Mahayuddin, Zainal Rasyid; Saif, A

Moving Object Detection Using Semantic Convolutional Features

Published in:
Journal of Information System and Technology Management

DOI:
10.35631/JISTM.729003

Published: 31/12/2022

Document Version
Publisher's PDF, also known as Version of record

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666

Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT (JISTM)
www.jistm.com

MOVING OBJECT DETECTION USING SEMANTIC CONVOLUTIONAL FEATURES

Zainal Rasyid Mahayuddin[1*], A F M Saifuddin Saif [2*]

1 Faculty of Information Science and Technology, University Kebangsaan Malaysia, Malaysia
Email: zainalr@ukm.edu.my
2 School of Engineering, Aalto University, Finland
Email: a.f.saif@aalto.fi
* Corresponding Author

Article Info:

Article history:
Received date: 20.07.2022
Revised date: 05.10.2022
Accepted date: 20.10.2022
Published date: 01.12.2022

To cite this document:

Mahayuddin, Z. R., & Saif, A. F. M. S. (2022). Moving Object Detection Using Semantic Convolutional Features. *Journal of Information System and Technology Management, 7* (29), 24-41.

DOI: 10.35631/JISTM.729003

Abstract:

Moving object detection from aerial images remains an unsolved problem in computer vision research domain. Detection results are not precise due to blurry aerial images, thin edges and noise. Various methods were previously proposed for moving object detection which could not provide robust results due many challenges, i.e., noise, motion detection, lack of appropriate features, lack of effective classification approach, complex background and variations in illumination. This research proposes an efficient method for moving object detection using convolutional semantic features from VGG-16 to use motion patterns to facilitate detection in each frame and provides smaller area as region of interest. Proposed method reduces probability motion intensity information getting lost in case of same coloured object in the background and thus minimizes background complexity. After that, proposed method performs semantic features distance measurement to calculate linear distances in each frame. In this context, if there is any frame loss due to noise or illumination variation, proposed method uses Kalman filter to process that frame by illuminating noise. Finally, decision for final moving detection is determined using random forest classifier from semantic convolutional feature vector by generating a set of probabilities for each class. Experimental results show that the proposed method can detect moving objects efficiently, which in turn will decrease the operating time and increase the detection rate compared to previous research methods.

Keywords:

Object Detection, Feature Extraction, Fourth Industrial Revolution

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

Introduction

Significant characteristic extraction from aerial images from unmanned aerial vehicles (UAVs) remains an unsolved issue in the field of computer vision, and efficient and significant pictorial information selections have not been adequately ad-dressed in previous research. However, in real environments, moving object extraction becomes challenging due to various constraint factors mentioned in figure1. This research proposes an efficient method to facilitate fast moving object detection in the frame achieved via a two-frame difference method and a standalone segmentation approach using adaptive threshold optimization.

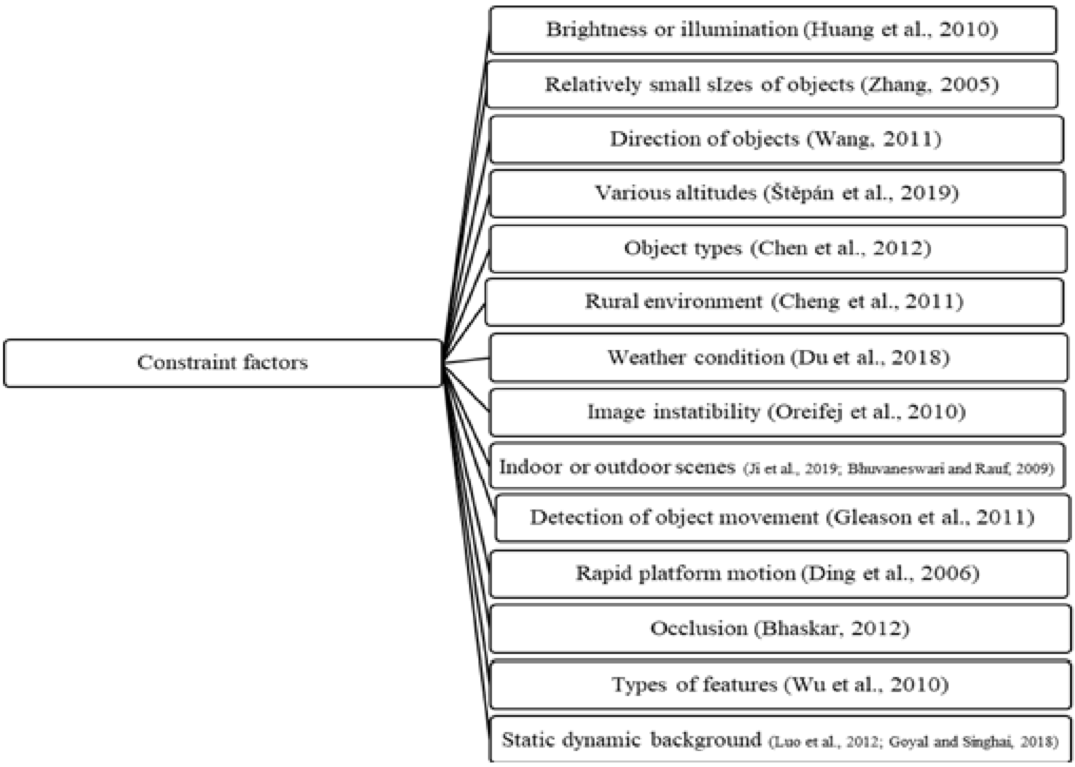

Figure 1: Constraint Factors For Moving Object Detection From Aerial Types Of Images.

Existing methodologies for moving object detection can be grouped into four categories, i.e., motion detection (Huang et al., 2010), deep learning techniques (Zhao et al., 2019; Lee et al., 2017; Benjdira et al., 2019) frame difference methods (Jiang et al., 2009; Cheraghi and Sheikh, 2012) and segmentation methods (Zhang, 2005; Xingbao et al., 2011). Motion is the core vital issue for moving object detection which needs to be accurately detected using efficient methods. However, existing motion detection methods are affected by several practical problems for aerial images, such as variation of object or camera platform, complex manipulation of motion parameters (Saif et al., 2014). Several deep learning techniques were recently proposed for moving object detection such as Faster R-CNN, You Only Look Once version 3 (YOLOv3) which are extremely computationally complex and depended on high-end Graphics Processing Units (GPUs) (Zhao et al., 2019; Lee et al., 2017; Benjdira et al., 2019). Frame difference-based approaches cause the moving object to be represented as pieces due to the objects colour homogeneity and grab the motion information. Frame difference detects the

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003
pixels with motion but cannot obtain the complete shape of the object. Moreover, existing segmentation methods have been used to determine the candidates of moving objects in each video frame. Image segmentation such as spatiotemporal segmentation (Zhang, 2005) and feature points extraction-based segmentation (Yang et al., 2012) extracts a more complete shape of the objects (sometimes the background is also included). However, the segmentation approach alone does not have the ability to distinguish moving regions from the static background. In previous research, various types of features were considered for the detection of moving objects, e.g., gradient measurement (Yang et al., 2012), Harris corners (Ibrahim et al., 2010), edge maps (Wang, 2011), and spatial edges (Wang et al., 2020). However, none of these previous studies has provided reliable validation, especially for significant feature selection to reduce errors. Most of these approaches were proven to yield good results, but they are not generally applicable to images that were captured from various altitudes with various constraint factors mentioned in Fig. 1 from moving cameras objects, such as UAVs, which demands an efficient method for moving object detection.

This research proposes an efficient method where frame difference is integrated with significant pixel intensity difference achieved from maximum direction with adjacent neighbouring pixels for each pixel as an improved segmentation methodology for optimal performance. Proposed research used two-frame difference method where the desired frame is passed through a denoising effect to achieve optimal detection performance and to reduce computation time. The basic differences between the pro-posed method and existing research methods is the usage of the frame difference method and local maximum pictorial intensity difference as an improved segmentation, which increases the detection rate and simultaneously decreases the false alarm rate and computation time. In addition, proposed method used adaptive threshold optimization approach to hold sharp edges and thin edges which play significant influential factors to improve overall performance for aerial types of images during validation. The remaining sections of this paper are organized as follows. The related work section presents a critical but comprehensive literature review of the previous methods. After a brief overview of existing research, proposed method is elaborated in the re-search methodology section. The performance of the proposed method is demonstrated via extensive experiments in the experimental results and analysis section. The concluding remarks are given with the detailed contribution of this research with future work in the conclusion section.

Background Study
Generally, most of the previous research for moving object detection can be categorized based on four aspects, i.e., motion detection, deep learning techniques, frame difference and segmentation, as shown in figure 2. Details comprehensive review in these aspects is illustrated below.

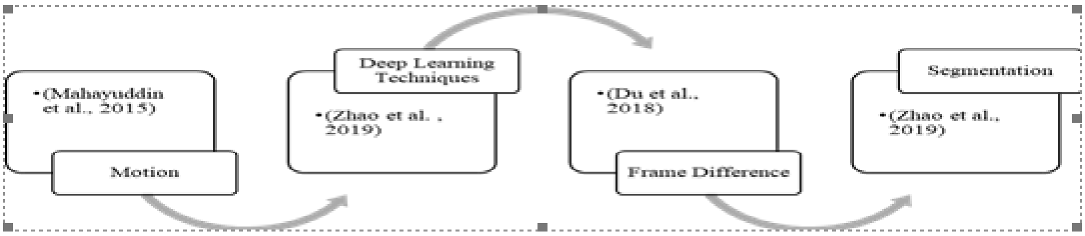

Figure 2: Three Core Aspects For Moving Object Detection In Previous Research.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

Motion Detection

Motion needs to be accurately detected using efficient methods, which can be affected by a number of practical problems for aerial images (Saif et al., 2014). Moreover, detection of motion and detection of moving objects are coupled. If proper motion detection is done, detection of moving objects from UAV aerial images becomes easier. Very few research studies have worked on adaptive robust handling of noise and unfixed motion change, as well as unfixed moving object direction (Mahayuddin et al., 2015). For that reason, adaptive motion detection is needed for better detection of moving objects from UAV aerial images. Jiang et al. (2009) detected moving objects by the illustration of aerial image characteristics and a motion vector processing method using camera motion estimation and compensation. However, under the same motion for more than one object, their research does not provide satisfactory validation. Luo et al. (2012) detected static vehicles along with moving vehicles by clustering single points obtained from motion estimation. However, their approach is not well suited for the complexity of shortening environments, real-time background changes and inconspicuous object features. Cheraghi and Sheikh (2012) detected moving object by distinguishing motion from the background using image registration. However, cluttered environments and appearance similarity make the detection more difficult. Chen et al. (2019) used robust sparse fractional ambiguity function (RSFRAF) dealing with high-order motions. However, their research depends on two-stage threshold processing rather than adaptive threshold. All of the motion analysis approaches from the previous research typically depend on numerous parameters, which increase the computational complexity.

Deep Learning Frameworks

Several deep learning techniques were recently proposed for moving object detection based on convolution neural network (CNN). The evolution of Graphic Processing Units (GPUs) also significantly contributed to the adoption of CNN in computer vision overcoming the problems of real-time processing of computation intensive tasks through parallelization for moving object detection from aerial images. In addition, latest trends in cloud robotics have also enabled offloading heavy computations using advanced deep learning algorithms in the context of moving object detection from aerial images (Lee et al., 2017). Benjdira et al. (2019) investigated Faster R-CNN and You Only Look Once version 3 (YOLOv3), in the context of car detection from aerial images where they claimed YOLOv3 outperforms Faster R-CNN in processing time. In the context of their research, Faster R-CNN misses some instances more than YOLOv3 in lieu with great performance gap between the two algorithms in processing one image per time. However, Dataset they used needs to be extended to add different lighting conditions, i.e., day, night, morning, evening and different environmental, i.e., factors urban, rural etc. Zhao et al. (2019) presents a framework for moving vehicle detection, tracking and geolocation using You Only Look Once version 3 (YOLOv3) to detect small vehicles from the airborne video. However, in their research target vehicle will be lost and target Geolocation is unable to work in the presence of long-time occlusion. In addition, due to the large number of sine, cosine, and matrix inversion calculations in the geolocation process, accuracy of the intermediate process was affected, resulting in error in the final geolocation result. Due to usage of extreme computational complexity and high-end Graphics Processing Units (GPUs) that require too much power and weight by CNNs based object detection, Lee et al. (2017) proposed moving the computation to off-board computing cloud in lieu with low level object detection and short-term navigation on-board. However, moving recognition to cloud introduces unpredictable lag from communication latencies. Besides, cloud based moving recognition introduces a number of variables that are beyond the control of the robot system. In addition,

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003
communicating with a remote cloud typically introduces unpredictable network delay, and the cloud computation time itself may depend on which compute resources are available and how many other jobs are running on the system at any given moment in time. This means that although the cloud may deliver near real time performance in the average case, latencies may be quite high at times, such that on board processing is still needed for critical tasks like stability control.

Distance Measurements
Another aspect, frame difference, provides potential hope for the estimation of motion information in subsequent frames in real-time videos. However, most previous research studies did not provide proper motion estimation to handle six uncertainty constraint factors (UCF) (Saif et al., 2013). Wang (2011) proposed a feature extraction framework using frame difference for vehicle detection using shadows with the combination of rotational invariant shape matching of corner features. They used a shape context descriptor by extracting sample points from the Harris corner response map instead of extracting them from the object edges. However, the shadow-based segmentation algorithm proposed by Wang (2011) could not identify object clocked shadows, and their research mostly depends on the objects having fixed sizes. Pollard and Antone (2012) considered different types of objects for reliable detection and tracking of low-resolution objects of varying size and shape in challenging wide-area video. However, their research considers only even planes or surfaces and did not provide validation for uneven brightness. Gaszczak et al. (2011) considered uneven brightness by providing an automated detection for humans using multiple trained cascaded Haar classifiers combined with additional multivariate Gaussian Shape Matching. Their approach facilitates real-time detection of both static and moving vehicles, but vehicles orientation, colour, type and configuration are invariant, which is the main weakness of their approach. Moreover, when humans and vehicles are located in the same space, their approach is not able to separate features of rigid and nonrigid objects. In this context, Štěpán et al. (2019) worked on accurate localization and illumination variation, for which they proposed reliable future position prediction of coloured objects. However, experimental validation of their proposed research needs to be more comprehensive for reliable validation. The constraint of extracting partial shape of an object existed in most of the previous research that used frame difference, although this method includes motion information extraction characteristics. In this context, segmentation methods have the capability to extract the complete shape of the object or attempt to segment images into several parts by assigning each pixel to subregions (Ji et al., 2019; Saif et al., 2013).

Segmentation Methods
Generally, most of the previous segmentation methods based on critical review can be categorized into two groups, i.e., region-based and various feature extraction-based methods. Segmentation using a region growing method uses each of the pixels in a region with respect to some property, and the more important factor is that it assumes that neighbouring pixels have the same gray-level intensity to be segmented (Wang, 2011; Saif et al., 2013). Moranduzzo and Melgani (2012) used region-based segmentation and introduced a new method to detect cars using SIFT (Scalar invariant feature transform) and classification based on some key points. However, they considered static backgrounds only and used high-resolution images by using a sensor, which is expensive and unrealistic in urban sensor environment situations. Chen et al. (2012) used region-based segmentation by introducing a new framework for robust on-road vehicle detection, which worked well with the complex

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

backgrounds of urban scenes. However, they used grayscale input images for real-time detection, which is unrealistic. For complex background issues, Mahalingam and Subramoniam (2018) proposed a two-phase background estimation (BE) module and used it to select the optimal background candidates for the generation of the updated background models. However, ignorance of various object sizes and motion parameter estimation indicates that their research requires further investigation. In most of the previous research, region-based segmentation requires huge computation time. Moreover, the presence of noise or variation in intensity causes holes or over-segmentation, which indicates the need for efficient selection or usage of features to improve the performance of moving object detection.

Xingbao et al. (2011) used colour features and attracted attention in human vision by detecting moving objects using a context-aware saliency detection algorithm and a Kalman filter, which were associated with the surrounding environments to segment points. Their approach was not affected by shape resolution and appearance, thus overcoming the shortcomings of the traditional segmentation algorithm and being suitable for aerial image segmentation. They applied their proposed methodology to a single-frame basis approach for moving object detection, which becomes the shortcomings for the lack of adequate motion estimation to achieve optimal detection performance. In addition, their research did not consider various directions of objects during experimentation, and they did not provide enough experimental evidence to prove the eligibility of their applied method. In this context, Kembhavi et al. (2010) considered different directions of objects and introduced a new feature selection method called ordered prediction selection using a larger and richer feature set referred to as a colour probability map (CPM), which was extracted from neighbouring pixels. A CPM is capable of capturing colour statistics of objects and their surroundings. Due to the inability to detect objects of the same colour in the given frame, their proposed methods did not provide expected segmentation methodology. Moreover, their research is only workable for structural vehicle shapes and does not work well for the presence of stark. Colour features used by previous research could not overcome the constraint of detecting objects of the same colour in rural or urban areas, along with structural shape issues. However, the problems of using colour features can be overcome by using edge features because these characteristics indicate local intensity variation among various directions for each pixel in frames or images (Saif and Mahayuddin, 2018; Chung and He, 2007; Cheng et al., 2011).

Jiang et al. (2009) used edge features and considered a motion vector processing method using camera motion estimation and compensation, and their experimentation was only for cluttered environments. However, due to the consideration of motion estimation, their approach depends on a huge number of parameters. Cheng et al. (2011) used canny edge features for rural and cluttered environments, and they extended a pixel-wise classification method by preserving the relation among neighbouring pixels in a region in the feature extraction process. However, because the method depends on pixel-based classification, in the case that any important pixels are missing, nonvehicle objects can be detected as vehicles. In the post processing step, Cheng et al. (2011) applied a fixed vehicle size and aspect ratio constraint. Thus, the proposed method could not overcome the weakness of symmetric property-based vehicle detection. Moreover, previous researchers also performed other edge feature-based detection, i.e., Sobel (Guo and Yu, 2012; Wang et al., 2019; MAHAYUDDIN and SAIF, 2020) and Prewitt (Raghuvanshi and Datar, 2013; Mahayuddin and Saif, 2020). However, these studies could not provide good edge detection results with thin and smooth edges. For the canny edge-based detection method, good performance depends on adjusting important parameters, which increases the computation time

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

and effectiveness of the method (Fu and Celenk, 2013; Teutsch and Krüger, 2012). The canny method uses a lot of memory during processing, which causes huge computation time and increases the computational complexity. All of these existing methods performed a 3 × 3 matrix multiplication operation on all of the pixels in the given frame, which produces complexity for manipulation of width and height along with dimension.

Based on the comprehensive reviews mentioned above, this paper proposes an efficient method where frame difference is integrated with pictorial intensity difference from maximum direction of a single pixel in lieu with the usage of a mathematical model refereed as adaptive threshold optimization for holding significant edges from aerial images. The proposed method is able to provide the expected detection performance by operating on all neighbouring pixels for each pixel instead of 3 × 3 matrix manipulation operations upon all of the pixels in the desired frame achieved from a two-frame difference method.

Proposed Research Methodology

Proposed Method

Proposed method mentioned in figure 3 carries the basic difference, which is to apply significant intensity differences in the frame achieved from a two-frame difference method, while most of the previous research applied only background subtraction after gray scaling, which is known as binarization. Another difference is that the proposed method is composed of a two-frame difference method and segmentation method together, while most of the previous research mainly depends on either frame difference or segmentation approaches. In addition, proposed research introduces adaptive threshold optimization approach to improve overall performance of the overall methodology by holding various edges from aerial types of images in lieu with handling of noise and unfixed motion change. Details of the proposed method are explained in detail below.

Let $I(m,n,t)$ be the original frame at time t, where (m,n) denotes a pixel position in the original frame, and $I_A(m,n)$ and $I_B(m,n,)$ are the two consecutive frames at times t and t-1 shown in figure 4. Because only the segmentation approach does not have the ability to differentiate moving region from basic static background region, and frame difference approach contains only single-pixel motion instead of overall object motion (Saif et al., 2013a; Saif et al., 2013b), this research includes frame difference and segmentation together, which is expected to yield optimal detection results from aerial images. The proposed research uses frame difference using 1 frame per second and 3 frames per second to find the difference between two consecutive frames to decide whether the next or previous frame contains any changes or not. Difference image $I_f(m,n,t)$ of two consecutive frames is obtained by equation (1).

$$I_f(m,n,t) = \mathrm{round}\left(I_B(m,n,t) - I_B(m,n,t-1)\right) \qquad (1)$$

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

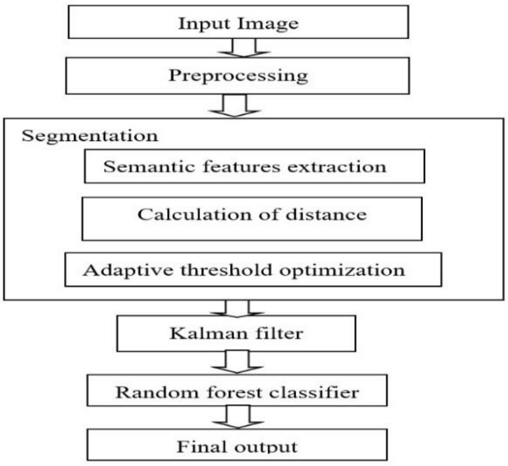

Figure 3: Proposed Method For Moving Object Detection From Aerial Images.

Looping over the pixels between two consecutive frames serves the main purpose of the frame difference approach. If the two corresponding pixels are equal, then the resultant area is shown as white, and otherwise, it is shown as red, as shown in figure 4 (c).

$I_B(m,n,t)$ can be considered for the next steps by satisfying the condition $I_f(m,n,t) > 0$. Let $I_f(m,n,t)$ be the median filtered result from $I_B(m,n,t)$, then $I_f(m,n,t)$ is converted into $I_g(m,n,t)$, which is considered as the grayscale frame. $I_f(m,n,t)$ and $I_g(m,n,t)$ are shown in figure 4 (d) and figure 4(e).

Neighbouring Features Distance Measurement

In gray scale frame denoted as $I_g(m,n,t)$, each pixel has eight neighbours. If pixel coordinates are x and y then the neighbouring pixels are $(i-1, j-1)$, $(i-1, j)$, $(i-1, j+1)$, $(i, j-1)$, (i, j), $(i, j+1)$, $(i+1, j-1)$, $(i+1, j-1)$, $(i+1, j)$, and $(i+1, j+1)$, as shown in Fig. 4(f). The edge difference, which has the same meaning as the colour difference of two pixels, means the RGB difference between two pixels. Consider two pixels $A(i_m, j_n)$ and $B(i_m, j_{n+k})$; the colour difference of $A(i_m, j_n)$ and $B(i_m, j_{n+k})$ can be defined as $K(i, j)$, which represents the value for red, $L(i, j)$, which represents the value for green, and $M(i, j)$, which represents the value for blue. Equations to extract the values of red, green and blue for two pixels $A(i_m, j_n)$ and $B(i_m, j_{n+k})$ are defined in equation (2), equation (3) and equation (4).

$$K(i, j) = |(i_m + j_{n+k}) - i_m j_n| \tag{2}$$

$$L(i, j) = |(i_m + j_{n+k}) - i_m j_n| \tag{3}$$

$$M(i, j) = |(i_m + j_{n+k}) - i_m j_n| \tag{4}$$

In gray scale frame denoted as $I_g(m,n,t)$, each pixel has eight neighbours.

Adaptive Threshold Optimization

Let the total number of pixels be N in a video frame. Threshold value denoted as is defined by the following proposed mathematical model.

$$\theta = \psi - \frac{\psi \times (\log_2(N) + 1)}{100}$$

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

Here, ψ denotes mean values of pictorial intensities.
Let, for all pixels N of $I_g(m,n,t)$ are kept in P (m, n), if $K(i,j)\,|\,L(i,j)\,|\,M(i,j) > \theta$, where θ is the adaptive threshold value used in this research. Based on the same threshold, object is identified using the difference edge detector method from P (m, n) that satisfy P (m, n) > θ .

Kalman Filter
Kalman filter is used to provide the best estimate of states in the presence of noise. Proposed method uses Kalman filter to optimally estimate distances and angles for higher accuracy rate. During distances and angles calculation, if there is any frame loss due to noise or illumination variation then Kalman filter is used by the proposed method to process that frame by illuminating noise. Although, median filter was applied during pre-processing step, some frames can be often still noise may cause deviation in performance.

Classification
Finally, decision for moving objects is determined using random forest classifier from single feature vector by generating a set of probabilities for each class. In this context, probabilities are estimated using mean predicted class probabilities of the trees in the forest where class probability of a single tree is the fraction of samples of the same class in the tree. Class with highest probability is the one that is assigned to the frame as the "decision". In this context, ratio of the highest probability to the second highest probability is referred to as "confidence" of the decision. In this regard, proposed method by this research used adaptive threshold using equation (5) (Wu et al., 2010; Yu and Medioni, 2009). Any decision with confidence more than threshold is considered as "violence" and others are "non violence".

$$T = V - \frac{V \times (log_2(I)+1)}{100} \qquad (5)$$

Here, total number of frames is denoted as I, mean value of pictorial intensities is denoted as V in a video frame. Threshold value is denoted as T.

Experimental Results and Discussion

Datasets
Three major modules were developed for the overall methodology, i.e., image acquisition, segmentation and classification. This research collected 88 and 131 frames at 1 frame per second and 3 frames per second from aerial datasets of the Center for Research in Computer Vision (CRCV) from the University of Central Florida (Saif et al., 2015; Tian et al., 2013). The collected data sets represent a diverse pool of action features at different heights and aerial viewpoints. The frame size in the experiment is 320 × 190.

Experimental Results
This section presents experimental results and analysis for the proposed method for moving object detection using aerial images and compared with previous research results. Moreover, the proposed method is compared with other manual feature-based moving object detection methods, such as Sobel, Prewitt and Canny edge-based detection to ensure the same hardware platform performance in terms of detection rate, false alarm rate and computational time. Figure 4 (a) and figure 4(b) show two consecutive frames (the 101st and 102nd frames), figure 4(e) shows gray scale resultant frame denoted as $I_g(m,n,t)$ after denoising, figure 4(f) shows

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

the pixel structures of gray scale frame denoted as $I_f(m,n,t)$, and finally figure 7(d) shows the resultant output denoted as $I_e(m,n)$ using the proposed method for moving object detection.

The measurement of detection rate (DR), false alarm rate (FAR) and computation time (CT) by estimating metrics such as true positive (TP), false positive (FP), and false negative (FN) (Saif et al., 2014; Chen et al., 2019; Saif et al., 2015; Mahayuddin and Saif, 2019) for the proposed method is shown in Table 1 using the 1-fps and 3-fps frame rates. Using 1 fps, proposed method achieved detection rate of 94.44%, false alarm rate of 6.07% and computation time of 208.91 ms, while using 3 fps, proposed research achieved detection rate of 98.28%, false alarm rate of 6.8% and computation time of 220.12 ms.

Comparison with Previous Research Results

A comparison with the previous states of the art in terms of detection rate is shown in figure 5. Comparison with previous research results is demonstrated based on four aspects, i.e., Motion based Detection (Huang et al., 2010), spatiotemporal segmentation (Zhang, 2005), features Point based segmentation (Yang et al., 2012), other distance calculation-based method (Saif et al., 2014), other deep learning techniques (Zhao et al., 2019; Lee et al., 2017) and combined approach based on features distance and segmentation (Štěpán et al., 2019). false alarm rate (FAR) and computation time, are measured. DR and FAR are measured based on parameters such as true positive (TP), false positive (FP) and false negative (FN).

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

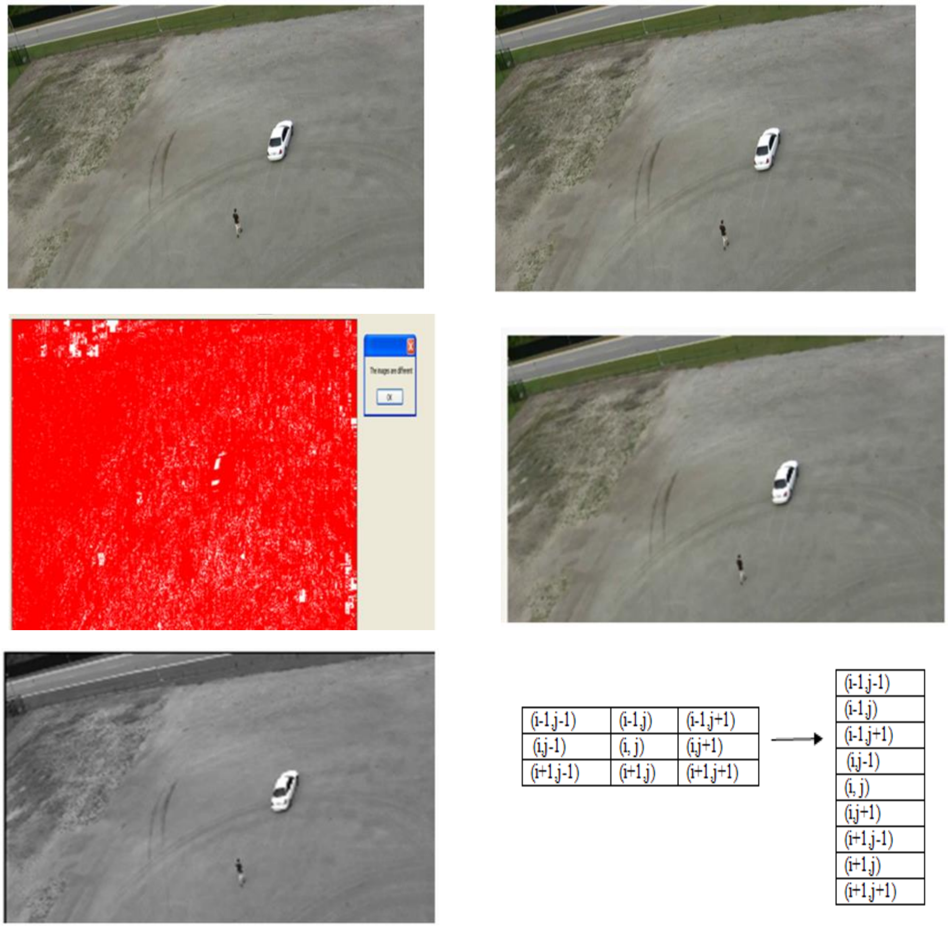

Figure 4: (A) Consecutive Frame At T Time, (B) Consecutive Frame At T-1 Time, (C) Sample Frame Difference Between Two Consecutive Frames, (D) Median Filtered Frame, (E) Converted Grayscale Frame, (F) Sample 8 Neighbouring Pixel Structure Of Grayscale Frame.

Zhao et al. (2019) received detection rate of 90.61% using You Only Look Once version 3 (YOLOv3). However, research by Zhao et al. (2019) is unable to work in the presence of long-time occlusion for detecting targeted object. Lee et al. (2017) investigated various CNNs based object detection methods, i.e., Faster R-CNN, YOLO and Fast YOLO where they introduced off-board computing cloud in lieu with low level object detection and short-term navigation on board. Lee et al. (2017) achieved detection rate of 83.9% using Faster R-CNN, 79.4% using YOLO and 78.3% using Fast YOLO. However, off-board computing cloud introduces unpredictable lag from communication latencies, huge number of variables that are beyond the control of the robot system, unpredictable network delay and cloud computation time itself may depend on which compute resources are available and how many other jobs are running on the system at any given moment in time. Huang et al. (2010) received detection rate of 56% using ego motion of airborne vehicle by feature-point based image alignment. However, research by Huang et al. (2010) mostly depends on shape of the objects which caused very low detection rate. In this context, Zhang (2005) received detection rate of 65% using spatiotemporal segmentation. Although in their experimentation, object size varies from several to thousands of pixels in different videos, their research did not provide reliable validation in case of low contrast perspectives. Yang et al. (2012) received detection rate of

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003
70.06% using feature point tracking method. However, their experimentation needs to be extended in case of missing motion features for reliable validation. Pollard and Antone (2012) received detection rate of 50% using frame difference approach. However, their research considers only ever plane or surface during experimentation by assuming constant platform velocity.

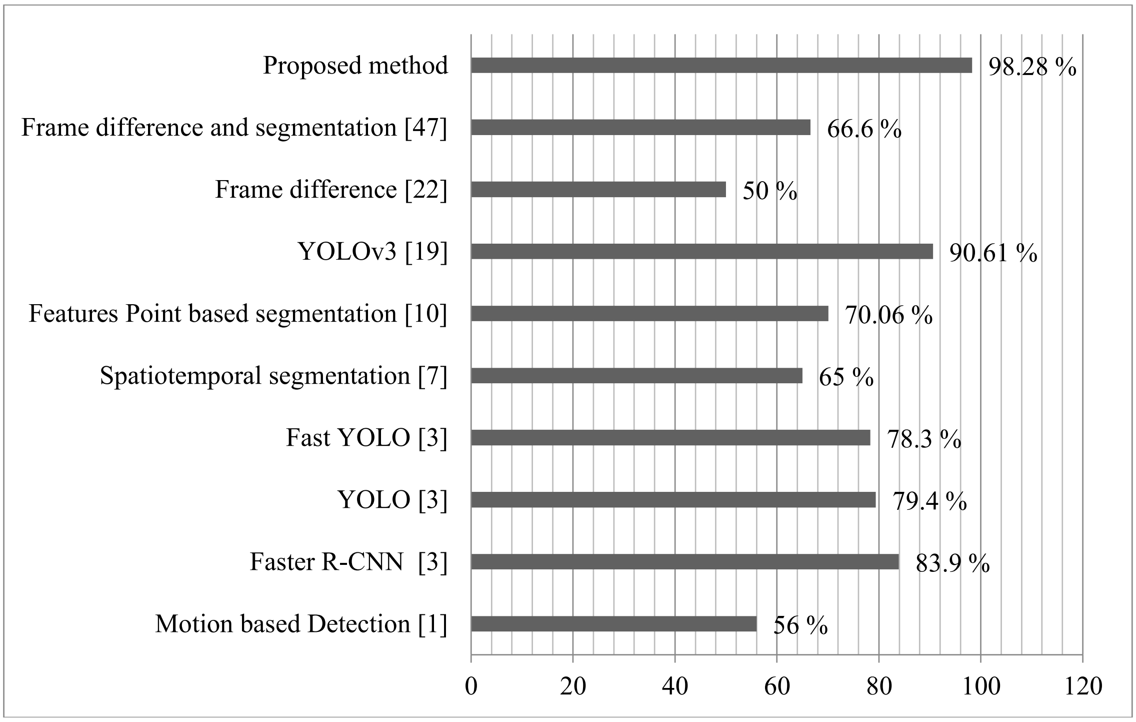

Figure5: Detection Rate Of The Proposed Method And States Of The Art.

Table 1: Experimental Results

FR	DR(%)	FAR(%)	CT(ms)
1 fps	94.44	6.07	208.91
3 fps	98.28	6.8	220.12

Oreifej et al. (2010) received detection rate of 66.6% by applying frame difference and segmentation together where in the segmentation part they used weighted voter-candidate formulation. However, their research assumed that the object position in the next frame should be close to its position in the current frame, while the recognition does not enforce any motion assumptions. Proposed method by this research provides highest detection rate of 98.28% comparing with previous research using semantic features distances. In this context, Kalman filter was applied to rectify frame loss due to noise or illumination variation. Besides, usage of random forest classifier ensured classification of single feature type in order to avoid complication like using multiple types of features causes lower processing time comparing with previous research methods.

Validation Under Same Hardware Platform

Proposed method is also evaluated under same hardware platform for three other manual features type, i.e., Sobel, Prewitt and Canny shown in Fig. 6. For each of the evaluation,

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

proposed research used 1 fps and 3 fps as frame rate. Compared with other edge-based methods using 1 fps, the proposed method provides a higher detection rate where detection rate for Sobel is 84.45%, that for Prewitt is 86.08%, and that for Canny is 87.23%, while the proposed method achieves 94.44%. In comparison with other edge methods using 3 fps, the proposed method also provides higher detection rate where the detection rate for Sobel is 85.46%, that for Prewitt is 86.17%, and that for Canny is 89.36%, while the proposed method achieves 98.28%.

Compared against other edge methods using 1 fps, the proposed method also provides the lowest false alarm rate, Sobel received 11.09%, that for Prewitt is 11.71%, and that for Canny is 14.41%, while the proposed method achieves 6.07%. The proposed method also provides the lowest false alarm rate at 3 fps which is 6.8%, that for Sobel is 11.78%, that for Prewitt is 11.22%, and that for Canny is 13.77%.

For 1 fps, the proposed method required a computation time of 208.91 ms (milliseconds), while Sobel, Prewitt, and Canny required computation times of 295.22 ms, 303.9 ms and 243.16 ms, respectively. For 3 fps, the proposed method required a computation time of 220.12 ms, while Sobel, Prewitt, and Canny required computation times of 363.64 ms, 365.29 ms and 271.19 ms, respectively. So, in case of computation time also, proposed method shows prominent performance comparing with other edge features-based detection. Sample outputs are shown in Fig. 7 for Sobel, Prewitt, Canny edge feature-based moving object detection.

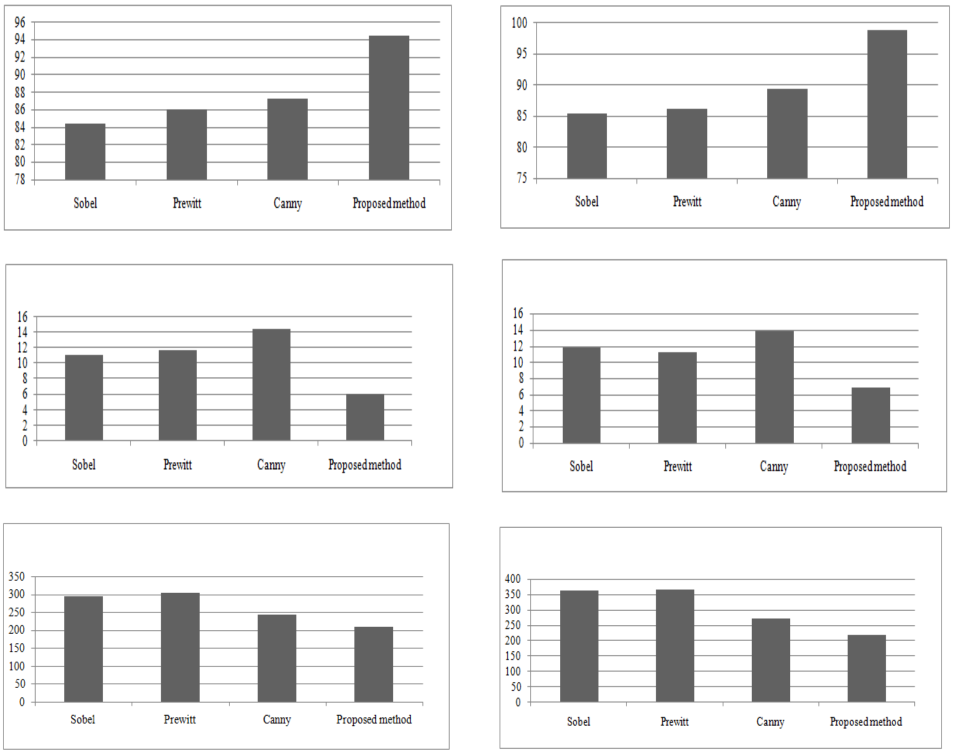

Figure 6: Same Hardware Platform Measurement Using (a) 1 Frame Per Second, (b) 3 Frame Per Second; False Alarm Rate Using (c) 1 Frame Per Second, (d) 3

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

Conclusion

Proposed method used motion patterns by extracting convolutional semantic features for uncertainty measurement of pictorial intensity distribution based on efficient scene interpretation. In case of similar coloured object in the background, semantic properties provide certain particular weighted average of pictorial intensities causes attractive interpretation which played vital role to minimize background complexity. After that, semantic features distances are calculated for each frame followed by Kalman filter to rectify frame loss due to noise or illumination variation which plays significant role to optimally estimate distance for higher accuracy rate. Adaptive threshold optimization was applied to hold sharp and thin edges in lieu with adaptive robust handling of noise and unfixed motion change which plays significant role to achieve reliable experimental results comparing with previous research results. Finally, proposed method used random forest classifier to classify single feature type in order to avoid complication like using multiple types of features causes lower processing time comparing with previous research methods. Proposed method achieved highest accuracy rate of 98.28% comparing with existing research results in lieu with lowest false alarm rate and computation time. However, when small-sized objects are placed very closely, blob or region of interest detection is still considered to be an unsolved issue for moving object detection from aerial images and will be further investigated using the proposed method in the future. The proposed method is expected to be used by UAV operators or related researchers for further research or investigation for areas where access is restricted or rescue areas, various object identification in specific areas, crowd flow analysis, anomaly detection, intelligent traffic management and so forth.

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

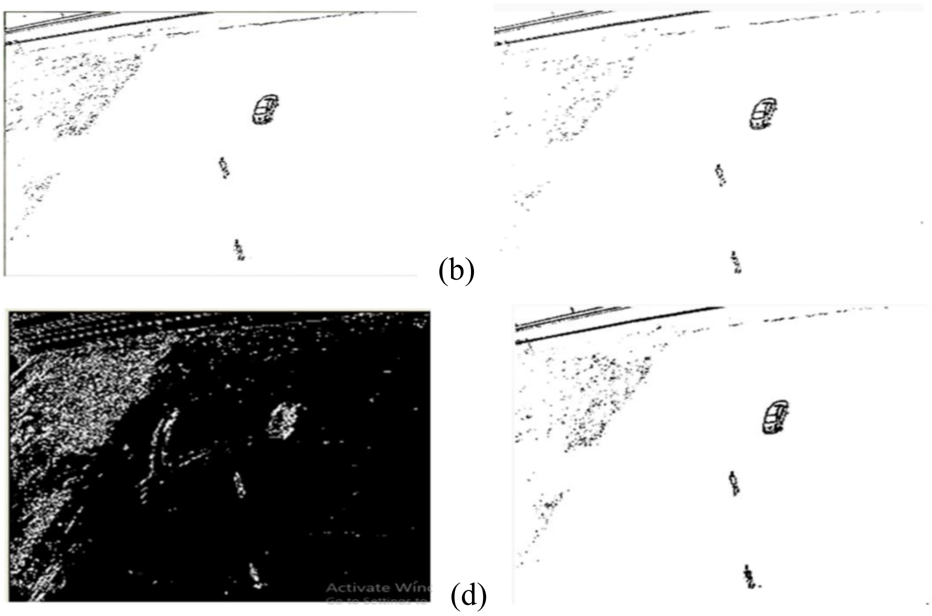

Figure 7: Resultant Frame Using (a) Sobel, (b) Prewitt, (c) Canny, (d)

Acknowledgement

The authors would like to thank Universiti Kebangsaan Malaysia for providing financial support under the “Geran Universiti Penyelidikan” research grant, GUP-2020-064.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

References

Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A., & Ouni, K. (2019). Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3. 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS),

Bhaskar, H. (2012). Integrated human target detection, identification and tracking for surveillance applications. 2012 6th IEEE International Conference Intelligent Systems.

Bhuvaneswari, K., & Rauf, H. A. (2009). Edgelet based human detection and tracking by combined segmentation and soft decision. 2009 International Conference on Control, Automation, Communication and Energy Conservation.

Chen, L., Jiang, Z., Yang, J., & Ma, Y. (2012). A coarse-to-fine approach for vehicles detection from aerial images. 2012 International Conference on Computer Vision in Remote Sensing.

Chen, X., Guan, J., Wang, G., Ding, H., & Huang, Y. (2019). Fast and refined processing of radar maneuvering target based on hierarchical detection via sparse fractional representation. *IEEE Access*, *7*, 149878-149889.

Cheng, H.-Y., Weng, C.-C., & Chen, Y.-Y. (2011). Vehicle detection in aerial surveillance using dynamic Bayesian networks. *IEEE transactions on image processing*, *21*(4), 2152-2159.

Cheraghi, S. A., & Sheikh, U. U. (2012). Moving object detection using image registration for a moving camera platform. 2012 IEEE International Conference on Control System, Computing and Engineering.

Chung, Y.-c., & He, Z. (2007). Low-complexity and reliable moving objects detection and tracking for aerial video surveillance with small uavs. 2007 IEEE International Symposium on Circuits and Systems.

Ding, W., Gong, Z., Xie, S., & Zou, H. (2006). Real-time vision-based object tracking from a moving platform in the air. 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems.

Dorfmuller-Ulhaas, K., & Schmalstieg, D. (2001). Finger tracking for interaction in augmented environments. Proceedings IEEE and ACM international symposium on augmented reality.

Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., Zhang, W., Huang, Q., & Tian, Q. (2018). The unmanned aerial vehicle benchmark: Object detection and tracking. Proceedings of the European Conference on Computer Vision (ECCV).

Fu, Q., & Celenk, M. (2013). Marked watershed and image morphology based motion detection and performance analysis. 2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA).

Gaszczak, A., Breckon, T. P., & Han, J. (2011). Real-time people and vehicle detection from UAV imagery. Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques.

Gleason, J., Nefian, A. V., Bouyssounousse, X., Fong, T., & Bebis, G. (2011). Vehicle detection from aerial imagery. 2011 IEEE International Conference on Robotics and Automation.

Goyal, K., & Singhai, J. (2018). Texture-based self-adaptive moving object detection technique for complex scenes. *Computers & Electrical Engineering*, *70*, 275-283.

Guo, C., & Yu, J. (2012). Boundary and shadow position-based moving objects detection. 2012 5th International Congress on Image and Signal Processing.

Huang, C.-H., Wu, Y.-T., Kao, J.-H., Shih, M.-Y., & Chou, C.-C. (2010). A hybrid moving object detection method for aerial images. Pacific-Rim Conference on Multimedia.

Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003
Ibrahim, A. W. N., Ching, P. W., Seet, G. G., Lau, W. M., & Czajewski, W. (2010). Moving objects detection and tracking framework for UAV-based surveillance. 2010 Fourth Pacific-Rim Symposium on Image and Video Technology.

Ji, Y., Zhang, H., Tseng, K.-K., Chow, T. W., & Wu, Q. J. (2019). Graph model-based salient object detection using objectness and multiple saliency cues. *Neurocomputing*, *323*, 188-202.

Jiang, Z., Ding, W., & Li, H. (2009). Aerial video image object detection and tracing based on motion vector compensation and statistic analysis. 2009 Asia Pacific Conference on Postgraduate Research in Microelectronics & Electronics (PrimeAsia).

Kembhavi, A., Harwood, D., & Davis, L. S. (2010). Vehicle detection using partial least squares. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(6), 1250-1265.

Lee, J., Wang, J., Crandall, D., Šabanović, S., & Fox, G. (2017). Real-time, cloud-based object detection for unmanned aerial vehicles. 2017 First IEEE International Conference on Robotic Computing (IRC).

Luo, P., Liu, F., Liu, X., & Yang, Y. (2012). Stationary vehicle detection in aerial surveillance with a UAV. 2012 8th International Conference on Information Science and Digital Content Technology (ICIDT2012).

Mahalingam, T., & Subramoniam, M. (2020). A robust single and multiple moving object detection, tracking and classification. *Applied Computing and Informatics*.

Mahayuddin, Z., & Saif, A. (2020). Efficient Hand Gesture Recognition Using Modified Extrusion Method based on Augmented Reality. *TEST Engineering and Management*, *83*, 4020-4027.

Mahayuddin, Z. R., & Saif, A. S. (2019). A COMPARATIVE STUDY OF THREE CORNER FEATURE BASED MOVING OBJECT DETECTION USING AERIAL IMAGES. *Malaysian Journal of Computer Science*, 25-33.

Mahayuddin, Z. R., & Saif, A. S. (2019). A Comprehensive Review Towards Appropriate Feature Selection for Moving Object Detection Using Aerial Images. International Visual Informatics Conference.

MAHAYUDDIN, Z. R., & SAIF, A. S. (2020). A COMPREHENSIVE REVIEW TOWARDS SEGMENTATION AND DETECTION OF CANCER CELL AND TUMOR FOR DYNAMIC 3D RECONSTRUCTION. *Asia-Pacific Journal of information technology and multimedia*, 28-39.

Mahayuddin, Z. R., Saif, A. S., & Prabuwono, A. S. (2015). Efficiency measurement of various denoise techniques for moving object detection using aerial images. 2015 International Conference on Electrical Engineering and Informatics (ICEEI).

Moranduzzo, T., & Melgani, F. (2012). A SIFT-SVM method for detecting cars in UAV images. 2012 IEEE International Geoscience and Remote Sensing Symposium.

Oreifej, O., Mehran, R., & Shah, M. (2010). Human identity recognition in aerial images. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Pollard, T., & Antone, M. (2012). Detecting and tracking all moving objects in wide-area aerial video. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.

Raghuvanshi, R. S., & Datar, A. (2013). Comparison of gradient operator based pseudocolored enhanced medical images. 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT).

Saif, A., & Mahayuddin, Z. R. (2018). Moving Object Segmentation Using Various Features from Aerial Images: A Review. *Advanced Science Letters*, *24*(2), 961-965.

Journal of Information System and Technology Management JISTM
EISSN: 0128-1666
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003
Saif, A., Prabuwono, A., & Mahayuddin, Z. (2013). Adaptive long term motion pattern analysis for moving object detection using UAV aerial images. *International Journal of Information System and Engineering*, *1*(1), 50-59.
Saif, A., Prabuwono, A. S., & Mahayuddin, Z. R. (2014). Moving object detection using dynamic motion modelling from UAV aerial images. *The Scientific World Journal*, *2014*.
Saif, A. S., Prabuwono, A. S., & Mahayuddin, Z. R. (2013). Adaptive motion pattern analysis for machine vision based moving detection from UAV aerial images. International Visual Informatics Conference.
Saif, A. S., Prabuwono, A. S., & Mahayuddin, Z. R. (2013). Real time vision based object detection from UAV aerial images: a conceptual framework. FIRA RoboWorld Congress.
Saif, A. S., Prabuwono, A. S., & Mahayuddin, Z. R. (2014). Motion analysis for moving object detection from UAV aerial images: A review. 2014 International Conference on Informatics, Electronics & Vision (ICIEV).
Saif, A. S., Prabuwono, A. S., & Mahayuddin, Z. R. (2015). Moment feature based fast feature extraction algorithm for moving object detection using aerial images. *PloS one*, *10*(6), e0126212.
Saif, A. S., Prabuwono, A. S., Mahayuddin, Z. R., & Himawan, H. T. (2013). A review of machine vision based on moving objects: object detection from UAV aerial images. *International Journal of Advancements in Computing Technology*, *5*(15), 57.
Štěpán, P., Krajník, T., Petrlík, M., & Saska, M. (2019). Vision techniques for on-board detection, following, and mapping of moving targets. *Journal of Field Robotics*, *36*(1), 252-269.
Teutsch, M., & Krüger, W. (2012). Detection, segmentation, and tracking of moving objects in UAV videos. 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance.
Tian, Y., Sukthankar, R., & Shah, M. (2013). Spatiotemporal deformable part models for action detection. Proceedings of the IEEE conference on computer vision and pattern recognition.
Wang, B., Chen, L., & Zhang, Z. (2019). A novel method on the edge detection of infrared image. *Optik*, *180*, 610-614.
Wang, Q., Zhang, L., Zou, W., & Kpalma, K. (2020). Salient video object detection using a virtual border and guided filter. *Pattern Recognition*, *97*, 106998.
Wang, S. (2011). Vehicle detection on aerial images by extracting corner features for rotational invariant shape matching. 2011 IEEE 11th International Conference on Computer and Information Technology.
Wu, C., Cao, X., Lin, R., & Wang, F. (2010). Registration-based moving vehicle detection for low-altitude urban traffic surveillance. 2010 8th World Congress on Intelligent Control and Automation.
Xingbao, W., Chunping, L., Gong, L., Long, L., & Shengrong, G. (2011). Pedestrian recognition based on saliency detection and Kalman filter algorithm in aerial video. 2011 Seventh International Conference on Computational Intelligence and Security.
Yang, Y., Liu, F., Wang, P., Luo, P., & Liu, X. (2012). Vehicle detection methods from an unmanned aerial vehicle platform. 2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012).

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 7 Issue 29 (December 2022) PP. 24-41
DOI: 10.35631/JISTM.729003

Yu, Q., & Medioni, G. (2009). Motion pattern interpretation and detection for tracking moving vehicles in airborne video. 2009 IEEE Conference on Computer Vision and Pattern Recognition.

Zhang, S. (2005). Object tracking in unmanned aerial vehicle (uav) videos using a combined approach. Proceedings. (ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.

Zhao, X., Pu, F., Wang, Z., Chen, H., & Xu, Z. (2019). Detection, tracking, and geolocation of moving vehicle from uav using monocular camera. *IEEE Access*, *7*, 101160