Robust Vision-based Multiple Moving Object Detection and Tracking from Video Sequences

Othman Omran Khalifa, Norun Abdul Malek*, Kazi Istiaque Ahmed, Farah Abdul Rahman
Department of Electrical and Computer Engineering, Kulliyyah of Engineering International Islamic University Malaysia

Article Info	ABSTRACT
	Detection of Moving Objects and Tracking is one of the most concerned issue and is being vastly used at home, business and modern applications. It is used to identify and track of an entity in a significant way. This paper illustrates the way to detect multiple objects using background subtraction methods and extract each object features by using Speed-Up Robust Feature algorithm and track the features through k-Nearest Neighbor processing from different surveillance videos sequentially. In the detection of object of each frame, pixel difference is calculated with respect to the reference background frame for the detection of an object which is only suitable for any ideal static condition with the consideration of lights from the environment. Thus, this method will detect the complete object and the extracted feature will be carried out for the tracking of the object in the multiple videos by one by one video. It is expected that this proposed method can commendably abolish the impact of the changing of lights.

Corresponding Author:

Norun Abdul Malek,
Department of Electrical and Computer Engineering,
Kulliyyah of Engineering
International Islamic University Malaysia
Email: norun@iium.edu.my

1. INTRODUCTION

In computer vision, the automatic visual detection of a thing, as well as tracking, is one most thought-provoking disputes at homes, businesses, and industries. The number of visual detection and tracking applications are voluminous while those extensive varieties of the applications can be found as well inreconnaissance system, vehicle tracking and aerospace applications are the names of a few. By using conventional motion estimation methods and pattern recognition, the resolutions for tracking and detection of abstract things specifically vehicles, in general, is a delinquent. In one hand, detection of moving things from the background image to the continuous video frames are treated as recognition of the moving targets and on the other hand, finding various locations of the moving things in a video is treated as tracking of the moving targets. To detect and track down those moving objects, it requires process to perform such kind of task. Subtraction of two consecutive frames, subtraction of background from frames and optical flow are the main of the well-known methods for the moving object detection [1]. In the Optical Flow moving object detection procedure [2], the flow field image is calculated and the distribution of the feature is done by cluster processing which is better. But this procedure is not suitable for real-time processing because of its large amount of calculation and its sensitivity to noise and lack of anti-noise performance [3]. On the other hand, in background subtraction procedure [3], the moving object is detected by subtracting the background from the current frame, is a simple procedure and in the case of already known background, this process could able to provide a complete information about the objects. For the feature extraction, there are many available methods such asSpeed-Up Robust Features (SURF) which is a speed-up Version Scale Invariant Feature Transforms (SIFT). Many authors ,[1], [4]–[6] utilize certain familiarity to provide the video advancement by

defining the relationship between objects with spatial, temporal, and co-occurrence. On the other hand, many authors highlight [3], [4], [7] the dispute of finding multiple objects from the continuous sequence of the frame as related frames. In [4] shows the tracking of the object and proposed a method to localize the object in the image and they showed that it works better with SURF feature extraction and improved Camshift algorithm by automatically adjusting the illumination to find out a lost target to track down of the features of the object. Moreover, the authors [8], shows the tracking strategy using the same Camshift procedure for the PTZ Camera. Nevertheless, this research work did not provide satisfactory output performance for the image without strong image texture and their features as shown in Figure 1.
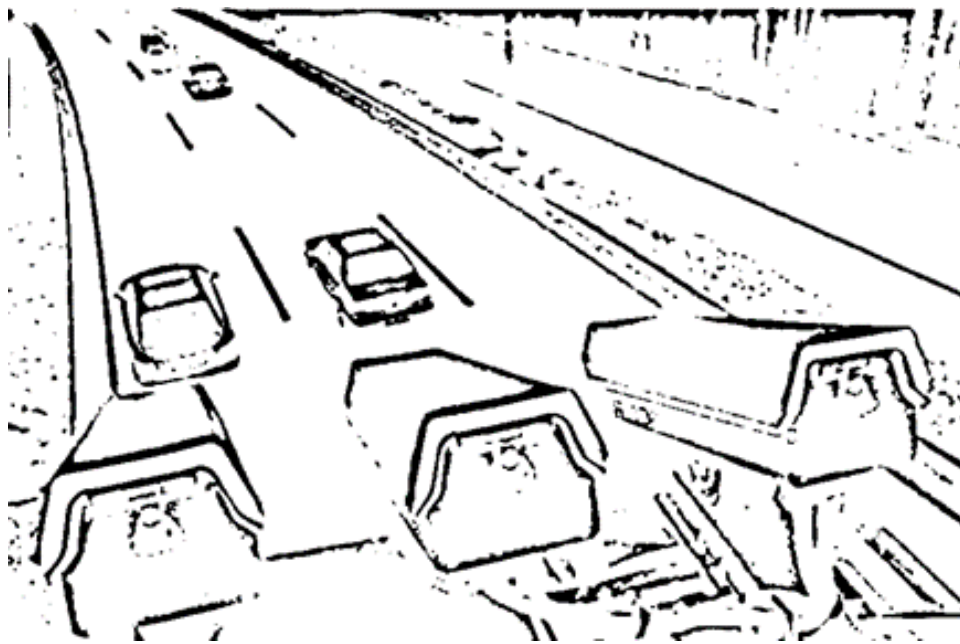

Figure 1. General Representation of the System

However [5] showed successfully for the same sort of image objects which also taken into consideration of multiple situations. It shows that it will work perfectly if the image object scaled down or up and change in rotation in multiple object detection research.However, it did not provide any clear information for the 2 or more distinct images where for the research it required different sort of object to be identified and it is also not clear on the video segmentation part. Similarly by using Improved KLT Tracking Classifier above SURF, [6] simulated tracking the detected object by extracting the features using SURF and track down the feature using KLT for multiple objects. In [2], the Demonstrated Kalman channel is utilized to scale the condition of an objective question. An optical stream can gauge the speeds of a protest and using deformable fragment method [9] confirmed an approach to perceive and track multiple object in CCTV by in view of spatial constrains among the objects through the benefits of HOG feature extraction. Conversely, [10] exemplifies adaptive fusion centered manifold features while comparison define that the fusion centered approach is more stout than the artifact rule and prejudiced sum rule. In [1], the authors analyzed their work and they utilizes a static camera for video and the main edge of frame considered as a foundation outline. At that point, the foundation was reduced from the current edge of the frame. However, those paper could not clearly present examples for reader's comprehension. Table 1 listed a summary of literature reviews that are being used in this research.

Table 1. Related Literatures Review Summary

Research Title	Approach	Strengths
Multiple Object Tracking using Kalman Filter and Optical Flow [2]	Simulation	Kalman channel is utilized to scale the condition of an object location. An optical flow can identify the speeds of an object.
Moving Object Detection and Tracking for Video Surveillance [1]	Simulation	It utilizes a static camera for video and the main edge of frame considered as a foundation outline. At that point, the foundation was reduced from the current edge of the frame.
Object tracking using improved Camshift with SURF method [7]	Analytical and Simulation	SURF feature extraction and improved Camshift algorithm by automatically adjusting the illumination to find out lost target to track down of the features of the object
Traffic Sign Recognition Using SURF : Speeded Up Robust Feature Descriptor and Artificial Neural Network Classifier [11]	Simulation	High rate accuracy to recognize, color-based segmentation along with the morphological and geometrical properties
Multiple object detection for smart TV shopping video using point to point feature based SURF method [5]	Simulation	Successfully simulated for the same sort of image objects which also taken into consideration of multiple situations i.e. if the image object scaled down or up and change in the rotation in multiple object detection it works fine
Multiple Object Tracking by Improved KLT Tracker Over SURF Features [6]	Analytical and Simulation	front facing feature extraction and recognition for multiple objects

2.   PROPOSED METHOD

The main objective of this system is to provide robustness and accuracy for the detection of the multiple moving objects through developing an algorithm and track the detected object features in the video frames.
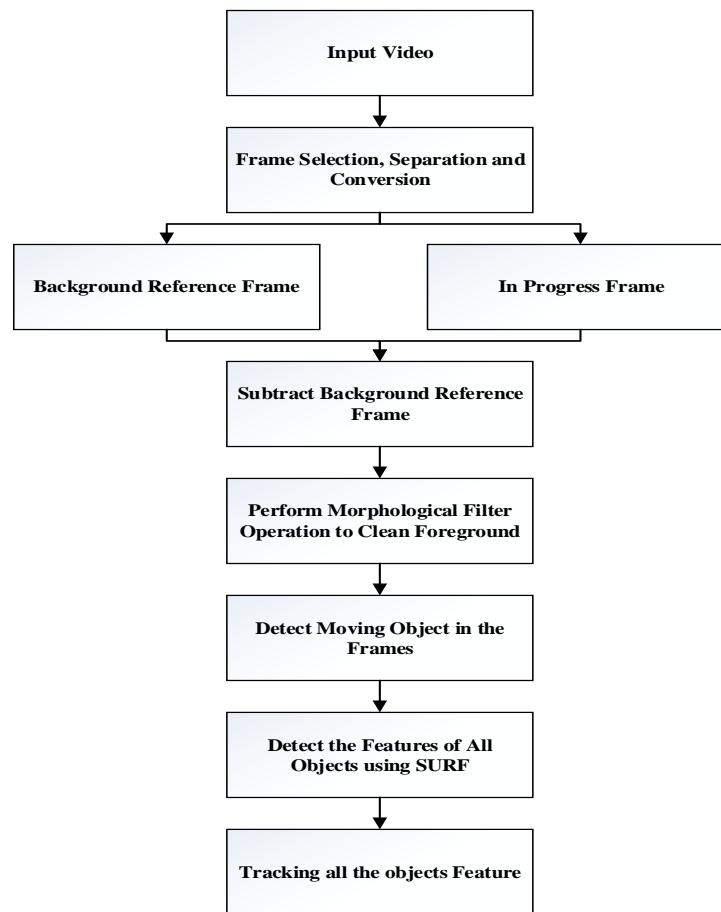

Figure 2. Overview of the Processing Steps

i. By using Static cameras it is required to take input video for processing and convert frames to images where the first 25frames will be treated as the Background.
ii. After the last Background training frame, next frame is treated as the in-progress frame and apply Background subtraction method through subtracting background reference frame.
iii. It may contain noise and it must require reducing noise. To reduce noise and to receive clear foreground objects, filter using morphological filters and moving objects are detected.
iv. From the detected objects, extract features each individual by using SURF.
v. Track the detected features in the video by the k-NN algorithm and in each frame tracking step the object features will be set as the old feature.

3. RESEARCH METHOD

A model of appearance, model of location]and a strategy for searching are the three major components in any tracking system. For the proposed multiple object detection and tracking we generalize the appearance model using for the detection using background subtraction method which is followed by the extraction of features by SURF and continuous tracking by the KNN.

3.1. Speed-Up Robust Features (SURF)

Scale-Invariant Feature Transform (SIFT) is an effective way to deal with highlight identification presented by [12]. The SURF-calculation depends on similar standards and steps, however, it uses an alternate plan and it ought to give better outcomes, quicker. With a specific end goal to recognize include focuses in a scale-invariant way, SIFT utilizes a falling separating approach whereasthe Difference of Gaussians, DoG, is ascertained on continuously downscaled pictures [13].

3.2. Keypoint Detection Using SURF

Generally, the method to accomplish scale invariance is to look at the picture at various scales, scale space, utilizing Gaussian pieces. Both SIFT and SURF partitions the scale space into levels and octaves. An octave compares to a multiplying of σ, and the octave is partitioned into consistently dispersed levels as shown in Figure 3.
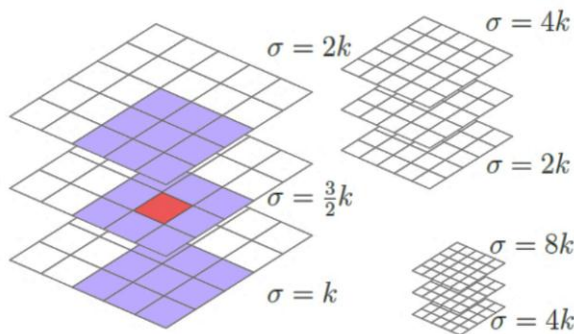

Figure 3: octaves with 3 levels, the area for the 3×3×3 non-most extreme concealment used to identify elements is highlighted from [13].
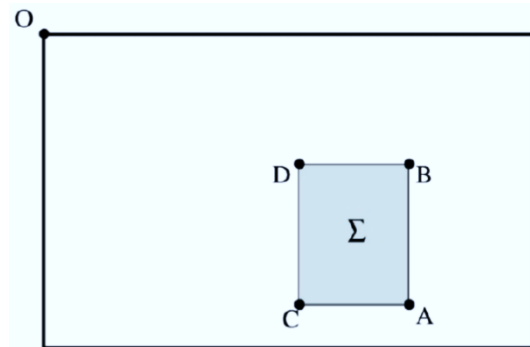
Figure 4: Integral images for Area computation from [7]

Both methodologies assemble a pyramid of reaction maps, with various levels inside octaves. A reaction guide is the consequence of an operation on the picture. The intrigue focuses are the focuses that are outrageous among 8 neighbors in the present level and its 2×9 neighbors in the level beneath or more. This is a non-greatest concealment in a 3×3×3 neighborhood, the connection between levels, octaves, and neighborhood is outlined in Figure 4 on top of this segment [13].

SURF is described by the utilization of essential pictures. It is described, the counts of the zone of an upright rectangular district are lessened to four operations, and the computation of first-request Haar wavelet reaction will be six operations. The central image of the image I(x, y) (0≤x≤M, 0≤y≤N) can be well-defined by the formula:

$$I(x) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(x,y) \tag{1}$$

In [14] displays how to achieve reckless pixel intensities, which can be considered by:

$$\sum(ABCD) = [\sum I(A) + \sum I(B)] - [\sum I(C) + \sum I(D)] \tag{2}$$

Gaussian pyramid, i.e., the picture scale space is principally used to discover intrigue focuses in various scales. Here, Gaussian parts can be changed in size to make the Gaussian pyramid. As taking after figure appears, Laplacian of Gaussian is approximated to the crate channel.
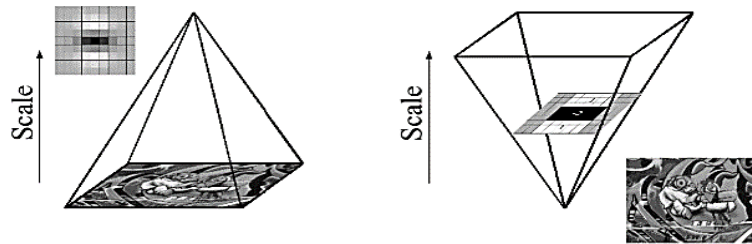

Figure 5. Intrigue Focus of Laplacian of Gaussian From [13]

Utilizing this strategy, different layers of the scale-space pyramid can be handled all the while and it invalidates the need to subsample the picture, accordingly having better execution. To figure out if a point is most extreme, the determinant of Hessian is utilized at the intrigue purposes of restriction. Assume f(x, y) is a persistent capacity with two factors, then the Hessian framework is:

$$\mathcal{H}(f(x,y)) = \begin{bmatrix} \frac{\delta^2 f}{\delta x^2} & \frac{\delta^2 f}{\delta x \delta y} \\ \frac{\delta^2 f}{\delta x \delta y} & \frac{\delta^2 f}{\delta y^2} \end{bmatrix} \tag{3}$$

and the determinant:

$$\begin{aligned} \mathcal{H}(f(x,y)) &= \det \mathcal{H} \\ &= \left[\frac{\delta^2 f}{\delta x^2} * \frac{\delta^2 f}{\delta y^2}\right] - \left[\frac{\delta^2 f}{\delta x \delta y} * \frac{\delta^2 f}{\delta x \delta y}\right] \end{aligned} \tag{4}$$

If det $\mathcal{H}$ <0, which means the Eigen values of $\mathcal{H}$ have different signs, and then the point is not a confined maximum. Otherwise it is a maximum and from [10], Replacing f(x, y) with I(x, y), the Hessian matrix of the image is:

$$\mathcal{H}(f(x,y)) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \tag{5}$$

and

$$I(x) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(x,y)$$

While the value of $L_{xx}(x,\sigma) = I(x) * \frac{\delta^2}{\delta x^2} g(\sigma)$ and $L_{xy}(x,\sigma) = I(x) * \frac{\delta^2}{\delta xy} g(\sigma)$

3.3. Interest point Detection Using SURF

SURF intrigue point descriptor ascertains the Haar reactions in both X and Y organizes in the circle locale focused at intrigue focuses with a sweep of 6σ. It depends on the predominant introductions of all the intrigue focuses. The span of Haar wavelet is 4σ, and the total of vectors is computed in each 60 degrees in the circle. At long last, the introduction with the biggest total of vectors is the overwhelming introduction. The procedure appeared in the accompanying figure.
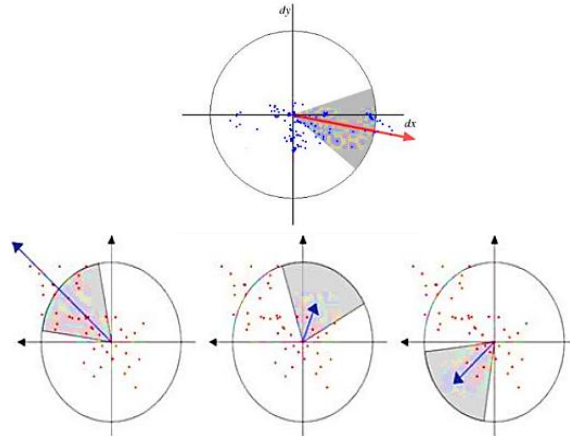
Figure 6. Assignment for Orientation From [7]

After the assurance of overwhelming introduction, [13] describes a square window is built which is focused at each intrigue point with a side length of 20σ. At that point, it is partitioned into 4×4 sub-district and the wavelet reaction is figured in both the overwhelming introduction and the introduction vertical to it. On the off chance that we characterize the wavelet of x and y as dx and dy, then there will be 4 values Σdx, Σdy, Σ|dx|, Σ|dy|, and absolutely it will be a 64-length vector for each intrigue point. In this manner, it is possible to acquire the descriptor segment by normalizing it.

3.4. k-Nearest Neighbor (k-NN) Classifier

K-Nearest Neighbor (KNN starting now and into the foreseeable future) is one of those calculations that are exceptionally easy to see however works unimaginably well practically speaking. Additionally, it is shockingly adaptable and its applications run from vision to proteins to computational geometry to charts et cetera. KNN is a non-parametric sluggish learning calculation. Authors in [15] explained when the method is non-parametric, it implies that it doesn't make any suppositions on the hidden information appropriation. This is quite valuable, as in this present reality, the vast majority of the down to earth information does not regard the ordinary hypothetical suspicions made (e.g. Gaussian blends, directly distinct and so on).

3.5. Assumptions in the KNN Classifier

KNN accept that the information is in an element space. All the more precisely, the information focuses are in a metric space. The information can be scalars or potentially even multidimensional vectors[16]Since the focuses are in highlight space, they have a thought of separation – This need not really be a Euclidean separation in spite of the fact that it is the one regularly utilized.

3.6. KNN for Density Estimation

In spite of the fact that order remains the essential utilization of KNN, we can utilize it to do thickness estimation too. Since KNN is non-parametric, it can do estimation for discretionary disseminations. The thought is fundamentally the same as utilization of Parzen window. Rather than utilizing hypercube and portion capacities, for evaluating the thickness at a point x, put a hypercube focused at x and continue expanding its size till k neighbors are caught. Presently appraise the thickness utilizing the equation,

$$p(x) = \frac{k}{nV} \qquad (6) \qquad \text{and} \qquad I(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(x, y)$$

Where n is the aggregate number of V is the volume of the hypercube. See that the numerator is basically consistent and the thickness is affected by the volume[13].

4. RESULTS AND ANALYSIS

In the following outputs, the results of the simulation for the moving object detection has been shown where we particularly used a still camera to record video frames. In the following figures, it is shown Background Reference Frame (Figure 7). From the in-progress frame, the background image subtracted to detect foreground multiple objects (Figure 8) which indicate the difference between the in-progress original frame and the reference background frame. The next image (Figure 9) indicates the morphologically filtered

frame of that respective foreground frame for a clear foreground and which is followed by the detection of objects. The detected objects are segmented and gather the features for strongest points using SURF. By continuously tracking the detected features in the video by the k-NN algorithm, in each step, the new object features will be set as the old feature in Figure 10-15.
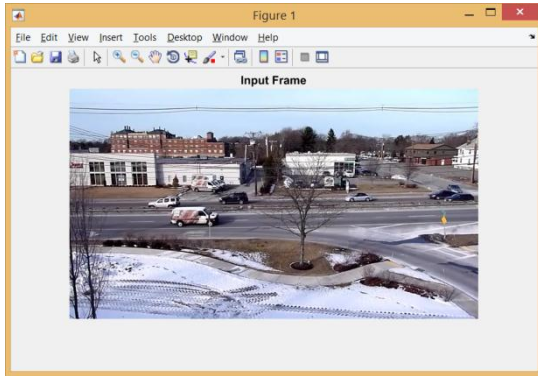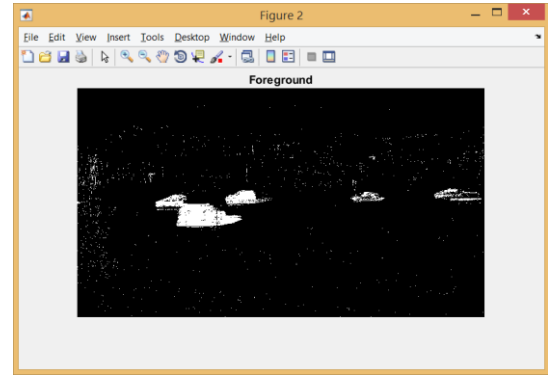

Figure 1. Reference Background [13]

Figure 2. Foreground Frame after Subtraction from Figure 7 (May Contain Noise)
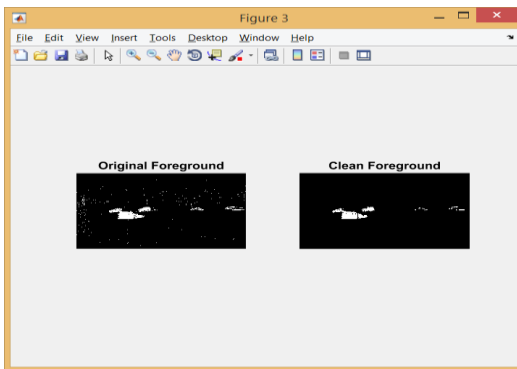

Figure 3. Side by Side Unclear and Clear image (after applying Morphological Filter)
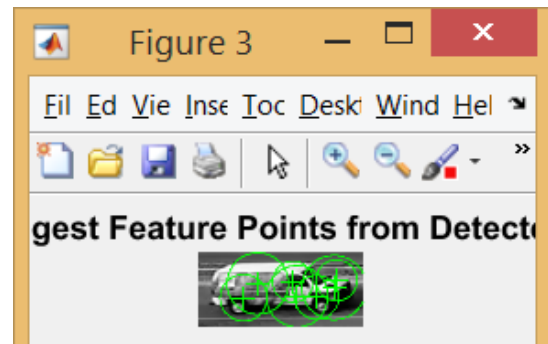

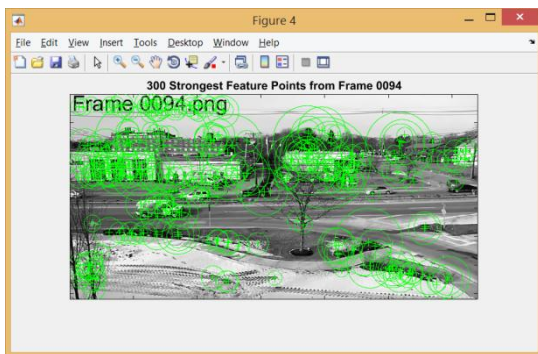Figure 4. A Moving Object with Strongest Feature Points of the Object (using SURF)

Figure 5. Strongest Feature Points from Frame 94 (using SURF).
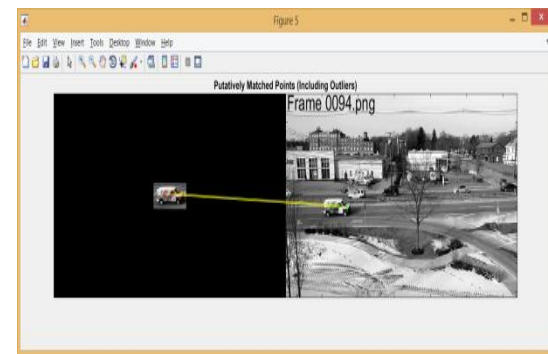

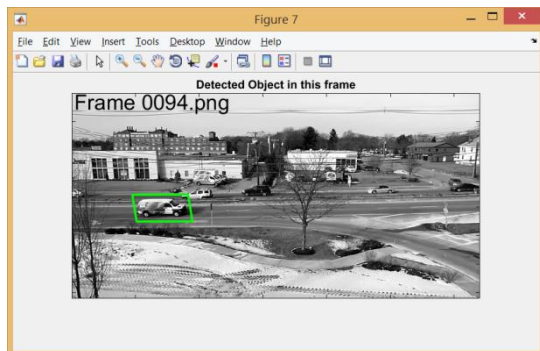Figure 6. Detected Object Feature Matching in Frame 94 (using SURF)
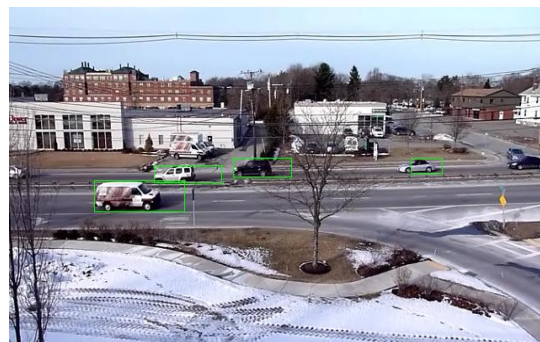
Figure 7. Detected Object on Frame 94

Figure 8. All Detected Objects in the Frame 94

Figure 95: After Tracking until Frame Number 134

5.   CONCLUSION

With the multiple occurrences, the maximum multiple object detection works for the same types of objects. However, for the multiple object detection and tracking in multiple video files, it is definitely required to detect different types of object in the same or the different video files. The objective of this work to detect multiple object with same or different types in the same or different video files one by one using point using feature point to feature point matching. One of the extreme plus point for using this proposed tactic that it can detect the objects notwithstanding a scale alteration or in-plane variation.

REFERENCES

[1]   M. J. J. Jadhav, "Moving Object Detection and Tracking for Video Survelliance," *Int. J. Eng. Res. Gen. Sci.*, vol. 2, no. 4, pp. 372–378, 2014.
[2]   S. Shantaiya, K. Verma, and K. Mehta, "Multiple Object Tracking using Kalman Filter and Optical Flow," *Eur. J. Adv. Eng. Technol.*, vol. 2, no. 2, pp. 34–39, 2015.
[3]   J. Zhang, S. Xu, K. Huang, and T. Luo, "Accurate Moving Target Detection Based on Background Subtraction and SUSAN," *Int. J. Comput. Electr. Eng.*, vol. 4, no. 4, pp. 436–439, 2012.
[4]   S. Joshi, S. Gujarathi, and A. Mirge, "Moving Object Tracking Method using Improved Camshift with Surf Algorithm, " *Int. J. Adv. Sci. Eng. Technol.*, vol. 2, no. 2, pp. 14–18, 2014.
[5]   D. Chhabra and A. Verma, *"Multiple object detection for smart TV shopping video using point to point feature based SURF method,"* 2016 Int. Conf. Inven. Comput. Technol. (ICICT), Coimbatore, pp. 1–6, 2016.

[6]   V. Buddubariki, "Multiple Object Tracking by Improved KLT Tracker Over SURF Features," 2015.
[7]   J. Li, J. Zhang, Z. Zhou, W. Guo, B. Wang, and Q. Zhao, *"Object tracking using improved Camshift with SURF method,"* International Workshop on Open-Source Software for Scientific Computation (OSSC), 2011, pp. 136-141.
[8]   W. Song, "Real-time Moving Object Tracking System Using Cam-shift Algorithm," *TELKOMNIKA Indones. J. Electr. Eng*, vol. 11, no. 11, pp. 6611–6617, 2013.
[9]   F. Teng and Q. Liu, "Robust Multiple Ship Tracking in Inland Waterway CCTV System," *TELKOMNIKA Indones. J. Electr. Eng.*, vol. 12, no. 11, pp. 7772–7777, 2014.
[10] J. Cao, L. Guo, J. Wang, and D. Wu, "Object Tracking Method Based on a New Multi-Feature Fusion Strategy," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 12, no. 9, pp. 6811–6818, 2014.
[11] M. Z. Abedin, P. Dhar and K. Deb, *"Traffic Sign Recognition Using SURF : Speeded Up Robust Feature Descriptor and Artificial Neural Network Classifier,"* International Conference on Electrical and Computer Engineering (ICECE), pp. 198–201, 2016.
[12] D. G. Lowe, "Distinctive image features from scale invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–11020042, 2004.
[13] J. T. Pedersen, "Study group SURF : Feature detection & description," Aarhus Universitet: Aarhus, Denmark, pp. 1–12, 2011.
[14] S. Emarose and B. Pinnamaneni, "Face Recognition and Person Localization using SURF for Automated Attendance System," *International Journal of Computer Applications*, No. 8, October, 2014.
[15] R. Muralidharan, "Object Recognition Using K-Nearest Neighbor Supported By Eigen Value Generated From the Features of an Image," *International Journal of Innovative Research in Computer and Communication Engineering,* Vol. 2 (8)*, pp. 5521–5528, 2014.
[16] R. Muralidharan and C. Chandrasekar, "Object recognition using SVM-KNN based on geometric moment invariant," *Int. J. Comput. Trends Technol.*, vol. 1, no. 3, pp. 215–220, 2011.

BIOGRAPHIES OF AUTHORS

Othman Omran Khalifa received his Bachelor's degree in Electronic Engineering from the Garyounis University, Libya in 1986. He obtained his Master degree in Electronics Science Engineering and PhD in Digital Image Processing from Newcastle University, UK in 1996 and 2000 respectively. He worked in industrial for eight years and he is currently a Professor at Electrical and Computer Engineering Department, International Islamic University Malaysia. His area of research interest is Communication Systems, Information theory and Coding, Digital image/video processing, coding and Compression, Wavelets, Fractal and Pattern Recognition. He published more than 450 papers in international journals and Conferences. He is SIEEE member, IEEE computer, Image processing and Communication Society member.

Norun Abdul Malek obtained her PhD degree from School of Electronic, Electrical and Systems Engineering, Loughborough University, UK in 2013. She has been appointed as an Assistant Professor in Electrical and Computer Engineering Department, Faculty of Engineering, International Islamic University Malaysia (IIUM). Her research interest includes to antenna and propagation, signal processing particularly of antenna arrays, algorithms and wireless communication systems. She is an active member of the IEEE, a registered member of the Board of Engineers Malaysia (BEM) and Institute of Engineers Malaysia (IEM).

Kazi Istiaque Ahmed pursuing his Master degree from Electrical and Computer Engineering, Kulliyah of Engineering, International Islamic University Malaysia, Malaysia. He obtained his Bachelor of Science Engineering in Computer Science and Engineering from Bangladesh University of Business and Technology, Bangladesh. He worked for the industry more than three years. His research interest area in image processing, Biomedical Image processing, Data Mining, Networking, Information theory and Coding, Digital image / video processing and Pattern Recognition.

Farah Diyana Abdul Rahman obtained her PhD degree from Department of Electrical and Electronic Engineering, University of Bristol, UK in 2015. She has been appointed as an Assistant Professor in Electrical and Computer Engineering Department, Faculty of Engineering, International Islamic University Malaysia (IIUM). Her research interest includes image and video processing, video quality evaluation, multimedia transmission and wireless communication systems. She is an active member of the IEEE, a registered member of the Board of Engineers Malaysia (BEM) and Institute of Engineers Malaysia (IEM).