Joint Tracking and Event Analysis for Carried Object Detection

Aryana Tavanai
fy06at@leeds.ac.uk
Muralikrishna Sridhar
scms@leeds.ac.uk
Eris Chinellato
e.chinellato@leeds.ac.uk
Anthony G. Cohn
a.g.cohn@leeds.ac.uk
David C. Hogg
d.c.hogg@leeds.ac.uk

School of Computing
University of Leeds
Leeds, UK

Motivation This paper investigates whether reasoning about events that objects may participate in can facilitate tracking. For example, in domains where objects are frequently carried, dropped and exchanged, the trajectories of objects and the people carrying them have prototypical spatio-temporal relationships. In this work, we exploit this relationship introduced by events in order to improve the outcome of object tracking. We particularly focus on event-based tracking for domains which tend to contain generic objects, for which it is not straightforward to train class specific object detectors. The task of object tracking becomes especially challenging in such domains, as false and missing detections are highly prevalent [4]. This leads to false tracks, and also tracks that are heavily fragmented. We observe this phenomenon when we apply state-of-the-art trackers [1, 3, 4]. Furthermore, the prevalence of false and fragmented tracks makes the goal of incorporating events particularly challenging as event based tracking is a circular problem. This problem arises as it involves inferring events using reasonable tracks, and then using the events to subsequently improve the tracks.

Approach Given a set of tracklets, the aim is to exploit the spatio-temporal structure of tracks induced by events in order to improve object tracking. More specifically, we exploit learned domain-specific temporal transitions between events in order to target true positive tracklets, which we then use to form meaningful whole tracks. Our approach is illustrated in Fig. 1 and is performed iteratively in an optimisation similar to [2]. Our framework is not constrained to using any particular tracker to build the set of tracklets and, in principle, can be applied to any tracker. To illustrate this point, we have applied our framework to three state-of-the-art trackers [1, 3, 4].

Evaluation We have evaluated our approach in terms of both tracking and event recognition which is performed on a newly created MINDSEYE2015 dataset (Fig. 2). This dataset contains a large number of events representing the changing relation between objects and people, captured from three different viewpoints. We have defined seven events (Carry, Static, Pickup, Putdown, Drop, Raise, Roll) to allow for a full description of the scene with regards to the state of the carried object, from the start of its appearance to its disappearance. Ground truth for person tracks, carried object tracks and events are fully annotated. This dataset has been made publicly available with all ground truth annotations, carried object detections, tracklets and final tracks at http://doi.org/10.5518/9. Relevant code including our carried object detector [4] can be found at: http://www.engineering.leeds.ac.uk/joint-tracking-and-event-analysis. We are currently extending our approach to include multi-person, multi-object events such as giving, exchanging or replacing objects.

[1] A. Andriyenko, K. Schindler, and S. Roth. Discrete-continuous optimization for multi-target tracking. In *CVPR*, pages 1926–1933, 2012.

[2] S. Oh, S. Russell, and S. Sastry. Markov chain monte carlo data association for multi-target tracking. *Transactions on Automatic Control,*, 54(3):481–497, March 2009.

[3] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In *CVPR*, pages 1201–1208, 2011.

[4] A. Tavanai, M. Sridhar, F. Gu, A. G. Cohn, and D. C. Hogg. Carried object detection and tracking using geometric shape models and spatio-temporal consistency. In *Computer Vision Systems*, volume 7963 of *LNCS*, pages 223–233. Springer, 2013.

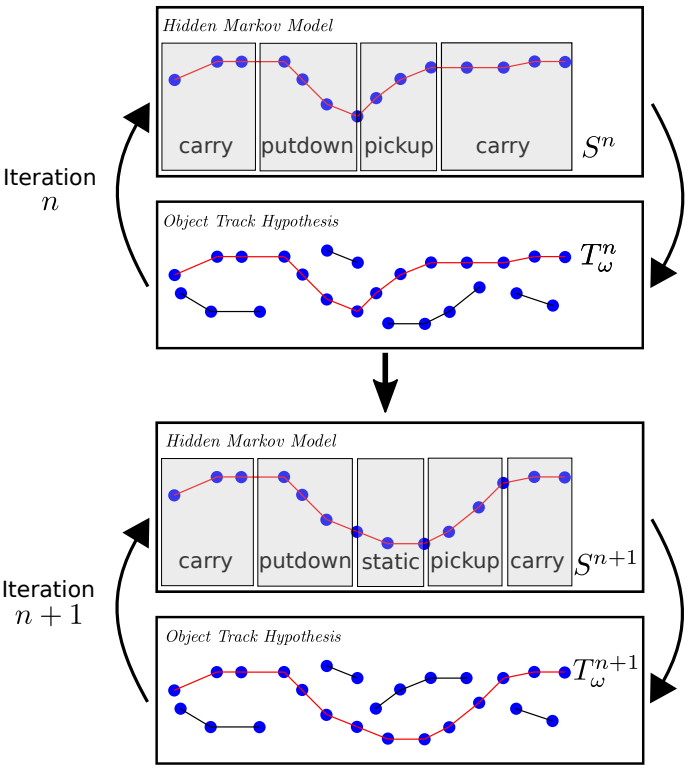

Figure 1: We illustrate two consecutive iterations within the Joint Tracking and Event Analysis optimisation. Given a set of tracklets $\mathcal{T}$ (on the left), at each iteration, a temporally-disjoint subset ω is selected and a contiguous track T_ω is produced by linearly interpolating across any gaps. The Viterbi path S^* of event labels in the HMM is inferred from T_ω (arrow up), leading to an HMM measure (arrow down) and combined with the spatio-temporal factors to give an overall probability. In the next iteration, a change to the subset ω is made and the overall probability re-computed. In this case, the new configuration is accepted since the probability is increased.

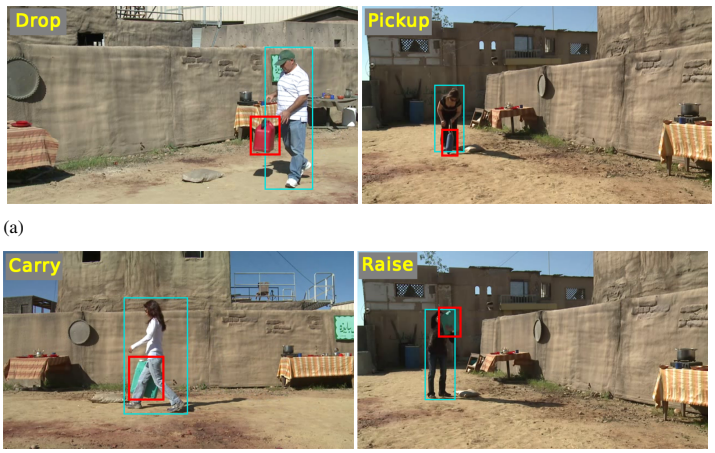

Figure 2: Sample images from the MINDSEYE2015 dataset illustrating various viewpoints, people (cyan), objects (red) and events (yellow text).