Occlusion handling in videos object tracking: A survey

B Y Lee[1,3], L H Liew[1], W S Cheah[2] and Y C Wang[2]

[1] Faculty of Mathematics and Computer Sciences, Universiti Teknologi MARA,
Malaysia
[2] Faculty of Computer Sciences and Information Technology, Universiti Malaysia
Sarawak, Malaysia

Email: bengyong@gmail.com

Abstract. Object tracking in video has been an active research for decades. This interest is motivated by numerous applications, such as surveillance, human-computer interaction, and sports event monitoring. Many challenges regarding tracking objects remain, this can arise due to abrupt object motion, changing appearance patterns of objects and the scene, non-rigid object structures and most significancly occlusion of tracked object (be it object-to-object or object-to-scene occlusions). Generally, occlusion in object tracking occurs under three situations: self-occlusion, inter-object occlusion by background scene structure. Self-occlusion most frequently arises while tracking articulated objects when one part of the object occludes another. Inter-object occlusion occurs when two objects being tracked occlude each other whereas occlusion by the background occurs when a structure in the background occludes the tracked objects. Typically, tracking methods handle occlusion by modelling the object motion using linear and non-linear dynamic models. The derived models will be used to continuously predicting the object location when a tracked object is occluded until the object reappears. Examples of these methods are Kalman filtering and Particle filtering trackers. Researchers have also utilised other features to resolved occlusion, for example, silhouette projections, colour histogram and optical flow. We will present some results from a previously conducted experiment when tracking single object using Kalman filter, Particle filter and Mean Shift trackers under various occlusion situations. We will also review various other occlusion handling methods that involved using multiple cameras. In a nutshell, the goal of this paper is to discuss in detail the problem of occlusion in object tracking and review the state of the art occlusion handling methods, classify them into different categories, and identify new trends. Moreover, we discuss the important issues related to occlusion handling including the use of appropriate selection of motion models, image features and use of multiple cameras.

1. Introduction

Object tracking in video has been an active research since more than decades ago. This interest is motivated by numerous applications, such as surveillance [1-4] human-computer interaction [5,6], and robots navigation [7,8]. Many of these applications of object tracking have the goal to have ability to automatically understand events happening at a site [9]. To have higher level of understanding of events requires that certain lower level computer vision tasks to be performed. These may include tracking objects, handling occlusion and detection of unusual motion.

Object tracking is a process of monitoring an object's spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc [10]. Many challenges still remains big on tracking objects, this can arise due to abrupt object motion, changing appearance patterns of objects and the scene, non-rigid object structures and most significantly is handling occlusion of tracked object.

---

[3] To whom any correspondence should be addressed.

Many studies have been conducted to improve object tracking technique reliability and accuracy. An article written by [11] summarises the techniques in widespread use and classifies them into 35 different algorithmic types as well as providing a comprehensive literature survey of object tracking. Another comprehensive was also produced by [12] has reviewed on the techniques used in object tracking and categorize the techniques on the basis of the object and motion representations used. The review also comprehensively provides detailed descriptions of representative methods in each category, and examines their strength and weaknesses.

However these previous articles discussed very little on the occlusion handling methods and it has been a long time since the last survey was conducted on object tracking. Therefore, the goal of this paper is to discuss the problem of occlusion in object tracking and review the state of the art occlusion handling methods, classify them into different categories, and identify new trends. Moreover, we discuss the important issues related to occlusion handling including the use of appropriate selection of motion models, image features and use of multiple cameras.

2. Occlusion in Object Tracking

In video object tracking, occlusion happened when a tracked object or its key attributes used to recognize its identity is not available for a camera sensor to keep tracking its spatial state while the object is still present at the scene. Occlusion can be caused by various reasons, occlusion occurs when two or more inter-object occlusion and occlusion by the background scene structure. Sometimes, occlusion occurs when one part of the object occludes another, this is known as self-occlusion. This situation most frequently arises while tracking articulated objects. Inter-object occlusion occurs when two objects being tracked occlude each other. Similarly, occlusion by the background occurs when a structure in the background occludes the tracked objects.

Handling occlusion in a single object environment is a straight forward task. If only a single object will appear within the scene, the tracking task will only involved detecting object appearance. However, for scene with multiple object appearance, handling occlusion becomes complicated. For inter-object occlusion, [13] view the multi-object trackers problem as a challenging problem, especially when the targets are "identical", in the sense that the same model is used to describe each target. Occlusion handling ability of a tracking techniques also depend on the severity of occlusion and the state of the occlusion. In the subsection below, we will discuss how occlusion severity can be classified and various states of occlusion were identified by [14].

2.1. Severity of Occlusion

Existing literature has identified a number of occlusion categories in the context of tracking in ad hoc manner. One way to categories occlusion is based on the severity of the occlusion. Occlusion can be categorized as non-occlusion, partial occlusion, full occlusion and long term full occlusion. In multi-camera setup, the interval between leaving the field of view of a camera and moving to another camera or return to the previous camera can also be considered as occlusion.

During non-occlusion, the tracked object appears as a single blob having all tracking features exposed to the camera sensor for tracking. Most of the existing tracking methods that are based on appearance attribute to track a target such as Template Matching and Mean Shift [15] will be able to tracked non occlusion object accurately.

Partial occlusion happened when some of the key features of the tracked object are hidden from the camera during tracking. Partial occlusion could happen when part of the tracked object is blocked by other object or background structure or during self-occlusion. Partial occlusion of an object by a foreground structure is hard to detect since it is difficult to differentiate between the object changing its shape and the object getting occluded. When partial occlusion occurs, tracker that only adopts simple template matching without consideration of the evolution of the object over time will fail. Therefore, many newly developed tracking methods such as Mean Shift [15] include adaptive step into the tracking method to handle partial occlusion.

Full Occlusion happen when a tracked object is completely not visible while knowing that the object has not left the area of view of the camera. When full occlusion occurs, no one method that rely on image appearance could continue to track the object because there will be no more appearance clue left in the scene for occlusion. To handle full occlusion, many of the state of the art tracking method

incorporated object spatial motion model into the tracking process. For instance, Kalman filter are used for estimating the location and motion of objects in [16,17,18].

2.2. State of Occlusion

In 2011, [14] formulate a set of occlusion cases by considering the spatial relations among tracked object and the detected foreground blob(s), to show that only 7 occlusion states are possible. Other than non occlusion and full occlusion, [14] define another five occlusion states based on fragmentation and grouping of blob(s) when partial occlusion occurs. For instance, when a tracked object is partially occluded by a foreground structure, the detected foreground blob of the tracked object could be broken into several individual small blobs. On the other hand, when a tracked object is occluded by other foreground objects, the foreground blobs of the tracked object could be grouped into a bigger blob with other foreground objects blob. The authors claimed to be the first to systematically analyze and formulate a complete set of occlusion cases by combining the cardinalities and modalities of object blob overlap.

3. Occlusion Handling

Occlusion handling is an inevitable problem in object tracking. The problem of occlusion occurs during the occlusion and after the occlusion. During the time of occlusion, two types of challenges may occur. First, when two foreground objects occlude one another, the foreground blob of the objects will group together and it will become challenging to classify the pixels in the blob to respective object accurately. Secondly, during occlusion, the actual location of a tracked object is difficult to determine since the visibility of the tracked object become limited or totally missing.

After the event of occlusion, especially the full occlusion, the challenge of associating the reappear object is tough. When an object reappear, it is difficult to decide if the object is a new object appear to the scene or it is an object reappear after occlusion. The problem becomes more complicated especially when tracking multiple objects with similar appearance. The solutions to this problem are normally known as occlusion recovery methods. To handle occlusions problem, various methods has been proposed. We have summarized a few state of the art methods and classify them based on the nature of the problem each of the methods tried to solve.

3.1. Depth Analysis

Solution to merged foreground blob problem is through determining the depth of the tracked objects such as [19,20], and [21]. By determining the depth of the object, when two or more object occlude one another creating a grouped blob, the depth model of the object can be use to separate the object blobs because the object nearer to the camera will have shorter depth compare to the object that is further. This makes segmenting the object blob during occlusion easier.

Various methods were used to estimate the depth of the tracked object and the structure in a scene. Previous researchers such as [19] and [21] used video sequences obtained with stereo cameras to calculate the depth of the objects in the video. Another method of generating the probability density functions for the depth of the scene at each pixel from a training set of detected blobs through projection of the 3D scene as 2D was proposed by [20] so that the occlusion landscape is made explicit such structures can defeat the process of tracking individuals through the scene. In [22] article, they proposed using Microsoft Kinect™ to recognize hand gesture.

3.2. Fusion methods

In a review article written by [12], they stated that previous researchers believe they can exploit the knowledge of the position and the appearance of the occluding object and occluded object to detect and resolve occlusion. A common approach to handle complete occlusion during tracking is to fuse the object motion with appearance model by linear dynamic models or by nonlinear dynamics and, in the case of occlusion, to keep on predicting the object location until the object reappears. For example, a Kalman filter is used for estimating the location and motion of objects in [16,18] and [23]. A nonlinear dynamic model are used in [24,25], and [26] where a particle filter employed for state estimation.

3.3. Optimal camera placement method

The chance of occlusion can be reduced by an appropriate selection of camera placement positions. For instance, if a 360 degree ultra wide angle camera is mounted on the ceiling of a room, which is, when a birds-eye view of the scene is available, occlusions between objects on the ground do not occur. However, ultra wide angle camera will distort images captured and objects captured in the video will have very identical appearance features. Other than that, installing more cameras in the surveillance scene can also help to solve occlusion problem [27,28,29]. With many cameras installed, when an object is obstructed from viewing by a camera, the object may still be visible from other camera, this reduce the chances from being occluded.

4. Data set and result analysis

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 [30] and ETISEO [31]. 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.

In order to explore the actual potential of the state of the art tracker, namely the Kalman Filter, Particle Filter and Mean Sift tracker, [32] introduced 64 simulation video sequences to experiment the effectiveness of each tracking methods on various occlusion scenarios. Tracking performances are evaluated based on Sequence Frame Detection Accuracy (SFDA) [33]. The results showed that Mean shift tracker would fail completely when full occlusion occurred. Kalman filter tracker achieved highest SFDA score of 0.85 when tracking object with uniform trajectory and no occlusion. Results also demonstrated that Particle filter tracker fails to detect object with non-uniform trajectory.

5. Conclusion

In this article, we present a survey of object handling methods and also give a brief review of related topics. We categorize the object occlusion methods into three categories, namely, depth analysis techniques, fusion techniques and camera placement techniques. The paper also provides detailed summaries of how occlusion was formed during object tracking and how each of these occlusion types affects object tracking result. Moreover, the paper also listed out some of the commonly use datasets for evaluation of occlusion handling techniques. We believe that, this article, a survey on occlusion handling technique in object tracking with comprehensive references, will be helpful to provide valuable insight into this research topic.

References

[1]   Huang K, Tao D, Yuan Y, Li X and Tan T 2011 Biologically Inspired Features for Scene Classification in Video Surveillance *Systems, Man, and Cybernetics, Part B: Cybernetics* **41** 307-313

[2]   I. Ali and M. N. Dailey, "Multiple human tracking in high-density crowds," *Image and Vision Computing,* vol. 30, no. 12, p. 966–977, 2012.

[3]   W. Nie, A. Liu and Y. Su, "Multiple Person Tracking by Spatiotemporal Tracklet Association," in *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, Beijing, 2012.

[4]   L. N. Rasid and S. A. Saudi, "Versatile Object Tracking Standard Database for Security Surveillance," in *Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on*, Kuala Lumpur, 2010.

[5]   S. S. Rautaray and A. Agrawal, "Design of gesture recognition system for dynamic user interface," in *Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on*, Kerala, 2012.

[6]   H. Tao and Y. Yu, "Finger Tracking and Gesture Recognition with Kinect," Chengdu, 2012.

[7]   P. Benavidez and M. Jamshidi, "Mobile robot navigation and target tracking system," in *System of Systems Engineering (SoSE), 2011 6th International Conference on*, Albuquerque, NM,

2011.

[8]   G. Capi, H. Toda and T. Nagasaki, "A vision based robot navigation and human tracking for social robotics," in *Robotic and Sensors Environments (ROSE), 2010 IEEE International Workshop on*, Phoenix, AZ, 2010.

[9]   A. Elgammal, R. Duraiswami, D. Harwood and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of IEEE,* vol. 90, no. 7, pp. 1151 - 1163, 2002.

[10] A. Ali and S. M. Mirza, "Object Tracking using Correlation, Kalman Filter and Fast Means Shift Algorithms," in *2nd International Conference on Emerging Technologies*, Peshawar, Pakistan, 2006.

[11] G. W. Pulford, "Taxonomy of multiple target tracking methods.," *Radar, Sonar and Navigation, IEE Proceedings,* vol. 152, no. 5, pp. 291 - 304, 2005.

[12] A. Yilmaz, O. Javed and M. Shah, "Object tracking: A survey," *ACM Computing Surveys (CSUR),* vol. 38, no. 4, pp. 1-45, December 2006.

[13] J. Maccormick and A. Blake, "A Probabilistic Exclusion Principle for Tracking Multiple Objects," *International Journal of Computer Vision,* vol. 39, no. 1, pp. 57 - 71, 2000.

[14] P. Guha, A. Mukerjee and V. K. Subramanian, "Formulation, Detection and Application of Occlusion States (Oc-7) in the Context of Multiple Object Tracking," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, Klagenfurt, 2011.

[15] D. Comaniciu and V. Ramesh, "Mean shift and optimal prediction for efficient object tracking," in *Image Processing, 2000. Proceedings 2000 International Conference on (Volume 3)* , Vancouver, BC, 2000.

[16] A. Ali and K. Terada, "A Framework for Human Tracking using Kalman Filter and Fast Mean Shift Algorithms," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, Kyoto, 2009.

[17] Z. Zhu, Q. Ji, K. Fujimura and K. Lee, "Combining Kalman filtering and mean shift for real time eye tracking under active IR illumination," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on (Volume 4)*, 2002.

[18] J. Zhao, W. Qiao and G.-Z. Men, "An approach based on mean shift and KALMAN filter for target tracking under occlusion," in *Machine Learning and Cybernetics, 2009 International Conference on (Volume 4)*, Baoding, 2009.

[19] M. Harville, G. Gordon and J. Woodfill, "Adaptive video background modeling using color and depth," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, Thessaloniki, 2001.

[20] D. Greenhill, J. Renno, J. Orwell and G. A. Jones, "Occlusion analysis: Learning and utilising depth maps in object tracking," *Image Vision Comput,* vol. 26, no. 3, pp. 430-441, March 2008.

[21] Y. Ma and Q. Chen, "Depth Assisted Occlusion Handling in Video Object Tracking," in *Advances in Visual Computing*, Berlin, Springer Berlin Heidelberg, 2010, pp. 449-460.

[22] J. Suarez and R. R. Murphy, "Hand Gesture Recognition with Depth Images: A Review," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on*, Paris, 2012.

[23] J. Ren and J. Hao, "Mean Shift Tracking Algorithm Combined with Kalman Filter," in *5th International Congress on Image and Signal Processing (CISP 2012)*, 2012.

[24] M. Isard and J. MacCormick, "BraMBLe: a Bayesian multiple-blob tracker," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001.

[25] Z. Duan, Z. Cai and J. Yu, "Occlusion detection and recovery in video object tracking based on adaptive Particle filters," in *Control and Decision Conference, 2009. CCDC '09.*, Guilin, 2009.

[26] D. Tang and Y.-J. Zhang, "Combining Mean-Shift and Particle Filter for Object Tracking," in

Image and Graphics (ICIG), 2011 Sixth International Conference on, Hefei, Anhui, 2011.

[27] J. P. Batista, "Tracking Pedestrians Under Occlusion Using Multiple Cameras," in *Image Analysis and Recognition*, Berlin , Springer Berlin Heidelberg, 2004, pp. 552-562.

[28] M. Mozerov, A. Amato, X. Roca and J. González, "Solving the multi object occlusion problem in a multiple camera tracking system," *Pattern Recognition and Image Analysis,* pp. 165-171, 2009.

[29] R. Raman, P. Sa and B. Majhi, "Occlusion prediction algorithms for multi-camera network," in *Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on*, Hong Kong, 2012.

[30] IEEE Computer Society, "Performance Evaluation of Tracking and Surveillance 2009," 2009. [Online]. Available: http://www.cvg.rdg.ac.uk/PETS2009/a.html.

[31] INRIA, "ETISEO - Home," 2007. [Online]. Available: http://www-sop.inria.fr/orion/ETISEO/.

[32] B. Y. Lee, L. H. Liew, W. S. Cheah and Y. C. Wang, "Simulation Videos for Understanding Occlusion Effects on Kernel Based Object Tracking," in *Computer Science and its Applications*, vol. 203, Netherlands, Springer Netherlands, 2012, pp. 139-147.

[33] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova and J. Zhang, "Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, no. 2, p. 319–336, 2009.

[34] G. Phadke, "Robust multiple target tracking under occlusion using fragmented mean shift and Kalman filter," in *Communications and Signal Processing (ICCSP), 2011 International Conference on*, Kerala, 2011.

[35] H. Lu, R. Zhang and Y.-W. Chen, "Head Detection and Tracking by Mean-shift and Kalman Filter," in *The 3rd Intetnational Conference on Innovative Computing Information and Control (ICICIC) 2008*, Dalian, Liaoning, 2008.

[36] E. Parvizi and Q. Wu, "Multiple object tracking base on adaptive depth segmentation," in *Computer and Robot Vision, 2008. CRV '08. Canadian Conference on*, Windsor, Ont., 2008.

[37] IEEE Computer Society, "Performance Evaluation of Tracking and Surveillance 2009," 8 3 2012. [Online]. Available: http://www.cvg.rdg.ac.uk/PETS2009/a.html.