Received: 29 November 2022 | Revised: 22 March 2023 | Accepted: 13 April 2023 | Published online: 24 April 2023

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 3(2) 123–130
DOI: 10.47852/bonviewAIA3202553

BON VIEW PUBLISHING

Spatio-Temporal FFT-Based Approach for Arbitrarily Moving Object Classification in Videos of Protected and Sensitive Scenes

Maryam Asadzadehkaljahi[1], Arnab Halder[2,*], Palaiahnakote Shivakumara[3] and Umapada Pal[2]

[1]*PromiseQ, Germany*

[2]*Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, India*

[3]*Computer Science and Information Technology, University of Malaya, Malaysia*

Abstract: Arbitrary moving object detection including vehicles and human beings in the real environment, such as protected and sensitive areas, is challenging due to the arbitrary deformation and directions caused by shaky camera and wind. This work aims at adopting a spatio-temporal approach for classifying arbitrarily moving objects. The proposed method segments foreground objects from the background using the frame difference between the median frame and individual frames. This step outputs several different foreground information. The mean of foreground images is computed, which is referred to as the mean activation map. For the mean activation map, the method employs the fast Fourier transform, which outputs amplitude and frequencies. The mean of frequencies is computed for moving objects in using activation maps of temporal frames, which is considered as a frequency feature vector. The features are normalized to avoid the problems of imbalanced features and class sizes. For classification, the work uses 10-fold cross-validation to choose the number of training and testing samples and the random forest classifier is used for the final classification of arbitrary moving and static videos. For evaluating the proposed method, we construct our dataset, which contains videos of static and arbitrarily moving objects caused by shaky cameras and wind. The results of the video dataset show that the proposed method achieves the state-of-the-art performance.

Keywords: moving objects detection, vehicles movements detection, shaky camera detection, subtraction approach, arbitrarily moving objects detection

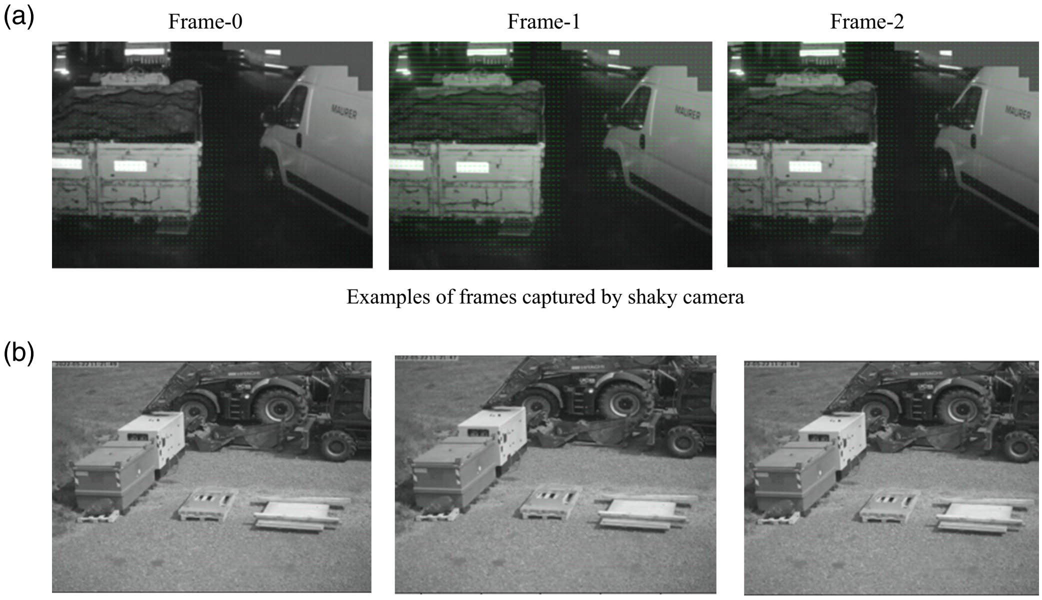
1. Introduction

Automation is common for all fields to make the system cost-effective and accurate and to prevent human error during the night, especially in protected and sensitive areas. For protecting and monitoring such areas from robbery, stealing, and tampering, there is a need for developing a powerful surveillance system. However, detecting intruders including humans and vehicles is not easy at night because of poor quality and lighting effects. In addition, arbitrary movements of objects, such as a tree leaf in the same scene due to wind and a shaky camera, make detecting moving objects (actual) more complex and challenging. The sample frames captured by a shaky and non-shaky camera are shown in Figure 1(a) and (b), respectively. It is observed from Figure 1(a) and (b) that the frames are suffering from poor quality and the objects are not visible properly including human movements. There are several methods proposed in the literature for static and moving object detection [1–8]. However, these methods focus on the video captured in the day with high quality. In addition, the methods are good for detecting objects which move in a particular direction and speed but not the video containing arbitrary movements and directions. Therefore, there is a dearth to develop a new method for the classification of static and arbitrary movements of objects, such as leaves in a tree and objects' movements due to shaky cameras in the video in real-time environments.

Hence, this work aims at developing a method for the classification of static video and arbitrary moving video. The proposed method works based on the fact that normal moving objects exhibit regular patterns like uniform direction, speed, and shapes while arbitrary moving objects exhibit irregular patterns like non-uniform direction, speed, and deformable shapes. These observations motivated us to explore the fast Fourier transform (FFT)-based approach to extract frequency feature vectors from the mean of foreground images in this work. This makes sense because the amplitude of frequencies changes according to the direction of objects and movements. Thus, the following are the contributions of the proposed work. (i) This is the first work for classifying arbitrary moving objects caused by shaky cameras and leaves caused by wind. (ii) The spatio-temporal information for extracting Fourier frequency vectors is new compared

*Corresponding author: Arnab Halder, Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, India. Email: Arnab.Halder@student.uts.edu.au

Figure 1
Illustrating the challenges of frames captured by shaky and non-shaky camera for classification



Examples of frames captured by non-shaky camera

to the existing methods. (iii) The proposed method is tested on a real-time environment like day and night videos.

The structure of the paper is as follows. The review of state-of-the-art methods for the classification of static and moving objects is presented in Section 2. In Section 3, the method for foreground and background separation, FFT for extracting frequency feature vector, and the approach for classification of static video and arbitrary moving video are described. Section 4 provides a discussion of several experiments to validate the proposed and existing methods. The conclusion and summary are presented in Section 5.

2. Related Work

Several methods are developed in the past for moving object detection and foreground object separation in the video. Since these methods are relevant to arbitrary object movement detection, we review the latest papers. Boufares et al. [2] propose a method for moving object detection using temporal difference and OTSU thresholding techniques. The approach uses input frame difference at the pixel level for detecting moving objects in the videos. Wang et al. [8] develop a model for moving object detection using frame difference and an algorithm for teaching video. The method uses the OTSU thresholding approach and median filter for moving object detection. Sadkhan et al. [9] aim to detect moving objects and track moving objects in the video. The approach uses a subtraction model for moving object detection, texture, shape, and color-based features are extracted for classification, and detected moving object tracking is done by the kernel and point-based approaches. Rahiminezhad et al. [10] explore adaptive coefficient and background subtraction for detecting moving objects in the video. The method focuses more on hardware implementation to fix it in a real-time environment. Sultana et al. [11] develop a model based on adversarial regularization for moving object detection in complex scenes. The approach is capable of handling partial occlusion and poor quality images.

Tang and Liu [12] propose a method for moving object detection using ghosts and shadows. The key idea of the method is to explore visual background extraction. The OTSU thresholding is used to detect moving objects in the video. Shu et al. [13] focus on developing a method for small-moving object detection in video. The method involves event-based moving object detection and for tracking, the work uses registration and foreground enhancement models. Wang et al. [14] present a model for multi-scale moving object detection based on spatio-temporal online matrix factorization. The approaach uses temporal difference with respect to frames, as well as motion difference and partial spatial motion information. Kim et al. [15] develop a model for moving object detection based on instant background modeling. The approach uses spatio-temporal information and instant background modeling which consists of inpainting and super pixels to enhance the fine details in the images. Huang et al. [16] develop a method for moving object detection based on independent component analysis. The approach uses a frame difference model for foreground moving object detection. The features extracted by frame difference are combined with spatial information to achieve the best results. Deng et al. [17] use super-resolution and optical image data for moving object detection. The approach uses Convolutional Neural Network (CNN) for feature extraction and fusion. Kovalenko et al. [18] use image sequences for moving object detection based on the thresholding technique. The approach estimates the deformation field using a stochastic gradient procedure.

It is observed from the above review that none of the methods aim at detecting the objects which move in an arbitrary direction in the video. Most methods focus on normal moving object detection-based subtraction models. Since the existing methods use specific properties of moving object detection, the methods may not be suitable for arbitrary moving object detection in the video. Furthermore, the scope of the existing methods is limited to day video but not night video where one can expect enormous degradations. Therefore, detecting arbitrarily moving object detection in both day and night video is an open challenge. Thus, this work aims at proposing a new method for classifying arbitrary moving video and static video.

3. Proposed Method

This work considers video capturing at the rate of 30 frames per second as input for the classification of video captured by the shaky camera and non-shaky camera. As mentioned in the previous section, the proposed work exploits the observation that as the direction of objects changes, it affects the motion of objects and their direction. To extract this observation, the proposed work separates the foreground from the background using a frame difference approach. The median of the 30 frames is computed, and it is considered as a reference frame of the input video. Each frame is subtracted from the reference frame, resulting in foreground regions. To study the effect of arbitrarily moving and static moving objects, the proposed method performs the FFT, which is an implementation of discrete Fourier transformation on the mean of foreground images (referred to as mean activation). This step outputs the amplitude of frequencies for all 30 frames. With the frequencies, the method obtains a single combined frequency by getting the average of all frequencies, which is a miniature representation of all the foreground motions in the video. We believe that for the frame with static objects, the range of combined frequency lies in a low range while the high range for the frame with moving objects. This makes a difference in classifying the frame with static and arbitrarily moving video. For each input video, the method outputs a frequency vector for the classification of arbitrary moving and static videos.

It is true that since there are no constraints on the number of objects in the frames, the size of the frames, and the size of the classes, there is a high chance of including features that do not contribute much to the classification. Therefore, for classification, the proposed work uses 10-fold cross-validation which provides the number of training and testing samples automatically. In this process, the dataset is partitioned into 10 random folds, and each fold is representing a miniature version of the overall dataset, every time the training is done on the 9 samples and evaluation is done upon the rest samples, it changes iteratively. When the extracted features are capable of discriminating shaky and non-shaky videos, motivated by the conventional machine learning algorithms which do not require large number of samples for successful classification, we propose to use random forest (RF) classifier for classification in this work [19–24]. The RF classifier is a well-known technique for classification, which can handle imbalanced features and noise features and can avoid overfitting problems. The pipeline of the proposed method can be seen in Figure 2.

3.1. FFT-based approach

For input of 30 frames of video, the proposed method obtains background by performing the median operation on 30 frames. The median of 30 frames is considered as the background of the input video as it is illustrated in Figure 3(a) for both shaky and non-shaky camera frames. It is noted from Figure 3(a) that image is degraded due to object movements and in the case of a non-shaky camera, the quality of the median image improves compared to the original frames. The individual frames are subtracted from the background information to obtain foreground information. The sample result is shown in Figure 3(b) for both shaky and non-shaky camera frames, where one can see the foreground image contains some edges of moving objects because the camera is not stable in this video. However, for the non-shaky frames, the foreground image contains nothing because there are no moving pixels in the video. For estimating the motion, we find the mean of foreground images, which contains the average motion information of frames, which is referred to as mean activation. The mean activation of input of shaky and non-shaky camera frames can be seen in Figure 3(c). To get the motion information, the proposed work employs the FFT on the mean activation as defined in Equations (1) to (3). This outputs the frequency of amplitudes of all the components (motions in the foreground), from which we can get the average of the frequencies. This process outputs a vector with frequencies for 30 frames of video, which is considered a frequency feature vector for classification as defined in Equations (1) to (3).

The FFT $y[k]$ of length N of the $length - N$ sequence $x[n]$ is defined as follows:

$$y[k] = \sum_{n=0}^{N-1} e^{-2\pi j \frac{kn}{N}} x[n] \tag{1}$$

And the inverse transform is as follows:

$$x[n] = \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi j \frac{kn}{N}} y[k] \tag{2}$$

The feature map of the proposed method is as follows:

$$\frac{1}{N} \sum y[k] = \frac{1}{N} \left(\sum_{n=0}^{N-1} e^{-2\pi j \frac{kn}{N}} x[n] \right)^2 \tag{3}$$

The effect of the above steps can be seen in Figure 4(a) and (b), where the combined frequencies lie in a small range for non-shaky camera video while the frequencies lie in a very high range in the case of shaky camera video. This is the advantage of the frequency domain for the classification of video containing moving objects and static objects. For the classification, the method is trained with samples provided by 10-fold cross-validation. At the same time, the same approach is used for choosing the testing samples. For the final classification, the method uses RF classifier which considers the features corresponding to training and testing samples for classification. The reason to use RF classifier is that this is a well-known technique for handling imbalanced feature vector sizes and if the feature vector contains noisy features.

The Random Forest method is a special type of decision tree, created by using randomly selected variables for a dataset, extracted by bootstrap sampling the classification is performed which is based on the majority of the decisions. Moreover, the contribution of each variable to data classification can be obtained using the created decision trees; furthermore, the importance of each variable can be determined. A combination of classifier trees represents a RF classifier. One of the finest approaches to representing input variables in the form of trees makes a forest-like structure. Input data are represented in trees and each tree specifies a class label. RF depends on its error rate. Error rate signifies in two directions. The first one is the correlation between trees, and the second one is the strength of the tree.

To show the effectiveness of the proposed frame difference over conventional frame difference, which computes consecutive frame difference for moving object detection, we estimate the amplitude range for the number of shaky and non-shaky camera frames as illustrated in Figure 4(b). In our method, the median frame is considered as a reference frame out of 30 frames of video. For the shaky video, one can expect more noise in the beginning and at the end of 30 frames because the shaky camera does not affect the middle frame much compared to the beginning and end frames. Therefore, the median frame provides stable content compared to the other temporal frames. However, in the case of the conventional approach, the consecutive frame difference is prone to error, which can attribute more as the number of frames passes. This is illustrated

Figure 2
Block diagram of the proposed method

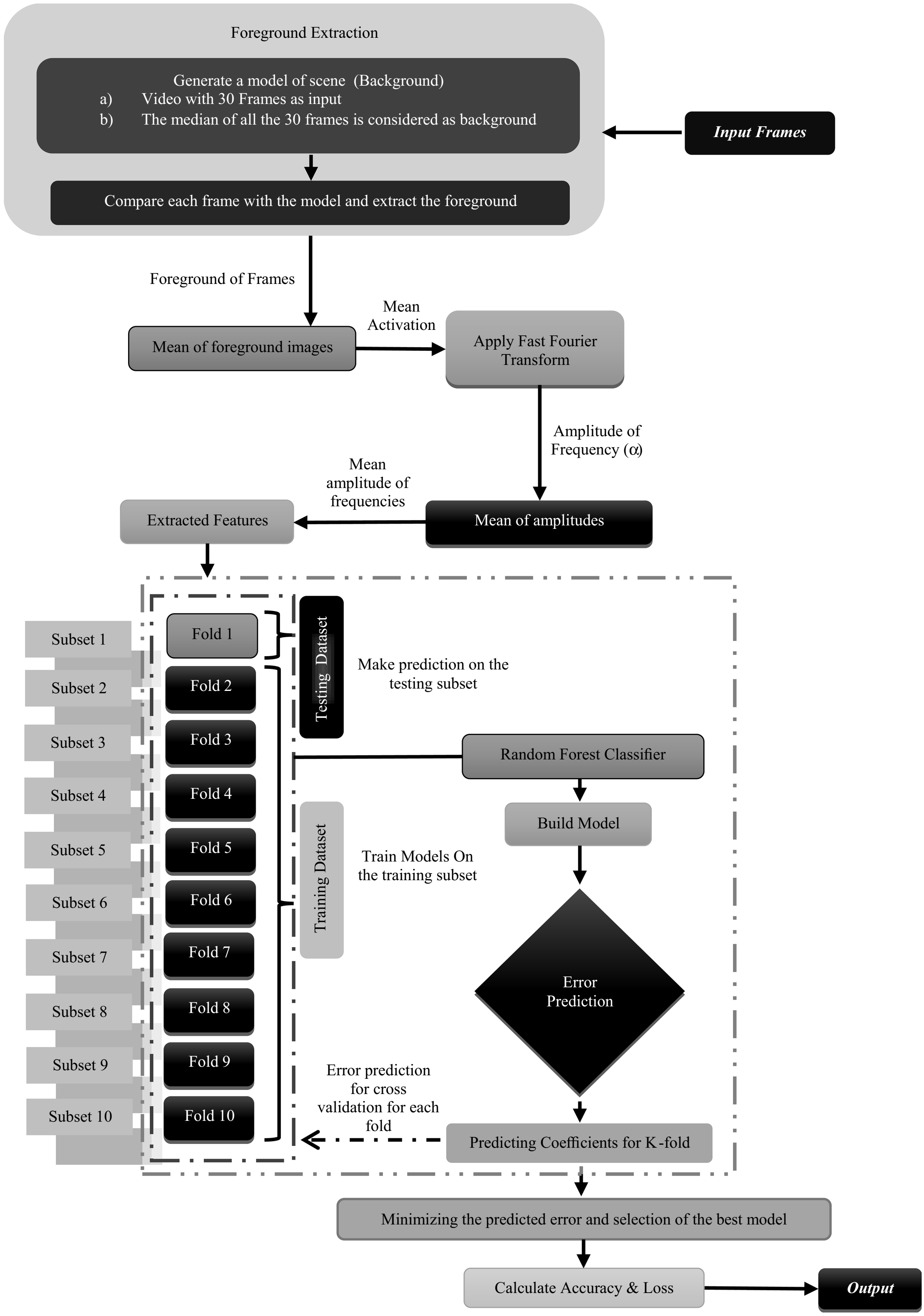
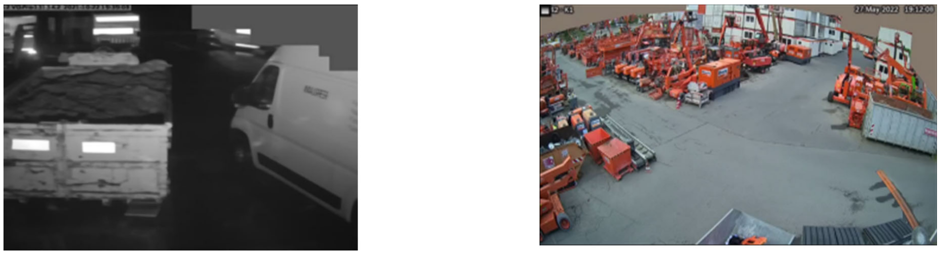


Figure 3
The steps for extracting features using the foreground and background information

Shaky camera Non-shaky camera

(a)



The results of background model (median of 30 frames)

(b)

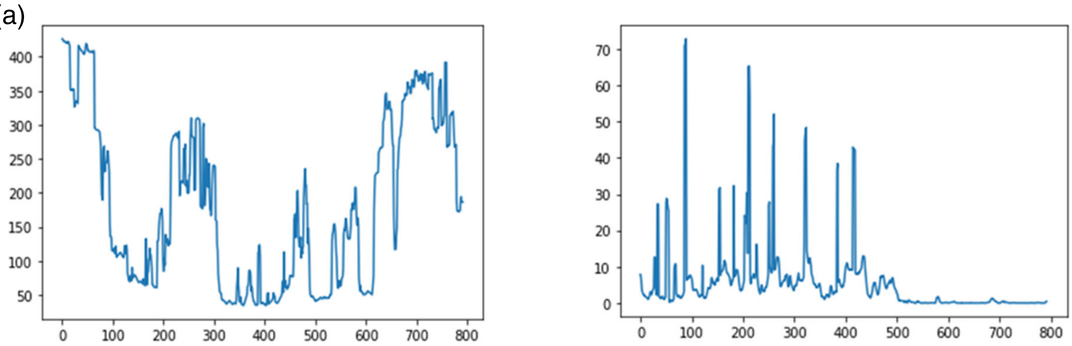
Foreground of shaky and non-shaky frames

(c)

Mean activation (Mean of foreground images of shaky and non-shaky video)

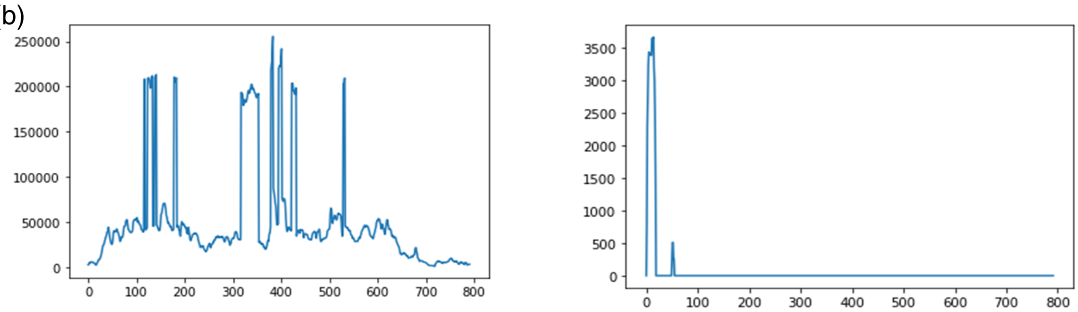
Figure 4
The extracted frequency from shaky and non-shaky camera

(a)



The combined frequency of amplitudes of mean activation of shaky and non-shaky

(b)



The combined frequency of amplitudes of mean activation of shaky and non-shaky using conventional frame difference method in Figure 4(b), where one can see more noise in the case of a shaky camera compared to the proposed method shown in Figure 4(a). For non-shaky video, the conventional approach lost vital information, as shown in Figure 4(a), while the proposed frame difference preserves the vital information. It is evident from the range of frequencies used on the Y axis, as shown in Figure 4(a) and (b), where a high range for the conventional frame difference approach while a small range for the proposed frame difference.

4. Experimental Results

Since there is no standard dataset for experimentation, we constructed our dataset to evaluate the proposed method. Our dataset consists of 2959 videos, of which 817 are shaky camera samples and 2142 non-shaky camera samples. The videos are captured at the rate of 30 frames per second. Furthermore, the dataset includes the video captured day and night of protected and sensitive areas, including indoor and outdoor scenes. Therefore, the dataset is complex and challenging for the classification of arbitrary moving video and static video.

To show the effectiveness of the proposed method, we implemented two state-of-the-art methods, namely Boufares et al. [2] and Wang et al. [8], that use temporal difference and OTSU thresholding for moving object detection in the video for comparative study. In addition, Rahiminezhad et al. [10] use subtraction method for moving object detection. The reason to choose the above methods is that the objective of the method is the same as the proposed method. In addition, the methods focus on detecting moving and non-moving objects for the classification of video, which is similar to the idea proposed in this work.

For evaluating performance of the proposed and existing methods, we consider the following standard measures: precision, recall, F1-score, and accuracy.

Accuracy: In a given dataset consisting of (TP + TN) data points, the accuracy is equal to the ratio of total correct predictions (TP + TN + FP + FN) by the classifier to the total data points. The model's accuracy can be calculated as defined in Equation (4).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}\ 0.0 < \text{Accuracy} < 1.0 \quad (4)$$

where TP is the true positive; TN is the true negative; FP is the false positive; and FN is the false negative.

Precision: This is equal to the ratio of the true positive (TP) samples to the sum of TP and false positive (FP) samples, which is defined as in Equation (5).

$$\text{Precision} = \frac{TP}{\text{TP} + \text{FP}} \quad (5)$$

Recall: Recall is the evaluation metrics equal to the ratio of the TP data samples to the sum of TP and false negative (FN) data samples, which is defined as in Equation (6).

$$\text{Recall} = \frac{TP}{\text{TP} + \text{FN}} \quad (6)$$

F1 Score: F1-score is equal to the harmonic mean of recall value and precision value. The F1-score gives the perfect balance between precision and recall, thereby providing a correct evaluation of the model's performance. F1-score can be calculated as defined in Equation (7).

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

4.1. Ablation study

In this work, we used the RF classifier for the classification of arbitrary moving video and static video. To test the contribution of the RF classifier, we compare the performance of the proposed method random classifier with other well-known classifiers. For this experiment, the proposed work calculates accuracy for replacing the RF classifier with the following classifiers on our dataset, and the results are reported in Table 1.

Decision tree classifier: A tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome. There are two nodes in decision trees, which are the decision node and leaf mode. Decision nodes are used to make any decision and can have multiple branches, whereas leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed based on the features of the given dataset and are a graphical representation to get all the possible solutions to a problem based on given conditions.

Gradient boosting classifier: It is one of the boosting algorithms used to minimize the bias error of the model. Gradient boosting algorithms can be used for predicting not only continuous target variable (as a regressor) but also categorical target variable (as a classifier). When it is used as a regressor, the cost function is the mean square error and when it is used as a classifier, then the cost function is log loss.

Support vector classifier (SVC): SVC is usually preferred for data analysis because of its computational capability within a very less time frame. This classifier works on the decision boundary concept recognized as a hyperplane. The hyperplane is used to classify the input data into the required target group. The SVC is not affected by overfitting problem and makes it more reliable.

Logistic regression classifier: This classifier is based on the probability of outcome from the input process data. Binary logistic regression is generally preferred in machine learning techniques for dealing with binary input variables. To categorize the class into specific categories, sigmoid function is utilized.

It is noted from Table 1 that the proposed method with a RF classifier reports the best accuracy compared to all other classifiers. Therefore, one can infer that the proposed method with a RF classifier is suitable for this work. When we compare the results of

Table 1
Accuracy for the proposed method with different classifiers and only FFT

Classifiers	Random forest	Decision tree	SVC	Logistic regression	Gradient boosting	Only FFT
Accuracy	0.82	0.688652	0.671846	0.697283	0.655512	0.76

Table 2
Performance of the proposed and existing methods for classification of arbitrary moving video

Methods	Proposed						Boufares et al. [2]						Wang et al. [8]						Rahiminezhad et al. [10]					
Classes	Shaky			Non-shaky			Shaky			Non-shaky			Shaky			Non-shaky			Shaky			Non-shaky		
Measures	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Results	0.80	0.71	0.83	0.85	0.77	0.81	0.60	0.65	0.63	0.63	0.58	0.60	0.59	0.73	0.65	0.65	0.50	0.57	0.58	0.64	0.60	0.62	0.56	0.59
Accuracy	**0.82**						0.62						0.62						0.60					

Note: The bold value signifies the accuracy of proposed method as compared to existings.

the proposed method with other classifiers, the logical regression approach is better than other classifiers. This is because the logical regression can cope with the imbalanced feature vectors, and it avoids the overfitting problems. However, other methods are good when the data are simple but not for non-linear data.

Only FFT: For this experiment, to show that FFT contributes significantly to classification, an average of 30 frames is supplied to FFT directly to extract features without separating the foreground from the median frame. The accuracy of the experiments is reported in Table 1. It is noted from Table 1 that, alone, FFT is not sufficient to achieve the best results (0.82 of the proposed method). However, one can infer that the features extracted from FFT contribute to significant classification.

4.2. Experiments on classification of arbitrarily moving objects

Quantitative results of the proposed and existing methods on our dataset are reported in Table 2, where it is noted that the proposed method is better than the three existing methods in terms of accuracy. The reason for achieving the best accuracy by the proposed method is that the steps of separating background and foreground regions in the images, FFT-based approach for frequency feature vector extraction, and the use of RF classifiers are generalized steps and hence work well for diversified videos. But the existing methods are developed for day images, the methods do not work well for night images. Therefore, the existing methods report poor results compared to the proposed method. In addition, the existing methods were developed to handle uniform moving objects in the video but not arbitrary moving objects in the video.

5. Conclusion and Future Work

For the classification of arbitrary moving video, we have proposed a new method based on background and foreground separation, FFT-based feature vector, and the RF classifier. For defining the background in the video frames, the proposed work uses a median of 30 frames as background for the input video. The frames are subtracted from the median image to obtain foreground information. The FFT has been employed on the mean of foreground frames. This step outputs a frequency feature vector for classification. For choosing the number of training and testing samples automatically, the proposed work uses 10-fold cross-validation. For classification, the proposed work uses the RF classifier by feeding the frequency feature vector as input. Experimental results on our dataset and comparative study with the existing methods show that the proposed method outperforms the existing methods in terms of accuracy. However, when the input video contains static, arbitrary moving objects and uniform moving objects, the performance of the proposed work degrades. This will be three-class classification problems, which is beyond the scope of the work. We plan to address this limitation in the near future by exploring the randomness of direction, speed, deformable shape, and the content of the video.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

Palaiahnakote Shivakumara is the Editor-in-Chief and Umapada Pal is an Advisory Board Member for *Artificial Intelligence and Applications*, and were not involved in the editorial review or the decision to publish this article. The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data available on request from the corresponding author upon reasonable request.

Author Contribution Statement

Maryam Asadzadehkaljahi: Conceptualization, Investigation, Resources. **Arnab Halder:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization. **Palaiahnakote Shivakumara:** Methodology, Validation, Writing – original draft, Supervision. **Umapada Pal:** Writing – review & editing, Supervision.

References

[1] Arjun, D., Indukala, P. K., & Menon, K. U. (2020). Development of a framework for effective surveillance and human intrusion detection in border regions covered with dry leaves. In *2020 Fourth International Conference on Inventive Systems and Control*, 106–109. https://doi.org/10.1109/ICISC47916.2020.9171198

[2] Boufares, O., Boussif, M., & Aloui, N. (2021). Moving object detection system based on the modified temporal difference and OTSU algorithm. In *2021 18th International Multi-Conference on Systems, Signals & Devices*, 1378–1382. https://doi.org/10.1109/SSD52085.2021.9429516

[3] Chandra, N., & Panda, S. P. (2021). A human intruder detection system for restricted sensitive areas. In *2021 2nd International Conference on Range Technology*, 1–4. https://doi.org/10.1109/ICORT52730.2021.9582099

[4] Kim, H., Kim, P., & Kim, H. J. (2020). Moving object detection for visual odometry in a dynamic environment based on

occlusion accumulation. In *2020 IEEE International Conference on Robotics and Automation*, 8658–8664. https://doi.org/10.1109/ICRA40945.2020.9196767

[5] Lenac, K., Cuzzocrea, A., & Mumolo, E. (2021). A novel genetic scan-matching-based registration algorithm for supporting moving objects tracking effectively and efficiently. *IEEE Access*, *9*, 91741–91753. https://doi.org/10.1109/ACCESS.2021.3091520

[6] Ramesh, B., Zhang, S., Yang, H., Ussa, A., Ong, M., Ochard, G., & Xiang, C. (2021). e-TLD: Event-based framework for dynamic object tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(10), 3996–4006. https://doi.org/10.1109/TCSVT.2020.3044287

[7] Vndana, G. S., Paradhasaradhi, B., & Srihari, P. (2022). Intruder detection and tracking using 77GHz FMCW radar and camera data. In *2022 IEEE International Conference on Electronics, Computing and Communication Technologies*, 1–6. https://doi.org/10.1109/CONECCT55679.2022.9865707

[8] Wang, Z., Wang, J., & Wang, N. (2021). Moving object detection and marking based on frame difference and train algorithm for teaching video. In *2021 IEEE 15th International Conference on Anti-Counterfeiting, Security, and Identification*, 61–65. https://doi.org/10.1109/ASID52932.2021.9651485

[9] Sadkhan, A. S. B., Talebiyan, S. R., & Farzaneh, N. (2021). An investigate on moving object tracking and detection in image. In *2021 1st Babylon International Conference on Information Technology and Science*, 69–75. https://doi.org/10.1109/BICITS51482.2021.9509887

[10] Rahiminezhad, A., Tavakoli, M. R., & Sayedi, S. M. (2022). Hardware implementation of moving object detection using adaptive coefficient in performing background subtraction algorithm. In *2022 International Conference on Machine Vision and Image Processing*, 1–5. https://doi.org/10.1109/MVIP53647.2022.9738764

[11] Sultana, M., Mahmood, A., & Jung, S. K. (2021). Unsupervised moving object detection in complex scenes using adversarial regularizations. *IEEE Transactions on Multimedia*, *23*, 2005–2018. https://doi.org/10.1109/TMM.2020.3006419

[12] Tang, M., & Liu, W. (2021). A moving object detection algorithm for removing Ghosting and shadow. In *2021 40th Chinese Control Conference*, 7207–7212. https://doi.org/10.23919/CCC52363.2021.9550513

[13] Shu, Y., Sui, Y., Zhao, S., Cheng, Z., & Liu, W. (2021). Small moving object detection and tracking based on event signals. In *2021 7th International Conference on Computer and Communications*, 792–796. https://doi.org/10.1109/ICCC54389.2021.9674247

[14] Wang, J., Zhao, Y., Zhang, K., Wang, Q., & Li, X. (2022). Spatio-temporal online matrix factorization for multi-scale moving object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(2), 743–757. https://doi.org/10.1109/TCSVT.2021.3066675

[15] Kim, W. J., Hwang, W., Lee, J., Woo, S., & Lee, S. (2022). AIBM: Accurate and instant background modelling for moving object detection. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 9021–9036. https://doi.org/10.1109/TITS.2021.3090092

[16] Huang, Y., Jiang, Q., & Qian, Y. (2021). A novel method for video moving object detection using improved independent component analysis. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(6), 2217–2230. https://doi.org/10.1109/TCSVT.2020.3023175

[17] Deng, Z., Cui, Z., & Cao, Z. (2022). Super resolution detection method of moving object based on optical image fusion with MMW radar. In *2022 IEEE International Geoscience and Remote Sensing Symposium*, 1900–1903. https://doi.org/10.1109/IGARSS46834.2022.9883395

[18] Kovalenko, R., Tashilinskii, A., & Tsaryov, M. (2021). Using threshold processing to moving object detection in the image sequence. In *2021 International Conference on Information Technology and Nanotechnology*, 1–4. https://doi.org/10.1109/ITNT52450.2021.9649334

[19] Bhardwaj, A., Dagar, V., Khan, M. O., Aggrwal, A., Alvarado, R., Kumar, M., . . . , & Proshad, R. (2022). Smart IoT and machine learning-based framework for water quality assessment and device component monitoring. *Environmental Science and Pollution Research*, *29*(30), 46018–46036. https://doi.org/10.1007/s11356-022-19014-3

[20] Chhabra, M., Ravulakollu, K. K., Kumar, M., Sharma, A., & Nayyar, A. (2023). Improving automated latent fingerprint detection and segmentation using deep convolutional neural network. *Neural Computing and Applications*, *35*(9), 6471–6497. https://doi.org/10.1007/s00521-022-07894-y

[21] Chithaluru, P., Stephan, T., Kumar, M., & Nayyar, A. (2022). An enhanced energy-efficient fuzzy-based cognitive radio scheme for IoT. *Neural Computing and Applications*, *34*(21), 19193–19215. https://doi.org/10.1007/s00521-022-07515-8

[22] Chittora, P., Chaurasia, S., Chakrabarti, P., Kumawat, G., Chakrabarti, T., Leonowicz, Z., . . . , & Bolshev, V. (2021). Prediction of chronic kidney disease-A machine learning perspective. *IEEE Access*, *9*, 17312–17334. https://doi.org/10.1109/ACCESS.2021.3053763

[23] Kaushik, K., Bhardwaj, A., Kumar, M., Gupta, S. K., & Gupta, A. (2022). A novel machine learning-based framework for detecting fake Instagram profiles. *Concurrency and Computation: Practice and Experience*, *34*(28), e7349. https://doi.org/10.1002/cpe.7349

[24] Solanki, Y. S., Chakrabarti, P., Jasinski, M., Leonowicz, Z., Bolshev, V., Vinogradov, A., . . . , & Nami, M. (2021). A hybrid supervised machine learning classifier system for breast cancer prognosis using feature selection and data imbalance handling approaches. *Electronics*, *10*(6), 699. https://doi.org/10.3390/electronics10060699

How to Cite: Asadzadehkaljahi, M., Halder, A., Shivakumara, P., & Pal, U. (2025). Spatio-Temporal FFT-Based Approach for Arbitrarily Moving Object Classification in Videos of Protected and Sensitive Scenes. *Artificial Intelligence and Applications*, *3*(2), 123–130. https://doi.org/10.47852/bonviewAIA3202553