background data from video sequences [3]. Our accuracy can be improved if we know anything about how the background will appear. In order to function correctly, the background subtraction technique relies on a stable background, posing significant challenges for immediate application. Due to the constant changing of the backdrop image, acquiring a static background in real applications is challenging, and background updates must be made carefully. Building an accurate representation of the backdrop scene has been the subject of numerous efforts in background modeling for moving object identification. The aim of background modeling is to create a reference model through temporal or running averaging. After the background model is established, the video frames undergo subtraction from a reference or background model. The moving object is thus assumed to be the pixels in the cur-rent new frame that differ from the backdrop model. Numerous researchers utilize backdrop modeling techniques for detecting moving objects. Table I, displays a synopsis of these articles' contents. M.F. Savas et al. introduced partitioned adaptive technique for identifying moving objects in dynamic surroundings in [4]. They have employed a counter structure similar to that utilized by Casares et al. [5] to arrange an adjustable limiting factor. This parameter was used to update the backdrop after foreground determination. Each 22 non-overlapping blocks is continuously includedin the background model update, the probability function model was normalized by integrating the area under it. R. Kalsotra et al. introduced a morphologically motion based approach object recognition in [6], together with background subtraction methodol- ogy and thresholding. In [7], Y. Zhou et al. proposed a method for identifying moving objects utilizing mono-camera visual motion tracking alongside foreground extraction and motion tracking correction. To compensate for motion inside the image, visual odometry was used.The background model update constantly includes each of the 22 non-overlapping blocks, and the integral of the probability function of the model was zero normalized. In[6], R. Kalsotra et al. proposed a background subtraction method, thresholding, and a morphologically based method for moving object recognition. An approach for identifying moving objects utilizing monocular visual odometry combined with background subtraction and motion correction was presented by Y. Zhou et al. in [7]. Visual odometry was employed to account for motion inside the image. J. Ye et al. introduced a background elimination approach in red, green, blue color space in [8], in which the background model is esti- mated using a metrically trimmed mean, and the scale estimation is carried out using mean absolute deviation. According to B. Karasulu et al.'s proposal in [9]. Additionally, track-ing involves updating a learned background model. Detection of moving objects is achieved through the frame differencing technique.

3. LITERATURE REVIEW BASED ON FRAME DIFFERENCE

The primary purpose of temporal or frame differencing algorithms is to identify the presence of moving objects against still backgrounds. The object's location in two successive frames differs because objects move about time frame by frame. As a result, it becomes possible to determine the precise location and movement of the object within a specific frame by computing the pixel variance between two successive incoming frames. The techniques for detecting moving objects through frame differencing gained rapid prominence with the introduction of the temporal/frame differencing method. In [10], Z. Xu et al Employed a three-frame comparison technique alongside GMM for background subtraction. They incorporated XOR and OR operations to replace the AND operation, enhancing the ability to differentiate foreground pixels. This augmentation bolsters the robustness of the proposed method, which relies on the three-frame comparison technique approach. Additionally, GMM extracts the entire foreground, including shadows, from the image. To create a comprehensive foreground mask, dual foreground masks are ultimately merged using logical conjunction. In [11], Y. Wang et al. suggested a pixel-wise non-parametric approach for moving object detection incorporating both spatial and temporal attributes. Using the starting n frames and selecting samples m times creates a background model while considering 33 neighborhood blocks. The created backdrop model can also be fitted on dynamic scenes thanks to a proposed new background update approach. Improved frame-difference and Gaussian mixture background subtraction are combined dynamic object identification in [12] by J. Guo et al. In the suggested method, image restoration and morphological procedures were applied for precise and reliable detection of moving objects. Thanks to a suggested novel background update method, the developed backdrop model can also be fitted on dynamic situations. In [12] by J. Guo et al Enhanced integration of frame differencing and GMBS methods is employed for detecting moving objects.

Table 1: Synopsis of the articles

Author/Year	Method Used	Dataset	Performance	Observations
M.F Sava et al. [4] 2018	Block-wise Adaptive Background Subtraction.	Wallflower and CD.net 2014 datasets	F-measure 0.8551	The technique produces noticeable effects at the gray level while slowing down frame processing.

R.Kalsotra et.al [6] 2017	Dynamic Background Removal with Morphological Filtering.	Own captured videos and Matlab videos.	Not Given	The suggested approach can withstand noisy and dynamic settings well, but it is unable to handle the influence of shadow and concealment.
Y. Zhou et.al. [7] 2017	Dynamic Object Detection via Monocular Visual Odometry-assisted Background Subtraction.	Real outdoor videos and simulated city videos.	Recall: 0.8290 Precision: 0.5675	The technique holds up well to motion camera footage with noticeable parallax.
J.M. Guo et.al. [29] 2013	Layered Codebook Background Modeling and Subtraction.	Campus 200, Water surface 480, Meeting room 1755, Indoor gttest1 342, and Intelligent Room 82	F-Measure: 0.9265	The suggested approach is capable of handling scenes' dynamic fluctuation to a large extent.
J. Ye et.al. [9]/ 2012	To find foreground pixels in the evaluation step, robust estimators and a quick test are used for background modeling.	Highway and Campus	Shadow Detection Rate: 84.15%	When applied to shadow problems, the approach yields observable results.
B. Karasulu et.al. [9] 2012	Simulated Annealing (SA) and entropy values are used for background subtraction.	CAVAIR Dataset	F-Measure: 0.6325	The technique adjusts to occlusions' dynamic variations.

The suggested solution used morphological techniques and picture restoration to accurately and con- sistently detect moving objects. S.S. Senga et al. suggested a method for detecting dynamic entity against static backgrounds [13]. This method begins by conducting preprocessing to eliminate noise and computes the disparity between the current frames and their consecutive predecessors. Following this, it identifies and retains pixels with the highest intensity values from the difference between frames, divides the separate into distinct categories, non-overlapping segments, and calculates the average and median of pixel intensities within each consecutive segment. As a result, foreground-background segmentation followed by object detection using morphological operations for post-processing the real-world video dataset analysis with successful object detection across diverse shapes and sizes [1].

4. LITERATURE REVIEW BASED ON VISUAL MOTION ESTIMATION

The perception of motion experienced by objects, surfaces, and edges within a visual scene due to the movement of the observer (such as an eye or a camera) is known as visual motion estimation [14]. It determines where a pixel will be in the following image by approximating the velocity of each pixel in the current images. The core principle of this approach suggests that moving objects display changes in pixel intensity, offering vital cues for their spatial positioning. The visual motion estimation field is thus approximatedin these methods, and the visual motion estimation distribution features of the image carry out clustering processing. Some moving object identification techniques based on visual motion estimation have been presented in the literature. TABLE III displays a description of these works. J. Huang et al. used homograph matrices manifested as visual motion estimation for simulate the backdrop [14]. A dual-mode evaluation system has been developed to identify moving pixels in the foreground to enhance the suggested method's performance in complex real-world scenarios. Zero visual motion estimation vectors were used to model the backdrop by L. Kurnianggoro et al. in [15]. Using a homography matrix, the prior frames are aligned in this method, and the density of visual motion estimation between the aligned result and the present frame is evaluated. Ultimately, a simple threshold based on visual motion estimation magnitude was utilized to detect foreground positions. L. Kurnianggoro et al.in [15] The background was represented by zero visual motion estimation vectors. Past frames were registered using a homography matrix, and the visual motion estimation density between the registered result and the current frame was evaluated. K.P. Risha et al. introduced a hybrid approach to moving object detection in [16] that combines morphological operations with the optic flow method. An approach for detecting moving objects utilizing TV-L visual motion estimation in conjunction with SLIC segmentation was put out by X. Li et al. in [17]. the incoming frame undergoes segmentation via SLIC, followed by the identification of flow vectors using the TV-L method. Pixels displaying notable variations in visual motion estimation gradients are detected through the application of an empirically established threshold, derived from the delineations of segmented superpixel boundaries. A technique for pedestrian detection from moving vehicles utilizing visual motion estimation in combination with HOG was proposed by J. Hariyono et al. in [18]. To ensure reliable human region detection, the altered objects within moving regions are ultimately identified through morphological post-processing. In addition, utilizing a Linear Support Vector Machine to classify Histogram of Oriented Gradients features extracted from the objects[1].

5. LITERATURE REVIEW BASED ON DEEP LEARNING

Applications in numerous facets of computer vision are now being developed funda- mentally differently thanks to deep learning-based methodologies that have modern- ized the computing landscape. Object detection is one area where Deep Learning has had a substantial impact [19]. Following the introduction of deep learning, moving object identification methods based on CNNs saw a quick development Introducing an CNN approach for Detecting Moving Objects in Dynamic Background Settings. The proposed methodology comprises two deep learning architectures: the motion network and the appearance network. A-Net aims to identify moving things by their look, while M-Net identifies their motion. To recognize moving objects, the two framework are finally integrated. To identify moving objects, Y. Chen et al. Proposed a deep sequential learning framework [20]. Through a deep encoder-decoder network, this technique first extracts pixel-wise semantic characteristics. Then, introduced a novel attention-based Long Short-Term Memory architecture tailored for pixel-level variations. In order to smooth the foreground boundaries. Utilizing a spatial transformer network alongside a conditional random field (CRF) layer. Before the Faster CNN model, as suggested by S. Ren et al. [22] in [21], E.D. Tejada et al. Utilized PCP as a preprocessing technique for video background modeling, aiming to enhance moving object detection and classification. In [23], M. Babaee et al. suggested a Background subtraction technique for moving object segmentation using deep CNN. A fresh method for backdrop modeling is also suggested, and the network outputs are then post-processed using spatial-median filtering. In [24], T.N. Le et al. introduced a framework for object detection based related to spatial and temporal features. A new spatio-temporal deep (STD) feature, including both local and global features, has been proposed in this method. Temporal segment-based global feature extraction utilizing a block-based CNN[26], Local feature extraction from temporal segments using a region-based CNN [25]. Additionally, a STCRF is introduced to infer spatial relationships among frame regions. A distinctive MFCN design tailored for background subtraction was proposed by D. Zeng et al. in [27]. Theyhave demonstrated that using the deep-level characteristics from MFCN can improve the precision of foreground identification. A background subtraction method leveraging spatial features learned through CNN was introduced by M. Braham et al. in [28]. The method employs a scaled-down backdrop model, adjusted to match the background scenery of a specific scene, to train the removal of the background from an input image patch [29].

6. CONCLUSION

This paper reviews various moving object identification methods that have been put forth in the literature. The effectiveness hinges on the alignment of visual scenes and moving object velocities, as indicated by studies, despite noticeable outcomes from approaches reliant on temporal/frame differences in static environments. On the other hand, noise and dim illumination brought on by atmospheric phenomena primarily affect background modeling and visual motion estimation-based techniques. Moreover, background modeling algorithms can unveil comprehensive object data when the background is accessible. Conversely, deep learning methodologies have garnered momentum in object detection, demonstrating promising outcomes across diverse challenging scenarios and issues. However, the models are computationallyexpensive because they require much training of data.

REFERENCES

[1] Roy SD, Bhowmik MK. A comprehensive survey on computer vision based approaches for moving object detection. In2020 IEEE Region 10 Symposium (TENSYMP) 2020 Jun 5 (pp. 1531-1534). IEEE.

[2] X. Ji, W. Zhiqiang and Y. Feng, ”Effective vehicle detection technique for traffic surveillance systems,” Journal of Visual Communication and Image Representa- tion, Elsevier, Vol. 17, No. 3, pp. 647-658, 2006.

[3] Yilmaz, O. Javed and M. Shah, ”Object tracking: A survey,” ACM computing surveys (CSUR), Vol. 38, No. 4, pp. 13, 2006.

[4] M.F. Savaş, H. Demirel and B. Erkal, ”Moving object detection using an adaptive background subtraction method based on block-based structure in dynamic scene,” Optik, Elsevier, Vol. 168, pp. 605-618, 2018.

[5] M. Casares, S. Velipasalar and A. Pinto, “Light-weight salient foreground detec- tion for embedded smart cameras,” Jounal of Computer Vision and Image Understanding, Vol. 114, pp. 1223–1237, 2010.

[6] R. Kalsotra and S. Arora, ”Morphological based moving object detection with background subtraction method,” 4th International Conference on Signal Process- ing, Computing and Control (ISPCC), IEEE, pp. 305-310, 2017.

[7] Y. Zhou and S. Maskell, ”Moving object detection using background subtraction for a moving camera with pronounced parallax,”. In Sensor Data Fusion: Trends, Solutions, Applications (SDF), IEEE, pp. 1-6, 2017.

[8] J. Ye, T. Gao and J. Zhang, ”Moving object detection with background subtraction and shadow removal,” 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, pp. 1859-1863, 2012.

[9] B. Karasulu and S. Korukoglu, ” Moving object detection and tracking by using annealed background subtraction method in videos: Performance optimization,” Expert Systems with Applications, Elsevier, Vol. 39, No. 1, pp. 33-43, 2012.

[10] Z. Xu, D. Zhang and L. Du, ”Moving Object Detection Based on Improved Three Frame Difference and Background Subtraction,” 2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), IEEE, pp. 79-82, 2017.

[11] Y. Yang, Q. Zhang, P. Wang, X. Hu, and N. Wu, ”Moving object detection for dynamic background scenes based on spatiotemporal model,” Advances in Multimedia, Hindwai, 2017.

[12] J. Guo, J. Wang, R. Bai, Y. Zhang and Y. Li, ”A New Moving Object Detection Method Based on Frame-difference and Background Subtraction,” IOP Conference Series: Materials Science and Engineering, IOP Publishing, Vol. 242, No. 1, pp. 012115, 2017.

[13] J. Zhang, J. Cao and B. Mao, ”Moving object detection based on non- parametricmethods and frame difference for traceability video analysis, Procedia Computer Science, Elsevier, Vol. 91, pp. 995-1000, 2016.

[14] J. Huang, W. Zou, J. Zhu and Z. Zhu, ”Visual motion estimation Based Real-time Mov-ing Object Detection in Unconstrained Scenes,” arXiv preprint arXiv:1807.04890, 2018.

[15] J. Son, I. Jung, K. Park and B. Han, “Tracking-by-segmentation with online gra- dient boosting decision tree,” IEEE International Conference on Computer Vision, pp. 3056-3064, 2015.

[16] K.P. Risha and A.C.Kumar, ”Novel method of detecting moving object in video,” Procedia Technology, Elsevier, Vol. 24, pp. 1055-1060, 2016.

[17] X. Li and C. Xu, “Moving object detection in dynamic scenes based on opti-cal flow and superpixels,” 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, pp. 84-89, 2015.

[18] J. Hariyono, V.D. Hoang and K.H. Jo. ”Moving object localization using visual motion estimation for pedestrian detection from a moving vehicle,” The Scientific World Journal 2014, 2014.

[19] K. Fragkiadaki, P. Arbelaez, P. Felsen and J. Malik, ”Learning to segment movingobjects in videos,” IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 4083-4090, 2015.

[20] Y. Chen, J. Wang, B. Zhu, M. Tang and H. Lu, ”Pixel-wise deep sequence learning for moving object detection,” IEEE Transactions on Circuits and Systems for Video Technology, 2017.

[21] E.D. Tejada and P.A. Rodriguez, ”Moving object detection in videos using prin- cipal component pursuit and convolutional neural networks,” 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, pp. 793-797,2017.

[22] S. Ren, K. He, R. Girshick and J. Sun, “Faster R-CNN: Towards Real- Time Object Detection with Region Proposal Networks,” Advances in neural information processing systems, pp. 91-99, 2015.

[23] M. Babaee, D.T. Dinh and G. Rigoll, ”A deep convolutional neural network for video sequence background subtraction,” Pattern Recognition, Elsevier, Vol. 76,pp. 635-649, 2018.

[24] T.N. Le and A. Sugimoto, ”Video salient object detection using spatiotemporal deep features,.”IEEE Transactions on Image Processing, Vol. 27, No. 10, pp. 5002-5015, 2018.

[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 580–587, 2014.

[26] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotem- poral features with 3D convolutional networks,” IEEE international conference on computer vision, IEEE, pp. 4489–4497, 2015.

[27] D. Zeng and M. Zhu, ”Background Subtraction Using Multiscale Fully Convolu- tional Network,” IEEE Access, Vol. 6, pp. 16010-16021, 2018.
[28] T.H. Lin and C.C. Wang, ”Deep learning of spatio-temporal features with geometric-based moving point detection for motion segmentation,” 2014 IEEE International Conference on Robotics and Automation (ICRA), IEEE, pp. 3058- 3065, 2014.
[29] Guo, Jing-Ming, Chih-Hsien Hsia, Yun-Fu Liu, Min-Hsiung Shih, Cheng-HsinChang, and Jing-Yu Wu. “Fast background subtraction based on a multilayer codebook model for moving object detection.” IEEE Transactions on Circuits andSystems for Video Technology 23, no. 10 (2013): 1809-1821.