Enhancing COCOA Framework for Tracking Multiple Objects in the Presence of Occlusion in Aerial Image Sequences

Vindhya P. Malagi, Vinuta V. Gayatri, Krishnan Rangarajan and D. R. Ramesh Babu

Abstract Multi object tracking in aerial image sequences is a topic of utmost importance in the field of computer vision for both military and civilian applications. In order to extract valid information about moving targets, it is required to detect and track these targets precisely in the input image sequences. Occlusion is one of the prominent problem areas that hinder efficient object tracking. Spatial reasoning literature fails to distinguish various analyses that are prominent to computer vision. However, recognizing valid occlusion states and mining their transition sequences help in analyzing the pose and motion of multiple interacting objects in the scene. In this paper, we propose an enhancement incorporating occlusion in the existing COCOA framework for tracking in aerial image sequence. The contribution of the paper is the novel idea of extracting occlusion cues as a pre-processing step to aid the tracker. We describe approaches to extract occlusion information in the scene and use it as a cue for efficient tracking.

Keywords Multi object tracking · COCOA framework · Occlusion sequence mining

V. P. Malagi (✉) · V. V. Gayatri · K. Rangarajan · D. R. Ramesh Babu
Computer Vision Lab, Dayananda Sagar College of Engineering, Bangalore, India
e-mail: vindhyapm@gmail.com

V. V. Gayatri
e-mail: vinuta06.gayatri@gmail.com

K. Rangarajan
e-mail: krishnanr1234@gmail.com

D. R. Ramesh Babu
e-mail: bobrammysore@gmail.com

P. P. Swamy and D. S. Guru (eds.), *Multimedia Processing, Communication and Computing Applications*, Lecture Notes in Electrical Engineering 213,
DOI: 10.1007/978-81-322-1143-3_18, © Springer India 2013

1 Introduction

Unmanned Air Vehicles (UAVs) play a critical role in surveillance, target tracking and reconnaissance in both urban and battlefield settings. One of the major challenges in tracking is handling occlusion of the objects being tracked. This paper gives a framework for mining various occlusion scenarios in aerial images from UAV and using them as cue to the tracker. Object tracking, in essence, deals with a set of image sequences that change over time. While the existing algorithms are able to track objects well in controlled environments, they usually fail in the presence of significant variation of the object's appearance or surrounding illumination. Occlusion has also been seen as one of the prominent problem areas in efficient object tracking. Here, the COCOA framework [1] is considered as the underlying framework for tracking moving objects from aerial images.

Tracking objects for long durations in aerial imagery is a challenging task as the objects are small in size, similar in appearance and tend to occlude one another while in motion. Also they may disappear at arbitrary points and then reappear after sometime. Such images invariably contain noise due to the relative movement of the camera with respect to the vehicle often called the 'jitter' and noise induced due to environmental conditions like brightness or glare, noise due to haze etc. Even shadows appearing in the image need to be removed as a preprocessing step. Though a lot of research has gone into addressing this problem of occlusion, a complete solution is still far from being achieved. In this paper we propose extensions to COCOA framework for tracking in the presence of occlusion.

2 Tracking Framework

We propose an extension that feeds various occlusion state cues to the tracking module in the existing COCOA framework. Figure 1 shows the proposed extension.

Here COCOA model forms the basis of the framework. We propose to enhance the working of this model by incorporating the occlusion cue module which takes in the appearance models (blobs) from the motion detector, compare them with the mined occlusion states and formulate appropriate occlusion related cues. The details of this processing are discussed in Sects. 3 and 4.

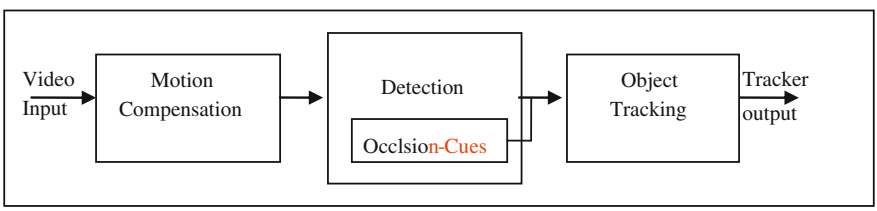

Fig. 1 Proposed framework with occlusion states as cue to the tracker

2.1 Motion Compensation

COCOA framework is an established framework for object tracking in aerial imagery. In this framework, the motion compensation model is used to compensate for the motion of the moving camera. By registering the whole video with respect to one global reference we are able to get a meaningful representation of the object trajectories which can be used to describe the input. Here, the image registration is performed by considering a small block of image set, extracting the features and matching them using SIFT [2]. The tentative matches are then filtered using RANSAC to find the correspondences that best fit a homography.

2.2 Motion Detection

After removing global camera motion, we detect local motion of objects moving in the scene. Motion detection provides a classification of the pixels in the video sequence into either foreground (moving objects) or background. A common approach used to achieve such classification is background removal, sometimes referred to as background subtraction [3], where each video frame is compared against a reference or background model, pixels that deviate significantly from the background are considered. Probabilistic modeling of the background has been popular for surveillance videos. However, these methods are found to be inapplicable to aerial image sequences, as they fail by either giving raise to large ghost effects or failing to distinguish between foreground and background pixel accurately. Therefore, we avoid probabilistic models in favor of simple median image filtering [4], which considers a background model with fewer artifacts using fewer frames. We use a 15 frame median image for the purpose with minimum ghost effect.

In surveillance applications, determining occlusion primitives is based on foreground blob tracking, and requires no prior knowledge of the domain or camera calibration. New foreground blobs are identified as putative objects which may undergo occlusions, split into multiple agents, merge back again, etc. Using temporal sequence mining, significant cues on the various occlusion states that can occur due to the above mentioned interaction can be generated that can serve as inputs to the tracker as explained in Sect. 4.

2.3 Tracking

In COCOA framework, tracking is critical for obtaining meaningful tracks that capture the motion characteristics over longer durations of time. Here, we have used appearance based tracking approach [5] to perform multi-target tracking. Regions of interest, or blobs, in successive frames are given by the motion

detection module. Each blob is represented by a distinct appearance and shape model.

In the proposed framework, the tracker makes use of the occlusion cues which are generated from the temporal sequence mining of occlusion states to refrain from loosing track of the moving objects in cases of partial or complete occlusion scenarios.

3 Occlusion Handling

In spatial reasoning, occlusion states are informative on the relative pose and motion of multiple objects. In the application of target tracking, occlusion seems to be inevitable. Former approaches of spatial reasoning are not adequate for such computer vision applications. The work here closely follows the work on occlusion by Guha et al. [6, 7].

Computer vision literature mostly talks about occlusion in terms of split and merge cases. However, OCS-14 paper of Guha [6] well defines 14 different occlusion states that cover almost all computer vision scenarios. But we observe that in particular, the occlusion states like ocSGP (static occlusion with grouping and partial visibility), ocSGF (static occlusion with grouping and fragmentation) and ocSG0 (static occlusion with grouping and no visibility) as stated in the paper cannot be justified as grouping which is interaction between two or more moving objects, brings in dynamic occlusion. Therefore it seems to be appropriate to consider occlusion states based on the practicality of the problem in hand.

Hence in this paper we first look into the simplified occlusion states in object tracking and then look into a way to mine these states. On extensive survey we found that very few researchers have worked explicitly on the problem of occlusion [6, 8, 9]. In the application of tracking, all the occlusions span over two important types of occlusion states—partial occlusion state and complete occlusion state. The rest of the occlusion states can be sub-spanned under these two states, based on visibility of the occluded object and static and dynamic nature of the occluder as follows:

1. Partial Occlusion
   (a) From a static occluder
   (b) From a dynamic occluder
   (c) Due to Split and Merge Scenarios
   (d) Entry/exit scenarios
2. Complete Occlusion
   (a) From a static occluder
   (b) From a dynamic occluder
   (c) Due to Split and Merge Scenarios
   (d) Entry/exit scenarios

Typical example of one of the above scenarios could be the case in which an object approaches a static occluder, undergoes a series of partial occlusions before completely getting occluded (considering their apparent sizes). Similarly at entry/exit points, an object may be partially visible in a sequence of images before it enters/exits the scene completely. Split, merge scenarios are most common in tracking problem and can occur from either static and/or dynamic occluder. A typical merge and split case gives rise to transitions from fully visible to partial or complete occlusion state and vice versa if the split of the occluded object from occluder happens after the merge. When an object is completely occluded over a long period of time before emerging back at a later time, it may be considered as a case of temporal occlusion.

In general, sequence of occlusion state transitions can be considered as important visual signatures of the interaction between objects. As an example, in a scenario where an object is entering a scene, it emerges with partial visibility before being completely visible. This means that the scene undergoes a sequence of transitions from one occlusion state (partial) to another before transitioning to complete visibility and can thus be termed as occlusion state transition. It is possible to use data mining techniques on the occlusion transitions as they emerge in a scene, and one can gain useful abstractions about the scene and the object behaviors.

4 Occlusion Mining

For learning the event signatures that can be fed as crucial cues to the tracker, we consider mining the spatio-temporal sequence patterns of the occlusion primitives and model these using substring trees [10]. A spatio-temporal sequence S is a list of locations, (x_1, y_1, t_1), (x_2, y_2, t_2),…, (x_n, y_n, t_n), where t_i represents the time-stamp of location (x_i, y_i) $(1 \leq i \leq n)$. The substring tree is a rooted directed tree whose root links to multiple substring sub-trees. Each node in a sub-tree consists of pattern element (occlusion state) and a counter, which counts the number of substrings (i.e., subsequences of elements) that contribute to the pattern formed by the path from the root to this node.

Querying the substring trees with the time ordered sequences of the occlusion states detects similarities with other events where this type of sequences arose, thereby detecting episodes of the transit-across-a-large-occlusion event.

The novelty of this paper is in the idea that these event signatures can be fed as occlusion cues to the tracker to better the overall tracking performance. For instance, the regions of frequent occlusion clearly reveal important information about the scene depth and also about object behaviors in the given scene context as shown in Fig. 2. Eventually such inferences can be picked up by the tracker to conclude on the upcoming visual scenario thus improving the precision and accuracy of tracking.

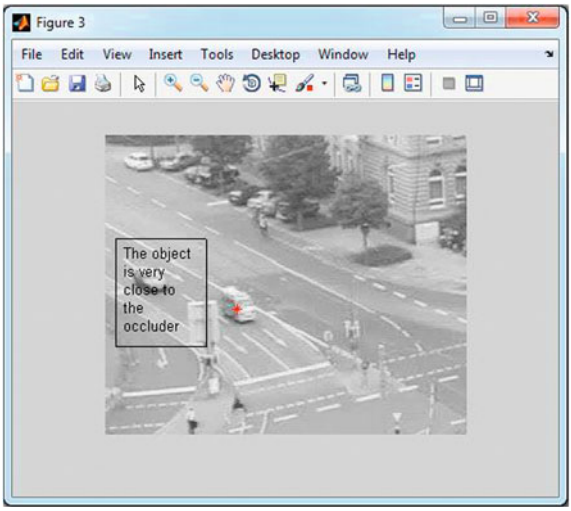

Fig. 2 **a**, **b** Entry with partial occlusion state (A car entering the scene from the right is partially visible at the image boundary)

5 Results

Preliminary implementation on this enhanced framework on aerial images shows promising results. All occlusion states mentioned in Sect. 3 could be identified and a few examples (Figs. 3, 4, 5, 6) are shown below. The video sequences considered for experimentation are of about 2,000 frames length. Frames with moving objects with different occlusion states are shown in figures (a) and their corresponding segmented output in (b).

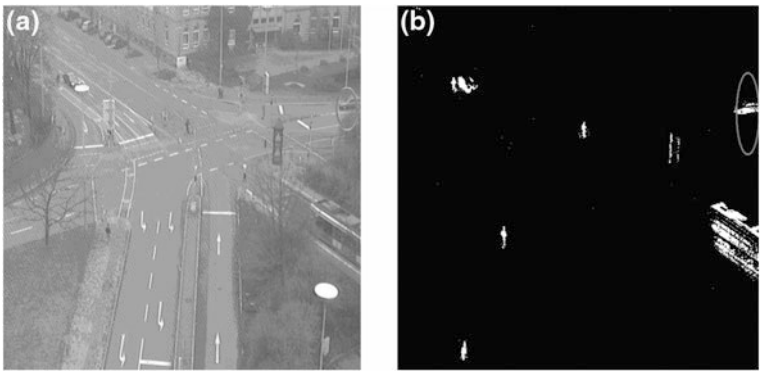

Fig. 3 **a**, **b** Entry with partial occlusion state (A car entering the scene from right is partially visible at the image boundary)

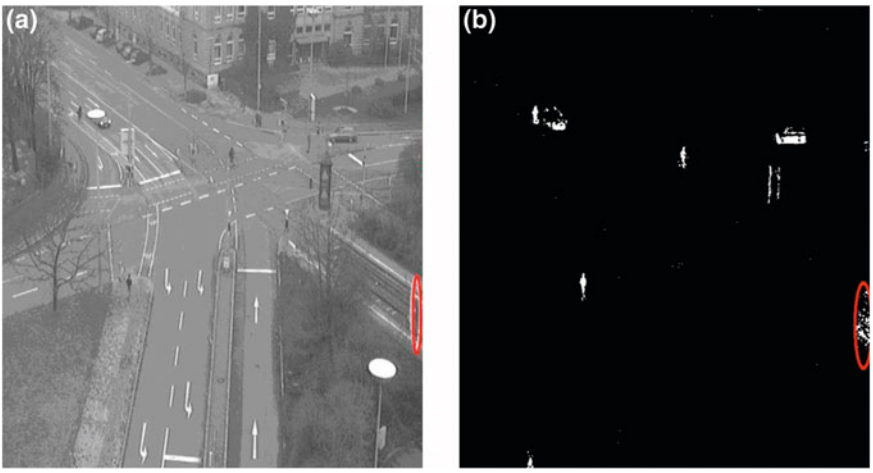

Fig. 4 **a**, **b** Exit with partial occlusion state (A car exiting the scene from right is partially visible at the image boundary)

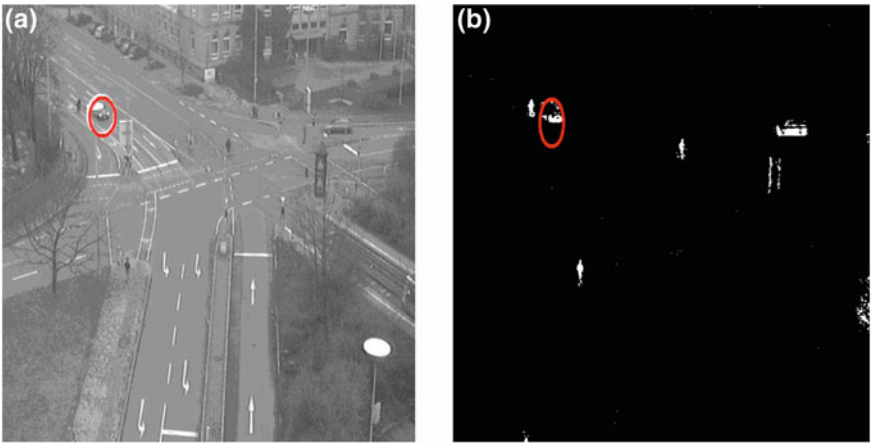

Fig. 5 **a**, **b** Partial occlusion with static occluder

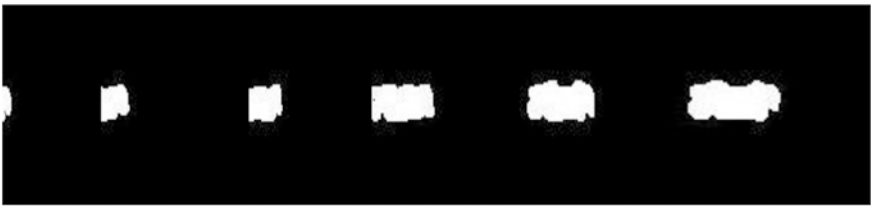

Fig. 6 Blobs from motion detection in a sequence of images showing the case of entry of a new object and revealing a series of occlusion state transitions of partial occlusion states before complete visibility

6 Conclusion

In this paper, a novel approach for efficient object tracking through occlusions has been proposed. This framework is based on the popular COCOA framework for object tracking in aerial images. The work has identified the simplified and valid occlusion states relevant to the tracking application in computer vision. It considers various occlusion state transitions, relevant to spatial–temporal reasoning that can be mined and given as cue to the tracker. The contribution of the paper is the idea of treating occlusion cues as the pre-processing step to the tracker. This proposed idea is likely to better the performance of the tracker. Work on building a strong mathematical model for the proposal and prototype implementations is in progress.

Acknowledgments This work is funded by ER & IPR, DRDO ref no: ERIP/ER/1104561/M/01/1353, India.

References

1. Ali S, Shah M (2006) COCOA—tracking in aerial imagery. SPIE Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications, Orlando
2. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Computer Vision 60(2):91–110
3. Stauffer C, Grimson WEL (1999) Adaptive background mixture models for real-time tracking. In: The proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2
4. Reilly V, Idrees H, Shah M (2010) Detection and tracking of large number of targets in wide area surveillance. In: The proceedings of 11th European Conference on Computer Vision, LNCS, vol 6313. pp 186–199, Springer
5. Senior A, Hampapur A, Tian Y, Brown L, Pankanti S, Bolle R (2006) Appearance models for occlusion handling. In: 2nd International Workshop on Performance Evaluation of Tracking and Surveillance Systems, Science Direct, Elsevier, Image and Vision Computing, vol 24. pp 1233–1243
6. Guha P, Mukerjee A, Venkatesh KS (2011) OCS-14: you can get occluded in fourteen ways. In: Proceedings of 22nd International Joint Conference on Artificial Intelligence, pp 1665–1675
7. Guha P, Mukerjee A, Venkatesh KS (2006) Appearance based multiple agent tracking under complex occlusions. PRICAI 2006: Trends in Artificial Intelligence LNCS, vol 4099. Springer, Heidelberg, pp 593–602
8. Galton A (1998) Modes of overlap. J Vis Lang Comput 9(1):61–79
9. Kohler C (2002) The occlusion calculus. In: Workshop on Cognitive Vision, Zurich, Switzerland
10. Cao H, Mamoulis N, Cheung DW (2005) Mining frequent spatio temporal sequence patterns. In: 5th IEEE International Conference on Data Mining, pp 82–89