International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 13, Number 5 (2017), pp. 745-765

http://www.ripublication.com

A Review of Detection and Tracking of Object from Image and Video Sequences

Mukesh Tiwari
Research Scholar, Department of Electronics & Communication, University Institute of Technology, Rajiv Gandhi Technical University, Bhopal – 462033 M.P. (India)

Dr. Rakesh Singhai
Professor, Department of Electronics & Communication, University Institute of Technology, Rajiv Gandhi Technical University, Bhopal – 462033 M.P. (India).

Abstract

Object detection and tracking is one of the critical areas of research due to routine change in motion of object and variation in scene size, occlusions, appearance variations, and ego-motion and illumination changes. Specifically, feature selection is the vital role in object tracking. It is related to many real time applications like vehicle perception, video surveillance and so on. In order to overcome the issue of detection, tracking related to object movement and appearance. Most of the algorithm focuses on the tracking algorithm to smoothen the video sequence. On the other hand, few methods use the prior available information about object shape, color, texture and so on. Tracking algorithm which combines above stated parameters of objects is discussed and analyzed in this research. The goal of this paper is to analyze and review the previous approach towards object tracking and detection using video sequences through different phases. Also, identify the gap and suggest a new approach to improve the tracking of object over video frame.

Keywords: Object tracking, object recognition, statistical analysis, object detection, background subtraction, performance analysis, optical flow

1. INTRODUCTION

Recently, there is an advance of miniaturization and lower the cost of cameras have preferred the implementation of large-scale networks of the camera. This increasing number of cameras could permit novel signal processing applications which employ multiple sensors in extensive areas. Object tracking is the novel procedure for discovering moving objects beyond time by utilizing the camera in video sequences

(Kothiya and Mistree, 2015). Their main aim is to relate the target objects as well as the shape or features, location of the objects in successive video sequences. Subsequently, the object classification and detection are essential for object tracking in computer vision application. Additionally, the tracking is the first step towards locating or detects the moving object in the frame. Followed by this, detected object could be divided as swaying tree, birds, human, and vehicles and so on. Though, in image processing approach object tracking using video sequences, is a challenging task. Furthermore, several issues appear ascribed to occlusion of the object to scene, object to object, complex object motion, real-time processing requirements as well as the improper shape of the object.

However, this tracking has a large number of benefits, few of them are traffic monitoring, robot vision, surveillance and security and video communication, public areas like underground stations, airports, mass events and animation (Kim, 2007; Lowe, 2004; Ojha and Sakhare, 2015; Yilmaz et al., 2006). Thus, the particular application needs optimal trade-off among computing, communication, and accuracy over the network. The revenue related to computing and communication relies on the amount and type of cooperation executed among cameras for data collection, dispensing and processing to confirm decisions and to reduce the estimation errors and ambivalence.

Subsequently, this tracking can be explained as the procedure of determining the orientation of object across the time as the object moves throughout a scene. This is posting importance in the arena of computer vision because of expansion of high-powered computers and the growing need for automated surveillance systems, and it is broadly applied for applications namely automated surveillance, robotics monitoring, human-machine interface, motion-based recognition, vehicle navigation, traffic monitoring and video indexing. A substantial number of such applications require reliable tracking methods which meet real-time restrictions and are challenging and complex with respect to changes of object movement, scale and appearance, illumination of scene and occlusion. The results of tracking could be impacted by the disparity of one among the parameters. Due to tackle the above-explained issues and others in object tracking numerous approaches have been proposed (Yilmaz et al., 2006). In this object tracking application, target object could be determined as anything which is engaging for analysis. In addition, moving objects tracking is one of the major tasks in computer vision and broadly applied in industrial vision, intelligent transport systems and visual surveillance (Comaniciu et al., 2003, 2000). In the recent years, Video surveillance has widely adopted to monitor the security sensitive areas include highways, borders, department stores, banks and crowded public places. The development in computing power, the infrastructure of high-speed network and accessibility of large-capacity storage devices cover the way for inexpensive, multi-sensor video surveillance systems. Keeping a track on the moving object is a critical task.

The capability of machines to identify the suspicious object and further identify their activities in a specific environment is an important part of permitting a machine to interact with humans in effective and easy manner. The current approach for analyzing and detecting the suspicious object usually needs exceptional markers

connected to the suspicious object that prevents the extensive technology application. In this paper, to study as wll as analyze the previous approach towards object tracking using video sequences through different phases.
Three key steps in video analysis are discussed as follows:

1. Identification of targeted object in moving sequence.
2. Object tracking based on one frame to another frame.
3. Tracking of the object from camera to camera.

2. LITERATURE REVIEW

In the previous study most of them have concentrated towards Object detection (Ben Ayed et al., 2015; Najva and Bijoy, 2016; Ramya and Rajeswari, 2016; Risha and Kumar, 2016; Shen et al., 2013; Soundrapandiyan and Mouli, 2015; Viswanath et al., 2015) ,Object tracking (Bagherpour et al., 2012; Coşkun and Ünal, 2016; Foytik et al., 2011; Lee et al., 2012; Poschmann et al., 2014; Weng et al., 2013; Yilmaz et al., 2006; Zhang et al., 2016) and Object recognition (Chakravarthy et al., 2015; Elhariri et al., 2015; Gang et al., 2010; Ha and Ko, 2015; Nair et al., 2011) for tracking the object using video sequences. These are discussed as follows. The basic flow diagram of an object tracking shown in figure 1.

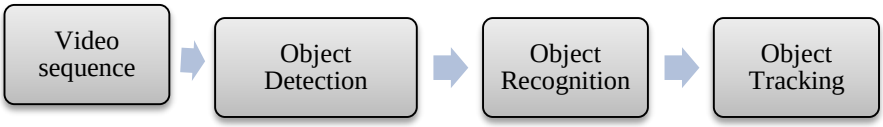

Figure 1 The Basic flow diagram of Object tracking

Studies related to object detection

The detection of an object in video sequence plays a significant role in many applications. Specifically as video surveillance applications (Amandeep and Goyal, 2015). The different types of object detection are shown in figure 2.

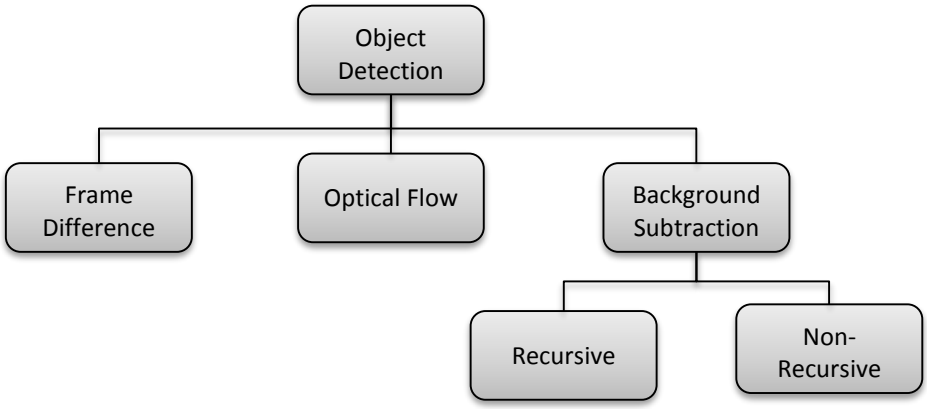

Figure 2 Types of object detection method

The previous studies related to object detection is discussed as follows:

Shen et al. (2013) proposed a novel hierarchical moving target detection method based on spatiotemporal saliency. Further, they achieved the refined detection results using temporal and spatial saliency information. The experimental results show that this approach identifies moving substances in the airborne video with high accuracy and efficiency. Additionally, this method does not have the effect of time delay when compared with HMI technique. However, this method evaluated object locations in all the video frames as self-sufficiently, false alarms which are inevitable.

Guo et al. (2012) suggested object detection approach for tracking the objects in video frames. The simulation result shows this technique was effective and accurateness, robust for generic object classes' detection with good performance. Further needs to focus towards increase classification accurateness in real-time object recognition.

Ben Ayed et al. (2015) proposed a method for detection of text data based on a texture in video frames by big data analytics. The video frames are decomposed into various fixed size blocks and these blocks are analyzed using har wavelet transform technique. Further, they used a neural network to classify the text and non-text blocks. However, this study needs to concentrate on extracting the regions towards remove the noisy regions as well as exclude the text like sections.

Viswanath et al. (2015) suggested and modeled the approach using non-panoramic background modeling. By the use of this approach, they modeled the entire picture element with one Spatio-temporal Gaussian. The simulations result shows this method able to identify the moving substances with fewer false alarms. However, this method fails once the adequate features are not obtainable from the section.

Soundrapandiyan and Mouli (2015) suggested a novel and adaptive method for pedestrian detection. Further, they separated the foreground objects from the background by image pixel intensities. Subsequently, they used high boost filter for enhancing the foreground edges. The efficacy of the proposed method is evident from the subject evaluation results as well as objective evaluation with around 90% of pedestrian's detection rate compared to the other single image existing methods. In future, they planned to improve the performance of the method with higher detection rate and low false positives on par with sequence image methods.

Ramya and Rajeswari (2016) suggested a modified frame difference method which uses the correlation between blocks of current image and background image to categorize the pixels as foreground and background. The blocks in the current image which are highly correlated with the background image are considered as background. For the other block, the pixel-wise comparison is made to categorize it as foreground or background. The experiments conducted proved this approach improves the frame difference method particularly as finding accuracy with speed. However, this study needs to focus towards other information available in the blocks such as shape and edge can be used to improve the detection accuracy.

Risha and Kumar (2016) suggested an optic flow with the morphological operation for object detection in video. Further applied morphological operation towards

obtaining clear moving target image. This study only concentrated on static camera. So need to focus on moving the camera as well as identify multiple objects in video frames.

Najva and Bijoy (2016) proposed a model for detection and classification of objects in videos by combining Tensor features with SIFT approach towards classifying the detected objects using Deep Neural Network(DNN). The DNN capable of handling large higher dimensional data with billions of parameters as like human brain. Simulation results obtained illustrate that the proposed classifier model produces more accurate results than the existing methods, which combines both SIFT and tensor features for feature extraction and DNN for classification.

Table 1 Comparative Study of Object Detection technique

Object Detection Method		**Basic Principle**	**Computational Time**	**Accuracy**	**Comments**
Temporal Differencing		Pixel-wise Subtraction of Current & Background frame	Low	High	Easy to implement (Chate et al., 2012; Mohan and Resmi, 2014)
					Sensitive to dynamic changes (Haritaoglu et al., 2000)
					Needs background frame with still objects (Mohan and Resmi, 2014)
Background Subtraction	Frame Differencing	Current frame is subtracted from background frame	Low to Moderate	Moderate to High	Simplest background Subtraction (Aldhaheri and Edirisinghe, 2014; Haritaoglu et al., 2000)
					Cannot be used for real-time applications (Mohan and Resmi, 2014)
	Approximate Median	Simple subtraction between median frame & test frame	Low to moderate	Moderate	No need for adequate background modeling (Aldhaheri and Edirisinghe, 2014)
					Requires a buffer with recent pixel values (Aldhaheri and Edirisinghe, 2014)
	Running Gaussian Average	Based on Gaussian probability density function of pixels	Moderate to high	Moderate	Much suitable for real-time applications (Aldhaheri and Edirisinghe, 2014)
					Statistical calculations consumes more time
	Mixture of Gaussian	Based on multimodal distribution	Moderate to high	Moderate to high	Low memory requirement (Zhiqiang et al., 2006)
					Cannot cope up with objects as well as noise(Tao Zhang et al., 2010)
Optical Flow		Uses optical flow distribution characteristics of pixels of object	Moderate to high	High	This approach offers entire moving data(Krishna et al., 2011) however require more calculations

Studies related to object tracking

The object tracking is the term which used to identify the moving object position as well as tracking them from video sequences (Balasubramanian et al., 2014). The tracking method is classified into three types such as kernel, point and silhouette based tracking (Yilmaz et al., 2006). Compared to silhouette method, existing most of them have focused on kernel-based method due to high accuracy with less computational cost. However, the point tracking method has less computational cost with reduce in accuracy (Weng et al., 2013). The various types of object tracking techniques are shown in figure 3.

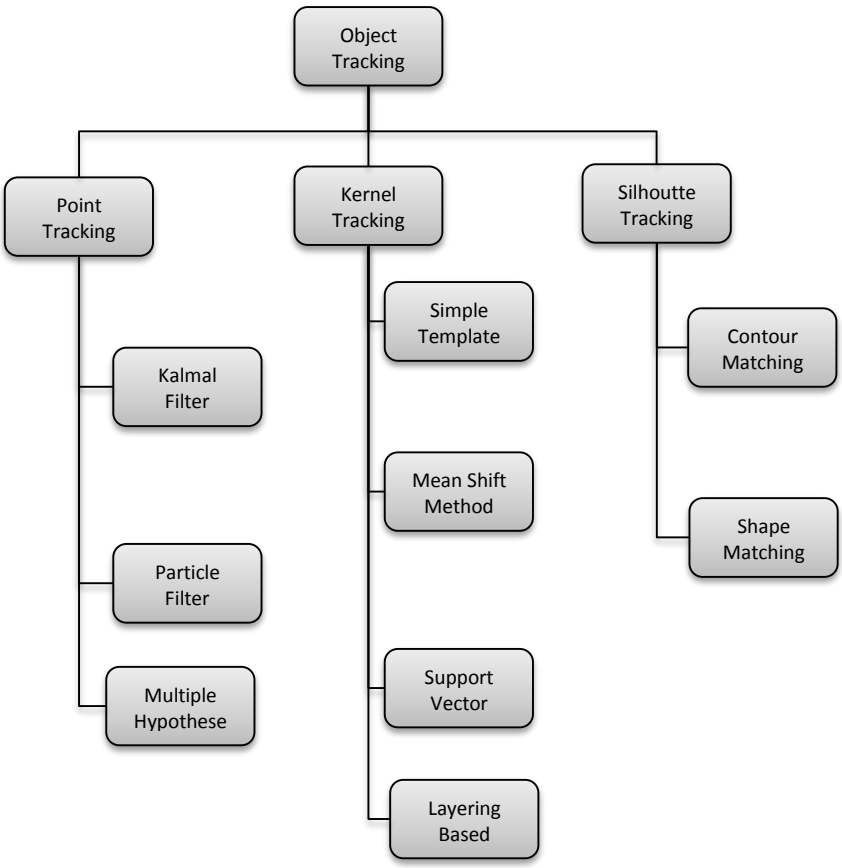

Figure 3 Types of object tracking approach

The previous studies related to object tracking method are discussed as follows:
Sarkar et al. (2012) developed color information method for identification of skin in image occurred from FERET especially for mouth and eye region detection. This technique requires less computational cost as well as applied to video sequences. However, this method incapable towards identifying the small face images which away from the camera.

Weng et al. (2013) proposed an algorithm for improving the performance of natural feature selection in the real world. Further, they used to speed up robust features (SURF) for features extraction from live mobile camera image and recognition. These extracted features are calculated using pose matrix through Homography approach. The simulation result shows this algorithm tracked and recognized the object from natural features in easy, speedy with suitable way. However, its speed and accuracy need to be improved.

Zhang et al. (2016) proposed an approach by combining frame difference and non-parametric method for video analysis traceability. The simulation result proved this approach performance was better than the traditional frame difference and GMM. Further, it can able to remove the noise from a background which gives us the ability to detect the moving object more precisely in the applications such as food and agriculture related product traceability analysis. However, this study needs to enhance the capability of traceability system and supply the visual supply chain for the common user to ensure the safety.

Coşkun and Ünal (2016) suggested camshift technique towards track the object from the video sequences. Further, they demonstrated this approach successful carried out in mobile platform even with the change in object size and illumination. The drawback observed in this proposed technique fails to perform for input video with full occlusion. Houssineau et al. (2016) suggested a parameterization based disparity space for non-rectified camera networks, extended to moving objects, and integrated into a Bayesian multi-object tracking with sensor calibration technique. The performance of the obtained framework has been demonstrated for camera calibration on simulated with real data, underlying the problems of single-object localization and tracking, as well as for multi-object tracking. Further, they planned to enhance the proposed method towards other multi-object filters and a comparative study of these approaches for camera-based tracking as well as for camera calibration.

Oiwa et al. (2016) suggested probabilistic background model towards tracking the object from video sequences. The simulation results show the accuracy and effectiveness of this method high compared to previous technique. However, this study needs to concentrate on higher speed as well as improve the accuracy of object tracking.

Mohammed and Morris (2014) suggested a color-based technique which was the combination of accruing and normalizing histograms towards object tracking under different conditions based on a mobile device. This technique was easy to use and robust against varying illuminations. However, this technique fails to identify the entire region of symbols due to the extreme camera view.

Aggarwal et al. (2006) suggested a novel technique which was the combination of motion estimation and background subtraction for object tracking using video sequences. The system mainly focused on four scenarios such as interpolation, identify the object, subtract the background and object selection. However, this study needs to focus more towards full occlusions of video sequences, multiple object tracking, fast camera motion and unsupervised object tracking.

Fatima et al. (2013) suggested image segmentation approach for track objects of interests through specifying the color intensities. A minimum distance classifier approach is used for object classification. They attain the object tracking through specifying the object centroids in all the video frames. The simulation results proved this technique was more efficient for contextual approximation. However, this study needs to focus towards implementing segmentation which would work with occluded images with multiple objects and computationally more efficient.

Lecumberry and Pardo (2005) proposed an algorithm for semi-automatic object tracking in videos by various features with probabilistic relaxation method. The proposed algorithm performs effectively for object tracking specifically for object borders is smooth and accurate. When combined with sophisticated methods it performs effectively for object tracking. The accuracy of the borders of the tracked objects depends on the power of discrimination of selected features as well as the appearance of new objects and/or background. However, this study needs to focus towards improving the accuracy of this algorithm, particularly as rigid objects.

Table 2 Comparative Study of Object Tracking Methods

Object tracking method		**Algorithm used**	**Computational time**	**Accuracy**	**Comments**
Point Tracking	Kalman Filter	Kalman filtering algorithm	Low to Moderate	Moderate	This approach applicable to track point even in noisy images (Javed and Shah, 2002)
					Distributed State variables (Gaussian) (Ali and Zafar, 2009; Javed and Shah, 2002)
	Particle Filter	Recursive Bayes filtering	Moderate to High	High	Good results for occultation and complex background (Hu et al., 2004)
					Not advisable for real-time applications due to big calculations (Athanesious and Suresh, 2012; Hu et al., 2004)
	Multiple Hypothesis tracking	MHT algorithm	Low	Low to moderate	Adapt new object as well as exists existing object(Blackman, 2004)
					High computation in memory and time(Blackman, 2004)
Kernel Tracking	Simple template matching	Matching region of interest in video	Low to moderate	Low	Capable of dealing with partial occlusion (Cucchiara et al., 2003)
					Require equivalent model for each region of interest for each image (Cucchiara et al., 2003)
	Mean shift method	Expression & location of object; optimal gradient decline	Low	Moderate	Can be used for real-time applications due to less calculations (Tao Zhang et al., 2010)
					Iterations get into local maximum easily (Tao Zhang et al., 2010)
	Support vector	Positive & negative training	Moderate	Moderate	Can handle single image and partial occlusions (Avidan, 2004)

	machine	values			Need physical initialization and training (Avidan, 2004)
	Layering based tracking	Shape representation using intensity	Moderate	Moderate to high	Track multiple objects and full occlusion (Wei and Qin, 2007)
					Require parametric models of each pixel (Wei and Qin, 2007)
Silhouette tracking	Contour matching	Gradient Descent Algorithm	Moderate	Moderate to high	Object Shape is Implicitly modeled (Karasulu, 2010)
					Requires time for state space estimation (Karasulu, 2010)
	Shape matching	Hough Transform	High	High	Less sensitive towards variance of appearance
					Need to enhance the performance (Karasulu, 2010)

Table 3 Comparative Study of Object Classification Methods

Object Classification Method	Computational Time	Accuracy	Comments
Shape Based	Low	Moderate to high	Pattern matching algorithm was applied. However, it's not worked at dynamic situation as well as unable towards control internal movements (Javed and Shah, 2002)
Motion Based	High	Moderate	struggles towards recognizing on-moving humanoid (Javed and Shah, 2002)
Color based	High	High	Delivers expenditure of added computation time with improved quality(Hu et al., 2004)
Texture Based	High	High	They described the color distribution using Gaussian Mixture Model in image sequence as well as segment background and objects of the image(Hu et al., 2004)

Studies related to object recognition

Gang et al. (2010) developed a kernel locality preserving projections (KLPP) towards improving the accuracy and separable description of the objects. The simulation results proved, this approach more appropriate for space object recognition mainly considering changes of viewpoints. However, this study needs to concentrate on improving the accuracy level towards recognizing with limited trained objects as reference. Also, test the effectiveness of this approach on images of other objects models.

Nair et al. (2011) developed a technique for a combination of face recognition, tracking and detection approach towards identifying the individual faces. The simulation result shows this method increases the level of accuracy in recognizing and tracking faces and applicable to the real-time application. However, it fails in varying illumination conditions as well as the existence of lesser background in face recognition which limits the system performance.

Zhang and Jiang (2014) suggested regression based kernel technique on behalf of identifying multiview objects as well as approximating their poses. The simulation results proved obtained results are improved recognition, while comparative analyses with state-of-the-arts. Further, they authenticated the robustness and efficiency of this approach. However, the existing research needs to focus more attention on resolving issues of object space recognition

Chakravarthy et al. (2015) suggested and estimated a technique for video sequence stream processing. Further, they demonstrated the various types of situations which were based on arable and relational illustrations. However, this study fails to express complex situations. So, needs to focus towards accepting feature vectors as well as bounding boxes towards identifying the temporal and spatial computations.

Ha and Ko (2015) proposed vision-based shadow gesture recognition method for interactive projection systems. This method only splits the shadow area through merging the binary image with an input image using a learning algorithm that isolates the background from the input image. The developed approach isolates or differentiates shadow of a hand in based on convex hull, moment and defect in each region. After that, isolated the hand shadows by the convex hull, defect, and moment of each region. However, the robustness of this processing is not always assured. It means that at some conditions the gesture recognition fails, and unexpected results would happen.

Elhariri et al. (2015) discussed Random Forests (RF), Linear Discriminant Analysis (LDA) classifiers, Support Vector Machines (SVMs).Furthermore, for edge detection segmentation approach has been utilized in this research. Subsequently extracted the features using morphological operations. The simulation result shows the LDA and SVM provide better performance. However, the input dataset as elderly needs to be involved.

2.1 Statistical methods of tracking

Feris (2000) suggested a statistical skin-color model towards detecting the various regions of candidate's face as well as applied template matching techniques to define presence or absence of a face in particular region. They tracked the face of the candidate by lip corners, the location of pupils and nostrils location in the tracked image sequence. However, this study needs to improve the system robustness and accuracy of detection as well as tracking the facial features.

Foytik et al. (2011) proposed a Kalman filter for low-level recognition of differentiating between various objects. Then they analyzed the face recognition at various subspace formed from the big database and analyzed through Adaptive MPCA (Modular Principal Component Analysis). The simulation results proved this approach has more accuracy and able to make recognized faces with average time.

Bagherpour et al. (2012) developed a novel framework for tracking of upper parts' of the body using KLT (Kanade-Lucas-Tomasi) and Kalman filter for motion estimation. In the developed model limbs and motion of the body is estimated using Kalman filter

for accurate prediction of limb position. The experimental result shows that the developed model is compatible with the indoor environment for variation in object pose, background clutter and for various scenes. This study can further concentrate on improvisation of tracking results.

Lee et al. (2012) investigated different occlusion scenario and performed tracking under six different video simulation methods. They evaluated the performance using SFDA (Sequence Frame Detection Accuracy). Furthermore, they demonstrated mean shift, particle and Kalman filtering for evaluating tracking performance. Additionally, they found that for arbitrary movement of the object Particle Filter (PF) fails to perform effectively.

Kim (2007) objects are randomly chosen by a user are tracked using SIFT features and a Kalman filter. Specifically, they concentrated on tracking human, car, or pre-learned objects. The objects are accumulated, exploited the learning to successfully track objects even when the objects missing for some frames. However, this study needs to focus on higher resolution with finding the location of stationary objects.

Nagendran et al. (2014) proposed a method for effectively tracking moving objects in videos. They used affine transformation for stabilizing the video. Then extract these features using frame selection. Further, they used Kalman filter and Gaussian mixture model for tracking the moving objects. However, this study needs to concentrate on reduction of computational time as well as increasing recognition for various categories.

Poschmann et al. (2014) developed a PF approach using fusion technique for increasing adaptive tracking robustness. This research comparatively analyzed the various variants and demonstrated the feasibility of applying a framework for a real-world scenario. The major difficulty identified in this research is the threshold for learning is crucial which will be either too high or too low. Another issue identified is based on video, threshold is not updated whether bad or none. The stated problem can be overcome by exploitation of adaptive threshold feasibility in proposed approach or else need to find an alternate way to this challenge.

Mei et al. (2015) proposed a LAD (least absolute deviation) learning method based on a multitasking and multiview technique for tracking. The proposed approach uses PF for effective object tracking. The proposed approach is implemented under four different features of objects like color histogram, intensity, LBP (Local binary patterns) and HOG (Histogram of Oriented Gradients). Further, this research is examined under several challenging situations like noise availability in real-world, synthetic noisy sequence, availability of sequence in public and complete tracking of available data sets. The simulated results demonstrate that proposed method was provided the advantage of multiview data handling and task outlier. Further, the proposed approach exhibits superior performance for comparative examination of existing tracking methods.

Chandrajit et al. (2016) suggested a feature based method towards track the multiple moving objects in surveillance video sequence is proposed. This proposed method is assessed quantitatively using the precision and recall accuracy metrics. Further,

comparative evaluation with related works has been carried out to exhibit the efficacy of the proposed method. However, this study needs to focus on occlusion handling strategy and shadow elimination for increase the tracking results.

Moreno-Garcia et al. (2010) developed a motion-based algorithm for accurate prediction of moving objects shapes in the video sequence. By using fuzzy logic sequence segmentation and tracking performance were enhanced which improves uncertainty of work based on permanency value caused by impairment in computer vision. However, this study needs to focus on improving the Analysis phase in order to make it much more descriptive. In order to conquer the previous image for local search, tracking phase will be modified in the fuzzy region or else need to create a method to fit with the fuzzy region.

3. MATERIALS AND METHODS

The present study reviewed 74 research papers sharing the same context of the present study. The papers are selected from the search of comprehensive literature performed using the IEEE transactions, reputed international journals such as Elsevier, Springer, ACM, and papers through Google search. With the review and reference section from previous research papers, individual papers were selected for the study of previous literature. The selected studies support the video-based object detection, tracking and recognition with performance analysis. After careful analysis of previous papers, 48 articles that serve to be potentially useful for the present study were selected for the research.

4. RESULTS AND DISCUSSION

Shen et al. (2013) proposed a novel hierarchical moving target detection method. However, the drawback of the proposed method is the unavoidable cases of false alarms since frame by frame detection of object location is estimated. Soundrapandiyan and Mouli (2015) suggested a method for improving the performance with higher detection rate and low false positives on par with sequence image methods. Ramya and Rajeswari (2016) suggested a modified frame difference method which uses the correlation between blocks of current image and background image to categorize the pixels as foreground and background. However, this study needs to focus towards other information available in the blocks such as shape and edge can be used to improve the detection accuracy. Sarkar et al. (2012) developed a technique for face detection using color information availability. The model is potential enough to detect faces from images that are of low-resolution or from surveillance video frames. Furthermore, the proposed model also identifies face alignment in the images. Hence, image registration is facilitated, and the registered images are used for future uses.

Weng et al. (2013) enhanced the performance of tracking for real-world objects using naturally formulated approach. However, its speed and accuracy need to be improved through replacing SURF algorithm with the different computationally efficient algorithm. Poschmann et al. (2014) incorporated particle fusion approach with other

existing techniques for enhancing tracking performance which facilitate the adaptive tracking of the objects. The threshold for learning faces is a crucial task wherein if the threshold is too high or low, then there arise chances of bad or no updates which depends on the characteristics of the video. Hence, a necessity persists to examine the possibility of possessing an adaptive threshold or identify new methods to resolve this issue. Zhang et al. (2016) developed a novel algorithm for improving video analysis traceability by combining frame difference and non-parametric techniques. However, this study needs to enhance the capability of traceability system and supply the visual supply chain for the common user to ensure the safety of food and agriculture related products. A conventional camshaft approach is used by Coşkun and Ünal (2016) to track objects. However, with full occlusion, the proposed model will fail. Gang et al. (2010) devised 3d Objects based identification based on kernel locality preserving projections (KLPP).In a similar manner, Nair et al. (2011) developed a system for detection and recognition of various pose of the face also it tracks face efficiently. The proposed model also extends the face detection algorithm to track and recognized individual profile faces. However, the study limits its scope to face detection algorithm whereas there is a dire need to focus on skin segmentation algorithm which is used to create a face mask. The face mask creation mechanism is an exceptional algorithm wherein with the creation of face based on the pose; background could be neglected from the face training system. Additionally, pose estimation module is required for the determination of pose and suitable Eigenspace should be selected rather than the use of cascades. Elhariri et al. (2015) proposed an object recognition system which is based on the use of SVMs, LDA classifiers, RF and Histogram of Oriented Gradients. However, the previous study lacks focus on certain features which when applied with other machine learning classifiers achieves better accuracy of recognition. Objects that are faced previously are included as the present datasets does not include them. Additionally, faces could be utilized for the extension of the proposed system's applicability.

Many researchers have concentrated namely on Particle filter, Mean shift, and Kalman filter. Several works have been proposed to combine the prospects of these trackers in order to achieve better results of tracking. For instance, previous studies of Li et al. (2010) and Zhao et al. (2009) have combined Mean-shift tracking and Kalman filtering whereas Tang and Zhang (2011) devised a combined model of PF with Mean Shift tracking. One of the most explored objects tracking issues in computer vision is the handling of occlusion. Existing literature works have argued that occlusion handling is achieved through Particle filter, Kalman filter and Mean Shift tracking methods. Though the fusion of these methods is accurate, they are tested only on particular videos (Lee et al., 2012). It is also deemed that face detection methods imparting neural network-based approaches are exceptionally accurate and are computationally intensive in nature. Hence, there is a lack of possibilities in implementing neural approaches in real-time applications (Sarkar et al., 2012).

Visual tracking is facilitated by the least squares (LSs) criterion which is the use of Euclidean distance to calculate approximately tracking of sparse representation (Xue Mei and Haibin Ling, 2011; Zhang et al., 2013). It is deemed that if Gaussian data distribution is involved, then the performance of LS criterion is efficient. The reason

is attributed to the properties of smoothness and differentiability and solutions are mostly based on gradient methods (Nesterov, 2013). However, real tracking instances are prone to noise which includes cases of Laplace noise, background clutter, and salt-and-pepper noise. In such cases, the methods of LS- criterion are into serious degradation and hence accurate estimation of noise could not be performed. Through the analysis of existing researchers, Guan et al. (2012) and Harter (1974) claim that least absolute deviation (LAD) has better efficiency in estimation of noise than LS, especially in cases where heavy-tailed noise prevails. On the contrary, tracking issues imply different data representations which are categorized based on intensity (Mei et al., 2007), color (Collins et al., 2005), edge (Kwon and Le, 2010), wavelet (Jepson et al., 2003), and texture. Tracking performance could be improved significantly if different sources of information are considered since with these characteristics, tracking could be facilitated (Badrinarayanan et al., 2007; Du and Piater, 2008; Kwon and Le, 2010; Weifeng Liu and Dacheng Tao, 2013). Though multiple views are to be considered, another issue is the ways of integrating data representations and a model to examine independence and mutual dependence. Research by Zhang et al. (2013) reported that for visual tracking applications features are not possible to implement as individual task hence it needs to be an exhibit in all sequential tasks. Furthermore, outlier tasks also exist very often. Outlier tasks do not share the features that are common with the majority of tasks. Similarly, Zhang et al. (2013) and Xue Mei and Haibin Ling (2011) utilized the feature of intensity to model target's appearance change. The intensity appearance model with L1 minimization shares robustness towards partial OCC and other tracking issues (Xue Mei and Haibin Ling, 2011). Even though this developed model is sensitive to track non-rigid objects since shape deformation is more evident with these objects.

The present study has reviewed 74 articles, from which 19 from IEEE digital library, 7 from an international conference, 8 were from springer digital library, 6 from Elsevier, 6 from Science direct, and 28 from Google scholar (Figure 4). Of these articles, the following aspects are addressed: Object detection, tracking, recognition and statistical methods of tracking specifically as a video sequence.

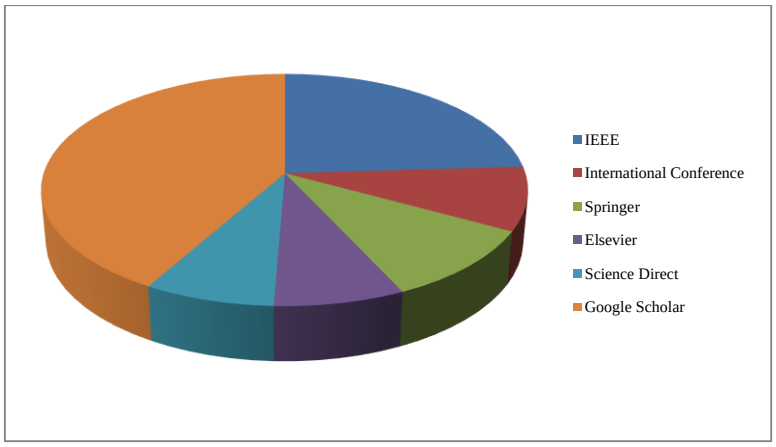

Figure 4 Number of research article referred

5. CONCLUSION

In this paper, review on different object detection, tracking, recognition techniques, feature descriptors and segmentation method which is based on the video frame and various tracking technologies. This approach used towards increase the object detection with new ideas. Furthermore, tracking the object from the video frames with theoretical explanation is provided in bibliography content. The bibliography content is the most significant contribution of research since it will lead to a new area of research. We have identified and discussed the limitation/future scope of various methods. Also, we have noted some methods which give accuracy but have high computational complexity. Specifically, the statistical methods, background subtraction, temporal differencing with the optical flow was discussed. However, these technique needs to concentrate towards handling sudden illumination changes, darker shadows and object occlusions (Susar and Dongare, 2015).

6. FUTURE SCOPE

- Design and simulation of complex video sequence and test them using same tracking algorithm. In the potential scenario, occlusion is used for an object with the same color for the moving objects or else using bigger occlusion with longer occlusion time. Increasing the number of the object help to identify the efficiency and functionality of the tracking algorithm.
- Weight parameters are needed to be added for individual intensity levels of each pixel. In an image, if an intensity value is assigned as foreground based on the current frame then it has less probability that foreground also has similar pixel coordinate so that BG weightage for the pixel is set to the minimum than the initial value. Through adding weightage lower than the initial value provides the advantage of removing the old pixel value with least probability rather than the evolved scene.
- Need to focus towards enhancing the variance data of each channel based on the Mahalanobis distance calculation. By this, can able to adopt a change in the rapid scene through Euclidean distance algorithm.

REFERENCES

[1] Aggarwal, A., Biswas, S., Singh, S., Sural, S. & Majumdar, A.K., 2006. Object Tracking Using Background Subtraction and Motion Estimation in MPEG Videos, in 7th Asian Conference on Computer Vision. Springer-Verlag Berlin Heidelberg, pp. 121–130. doi:10.1007/11612704_13

[2] Aldhaheri, A.R. & Edirisinghe, E.A., 2014. Detection and Classification of a Moving Object in a Video Stream, in: Proc. of the Intl. Conf. on Advances in Computing and Information Technology. Institute of Research Engineers and Doctors, Saudi Arabia, pp. 105–111. doi:10.3850/ 978-981-07-8859-9_23

[3] Ali, S.S. & Zafar, M.F., 2009. A robust adaptive method for detection and tracking of moving objects, in: 2009 International Conference on Emerging Technologies. IEEE, pp. 262–266. doi:10.1109/ICET.2009.5353164

[4] Amandeep, Goyal, M., 2015. Review: Moving Object Detection Techniques. Int. J. Comput. Sci. Mob. Comput. 4, 345 – 349.

[5] Athanesious, J. & Suresh, P., 2012. Systematic Survey on Object Tracking Methods in Video. Int. J. Adv. Res. Comput. Eng. Technol. 1, 242–247.

[6] Avidan, S., 2004. Support vector is tracking. IEEE Trans. Pattern Anal. Mach. Intell. 26, 1064–1072. doi:10.1109/TPAMI.2004.53

[7] Badrinarayanan, V., Perez, P., Le Clerc, F. & Oisel, L., 2007. Probabilistic Color and Adaptive Multi-Feature Tracking with Dynamically Switched Priority Between Cues, in: 2007 IEEE 11th International Conference on Computer Vision. IEEE, pp. 1–8. doi:10.1109/ICCV.2007.4408955

[8] Bagherpour, P., Cheraghi, S.A. & Bin Mohd Mokji, M., 2012. Upper body tracking using KLT and Kalman filter. Procedia Comput. Sci. 13, 185–191. doi:10.1016/j.procs.2012.09.127

[9] Balasubramanian, A., Kamate, S., & Yilmazer, N., 2014. Utilization of robust video processing techniques to aid efficient object detection and tracking. Procedia Comput. Sci. 36, 579–586. doi:10.1016/j.procs.2014.09.057

[10] Ben Ayed, A., Ben Halima, M., & Alimi, A.M., 2015. MapReduce-based text detection in big data natural scene videos. Procedia Comput. Sci. 53, 216–223. doi:10.1016/j.procs.2015.07.297

[11] Blackman, S.S., 2004. Multiple hypotheses tracking for multiple target tracking. IEEE Aerosp. Electron. Syst. Mag. 19, 5–18. doi:10.1109/MAES.2004.1263228

[12] Chakravarthy, S., Aved, A., Shirvani, S., Annappa, M., & Blasch, E., 2015. Adapting Stream Processing Framework for Video Analysis. Procedia Comput. Sci. 51, 2648–2657. doi:10.1016/j.procs.2015.05.372

[13] Chandrajit, M., Girisha, R., & Vasudev, T., 2016. Multiple Objects Tracking in Surveillance Video Using Color and Hu Moments. Signal Image Process. An Int. J. 7, 15–27. doi:10.5121/sipij.2016.7302

[14] Chate, M., Amudha, S., & Gohokar, V., 2012. Object Detection and tracking in Video Sequences. ACEEE Int. J. Signal Image Process. 3.

[15] Collins, R.T., Yanxi Liu, & Leordeanu, M., 2005. Online selection of discriminative tracking features. IEEE Trans. Pattern Anal. Mach. Intell. 27, 1631–1643. doi:10.1109/TPAMI.2005.205

[16] Comaniciu, D., Ramesh, V. & Meer, P., 2003. Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. 25, 564–577. doi:10.1109/TPAMI.2003.1195991

[17] Comaniciu, D., Ramesh, V. & Meer, P., 2000. Real-time tracking of non-rigid objects using mean shift, in: IEEE International Proceedings of the Computer Vision and Pattern Recognition. IEEE Computer Society, DC, pp. 142–149.

[18] Coşkun, M. & Ünal, S., 2016. Implementation of Tracking of a Moving Object Based on Camshift Approach with a UAV. Procedia Technol. 22, 556–561. doi:10.1016/j.protcy.2016.01.116

[19] Cucchiara, R., Grana, C., Piccardi, M. & Prati, A., 2003. Detecting moving objects, ghosts, and shadows in video streams. IEEE Trans. Pattern Anal. Mach. Intell. 25, 1337–1342. doi:10.1109/TPAMI.2003.1233909

[20] Du, W. & Piater, J., 2008. A Probabilistic Approach to Integrating Multiple Cues in Visual Tracking, in: Computer Vision – ECCV 2008. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 225–238. doi:10.1007/978-3-540-88688-4_17

[21] Elhariri, E., El-Bendary, N., Hassanien, A.E. & Snasel, V., 2015. An Assistive Object Recognition System for Enhancing Seniors Quality of Life. Procedia Comput. Sci. 65, 691–700. doi:10.1016/j.procs.2015.09.013

[22] Fatima, H., Shah, S.I.A., Jamil, M., Mustafa, F. & Nadir, I., 2013. Object Recognition, Tracking and Trajectory Generation in Real-Time Video Sequence. Int. J. Inf. Electron. Eng. 3, 639–642. doi:10.7763/IJIEE.2013.V3.395

[23] Feris, R.S., 2000. Detection and Tracking of Facial Features in Video Sequences, in: Lecture Notes in Artificial Intelligence. Springer-Verlag Berlin Heidelberg, pp. 197–206.

[24] Foytik, J., Sankaran, P. & Asari, V., 2011. Tracking and recognizing multiple faces using Kalman filter and ModularPCA. Procedia Comput. Sci. 6, 256–261. doi:10.1016/j.procs.2011.08.047

[25] Gang, M., Zhiguo, J., Zhengyi, L., Haopeng, Z. & Danpei, Z., 2010. Full-viewpoint 3D space object recognition based on kernel locality preserving projections. Chinese J. Aeronaut. 23, 563–572. doi:10.1016/S1000-9361(09)60255-7

[26] Guan, N., Tao, D., Luo, Z. & Shawe-Taylor, J., 2012. MahNMF: Manhattan Non-negative Matrix Factorization.

[27] Guo, L., Liao, Y., Luo, D. & Liao, H., 2012. Generic Object Detection Using Improved Gentleboost Classifier. Phys. Procedia 25, 1528–1535. doi:10.1016/j.phpro.2012.03.272

[28] Ha, H. & Ko, K., 2015. A method for image-based shadow interaction with virtual objects. J. Comput. Des. Eng. 2, 26–37. doi:10.1016/j.jcde.2014.11.003

[29] Haritaoglu, I., Harwood, D. & Davis, L.S., 2000. W/sup 4/: real-time surveillance of people and their activities. IEEE Trans. Pattern Anal. Mach. Intell. 22, 809–830. doi:10.1109/34.868683

[30] Harter, W.L., 1974. The Method of Least Squares and Some Alternatives: Part I. Int. Stat. Rev. / Rev. Int. Stat. 42, 147. doi:10.2307/1403077

[31] Houssineau, J., Clark, D.E., Ivekovic, S., Lee, C.S. & Franco, J., 2016. A Unified Approach for Multi-Object Triangulation, Tracking and Camera Calibration. IEEE Trans. Signal Process. 64, 2934–2948. doi:10.1109/TSP.2016.2523454

[32] Hu, W., Tan, T., Wang, L. & Maybank, S., 2004. A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE Trans. Syst. Man Cybern. Part C (Applications Rev. 34, 334–352. doi:10.1109/TSMCC.2004.829274

[33] Javed, O. & Shah, M., 2002. Tracking and Object Classification for Automated Surveillance, in: Proceeding ECCV '02 Proceedings of the 7th European Conference on Computer Vision-Part IV. Springer-Verlag London, UK, pp. 343–357.

[34] Jepson, A.D., Fleet, D.J. & El-Maraghi, T.F., 2003. Robust online appearance models for visual tracking. IEEE Trans. Pattern Anal. Mach. Intell. 25, 1296–1311. doi:10.1109/TPAMI.2003.1233903

[35] Karasulu, B., 2010. Review and Evaluation of Well-Known Methods for Moving Object Detection and Tracking in Videos. J. Aeronaut. Sp. Technol. 4, 11–22.

[36] Kim, Y.M., 2007. Object Tracking in a Video Sequence.

[37] Kothiya, S. V. & Mistree, K.B., 2015. A review on real-time object tracking in video sequences, in: 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO). IEEE, pp. 1–4. doi:10.1109/EESCO.2015.7253705

[38] Krishna, M.T.G., Ravishankar, M. & Babu, D.R.R., 2011. Automatic detection and tracking of moving objects in complex environments for video surveillance applications, in: 2011 3rd International Conference on Electronics Computer Technology. IEEE, pp. 234–238. doi:10.1109/ICECTECH.2011.5941596

[39] Kwon, J. & Le, K.M., 2010. Visual tracking decomposition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Portland, OR, USA, pp. 1269–1276.

[40] Lecumberry, F. & Pardo, A., 2005. Semi-Automatic Object Tracking in Video Sequences, in: XI Congreso Argentino de Ciencias de La Computación.

[41] Lee, B.Y., Liew, L.H., Cheah, W.S. & Wang, Y.C., 2012. Measuring the effects of occlusion on kernel based object tracking using simulated videos. Procedia Eng. 41, 764–770. doi:10.1016/j.proeng.2012.07.241

[42] Li, Y.H., Pang, Y.G., Li, Z.X. & Liu, Y.L., 2010. An intelligent tracking technology based on Kalman and Mean shift algorithm. ICCMS 2010 - 2010 Int. Conf. Comput. Model. Simul. 1, 107–109. doi:10.1109/ICCMS.2010.76

[43] Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vis. 60, 91–110. doi:10.1023/B:VISI.0000029664.99615.94

[44] Mei, X., Hong, Z., Prokhorov, D. & Tao, D., 2015. Robust Multitask Multiview Tracking in Videos. IEEE Trans. Neural Networks Learn. Syst. 26, 2874–2890. doi:10.1109/TNNLS.2015.2399233

[45] Mei, X., Zhou, S.K. & Porikli, F., 2007. Probabilistic Visual Tracking via Robust Template Matching and Incremental Subspace Update, in: Multimedia and Expo, 2007 IEEE International Conference on. IEEE, pp. 1818–1821. doi:10.1109/ICME.2007.4285026

[46] Mohammed, A.D. & Morris, T., 2014. A Robust Visual Object Tracking Approach on a Mobile Device, in: IFIP International Federation for Information Processing. pp. 190–198.

[47] Mohan, A.S. & Resmi, R., 2014. Video image processing for moving object detection and segmentation using background subtraction, in: 2014 First International Conference on Computational Systems and Communications (ICCSC). IEEE, pp. 288–292. doi:10.1109/COMPSC.2014.7032664

[48] Moreno-Garcia, J., Rodriguez-Benitez, L., Fernández-Caballero, A. & López, M.T., 2010. Video sequence motion tracking by fuzzification techniques. Appl. Soft Comput. 10, 318–331. doi:10.1016/j.asoc.2009.08.002

[49] Nagendran, A., Dheivasenathipathy, N., V. Nair, R. & Sharma, V., 2014. Recognition and Tracking Moving Objects Using Moving Camera in Complex Scenes. Int. J. Comput. Sci. Eng. Appl. 4, 31–36. doi:10.5121/ijcsea.2014.4203

[50] Nair, B.M., Foytik, J., Tompkins, R., Diskin, Y., Aspiras, T. & Asari, V., 2011. Multi-pose faces recognition and tracking system. Procedia Comput. Sci. 6, 381–386. doi:10.1016/j.procs.2011.08.070

[51] Najva, N. & Bijoy, K.E., 2016. SIFT and Tensor-Based Object Detection and Classification in Videos Using Deep Neural Networks. Procedia Comput. Sci. 93, 351–358. doi:10.1016/j.procs.2016.07.220

[52] Nesterov, Y., 2013. Gradient methods for minimizing composite functions. Math. Program. 140, 125–161. doi:10.1007/s10107-012-0629-5

[53] Oiwa, D., Fukui, S., Iwahori, Y., Nakamura, T. & Bhuyan, M.K., 2016. Tracking with Probabilistic Background Model by Density Forests, in: Proc. of IEEE/ACIS. IEEE, Okayama, Japan, pp. 391–396.

[54] Ojha, S. & Sakhare, S., 2015. Image processing techniques for object tracking in video surveillance- A survey, in: 2015 International Conference on Pervasive Computing (ICPC). IEEE, pp. 1–6. doi:10.1109/PERVASIVE.2015.7087180

[55] Poschmann, P., Huber, P., Rätsch, M., Kittler, J. & Böhme, H.-J., 2014. Fusion of Tracking Techniques to Enhance Adaptive Real-time Tracking of Arbitrary Objects. Procedia Comput. Sci. 39, 162–165. doi:http://dx.doi.org/10.1016/j.procs.2014.11.025

[56] Ramya, P. & Rajeswari, R., 2016. A Modified Frame Difference Method Using Correlation Coefficient for Background Subtraction. Procedia Comput. Sci. 93, 478–485. doi:10.1016/j.procs.2016.07.236

[57] Risha, K.P. & Kumar, A.C., 2016. Novel Method of Detecting Moving Object in Video. Procedia Technol. 24, 1055–1060. doi:10.1016/j.protcy.2016.05.235

[58] Sarkar, R., Bakshi, S. & Sa, P.K., 2012. A Real-time Model for Multiple Human Faces Tracking from Low-resolution Surveillance Videos. Procedia Technol. 6, 1004–1010. doi:http://dx.doi.org/10.1016/j.protcy.2012.10.122

[59] Shen, H., Li, S., Zhu, C., Chang, H. & Zhang, J., 2013. Moving object detection in aerial video based on spatiotemporal saliency. Chinese J. Aeronaut. 26, 1211–1217. doi:10.1016/j.cja.2013.07.038

[60] Soundrapandiyan, R. & Mouli, P.V.S.S.R.C., 2015. Adaptive Pedestrian Detection in Infrared Images Using Background Subtraction and Local Thresholding. Procedia Comput. Sci. 58, 706–713. doi:10.1016/j.procs.2015.08.091

[61] Susar, R. & Dongare, M., 2015. Moving Object Detection, a Succinct Review. Int. J. Adv. Res. Comput. Commun. Eng. 4, 334–336. doi:10.17148/IJARCCE.2015.41277

[62] Tang, D. & Zhang, Y.J., 2011. Combining mean-shift and particle filter for object tracking, in: Proceedings - 6th International Conference on Image and Graphics, ICIG 2011. IEEE, pp. 771–776. doi:10.1109/ICIG.2011.118

[63] Tao Zhang, Zaiwen Liu, Xiaofeng Lian & Xiaoyi Wang, 2010. Study on moving-objects detection technique in video surveillance system, in: 2010 Chinese Control and Decision Conference. IEEE, pp. 2375–2380. doi:10.1109/CCDC.2010.5498797

[64] Viswanath, A., Kumari, R. & Senthamilarasu, V., 2015. Background Modelling from a Moving Camera. Procedia - Procedia Comput. Sci. 58, 289–296. doi:10.1016/j.procs.2015.08.023

[65] Wei, L. & Qin, W.J.L., 2007. Study on Moving Object Tracking Algorithm in Video Images, in: IEEE Conference on Electronic Measurement and Instruments, ICEMI 8th International Conference. IEEE, pp. 810–816.

[66] Weifeng Liu & Dacheng Tao, 2013. Multiview Hessian Regularization for Image Annotation. IEEE Trans. Image Process. 22, 2676–2687. doi:10.1109/TIP.2013.2255302

[67] Weng, E.N.G., Khan, R.U., Adruce, S.A.Z., & Bee, O.Y., 2013. Objects Tracking from Natural Features in Mobile Augmented Reality. Procedia - Soc. Behav. Sci. 97, 753–760. doi:10.1016/j.sbspro.2013.10.297

[68] Xue Mei & Haibin Ling, 2011. Robust Visual Tracking and Vehicle Classification via Sparse Representation. IEEE Trans. Pattern Anal. Mach. Intell. 33, 2259–2272. doi:10.1109/TPAMI.2011.66

[69] Yilmaz, A., Javed, O., & Shah, M., 2006. Object tracking: A survey. ACM Comput. Surv. 38, 13–es. doi:10.1145/1177352.1177355

[70] Zhang, H. & Jiang, Z., 2014. Multi-view space object recognition and pose estimation based on kernel regression. Chinese J. Aeronaut. 27, 1233–1241. doi:10.1016/j.cja.2014.03.021

[71] Zhang, J., Cao, J. & Mao, B., 2016. Moving Object Detection Based on Non-parametric Methods and Frame Difference for Traceability Video Analysis. Procedia Comput. Sci. 91, 995–1000. doi:10.1016/j.procs.2016.07.132

[72] Zhang, T., Ghanem, B. & Liu, S., Ahuja, N., 2013. Robust Visual Tracking via Structured Multi-Task Sparse Learning. Int. J. Comput. Vis. 101, 367–383. doi:10.1007/s11263-012-0582-z

[73] Zhao, J., Qiao, W. & Men, G.Z., 2009. An approach based on mean shift and Kalman filter for target tracking under occlusion. Proc. 2009 Int. Conf. Mach. Learn. Cybern. 4, 2058–2062. doi:10.1109/ICMLC.2009.5212129

[74] Zhiqiang, W., Xiaopeng, J. & Peng, W., 2006. Real-time is moving object detection for video monitoring systems. J. Syst. Eng. Electron. 17: 731–736. doi:10.1016/S1004-4132(07)60007-3.