USING SIMULATED VIDEOS TO ASSESS THE EFFECTS OF OCCLUSION ON KERNEL- DEPENDENT OBJECT TRACKING

SUBODH KUMAR MOHANTY, *Gandhi Institute of Excellent Technocrats, Bhubaneswar, India*
SUBHASRI RAY, *Jagannath Institute of Engineering and Technology, Cuttack, Odisha, India*

Abstract
Occlusion handling is one among the foremost studied problems for object tracking in computer vision. Many previous works claimed that occlusion may be handled effectively using Kalman filter, Particle filter and Mean Shift tracking methods. However, these methods were only tested on specific task videos. so as to explore the particular potential of those methods, this paper examined the tracking methods with six simulation videos that consider various occlusion scenarios. Tracking performances are evaluated supported Sequence Frame Detection Accuracy (SFDA). The results show that Mean shift tracker would fail completely when full occlusion occurs as claimed by many previous works. In most cases, Kalman filter and Particle filter tracker achieved SFDA score between 0.3 and 0.4. It demonstrates that Particle filter tracker fails to detect object with arbitrary movement in one among the experiments. The effect of occlusion on each tracker is analysed with Frame Detection Accuracy (FDA) graph
Keywords: Computer Vision, Object Tracking, Occlusion Handling, Kalman filter, Mean shift tracking methods

I. INTRODUCTION

Occlusion handing is a major challenge for object tracking in computer vision. Occlusion occurred when an object of interest is temporary disappeared from camera views during tracking while the object has not exit the region of interest (ROI). Region of interest in video surveillance is the viewable area in a video frame that is concerned with the user interest.
Occlusions happen due to three reasons. Firstly, an object can be occluded when it is blocked by the background structure such as the building pillar or furniture in a room [9]. Second, occlusion can happen when other moving foreground objects overlap the tracked object [2, 7]. Lastly, occlusion happens when track features are blocked from camera view when the tracked objects turn away from camera [7]. This is a common problem when tracking human using face recognition and face is occluded when a person turns his face away from the camera. This is known as self occlusion.

Many tracking methods have been proposed in handling occlusion using selected video samples. The selected video samples are usually obtained from actual recording by the authors or from benchmark dataset such as PETS [14] and ETISEO [5]. These video dataset provide a good impression on the performance of the proposed tracking method in real world. However, the complex scenario in the video such as shadow, illumination changes and moving background could obscure the evaluation of the actual performance of the tracking methods.

Therefore, we propose to run the experiments on simulation videos sequences. According to Taylor et al. [16], simulated video data is ideal to provide a good indication of which algorithms work well in a given scenario. In addition, simulated video can provides accurate ground truth for performance evaluation. In simulation videos, an ideal environment without any noise and distraction can be created. The results contained under such environment could reflect the actual performance of the tested tracking methods. It will also be easier to analyse the effect of occlusion in simulation videos since the environment and interaction between objects can be controlled.

The remainder of the paper is organized as follows. Section 2 gives a review of previous related works. Observation from the previous works suggested three tracking methods that are used in this paper. In section 3, the simulation videos used in the experiment are described. Tracking measurement and experimental results are reported in section 4 and 5 accordingly. The future work is discussed in section 6. Section 7 concludes the paper.

II. PREVIOUS WORKS

According to Yilmaz et al. [16], tracking methods can be classified into three categories based on the features used in tracking, namely the point tracking, silhouette tracking and kernel based tracking. For point tracking, computational cost is minimal but with scarification of accuracy. For silhouette tracking, the accuracy is high and it can handle transformable tracking object, but the computational cost is much higher. The kernel based tracking is widely used because it could provide high accuracy and the computational cost is lower than silhouette based tracking.

Many methods have been proposed in kernel based tracking. Comaniciu et al. [4] has reviewed comprehensively on kernel based tracking. Among the popular kernel based tracking methods, most of the current works focused on Kalman Filter, Particle Filter and Mean Shift tracking. Comaniciu et al. [5] and Yilmaz [22] has developed tracker based on Mean Shift to derive target object candidate based on appearance model similarity. The results from their research show that the Mean Shift tracker is robust to partial occlusion, background clutter, target scale variations and rotations in depth. However, Comaniciu et al. [5] and Yilmaz [22] did not discussed on how the Mean Shift tracker will performed in the situation of full occlusion and object with arbitrary trajectory. Hence, these three tracking methods are tested in the experiments conducted in this paper.

Mirabi & Javadi [15] and Wang et al. [20] experimented Kalman Filter tracker on real-world video sequences. Both experiments their system can deal with difficult situations such as noise, shadow and illumination changes. Kalman Filter tracker is also claimed as computationally cost effective in tracking object. These experiments do not consider tracking moving object with arbitrary trajectory.

Particle filters provide robust tracking of moving objects in a cluttered environment especially in tracking moving object that move in non-linear and non-Gaussian trajectory [3]. Chuo et al. [3] has used Particle Filter tracker with images' grey level model to track the moving object while Liang et al. [14] combined Particle Filter tracker with colour and shape model to track objects in video sequences.

Based on the reviews of these previous works, many of them [3, 5, 14, 15, 20, 22] only presented their results as images of video sequences that show the incidences of successful tracking. The accuracy of the tracking experiments was not discussed nor compare statistically with other tracking methods. Hence, in this paper, we will conduct a series of experiment that measure the accuracy of each tracker using a set of simulated video sequences so that the performance of each tracker can be analysed systematically.

Besides, authors that used own recorded video for testing usually did not extensively discuss the limitation of their proposed method. It is believed that their proposed methods are application based and the video samples used in their experiment are too complex to be analysed. Recorded video are normally affected by background noise, shadow and illumination changes. Therefore, to reduce complexity in video samples, we use simulated video sequences to control the video scene. With controlled scene setup, the results obtained from the experiments can reflect the actual ability of trackers in handling occlusion.

III. VIDEO DATA

A set of six simulation video sequences with moving object are generated. The label of the videos and their description are shown is Table 1. The simulation was created using OpenGL and Visual C++.

The controlled video sequence of video label A1 is shown in Figure 1, which simulates a single colour ball moves from left, rolls over to the right and exits the video frame. The ball in the video is simulated to move in constant speed and direction.

Figure 2 shows another simulated video sequence with a single colour ball moving (from left to right direction) towards the middle of the video frame and then moving backwards to the left and exit the frame. This video sequence is designed to test how tracking methods react to arbitrary movement which is labelled as A2.

Table 1: Video label and description

Label	Description
A1	Moving ball with constant speed and direction
A2	Moving ball with constant speed and arbitrary direction
A3	Moving ball with constant speed and direction with full occlusion
A4	Moving ball with constant speed and direction with partial occlusion
A5	Moving ball with constant speed and arbitrary direction with full occlusion
A6	Moving ball with constant speed and arbitrary direction with partial occlusion

An obstacle is placed in the middle of video frame to represent occlusion in simulation video. In the experiments, two types of occlusion are concerned which included full occlusion and partial occlusion. Figure 3 shows a big rectangle is placed in the middle of the video frame in a video with a ball moves at constant speed and direction. This video labelled as A3, is used to test how an object tracking method could handle moving object after full occlusion. Figure 4 shows some video frames of a video sequence of video label A4 for testing partial occlusion at constant speed and direction. Video with partial and full occlusion are also created for moving object with arbitrary direction change as described of video label A5 and A6.

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

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

Fig. 1. Video Frame (a) Frame no. 1, (b) Frame no 10, (c) frame no. 22, (d) frame no. 30 and (e) frame no. 50 of simulation video A1.

Fig. 2. Video Frame (a) Frame no. 1, (b) Frame no 10, (c) frame no. 22, (d) frame no. 30 and (e) frame no. 50 of simulation video A2.

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

Fig. 3. Video Frame (a) Frame no. 1, (b) Frame no 10, (c) frame no. 22, (d) frame no. 30 and (e) frame no. 50 of simulation video A3.

Fig. 4. Video Frame (a) Frame no. 1, (b) Frame no 10, (c) frame no. 22, (d) frame no. 30 and (e) frame no. 50 of simulation video A4.

IV. TRACKING PERFORMANCE MEASUREMENT

Two tracking performance measurement methods are used in this paper. Both measurement methods are based on the framework by Kasturi et al. [10], which are highly cited protocol for performance evaluation of object detection and tracking in video sequences (other papers agree such statement). The fist method is the Sequence Frame Detection Accuracy (SFDA) as denoted in equation (1) and the second method is ~~Frame Detection~~ Accuracy as expressed in equation (2). The measurements measure the number of object detected and missed detection, false positives and spatial alignment of the system output and ground-truth object.

To calculate the result for both mentioned measurements, ground through is generated using object detection algorithm based on background subtraction [8]. The path of the moving object in video sequence A1, A3 and A5 are identical while A2, A4 and A6 share another similar path. Therefore, only two object movement ground truths are generated for verification of the tracking results.

V. EXPERIMENTAL RESULTS

Experiments have been carried out to evaluate the performance of occlusion handling of the Kalman filter (KF) tracker, Particle filter (PF) tracker and Mean Shift (MS) tracker. The tracker algorithms in MATLAB script are modified and customized based on available sources to suit the experiments . The spatial information of the tracked object is written to text files. Tracking results of various tracker used for the experiments are shown in Table 2.

Kalman filter (KF) tracker used in the experiments is modified from Kashanipour [9]. The SFDA obtained from the tracking experiment using KF tracker is in between 0.3434 and 0.4677. The lowest SFDA score was obtained in video sequence A3 where full occlusion occurred.

Table 2: Tracking result (SFDA) for six simulation videos

SFDA

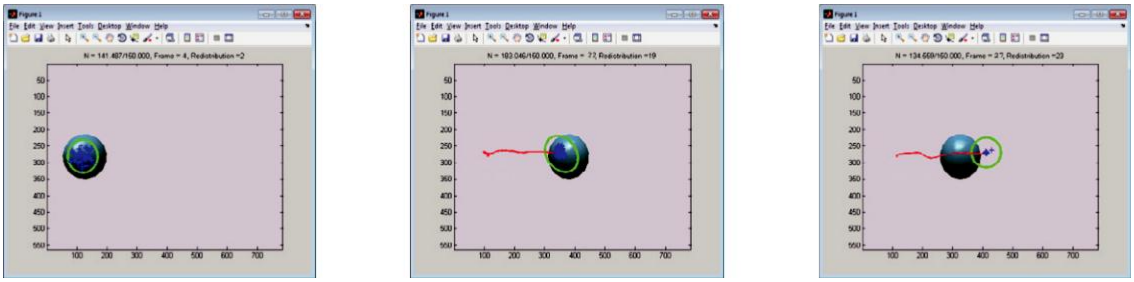
Video Sequence			

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

	KF	PF	MS
A1	0.4628	0.4725	0.5196
A2	0.4583	0.1606	0.6071
A3	0.3434	0.3473	0.0912
A4	0.3732	0.3567	0.4825
A5	0.4568	0.3701	0.3781
A6	0.4677	0.2616	0.5370
Average	**0.4270**	**0.3281**	**0.4359**

Particle filter (PF) tracker is used to track object in the same set of video sequences. The PF tracker used is based on Paris [16]. Based on the SFDA score in Table 2, the result of PF tracker is poorer than KF tracker. The lowest SFDA is achieved when performing PF tracker on video sequence A2.

Based on observation, the PF tracker fails to detect the moving object in the video sequence A2 after frame number 22. A close examination found that the lost track of the object is due to a long period of consistent trajectory of the moving object before frame 22. Stretched consistent trajectory caused the distribution area of the particle become contracted and cover only a small area in the video frame. Therefore, when the trajectory of the object changed suddenly, the PF fail to track the moving object as shown in Figure 5.



(a) (b) (c)

Fig. 5. Particle distribution: a) particles cover a large area at the initial state; b) when object trajectory remains consistent between frames, the particle area shrunk; c) particle area become so small and fail to detect the moving object change direction

Lastly, Mean Shift (MS) tracker by Bernhard [1] is used to test the performance of MS tracker. The result of the MS tracker is the best when no occlusion occurred. This can be identified by having the highest SFDA for both video sequence A1 (0.5196) and A2 (0.6071).

When full occlusion occurred in video sequence A3, the SFDA of MS tracker dropped drastically to 0.0912 only. For video sequence A4, full occlusion also occurred at frame number 22. However, the tracker manages to re-pick up the moving object because the object turns round to the location before full occlusion occurred and reappeared at frame 23. For partial occlusion,

the MS tracker has slightly poorer SFDA result when compared to non-occlusion video sequences, A1 and A2.
In general, the average result obtained in this paper are poorer than some improved tracking method such as Conte et al.[6] work, which used similarity measurement of matrix representation and appearance model to track moving object in dynamic scene. Conte et al. [6] tested their tracking method on PETS2009 S2.L1 video sequences, and obtained average SFDA of 0.505 while average SFDA obtained by this paper are 0.427 (KF), 0.328 (PF) and 0.436 (MS) accordingly. It is important to note that comparing results between different papers is not fair because the object, the background scene and the occlusion situation in the video sequences used are different. The results from this paper are suitable for understanding the effect of occlusion which be discussed in next section.

VI. EFFECT OF OCCLUSION

The SFDA in Table 2 generally provides an impression of the performance of three different trackers. However the SFDA only provides the average performance of each tracker. In order to closely view the effect of occlusion, the Frame Detection Accuracy from frame number 12 to frame number 31 is collected and analysed.

Frame number 12 is the frame where the complete moving object is still visible in video sequence A3 and A4. Occlusion begins at frame number 13 for video sequences A3 and A4 while for video sequence A5 and A6, occlusion begins at and frame number 15.

In video sequence A3, as shown in Figure 6, after a full occlusion, MS tracker failed to detect the moving object while KF and PF tracker could recover after the full occlusion is over. The PF tracker consistently maintains a higher FDA along the video frames even when occlusion occurred.

In video sequence A4, the path of the moving object in the video sequence is identical to the path of the moving object in video sequence A2. In these video sequences, the foreground object moved at a consistent speed to the middle of the video frame. At frame number 22, the object changed its direction and moved backward.

Figure 7 shows that the PF tracker could track the moving object better in video sequence A4 (with occlusion) compared to in video sequence A2 (without occlusion). The reason for PF tracker to be able to track better in video sequence A4 is because the occurrence of occlusion that happens gradually allows the PF tracker to distribute the particle into a bigger area before occlusion occurred. Therefore, when the object change its moving direction, the particle area is big enough to track the moving object and thus did not lost track of the moving object as in video sequence A2.

In video sequence A3, the tracking result of MS tracker deteriorates drastically after occlusion. From the result, it is rational to say that MS tracker has the worst tracking ability after full occlusion as claimed by many previous works [4, 12, 13, 18, 23].

For video sequences A5 where moving object is only partially occluded, all the three trackers show FDA score between 0.3 and 0.5 during occlusion and between 0.38 and 0.48 after occlusion is over at frame number 31 as shown in Figure 8.

For foreground object that has arbitrary trajectory with partial occlusion, both KF tracker and MS tracker could successfully recover quickly after occlusion as shown in Figure 9. PF tracker requires longer time to recover in video sequence A6 due to arbitrary trajectory after partial occlusion occurred.

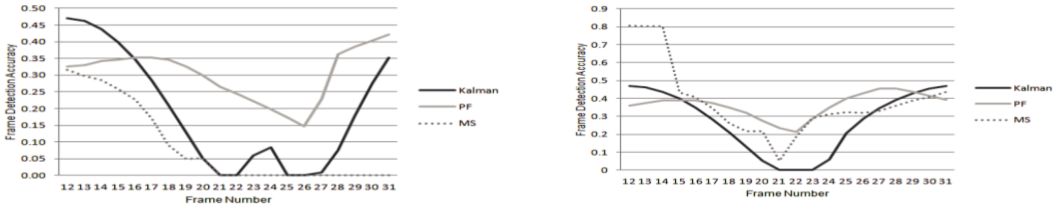


Fig. 6. Frame Detection Accuracy of Frame No. 12 to Frame no or Video Sequence A3

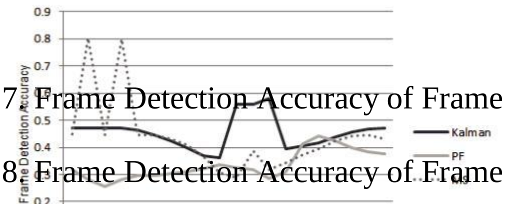


Fig. 7 Frame Detection Accuracy of Frame No. 12 to Frame no. 31 for Video Sequence A4

Fig. 8 Frame Detection Accuracy of Frame No. 12 to Frame no.31 for Video Sequence A5

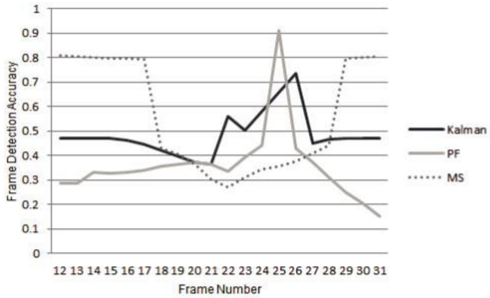


Fig. 9. Frame Detection Accuracy of Frame No. 12 to Frame no.31 for Video Sequence A6

VII. FUTURE WORK

The experiments in this paper are only conducted on trackers that use single predictor, namely the Mean shift, Particle filter and Kalman filter. Recently, many works have proposed to fuse these trackers to achieve a better tracking result. For instance, Li et al. [12] and Zhao et al. [23] have combined Kalman filtering and Mean-shift tracking and Tang and Zhang [18] proposed fusion of Particle filter with Mean Shift tracking.

Therefore, a set of experiment should be carried out with this fusion tracker in the future to identify their actual performance.Another possible future work would be to design more simulation video sequence with complex scenario and test them on the same trackers. Potential scenario would be to use occlusion object with similar colour to the moving object or to use bigger occlusion object to create longer occlusion time. Adding more moving objects would also be useful to study the capability of the trackers.

VII. FUTURE WORK

The experiments in this paper are only conducted on trackers that use single predictor, namely the Mean shift, Particle filter and Kalman filter. Recently, many works have proposed to fuse these trackers to achieve a better tracking result. For instance, Li et al. [12] and Zhao et al. [23] have combined Kalman filtering and Mean-shift tracking and Tang and Zhang [18] proposed fusion of Particle filter with Mean Shift tracking.

Therefore, a set of experiment should be carried out with this fusion tracker in the future to identify their actual performance.Another possible future work would be to design more simulation video sequence with complex scenario and test them on the same trackers. Potential scenario would be to use occlusion object with similar colour to the moving object or to use bigger occlusion object to create longer occlusion time. Adding more moving objects would also be useful to study the capability of the trackers.

VII. CONCLUSION

In this paper, a set of simulated video was design and generated to test the tracking capability of three popular trackers observed from review of previous works. Experiments are conducted using the Kalman filter, Particle filter and Mean Shift tracker.

Sequence Frame Detection Accuracy was used to evaluated the tracking performance of each tracker. Most result confirmed claimed of previous work but PF tracker was surprisingly fail to detect object with arbitrary movement.

Moreover, the effects of the occlusion on every tracking method are discussed in detail based on Frame Detection Accuracy. The capability of each tracker to recover from occlusion is also analysed using graphs.

REFERENCES

[1] Bernhard, S.: Mean-Shift Video Tracking, MATLAB Central, http://www.mathworks.com/matlabcentral/fileexchange/355, Mar 2012
[2] Di Caterina, G., Soraghan, J. J., 2011. “An Improved Mean Shift Tracker with Fast Failure Recovery Strategy after Complete Occlusion,” In *IEEE* Int. Conf. Advanced Video and Signal-Based Surveillance, pp. 130-135.
[3] Cho, J.U., Jin, S.H., Pham, X.D. and Jeon, J.W. 2006. “Object Tracking Circuit using Particle Filter with Multiple Features,” In SICE-ICASE, 2006. International Joint Conference, pp. 1431-1436.
[4] Comaniciu, D., Ramesh, V. and Meer P., 2003. “Kernel-based object tracking,” In IEEE Trans. on Pattern Analysis and Machine Intelligence, 25(5),564-577.
[5] Comaniciu, D., Ramesh, V. and Meer P. 2000. “Real-time tracking of non-rigid objects using mean shift,” In Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, pp. 142-149.
[6] Conte, D., Foggia, P., Percannella, G. and Vento, M. 2010. “Performance Evaluation of a People Tracking System on PETS2009 Database,” In Proceedings of 2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance, pp.119-126.
[7] ETISEO: Evaluation for video understanding, http://www-sop.inria.fr/orion/ETISEO/, Nov 2011
[8] Gao, J., 2003. “Self-Occlusion Immune Video Tracking of Objects in Cluttered Environment,” In IEEE Int. Conf. Advanced Video and Signal-Based Surveillance, pp. 79–84.
[9] Kashanipour, A.: 2D Target tracking using Kalman filter, MATLAB Central, http://www.mathworks.com/matlabcentral/fileexchange/14243-2d-target-tracking-using-kalman-filter, Mar 2012
[10] Kasturi, R., Goldgof, D., Soundararajan, P., Manohar, V., Garofolo, J., Bowers, R., Boonstra, M., Korzhova, V. and Zhang, J., 2009. “Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol,” In IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2), pp. 319–336.

[11] Li, G., Jean-Sébastien, F. and Marc, P., 2010. Multi-view Occlusion Reasoning for Probabilistic Silhouette-Based Dynamic Scene Reconstruction. International Journal of Computer Vision, 90 (3), pp.283-303.
[12] Li, Y., Pang, Y., Li, Z. and Liu, Y., 2010. "An Intelligent Tracking Technology Based on Kalman and Mean Shift Algorithm," In Second Int. Conf. on Computer Modelling and Simulation, 1, pp. 107-109.
[13] Li, Z., Tang, Q.L. and Sang N., 2008. Improved mean shift algorithm for occlusion pedestrian tracking, Electronic Letters, 44 (10), pp. 622-623.
[14] Liang, N., Guo, L. and Wang Y., 2012. "An Improved object tracking method based on particle filter," In Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on, pp. 3107-3110.
[15] Mirabi, M. and Javadi, S., 2012. "People Tracking in Outdoor Environment Using Kalman Filter," In Proceedings of Intelligent Systems, Modelling and Simulation (ISMS), 2012 Third International Conference on, pp. 303-307.
[16] S. Paris: Particle Filter Color Tracker, MATLAB Central, http://www.mathworks.com/matlabcentral/fileexchange/17960-particle-filter-color-tracker, Mar 2012.
[17] PETS2009: Benchmark data, http://www.cvg.rdg.ac.uk/PETS2009/a.html, Nov 2011.
[18] Tang D. and Zhang, Y.J., 2011. "Combining Mean-Shift and Particle Filter for Object Tracking," In Sixth Int. Conf. on Image and Graphics (ICIG),
pp. 771-776.
[19] Taylor, G. R., Chosak, A. J. and Brewe, P. C., 2007. "OVVV: Using virtual worlds to design and evaluate surveillance systems," In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
[20] Wang, H., Huo, L. and Zhang, J., 2011. "Target tracking algorithm based on dynamic template and Kalman filter," Communication Software and Networks (ICCSN), 2011 IEEE 3rd, pp. 330-333.
[21] Yilmaz, A., Javed, O. and Shah, M., 2006. "Object tracking: a survey," ACM Computing Surveys, 38(4).
[22] Yilmaz, A., 2007. "Object Tracking by Asymmetric Kernel Mean Shift with Automatic Scale and Orientation Selection," In Proceedings of Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, pp. 1-6,
[23] Zhao, J., Qiao, W.and Men, G. Z., 2009. "An Approach Based on Mean Shift and Kalman Filter for Target Tracking Under Occlusion," In Int. Conf. Machine Learning and Cybernetics, pp. 2058–2062.