A Comprehensive Survey of the Vehicle Motion Detection and Tracking Methods for Aerial Surveillance Videos

Hussam Saleh Abu Karaki[1], Saleh Ali Alomari[2] and Mohammed Hayel Refai[3]

[1]Faculty of Information Technology at Al Hussein Bin Talal University, Ma'an, Jordan
[2]Faculty of Science and information technology, Jadara University, Irbid, Jordan
[3]Department of Information Systems and Technology, Sur University College, Sur, Oman

Summary
Currently, object tracking and motion detection are considered challenging areas in the domain of computer vision where the complexity of tracking and object detection is increased by several factors. For examples, the illumination modification due to external factors, the parallax motion that is based on objects' movements including the camera motion produces a complicated overview within the sight including the forecast pertaining to the 3D sights through the 2D image. Hence, this leads to reduce the image quality and loss of video details. Other factors that include real-time video processing, distance, imagery noise, occlusions, viewpoint, surrounding context overlapping and several other factors limits the ability to precisely identify and track a detected object. Therefore, this paper highlights and explains a different method of identifying the movements of objects in captured frames and the suitability, limitations, and possible techniques for the methods of background subtracting temporal differencing, statistical models and optical flow, where the optical flow method is selected as the most appropriate for aerial videos in this paper. Three main issues are identified in this paper for tracking, categorization and object detection within aerial videos that form the parallax motion, which is created by the camera's dual movement and the objects over the ground, the impact of the camera altitude modification over the object representation in the captured frame and the impact of the illumination modification over the optical flow field.

Key words:
Vehicle detection, object motion detection, object tracking, aerial surveillance videos

1. Introduction

The need for video and aerial imagery has increased for its significance through various applications with surveillance as an essential concept. In order to reduce the meaningful information of the videos, a study that is based on detecting a moving body is conducted when tracking has witnessed an enormous increment during the previous eras. The attempts of computer visualization within video analysis and image processing methods are based on exploring, identifying and tracing particular objects from a sequence of image frames or an image itself. Security and safety requirements of military and society lead to move different objects through several applications (e.g. vehicles' and humans' tracking and detection are based on the use of aerial surveillance videos). Such applications consider public places as banks, department stores, parking lots, universities, shopping centers, sports stadiums, hospitals, private places (e.g. police centers, homes, martial bases, prisons and further particular services [1]), applications of transportation surveillance (e.g. transit through railways, expressways and cities [2], undergrounds [3] including airports [4]; martial implementations (e.g. rescue, searching and security expeditions, and boundary surveillance [5]; urban implementations (e.g. maritime surveillance, forest fires and the investigation of a catastrophe region and surveillance for people behavior like jogging, clapping, walking or other behaviors [6]. Figure 1 illustrates four scenarios related to distinct surveillance systems.

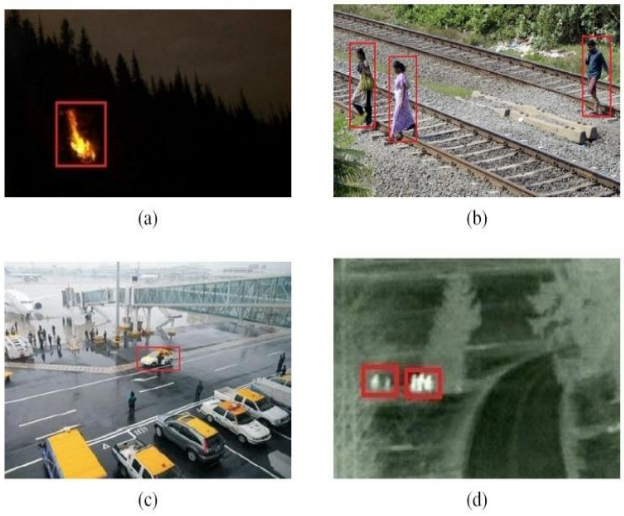

Fig. 1 Surveillance Schemes Paradigms (a) Forest fires discovery (b) Railway passage discovery (c) Prohibited parking discovery (d) Night visualization camera traces immigrants crossing limit

The applications indicated above on object tracking and motion detection is inspiring the interests of researchers throughout the world. For instance, the IEEE sponsored the international Conference on Advanced Video and Signal Based Surveillance (AVSS) in sixteenth occasions from a period of 2003 to 2019. Many international studies

Manuscript received January 5, 2019
Manuscript revised January 20, 2019

begin improving precise and strong visualization schemes [7] [8]. The Deep-learning-based method for a strong outdoor vehicle tracking is proposed in [9]. Within this domain, surveys serve as a pointer to the improvement that is encountered throughout the time. A research issued in 2018 regarding visualization surveillance for moving bodies including behaviors in [10] [18] and an overview of smart shared surveillance schemes in 2017 show the tendencies regarding the 2-Dimensional procedure [11].

In 2014, a conducted research of the common visual tracing that is related to the object representation studies various paths and considerations in tracking [12]. Different schemes of contemporary remote surveillance systems and intelligent visual surveillance pertaining to the public safety are proposed in 2010 according to two studies in [1] [13]. A study on vision-based methods for recognition, segmentation and action representation is conducted in 2011 [14]. In addition, studies in 2015 carries out up-to-date techniques in the behavior understanding and activity recognition within the video surveillance domain [15] and vehicle detection techniques within the aerial surveillance domain [16]. The entire previous surveys are considered proofs of the rising interest in visual surveillance system and computer vision for tracking and motion detection applications.

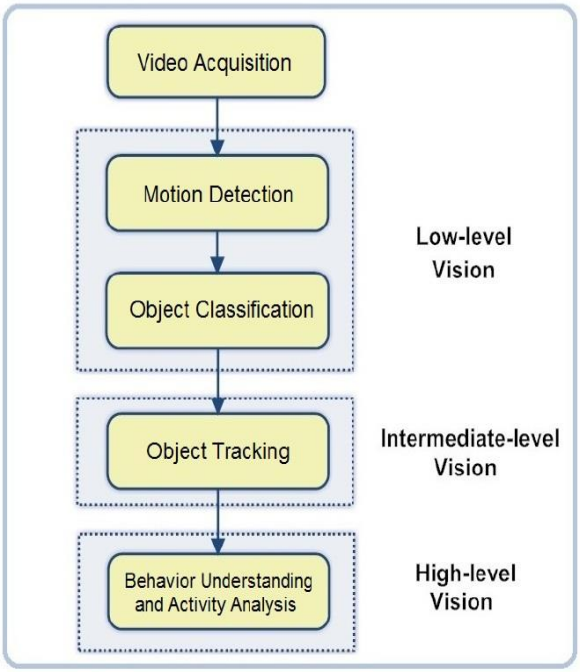

Fig. 2 General Framework of Visual Surveillance System

Due to the various considerations and settings for every application, there is no an existing standard that is suitable for the entire efficient tracking and motion detection methods. However, a general framework delivers the most appropriate consistency in improving such implementations. The broad method pertaining to the visualization surveillance scheme illustrates that three main stages for handling videos exist within the computer vision, called, low-stage, intermediate-stage and high-stage visualizations [10] [17] (see Figure 2). The following stages of the framework comprise: activity analysis, object classification, aerial video acquisition, motion detection, behavior understanding and object tracking. The stages in Figure 2 are further explained in the previous studies as introduced in [10] [12-13].

This paper represents a significant knowledge of the video-based surveillance. This is based on a succinct explanations of methods related to object tracking, object categorization and object detection. The paper concludes with an explanation on the problems and challenges that are related to the roadway vehicle tracking, which is faced by video researchers and analysts.

2. Frame Registration

Most of the pre-processing functions are seen as direct functions of manipulation and conversion replacing the frame registration that copes with minimising the impact of dual motion and its issues with developing and conserving the feature of the captured frames [19]. When moving camera captures a frame sequence, a limited space is captured by the frame from the whole sight. The space is moving because of the camera's shift such that a current space pertaining to the sight accesses all borders of the frame including an equivalent space gets off the recorded boundaries yielding in pixels that provide similar objects, which are located through many coordinates of the sequential frames. Hence, both foreground and background reflect a clear movement through to the system within the entire scene. In order to differentiate the camera's movements and independent objects within the scene from the background, a frame registration is needed to modify the whole pixels of a single frame in an equal move to their related locations according to their precedence frame. Methods for frame registration are based on their suitability for the properties of the captured video, such as the viewpoint, altitude and change of lighting and resolution during the capturing process of the video. Mainly, there exist two basic methods, which comprise: the intensity-based and the feature-based methods [19]. The feature-based method is based on the apparent contrast in image pixels that simplifies the matching of pixels within a frame to its particular location in a previous frame. The intensity-based method is based on the intensity disseminative patterns within a frame when the features of the pixels are not reflecting an apparent diversity [20]. Furthermore, hybrid algorithms can be proposed for the creating a frame registration.

There exist four main methods that are proposed in the literature based on three methods, which are called, the Robust-Data Alignment (RDA) (hybrid) approach, the correlation-based (intensity-based") algorithm, the Scale Invariant Feature Transform (SIFT) (feature-based) algorithm and the Lucas-Kanade (LK) (intensity-based) registration algorithm [21-23]. Intensity-based methods are normally based on the cross-correlation function between two consecutive frames that make it sensitive to lighting changes. The cross-correlation algorithm applies a Gaussian image pyramid method in order to down-sample the frames and conducts the registration within the coarsest resolution for increasing the computational efficiency of the algorithm. The entire images are evenly divided into different grids and different points within grid intersections are identified. For every pixel region on a grid intersection of a frame, a projective transformation with the previous frame is computed and the same measurement of normalised cross-correlated frames is calculated in order to acquire the most appropriate value for transformation. The LK algorithm utilises a Gauss-Newton gradient optimisation to line up images based on the use of image intensity differences including the intensity gradient information for enhancing the transformation's estimation [24]. The Feature-based registration extracts noticeable characteristics such as areas, corners, edges and points from the images and lines up the images by aligning these characteristics that are strong to any modifications in lighting and viewpoint compared to the intensity-based registration. The SIFT is a popular feature-based method, which applies the frame registration in video processing [22]. The SIFT produces a scale-space for the frames including many images related to the Difference of Gaussian (DoG) and searches for a local extreme within the improved model. By utilising a Gaussian kernel, the SIFT involves the region of a key point where the results are utilised to compute the local image gradients that are utilised to allocate the orientation of the key point. The descriptor is calculated by involving the region of a key point along with a Gaussian kernel with a sigma that is equal to 1.5 times of the key point sigma. After that, the SIFT implements the RANSAC in order to search for the ideal affine transformation among the image pairs that are related to the matching SIFT descriptors based on the use of a linear least squares fit. The hybrid registration methods combine intensity-based and feature-based methods to improve the registration's quality when both concepts are derived from the images. A common hybrid method, the RDA method [23], produces a prototype for the comparable frames by utilising explored characteristics from the target frame and intensity data from the reference frame based on applying a probability distribution function. After that, the algorithm measures the likeness among the frames by aligning the characteristics of the images in a manner that provides a least distance for the probability distribution function. Nonetheless, this method suffers from the complexity of defining the most effective distance measure equation depending on the distinctive characteristics of the frames, but could produce high quality results when the characteristics of the images are distinctive. An experiment in [21] [22] for video registration that the SIFT algorithm outdoes other relating methods according to a partial image overlap for aerial videos. The Speeded Up Robust Features (SURF) is defined as a local feature descriptor and detector, which is utilised for particular tasks such as 3D reconstruction, categorisation, registration and object recognition. It is to some extent stimulated by the Scale-Invariant Feature Transform (SIFT)" descriptor [25].

3. Video Surveillance

Surveillance refers to the performance of checking behaviours, actions or different modified information for different items including the intention of protecting, guiding, managing or influencing [11]. Currently, Vehicle Surveillance Videos (VSV) attracted many video processing researchers for vehicle, animal or people motion detection as given in [26] [27] [28] [29], classification [30] [31] [32], tracking [33] [34] [35] and behaviour understanding and activity analysis [6] [15] [10]. Based on the traffic monitoring applications, the background basically concentrates on methods that are in relation to VSV. For the vehicle surveillance, traffic administration implementations include device visualisation schemes for controlling and monitoring the traffic for ensuring security and safety measures over the roadway by implementing incident detections [36], vehicle tracing [83] and irregular behaviour discovery within gathering sights [37]. Traffic administration surveillance videos differ in their resource sensors, (e.g. portable/fixed), according to the intention of performing the implementation. Portable and fixed sensors are elaborated in Sections 3.1 and 3.2. Additionally, studies of tracking and motion detection [10] [16] [12] show that the field integrates several different tools and methods based on the platform depiction, the features and standards of the set of data and the implementation's purpose for item discovery schemes [12], object tracking methods [12], segmentation and motion detection methods [38] and object classification methods [10]. The next subsections explain the input video sources with their various platform representations (see Section 3.3). In this paper, a method is produced in order to automatically identify and track several moving objects excluding any learning phase. Further, a novel real time method is produced according to the particle filter and background subtraction. This method could automatically identify and track several moving

objects excluding any learning phase or previous knowledge regarding the initial position, nature or size. An empirical research is carried out through various video test sets [39].

3.1 Stationary (Fixed) Cameras

Fixed cameras in surveillance are utilised either outdoors or indoors. In computer vision, indoor cameras are basically utilised to check public buildings, banks and stores for identifying human, tracking movement, motion and recognising abnormal behaviours [40]. On the other hand, outdoor fixed cameras are utilised in traffic monitoring, city surveillance and university campuses in order to identify the behaviour of various objects and patterns of motion. The objects are categorised according to its features in order to modify the procedure of the video through to particular objects, such as vehicles or humans. In this paper, fixed cameras indicate to surveillance cameras, which are being used for the purpose of monitoring traffics. Videos that are handled by different device visualisation schemes indicate to the resource being a fixed sensor through two situations. In the first situation, fixed sensors are being used for checking the roadway's portions temporarily by utilising a fixed camera that is straddled over a non-shifting container, such as a parked vehicle. In the second situation, the camera is straddled on a fixed container (e.g. a light pole). A video registers a fixed sight in both situations such that a background is not able to modify. The shifting items and foreground are seamlessly contrasted and explored from the background in this case. Innovative device visualisation schemes that rely on fixed cameras controls the situations of several cameras, which cover broader regions of the scene such that frames overlap to create a single view [41] [42].

3.2 Non-Stationary (Shifting) Cameras

In contrast to fixed cameras, the frame's foreground and the background are modified. Shifting cameras are straddled on a handheld or a shifting car. The videos that are registered via these cameras during their movements exemplify a modification on the sight vision such that the background can moderately move from a single frame through to the following frame including some portions related to the background, which are vanishing including current portions accessing a boundary of a frame. The camera shift contains a movement through vertical and horizontal directions. Meantime, what forms the foreground as shifting cars over the street can bring a movement in which video's process is complicated. This process is based on the position of the shifting camera including its distance and vision from the items including the sight region it completely overlays.

In terrestrial cameras straddled over a car, the identification, categorisation and tracing more shifting cars is achieved from the front view or backside of different cars and from various spaces. Aerial cameras deliver a vertical-vision, and sometimes, it delivers an incomplete side-vision such that every situation provides methods and issues for its procedures. Shifting cameras bring various issues by removing different impacts of a dual movement and jitter impact yielding from its instable camera location [10] [43]. Having a trade-off among shifting and fixed cameras, the Pan-Tilt-Zoom (PTZ) cameras are considered fixed cameras, which have the ability of directing the control and to remote the zoom by capturing additional regions from the ground truth by shifting through to its direction and by concentrating on partial regions from the scene based on partially increasing the zoom of the frame. The PTZ remains limited with its flexibilities compared to the free shift of the shifting cameras [44]. Furthermore, when the PTZ videos are being processed, the same limitations and issues of the shifting and fixed cameras are yet encountered since the resultant video can be either the same to the motionless camera's videos once the PTZ does not perform some zooming and shifting over the taken sight, or is the same to non-motionless cameras once it performs shifting and out or in zooming. According to previous situations, the portions of a video are handled in combination to motionless cameras' methods, while for the other situation; videos must be handled in combination to non-motionless cameras' methods. However, in situations when a camera moves around, the procedure of a video seems to be easier because of the inadequacy rapidity for the camera shifting including the comparatively inadequate space from the actual land such that the zooming impact is seamlessly calculated because a camera contains a stable location. A number of cameras that are utilised based on fixed, shifting and PTZ-cameras, raises the significance of the performance optimisation over the processing level, such as the real time processing when the footage and scenes of those cameras are synchronized, and when cameras are represented as mobiles [45].

3.3 Cameras Platform Representation

The objectives of using the surveillance video involve handicapped aiding tools, military or domestic surveillance and professional or personal photography. The main purposes pertaining to the implementation identify the needs for being involved through the entire scheme (e.g. processing and machinery abilities related to the algorithms, frame rate and video quality. Another factor that is taken into account is the required quality of the video. The video position and feature of cameras, which are used in the procedure supports the quality of the produced videos or images, such as aerial surveillance cameras and indoor stationary cameras. Cameras are straddled on an aerial vehicle by checking massive regions

of a scene (see Sections 3.1 and 3.2). Nonetheless, they suffer from further issues compared to stationary cameras based on the modifications made on the parallax motion and the context of the background [46]. In terms of the video types, the output properties, input properties and cameras' resolution identify the video's quality. The camera resolution is explained in many measures pertaining to a taken video. The resolution identifies how many pixels for every square inch (ppi) in order to display the captured image frame and to provide further details that are captured by raising the ration of the ppi. The spatial resolution highlights the number of separate pixel quantity for every length of unit. The spectral resolution, which highlights several colours of an appearance including chronological resolution, identifies the number of related frames, which contains a single second of the recording where it is measured in frame for every second (frame per second). There is not any available new quality identified through the related research, which can classify the rate of a frame. Nevertheless, it is split into four classifications, which comprise: 1) small frame rate ranging from 1 to 10, 2) regular frame rate ranging from 11 to 71, 3) high frame rate ranging from 72 to 300, and 4) extremely great frame rates (301-"200,000,000) fps (Imaging, 2013). The greatest popular used rate of frame that is seamlessly processed by motion detection systems refers to the normal frame rate [47] [31] [44]. It is important to indicate that the frame rate, greater resolutions and colours raise the video's size efficiently. Input features within cameras select the zooming impact that minimises the space within the recorded scene including the camera through different image clearness influences that create an image noise (random dot pattern, which is superimposed over an image or blurred images). The zooming feature is affected through a focal length capability including the focus setting, which puts the space through two lenses within a camera (in millimetres) in order to check the ratio pertaining to the recording and embodiment for a portion of the sight through to a portion of the frame. The principal length that is called the optical zoom is considered to be a measurement that ensures the robustness of the scheme diverges or converges light from the sight through to a taken image frame. The optical zoom makes an incomplete distance of the seen sight via a camera that is taken within a frame. When an optical power gets higher, the optical zoom gets shorter. This implies that this type of zooming extends some portions of a sight over the video clearness. Additionally, the input is influenced by a direction's modification once registering the type of findings that are ever being changed for the sight without putting an impact on the quality [48]. The output properties indicate to the last illustration that is provided for the captured video. In the platform depiction, the optical surveillance output depiction is classified within two schemes, which comprises: 2D and 3D schemes, and is created into several colouring choices. For 2D images, image pixels are built onto 2-plane spaces including X and Y coordinates where 3D images are built onto 3-plane spaces including of X, Y and Z coordinates. The 3D images are displayed through 2 coordinates by utilising a 2D sketched system. Nonetheless, to effectively visualise the turning of an entity from various visions, it should be formed over 3 coordinates within a 3D sketched system [49]. In terms of video recordings, unique frames can visualise an image in either 3D or 2D according to different abilities of the cameras. Nevertheless, not only can the entire programs can handle 3D images, but can also order large sources for handling the frames efficiently including different effective performances, particularly, for an actual time procedure. When minimising 3D image schemes into 2D image schemes (named flattening images), it leads to loose many characteristics (information) regarding an image (e.g. minimising an image pixel or modifying many pixels' clarification quantity once they are compressed into 2 coordinates starting from the 3rd coordinate [50] [51]. Irrespective of a dimensional scheme, the colour of an image is acquired by allocating a colouring quantity for every pixel of an image, a broader a range for which those quantities refer to, the greater the size of an image. That can implemented by eliminating the colours from the image findings through grey scaling images that consist of its quantity group for 0 to 255, shades of grey ranging from (0) black and (255) white highlighting various shades (28 shades) of grey [52]. Once colours are important for videos, a scheme must be capable of using a colouring difference in an effective manner. Otherwise, the system develops the performance by utilising grey scale images representing the output of the video. The processed videos create a set of sequential images, which with its properties based on the sensor that is being utilised along with its illustrated format. The properties of the created images are different from each other in terms of the characteristics, such as in their location, noise, dimensional model, colour and frame rate (according to the distance and viewpoint from the objects to be explored) [12]. In fact, these videos are simply downgraded to lower levels in order to minimise the size and to increase the performance of the processing algorithms (e.g. 3D to 2D, or coloured images to grey scale or lower frame rate, and so on). However, improving a video to higher features is not simply conducted. In the domain of computer vision, the necessary presentation pertaining to the discovery and tracing identifies the practical fundamentals pertaining to an methods based on a represented format, such as the numbers and the kinds of the utilised cameras, processing capabilities, colour, location mobility, frame rate and so forth. The whole of these factors must be taken into account when improving a discovery and tracing scheme [45]. In the next subsections,

an elaboration of device visualisation schemes for an item discovery, tracing and categorisation is given.

4. Item Discovery

Item discovery in device visualisation controls the mission for exploring items including particular features within a video frame sequence. While the detection process is based on the type of an object, the object recognition (classification) fulfils this task by determining the object where in few advanced scenarios it determines its actions. The features of an object are related to the action patterns, features, appearance or a mixture of these features. Feature-based and appearance-based approaches are based on the external presentation of an object. On the other hand, action patterns are more based on the internal features or objects' actions, such as creating a motion. In Sections 5 and 6, motion detection methods and object classification are discussed, respectively.

5. Motion Detection

In the computer vision and image processing domains, allocating items into digitised images is not considered a hard mission. A picture frame is defined as a gathering of pixels from various quantities where every group of pixels gathers an item through to a frame and checks if it is a portion of a foreground object or an image background. In such situations of movement discovery, items form an area that is related to a frame by indicating a modification within their location in comparison with the previous frame based on other different areas in both frames with comparatively static locations [18] [38]. The most popular utilised methods for motion detection comprise: temporal differencing, optical flow, statistical models and background subtraction [10] [38]. These methods are highlighted in the subsections below.

5.1 Background Subtraction

Background subtraction is defined as a common method for motion detection that is utilised through a large range of different surveillance video applications. Conventional background subtraction methods provide comparisons between current frames which including a static background frame of consecutive pictures by utilising a pixel by pixel evaluation [53]. The aim of the background deduction is to provide segmentation for the foreground (e.g. shifting items) from the background (e.g. the remaining of the sight). Figure 3 shows the method.

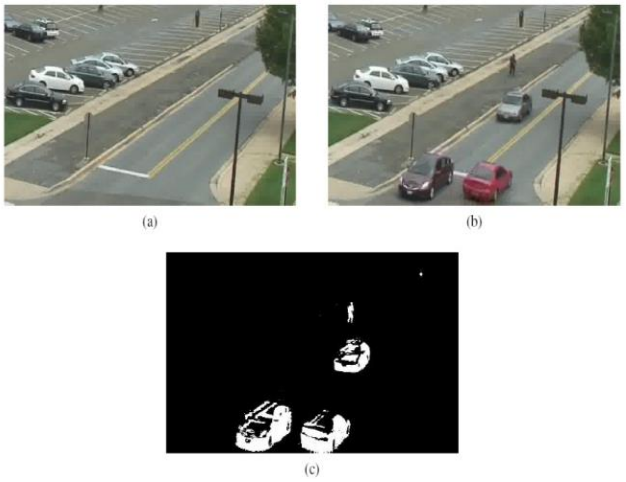

Fig. 3 Background deduction example (a) An example of the background scheme (b) The new frame (c) The background deduction findings

The foreground is processed and calculated by computing the separated pixels once providing a threshold for the absolute difference among a new frame It (x , y) and background scheme Bt (x , y) as in the equation 1:

$$| It(x,y) - Bt(x,y) | > t \quad (1)$$

Where, It denotes the new frame, Bt denotes the background scheme and t denotes the predefined threshold. The constraint of the background deduction method is based on identifying the movement within a tiny background including its sensitivity along through the illuminated modification [54]. Many methods are produced over the past years ago in order to tackle this limitation. For instance, a combination of Gaussians-based is considered to be a common method for estimating the modification of the illumination and the small object movement within the background [55] [56] [57]. The Codebook-based background subtraction model is utilised for sudden illumination modifications and dynamic backgrounds [58] [59]. The use of utilization of kernel density estimation-based approaches [60] [61] [62] and describing the pixel scheme according to a sample quantity for starting the background scheme from the 1st frame only and modifying the background scheme through the time instead of utilising a consecutive of frames to build the model [53] [63].

5.2 Chronological Change

Temporal difference computes the difference between two or three sequential frames within a video sequence in order to explore shifting areas [64]. Temporal differencing is calculated by providing a threshold for the absolute difference among the corresponding pixels within the new

frame It (x, y) including the previous frame It-1(x, y) as shown in the equation 2.

$$| It(x, y) - It-1(x, y) | > t \quad (2)$$

Where It denotes the current frame, It-1 denotes the previous frame and t denotes a predefined threshold. The temporal differencing method is computationally efficient and highly adaptive to dynamic environments. Nonetheless, it performs a weak task when exploring sufficient and applicable characteristics for pixels. Further, it is not capable of exploring the objects, which become stationary over the scene. Additionally, this method fails to explore the entire pixels of a shifting object (e.g. vehicle and person) [13][43] (see Figure 4).

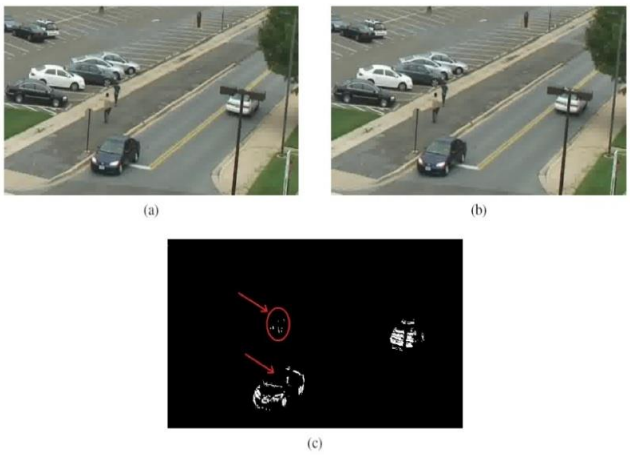

Fig. 4 Temporal Differencing Sample. (a) Frame t-1. (b) Frame t (c) Absolute frame difference result.

The findings from the changed frame illustrate some portions of the whole item. To show the faintness of the chronological change method, different mixed schemes are utilized to enhance the discovery, such as joining a frame change along with a background deduction [65] or with edge discovery methods in order to identify these edges at the beginning. Additionally, these methods calculate the variance of 2 frames, split a picture to smaller chunks, differentiate non-zero pixels with each other by including the provided threshold for reducing a noise impact and labeling every identified object through the blocks [66] [67] (see Figure 4).

5.3 Statistical Models

The behaviour of an object is highlighted in many concepts. Every concept is characterised as a mathematical equation based on a single or multiple variables. The relationships among various variables including the disseminations of their combined probability are formed in a model including the entire mathematical equations. Such models indicate to statistical models [68]. There exist many developed methods that utilise the statistical features in order to solve the main issue of major background subtraction methods [69] [51] [55] [70]. In order to improve the background subtraction, the background image is first designed by utilising a default model. Second, noises estimations are provided. Third, foreground pixels are explored. Fourth, an integration of a statistical tool and a geometrical constraint should be conducted in order to explore and eliminate shadows. Statistical models are defined as common methods in the motion detection domain due to their reliability in scenes and noise including the illumination modifications and shadows [71]. Shadow elimination and an effective background subtraction method are performed by implementing appropriate and robust statistical methods [51].

5.4 Optical Flow

Optical flow indicates to the apparent motion (i.e. rapidity of shifting) of illumination patterns within a picture structure. However, the seeming motion is produced through illuminative modifications. The motion feature regarding the visual stream refers to a 2D projection for 3D motion exteriors worldwide [72]. A direction and scale of a visual stream are formed into a 2D chart in order to categorise the shifting item based on the use of an Eigen value analysis. An optical flow is utilised for many different implementations of various domains, such as object segmentation, stereo disparity measurement, motion detection, motion investigation in videos including different types of videos. Figure 5 illustrates a picture including its visual stream vectors. This stream is effectively utilised in object tracking and motion detection applications [73] [43] [74] [54] [65]. The majority of optical flow researches are robustly related to the new design in Horn and Schunck (1981) and Lucas and Kanade (1981). The two schemes rely on differential algorithms that are implemented for two frames in order to measure the quantity of a modification between the frames according to the optical flow velocity of the pixels and delivers comparatively precise time derivatives [25].

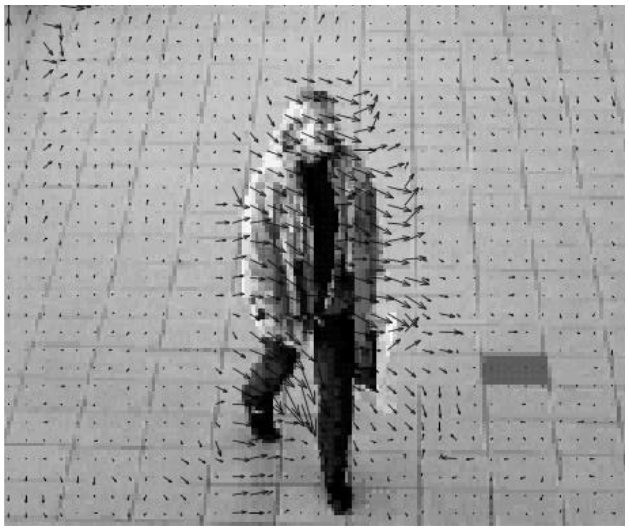

Fig. 5 Optical Flow Vectors in Moving Object

The Lucas and Kanade method is known to be faster due to the creation of the lower frame flow information based on the use of local smoothness constraints that gives "error prone" border approximations. The Horn and Schunck scheme creates a dens seamless stream feature (100% features) that offers further information of a particular frame based on the use of global smoothness constraints. Hence, it is capable of identifying a slight motion of objects, however, including lower speeds due to its iterative method. By taking into consideration the relatively low quality data that is obtained by most outdoor aerial videos and videos, the dense flow fields are needed to explore more effective characteristics for exploring an object through a motion, and thus, the concentration on the following sections of the remaining sections pertaining to this paper is based on the Horn and Schunck method. In order to provide a summary for the four motion detection methods, which are indicated within this section, Table 1 represents an overview of the disadvantages and advantages of every method including segmentation and motion detection methods. The first column involves the advantages where the second column involves the disadvantages. By comparing between the background subtraction and the motion detection methods, the statistical models and the temporal differencing are shown to be more appropriate for fixed cameras as they are capable of exploring the foreground from the background with less computational complexity in comparison with the optical flow. Nonetheless, more complicated operations are needed for non-fixed cameras in order to select the objects within the foreground with its motion, and to differentiate it among a background and camera's shifting. Consequently, visual stream schemes are seen to be further appropriate for unfixed cameras [75].

6. Item Categorization

In the device visualisation domain, object recognition methods control the tasks of extracting a provided object through a video or an image sequence by a computer system [10]. Categorising and/or determining a set of sequential images or an object within an image is not considered a trivial task as it is based on many factors, which are basically remaining from the dataset type and quality. Additionally, this case relies on the efficiency of the categorisation algorithm [76]. The purpose of the recognition algorithms is to identify whether particular objects are available within an image or not. This is performed by connecting objects through to the database of a system (training set) along to the regions of a picture (examining group) for searching for likenesses or to easily identify a group of pixels within a picture forming a separate item. Recognition encounters many common issues (e.g. the size and number of objects that appear within an image, the 3D-2D environment, object viewing point and available context. Furthermore, there exist many methods for categorising objects through an image (e.g. interpretation trees) [77] [25] and image feature-based methods (e.g. histograms and intensities).

Table 1: Advantages and Disadvantages of Motion Discovery Schemes

Schemes	Advantages	Disadvantages
Background Deduction	- Explores the furthermost accurate foreground through modelling a background (Valera and Velastin, 2005; Shuigen et al., 2009;). - Easy to be applied and rapid with static backgrounds (Withagen, 2006; Kim, Choi, Yi, Choi and Kong, 2010). (Yu Hang , 2018	-Sensitive to illumination modifications and small movements within the scene (weather, trees, etc.) (Chaohui et al., 2007; Shuigen et al., 2009; Yokoyama and Poggio, 2005). Complicated and time-consuming in estimating background models (Yokoyama and Poggio, 2005; Dewan et al., 2008).
Temporal Differencing	-Highly adapts to dynamic environments (Shuigen et al., 2009; Hu et al., 2004; Dedeoglu, 2004). -Computationally attractions (Withagen, 2006) and (Peng et. Al., 2018) -Widely using, simple , easy to implement (Chaohui et al., 2007).	-Weak at exploring the entire relevant characteristics of pixels (Shuigen et al., 2009; Hu et al., 2004; Dedeoglu, 2004; Chaohui et al., 2007). - Is not capable of controlling homogeneously coloured objects (Withagen, 2006). - is unsuccessful in finding stopped objects within the scene (Dedeoglu, 2004).
Statistical Model	- Accuracy in exploring an object to modify the illumination. -Insensitive to shadow and noise (Dedeoglu, 2004).	-Requires many training samples and computational complexities. -Is inappropriate for real-time processing (Chaohui et al., 2007).
Optical Flow	-Provide a motion of information for motion analysis (Li and Yu, 2008; Shuigen et al., 2009; Shafie et al., 2009). -Segmenting images into areas corresponding to various objects according to their individual velocities (Denman et al., 2009; Shafie et al., 2009). -Utilising an active camera (i.e. shifting cameras) (Valera and Velastin, 2005; Kim, Choi, Yi, Choi and Kong, 2010). (G. Jemilda net al., 2018)	-Is a complicated calculative method, time-consuming (Chaohui et al., 2007; Shuigen et al., 2009). -Sensitive to noise , need specialized hardware for real time video (Hu et al., 2004; Li and Yu, 2008). (G. Jemilda net al., 2018)

In order to connect with an object to an instance, every instance within the training set is initially characterised as a vector measurement (characteristic) in a characteristic extraction procedure. Characteristics might be categorical, integer or real values. The integration of the entire characteristics is called the feature space. The characteristics might express pixel intensity or might be minimised for simpler matching procedures based on the use of reduction methods, such as the Principal Component Analysis [13]. Categorising an object needs an extensive database (training set) for various objects through several viewpoints showing. In several cases, apparent characteristics identify the matching ratio. In the simple recognition domain, the system is not based on the object's nature or what it gathers, but it segments the image through to a set of various candidate objects according to the contrast of their characteristics. Additionally, the system identifies the class of candidate objects according to the training set matches or otherwise might remove unneeded objects by coupling the categorisation along with other particular algorithms. For instance, by coupling a categorisation to motion detection techniques through videos, objects having no motion or motion less than a predefined threshold are not identified. There exist many methods that determine and demonstrate an object through an image, such as silhouette, elliptical patches, multiple points, centroid, rectangular, object skeleton and contour [10].

Object categorisation is distributed in two main categories, which comprise: non-probabilistic and probabilistic classifiers. Probabilistic classifiers permit predicting more than a single class for a few entries taken from an input space, while non- probabilistic or deterministic classifiers sort entries into a single class per entry [78]. Non-probabilistic methods are the most popularly utilised in computer vision applications due to its suitable level of classification and ease of use [79]. Furthermore, the recognition purpose in this paper only includes the determination of whether an object within a frame refers to the class (vehicle) or not. Hence, non-probabilistic methods are more appropriate for delivering the answer to that question.

7. Object Tracking

Object tracking in computer vision refers to the depiction of the movement path for a shifting object through a frame sequence [80]. Tracking tracks various methods of finding and identifying shifting objects in order to control the traced object [12]. When a shifting object is found, the tracing process identifies the path of objects within the subsequent frames in the prediction techniques or the path alignment by easily determining its position and route of movement through every slide. Advanced tracking

techniques are needed to ensure that every object is effectively connected with a similar object in the following frames, and hence, calls for determining every object by a set of features for producing a particular track. When an object is found and categorised as a vehicle, the following phase involves tracking this object. Tracking is defined as the procedure of estimating the motion locations and parameters related to the item beginning from the initial frame (initialisation location) including the following frames [71]. A tracking method contains three main components, called Object Representation, Dynamic Model and Search Mechanism. The appearance of an object is influenced by many factors regarding the environmental method (i.e. the dataset features). An object representation selects the most appropriate task for finding a target object within a frame, such as demonstrating the object as an area of combined pixels within a frame. Further, advanced adaptive representation schemes are utilised according to discriminative or generative formulations. Since the representation is slightly different among the frames, a dynamic model must be given by predefining the model or by permitting the system to be aware of it based on a training data. The purpose of this model is to minimise the computational load and search space in every processed frame by estimating possible target states related to the object [39]. An object tracking for a mobile robot that is transferring throughout crowded urban environments is built on the previously proposed Deep Tracking Framework [81] [82].

The search mechanism investigates whether the tracking algorithm is stochastic or deterministic [80]. Deterministic schemes are designed and resolved accompanied by differential schemes where stochastic schemes optimise an objective task when taking into account interpretations on many frames through a Bayesian design. An item illustration defines the pixels within a particular characteristic space. A searching scheme solves an optimisation issue that minimises a searching function through particular characteristics within a characteristic space within the similarity, distance or categorisation measurements.

8. Common Issues in Aerial Surveillance

Videos of aerial surveillance encounter many popular problems (e.g. altitude change, illumination change and parallax motion). These problems that are encountered with the dataset are highlighted and explained in this section.

8.1 Parallax Motion

Shifting objects are controlled on the ground within the aerial videos by creating a movement, which is formed by a modification pertaining through to its seeming location among an initial location and an end location within 2 subsequent times (frames) including a time change that is measured via a frame rate such that every frame is similar to a time instance. Every movement consists of a magnitude (amount). The direction and object's motion on the ground indicates to a local motion. Meantime, the camera that represents a portion of an aerial car, creates a shift including a direction and magnitude of itself. The shift of a camera produces an opposite sense of motion when a ground is shifting with the opposite direction and same magnitude, while the camera is emerged as a motionless. The sense of motion indicates to a global motion. The grouping of local and global shifts indicates to a parallax motion. The capability of exploring a local motion distinguishes it based on a global motion and is considered to form an issue of tracking and aerial surveillance motion discovery [46].

8.2 Brightness Transformation

A visual stream is based on illumination patterns within a picture structure so that the seaming motion for an item relying on illumination modification could be easily defined [72]. Optical flow estimation methods are based on motion smoothness and/or brightness constancy for exploring shifting items between 2 subsequent frames. In actual situations, such restraints are uncontrolled since a modification occurs on the illumination through the time affected by the location of the illumination source and/or volume change regarding the camera and the location of the ground object, or the location of the object itself in the frame with regards to the source of illumination. According to the modification of illumination, each pixel contains a brightness value that is modified from a frame along to the subsequent frame that produces various findings through a visual steam method. The main issue of motion discovery within aerial videos, which utilises a visual steam is to propose different schemes, which are strong enough and efficient for modifying an illumination [29] [35]".

8.3 Height Modification

When the brightness of a location is not modified, many proportions of selected frames are modified based on the height of aerial cameras. When having a similar zooming proportion, the size of a taken region within a frame is increased for greater heights. Hence, the items' sizes that appear through a particular frame are minimized via a similar ratio. An impact on the height modification causes an issue in selecting the suitable size of the shifting objects, which are taken into account to be relevancies to the detection, tracking and categorization methods [21].

9. Conclusions

In this paper, the features of the implemented videos in computer visions are introduced, including details of their camera positions, uses, functionalities and demonstrations according to an object tracking and a difference between fixed-, moving- and PTZ cameras. This paper presented various methods of detecting shifting objects in captured frames and the appropriateness, restrictions and possible methods for the techniques of a background subtracting temporal differencing, optical flow and statistical approaches where the optical flow method is highlighted in this paper as the most appropriate method for aerial videos. Thereafter, the definition and need of identifying objects of interest within a frame sequence are all detailed in the object categorization section. The section compares between non-probabilistic and probabilistic classifiers, and highlights general categorization models and purposes. Subsequently, an object tracking in computer vision is drawn throughout the research general model and its tasks for tracking the components of different models related to an object representation where the search mechanism and dynamic model are also determined. In conclusion, the paper presents three main issues for object detection, categorization and tracking within aerial videos, which form the parallax motion that is created by the camera's dual movement and their objects on the ground, the influence of the camera altitude modification over the object representation through the captured frame and the influence of the illumination modification over the optical flow domain.

References

[1] Raty, T. (2010). Survey on Contemporary Remote Surveillance Systems for Public Safety, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on Vol.40 ,No.5, pp: 493–515.

[2] Oh, S., Park, S. and Lee, C. (2007). A Platform Surveillance Monitoring System Using Image Processing for Passenger Safety in Railway Station, Control, Automation and Systems, 2007. ICCAS'07. International Conference on, IEEE, pp. 394–398.

[3] Cavallaro, A. (2005). Event Detection in Underground Stations Using Multiple Heterogeneous Surveillance Cameras, Advances in Visual Computing pp. 535–542.

[4] Galati, G., Leonardi, M., Cavallin, A. and Pavan, G. (2010). Airport Surveillance Processing Chain for High Resolution Radar, Aerospace and Electronic Systems, IEEE Transactions on. Vol.46 ,No.3, pp: 1522–1533.

[5] O'Hara, S. and Fischer, A. (2009). Detecting People in IR Border Surveillance Video Using Scale Invariant Image Moments, Proceedings of SPIE, Vol. 7340, p. 73400L.

[6] Xu, X., Tang, J., Liu, X. and Zhang, X. (2010). Human Behavior Understanding for Video Surveillance: Recent Advance, Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on, IEEE, pp. 3867–3873.

[7] D'Orazio, T. and Leo, M. (2010). A Review of Vision-Based Systems for Soccer Video Analysis, Pattern Recognition Vol.43, No.8, pp: 2911–2926.

[8] Poppe, R. (2010). A Survey on Vision-Based Human Action Recognition, Image and vision computing Vol.28, No.6, pp: 976–990.

[9] Yu Hang , Chen Derong, Gong Jiulu, (2018)., Object tracking using both a kernel and a non-parametric active contour model, Neurocomputing. Vol.29 ,No.5 (2018),pp: 108–117.

[10] Hu, W., Tan, T., Wang, L. and Maybank, S. (2018). A Survey on Visual Surveillance of Object Motion and Behaviors, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 34(3): 334–352.ISSN 1841-9836, Vol.13, No.2, pp: 162-174, April 2018.

[11] Galič, M., Timan, T., & Koops, B. J. (2017). Bentham, Deleuze and beyond: an overview of surveillance theories from the panopticon to participation. Philosophy & Technology, 30(1), 9-37.

[12] Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Zhao, X., & Kim, T. K. (2014). Multiple object tracking: A literature review. arXiv preprint arXiv:1409.7618.

[13] Kim, I., Choi, H., Yi, K., Choi, J. and Kong, S. (2010). Intelligent Visual Surveillance - A Survey, International Journal of Control, Automation and Systems. Vol.8 ,No.5, pp: 926–939.

[14] Weinland, D., Ronfard, R. and Boyer, E. (2011)., A Survey of Vision-Based Methods for Action Representation, Segmentation and Recognition, Computer Vision and Image Understanding. 115(2): 224–241.

[15] Taha, A., Zayed, H. H., Khalifa, M. E., & El-Horbaty, E. S. M. (2015). Human activity recognition for surveillance applications. In Proceedings of the 7th International Conference on Information Technology (pp. 577-586).

[16] Mukhtar, A., Xia, L., & Tang, T. B. (2015). Vehicle Detection Techniques for Collision Avoidance Systems: A Review. IEEE Trans. Intelligent Transportation Systems, 16(5), 2318-2338.

[17] Wang, L., Hu, W. and Tan, T. (2003). Recent Developments in Human Motion Analysis, Pattern recognition 36(3): 585–601.

[18] Abu Karaki, H & Alomari, S. (2018). An interactive review of object motion detection, classification and tracking algorithms related to video analysis in computer vision. International Journal of Engineering Research and Technology. 11. 771-790.

[19] Goshtasby, A. A. (2005). 2-D and 3-D Image Registration: for Medical, Remote Sensing, and Industrial Applications, Wiley-Interscience. Shah, M. and Kumar, R. (2003). Video Registration, Vol. 5, Springer.

[20] Rao, C., Guo, Y., Sawhney, H. and Kumar, R. (2006). A Heterogeneous Feature-Based Image Alignment Method, Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, Vol. 2, IEEE, pp. 345–350.

[21] Mendoza-Schrock, O., Patrick, J. and Blasch, E. (2009). Video Image Registration Evaluation for a Layered Sensing Environment, Aerospace & Electronics Conference (NAECON), Proceedings of the IEEE 2009 National, IEEE, pp. 223–230.

[22] E. Karami, M. Shehata, A. Smith (2015), "Image Identification Using SIFT Algorithm: Performance Analysis Against Different Image Deformations," in Proceedings of the 2015 Newfoundland Electrical and Computer Engineering Conference,St. john's, Canada, November, 2015.

[23] Jwa, S., Tang, Z. and Ozguner, U. (2006). Robust Data Alignment Based on Information Theory and its Applications in Road Following Situation, Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE, IEEE, pp. 1328–1333.

[24] Sankaranarayanan, K. and Davis, J. W. (2008). A Fast Linear Registration Framework for Multi-Camera GIS Coordination, Advanced Video and Signal Based Surveillance, 2008. AVSS'08. IEEE Fifth International Conference on, IEEE, pp. 245–251.

[25] G. Jemilda, S. Baulkani, (2018)., Moving Object Detection and Tracking using Genetic Algorithm Enabled Extreme Learning Machine, International Journal of Computers Communications & Control Issn 1841-9836, 13(2), 162-174, April 2018.

[26] Burghardt, T. and Calic, J. (2006). Analysing Animal Behaviour inWildlife Videos using Face Detection and Tracking, Vision, Image and Signal Processing, IEE Proceedings-, Vol. 153, IET, pp. 305–312.

[27] Schwartz, W. (2011). Human Detection Based on Large Feature Sets Using Graphics Processing Units, Informatica: An International Journal of Computing and Informatics 35(4): 473–479.

[28] Jazayeri, A., Cai, H., Zheng, J. Y. and Tuceryan, M. (2010). Motion Based Vehicle Identification in Car Video, Intelligent Vehicles Symposium (IV), IEEE, pp. 493–499.

[29] Xiao, J., Cheng, H., Sawhney, H. and Han, F. (2010). Vehicle Detection and Tracking in Wide Field-of-View Aerial Video, Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, pp. 679–684.

[30] Liu, X., Dai, B. and He, H. (2011). Real-Time On-Road Vehicle Detection Combining Specific Shadow Segmentation and SVM Classification, Digital Manufacturing and Automation (ICDMA), 2011 Second International Conference on, IEEE, pp. 885–888.

[31] Dewan, M., Hossain, M. and Chae, O. (2008). Moving Object Detection and Classification Using Neural Network, Proceedings of the 2nd KES International conference on Agent and multi-agent systems: technologies and applications, Springer-Verlag, pp. 152–161.

[32] Lai, J., Huang, S. and Tseng, C. (2010). Image-Based Vehicle Tracking and Classification on the Highway, Green Circuits and Systems (ICGCS), 2010 International Conference on, IEEE, pp. 666–670.

[33] Kaiser, R., Thaler, M., Kriechbaum, A., Fassold, H., Bailer, W. and Rosner, J. (2011). Real- Time Person Tracking in High-Resolution Panoramic Video for Automated Broadcast Production, Visual Media Production (CVMP), 2011 Conference for, IEEE, pp. 21–29.

[34] Mohedano, R., Del-Bianco, C., Jaureguizar, F., Salgado, L. and Garcia, N. (2008). Robust 3D People Tracking and Positioning System in A Semi-Overlapped Multi-Camera Environment, Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, IEEE, pp. 2656–2659.

[35] Babenko, B., Yang, M. and Belongie, S. (2011). Robust Object Tracking with Online Multiple Instance Learning, Pattern Analysis and Machine Intelligence, IEEE Transactions on 33(8): 1619–1632.

[36] Trivedi, M., Mikic, I. and Kogut, G. (2000). Distributed Video Networks for Incident Detection and Management, Intelligent Transportation Systems, 2000. Proceedings. 2000 IEEE, IEEE, pp. 155–160.

[37] Popoola, O. and Ma, H. (2012). Detecting Abnormal Behaviors in Crowded Scenes, Research Journal of Applied Sciences 4.

[38] Withagen, P. (2006). Object Detection and Segmentation for Visual Surveillance, PhD thesis, University of Amsterdam.

[39] Zhao, Y., Gong, H., Jia, Y. and Zhu, S. (2012). Background Modeling by Subspace Learning on Spatio-Temporal Patches, Pattern Recognition Letters Vol.33 ,No. 9, pp: 1134–1147.

[40] Alexandre, L. and Campilho, A. (1998). A 2D Image Motion Detection Method Using a Stationary Camera, RECPAD98, 10th Portuguese Conference on Pattern Recognition, Lisbon, Portugal, Citeseer.

[41] Kang, J., Cohen, I. and Medioni, G. (2004). Tracking Objects from Multiple Stationary and Moving Cameras, Intelligent Distributed Surveilliance Systems, IEE, IET, pp. 31–35.

[42] Primdahl, K., Katz, I., Feinstein, O., Mok, Y., Dahlkamp, H., Stavens, D., Montemerlo, M. and Thrun, S. (2005). Change Detection from Multiple Camera Images Extended to Non-Stationary Cameras, Proceedings of Field and Service Robotics, Vol.2 ,No.1, pp: 15–23.

[43] Kim, J., Ye, G. and Kim, D. (2010). Moving Object Detection Under Free-Moving Camera, Image Processing (ICIP), 2010 17th IEEE International Conference on, IEEE, pp. 4669– 4672.

[44] Suhr, J. (2011). Moving Object Detection for Static and Pan-Tilt-Zoom Cameras in Intelligent Visual Surveillance, PhD thesis, Yonsei University.

[45] Porikli, F. and Tuzel, O. (2005). Object Tracking in Low-Frame-Rate Video, SPIE Image and Video Communications and Processing, Vol. 5685, Citeseer, pp. 72–79.

[46] Kang, J., Cohen, I., Medioni, G. and Yuan, C. (2005). Detection and Tracking of Moving Objects from a Moving Platform in Presence of Strong Parallax, Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, Vol. 1, IEEE, pp. 10–17.

[47] Somhorst, M. (2012). Multi-Camera Video Surveillance System, Master's thesis, Delft University of Technology.

[48] Wick, D. and Martinez, T. (2004). Adaptive Optical Zoom, Optical Engineering 43(1): 8–9.

[49] Sonka, M., Hlavac, V. and Boyle, R. (1999). Image Processing, Analysis, and Machine Vision, PWS Pub.

[50] Harman, P., Flack, J., Fox, S. and Dowley, M. (2002). Rapid 2D to 3D Conversion, Proc. of the SPIE, Vol. 4660.

[51] Jung, C. (2009). Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences, Multimedia, IEEE Transactions on 11(3): 571–577.

[52] Gonsales, R., Woods, R. and Eddins, S. (2004). Digital Image Processing Using MATLAB, Pearson Education Inc., publishing as Prentice Hall.

[53] Barnich, O. and Van Droogenbroeck, M. (2011). ViBe: A Universal Background Subtraction Algorithm for Video

Sequences, Image Processing, IEEE Transactions on (99): 1–1.

[54] Yokoyama, M. and Poggio, T. (2005). A Contour-Based Moving Object Detection and Tracking, Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on, IEEE, pp. 271–276.

[55] Bouwmans, T., El Baf, F. and Vachon, B. (2008). Background Modeling Using Mixture of Gaussians for Foreground Detection - A Survey, Recent Patents on Computer Science. Vol.1 ,No.3, pp: 219–237.

[56] Zhang, X., Hu, W., Luo, G. and Maybank, S. (2007). Kernel-Bayesian Framework for Object Tracking, Proceedings of the 8th Asian conference on Computer vision-Volume Part I, Springer-Verlag, pp. 821–831.

[57] Lee, K., Chuang, Y., Chen, B. and Ouhyoung, M. (2009). Video Stabilization using Robust Feature Trajectories, Computer Vision, 2009 IEEE 12th International Conference on, IEEE, pp. 1397–1404.

[58] Kim, K., Chalidabhongse, T., Harwood, D. and Davis, L. (2005). Real-Time Foreground- Background Segmentation Using Codebook Model, Real-time imaging 11(3): 172–185.

[59] Sun, I.-T., Hsu, S.-C. and Huang, C.-L. (2011). A Hybrid Codebook Background Model for Background Subtraction, Signal Processing Systems (SiPS), 2011 IEEE Workshop on, IEEE, pp. 96–101.

[60] Baradaran Kashani, H. and Seyedin, S. (2011). Background Estimation in Kernel Space, International Journal of Pattern Recognition and Artificial Intelligence 25.

[61] Mittal, A. and Paragios, N. (2004). Motion-Based Background Subtraction Using Adaptive Kernel Density Estimation, Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Vol. 2, IEEE, pp. II–302.

[62] Comaniciu, D., Ramesh, V. and Meer, P. (2003). Kernel-Based Object Tracking, Pattern Analysis and Machine Intelligence, IEEE Transactions on Vol.25, No.5, pp: 564–577.

[63] Barnich, O. and Van Droogenbroeck, M. (2009). ViBe: A Powerful Random Technique to Estimate the Background in Video Sequences, Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, pp. 945–948.

[64] [Zhao, S., Zhao, J., Wang, Y. and Fu, X. (2006). Moving Object Detecting Using Gradient Information, Three-Frame-Differencing and Connectivity Testing, AI 2006: Advances in Artificial Intelligence pp. 510–518.

[65] Zhang, P., Cao, T. and Zhu, T. (2010). A Novel Hybrid Motion Detection Algorithm Based on Dynamic Thresholding Segmentation, Communication Technology (ICCT), 2010 12th IEEE International Conference on, IEEE, pp. 853–856.

[66] Chaohui, Z., Xiaohui, D., Shuoyu, X., Zheng, S. and Min, L. (2007). An Improved Moving Object Detection Algorithm Based on Frame Difference and Edge Detection, Image and Graphics, ICIG 2007. Fourth International Conference on, IEEE, pp. 519–523.

[67] Wu-Chih Hua, Chao-Ho Chen b, Tsong-Yi Chen b, Deng-Yuan Huang c, Zong-Che Wud, (2015)., Moving object detection and tracking from video captured by moving camera, J. Vis. Commun. Image, Elsevier 164 180, http://dx.doi.org/10.1016/j.jvcir.2015.03.003.

[68] Kaplan, D. (2009). Statistical Modeling: A Fresh Approach, 2 edn, CreateSpace.

[69] Tsai, D. and Lai, S. (2009). Independent Component Analysis-Based Background Subtraction for Indoor Surveillance, Image Processing, IEEE Transactions on Vol.18 ,No.1, pp: 158–167.

[70] Woo, H., Jung, Y., Kim, J. and Seo, J. (2010). Environmentally Robust Motion Detection for Video Surveillance, Image Processing, IEEE Transactions on Vol.19 ,No.11,pp: 2838–2848.

[71] Dedeoglu, Y. (2004). Moving Object Detection, Tracking and Classification for Smart Video Surveillance, Master's thesis, Bilkent University.

[72] Horn, B. and Schunck, B. (1981). Determining Optical Flow, Artificial intelligence, Vol.17 ,No.1,pp: 185– 203.

[73] Denman, S., Fookes, C. and Sridharan, S. (2009). Improved Simultaneous Computation of Motion Detection and Optical Flow for Object Tracking, Digital Image Computing: Techniques and Applications, DICTA 2009., IEEE, pp. 175–182.

[74] Shafie, A., Hafiz, F. and Ali, M. (2009). Motion Detection Techniques Using Optical Flow, World Academy of Science, Engineering and Technology 56.

[75] Lucas, B. and Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision, Proceedings of the 7th international joint conference on Artificial Intelligence.

[76] Xiao, J., Cheng, H., Feng, H. and Yang, C. (2008). Object Tracking and Classification in Aerial Videos, Proceedings of SPIE, the International Society for Optical Engineering, Society of Photo-Optical Instrumentation Engineers, pp. 696711–1.

[77] Kohavi, R. and Quinlan, J. R. (2002). Data Mining Tasks and Methods: Classification: Decision-Tree Discovery, Handbook of data mining and knowledge discovery, Oxford University Press, Inc., pp. 267–276.

[78] Prince, S. (2012). Computer Vision: Models, Learning, and Inference, Cambridge University Press.

[79] Gower, J. C. and Ross, G. J. (1998). Non-probabilistic Classification, proceedings of the 6th Conference of the International Federation of Classification Societies (IFCS-98), Springer.

[80] Gu, Y., Li, P. and Han, B. (2009). Embedding Ensemble Tracking in a Stochastic Framework for Robust Object Tracking, Journal of Zhejiang University-Science Vol. 10 ,No.10, pp: 1476–1482.

[81] P. Ondr´uˇska and I. Posner, (2016)., "Deep tracking: Seeing beyond seeing using recurrent neural networks," in The Thirtieth AAAI Conference on Artificial Intelligence (AAAI), Phoenix, Arizona USA, February 2016.

[82] P. Ondr´uˇska, J. Dequaire, D. Z. Wang, and I. Posner, (2016)., "End-to-end tracking and semantic segmentation using recurrent neural networks," arXiv preprint arXiv:1604.05091, 2016.

[83] Jun, G., Aggarwal, J. and Gokmen, M. (2008). Tracking and Segmentation of Highway Vehicles in Cluttered and Crowded Scenes, Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on, IEEE, pp. 1–6.

Hussam Saleh Abu Karaki obtained his MSc in Computer Science from Universiti Sains Malaysia (USM), Pulau Penang, Malaysia in 2013. His research interests in Machine Learning, Pattern Recognition, Deep Learning Algorithm, Computer Vision, Visual Tracking, Object Recognition and Medical Image Analysis

Saleh Ali Alomari obtained his MSc and Ph.D. in Computer Science from Universiti Sains Malaysia (USM), Pulau Penang, Malaysia in 2008 and 2013 respectively. He is a lecturer at the Faculty of Science and Information Technology, Jadara University, Irbid, Jordan. He is Assistance Professor at Jadara University, Irbid, Jordan. He was the head of the Computer Network Department at Jadara University from 2014 until 2016. He is the candidate of the Multimedia Computing Research Group, School of Computer Science, USM. He is research assistant with Prof. Dr. Putra, Sumari. He is managing director of ICT Technology and Research and Development Division (R&D) in D&D Professional Consulting Company, Malaysia. He has published over 45 papers in international journals and refereed conferences at the same research area. He is a member and reviewer of several international journals and conferences. His research interest is in the area of multimedia networking, video communications system design, multimedia communication specifically on Video on Demand system, P2P media streaming, MANETs, caching techniques, overlay Network and for advanced mobile broadcasting networks as well.

Mohammed Hayel Refai obtained his MSc in Computer Science from Universiti Sains Malaysia (USM), Pulau Penang, Malaysia in 2007, and obtained his Ph.D. in Computer Science from Universiti Utara Malaysia, Kehadh, Malaysia in 2016. He is working as assistance professor at department of Information Technology in Sur University College in Oman