Efficient Second Order Multi-Target Tracking with Exclusion Constraints

Chris Russell †[1]
http://www.eecs.qmul.ac.uk/~chrisr/
Francesco Setti †[2]
settif@eecs.qmul.ac.uk
Lourdes Agapito[1]
http://www.eecs.qmul.ac.uk/~lourdes/

[1] School of Electronic Engineering and Computer Science
Queen Mary University of London
[2] University of Trento

Current state of the art multi-target tracking (MTT) exists in an "either/or" situation. Either a greedy approach can be used, that can make use of second-order information which captures object dynamics, such as "objects tend to move in the same direction over adjacent frames", or one can use global approaches that make use of the information contained in the entire sequence to resolve ambiguous sub-sequences, but are unable to use such second order information. However, the accurate resolution of ambiguous sequences requires both a good model of object dynamics, and global inference.

In this work we present a novel approach to MTT that combines the best of both worlds. By formulating the problem of tracking as one of global MAP estimation over a directed acyclic hyper-graph, we are able to both capture long range interactions, and informative second order priors. In practice, our algorithm is extremely effective, with a run time linear in the number of objects to be tracked, possible locations of an object, and the number of frames. We demonstrate the effectiveness of our approach, both on standard MTT data-sets that contain few objects to be tracked, and on point tracking for non-rigid structure from motion, which, with hundreds of points to be tracked simultaneously, strongly benefits from the efficiency of our approach.

The cost function that we optimise in tracking is substantially more informative than those used in existing efficient frameworks. Alongside the more usual cues that objects tend to appear in similar locations in adjacent frames, exclusion constraints, and the previously mentioned velocity cues, it also captures the persistent and temporary appearance of objects. This cost function takes the form:

$$\min_{\mathbf{x}\in L^{N\cdot F}} C(\mathbf{x}) = \sum_{o\leq N}\Big[\sum_{t\leq F} U_{o,t}(x_{o,t}) + \sum_{t\leq F-1} P_{o,t}(x_{o,t},x_{o,t+1}) + \sum_{t\leq F-2} T_{o,t}(x_{o,t},x_{o,t+1},x_{o,t+2})\Big] \quad (1)$$

$$\text{such that } \sum_{o\leq N}\Delta(x_{o,t}=l)\leq 1 \quad \forall t,\ \forall l\in L,\ l\notin O, \quad (2)$$

where (2) are the exclusion constraints, $U(\cdot)$ captures persistent object appearance, $P(\cdot,\cdot)$ captures location and temporary appearance cues, and $T(\cdot,\cdot,\cdot)$ captures velocity cues.

The key insight underlying our approach is that *belief propagation* can be used to solve cost functions containing these higher-order constraints. The higher-order exclusion constraint that at most one object may appear in any one location, at any one time can be decomposed into pairwise constraints over a directed acyclic graph, while the soft constraint that velocities should vary smoothly, can be solved using second-order belief propagation [2].

On top of this framework we propose efficient message updating strategies which makes our method the first global approach guaranteed to run in linear time.

Our approach is the first global method to make use of second order cues which describe the acceleration of points and our experimental results convincingly demonstrate their importance. Compared to other efficient algorithms [1] our approach is over a hundred times faster, and exhibits better worse case performance. This improved efficiency allows us to track hundreds of points easily, and to estimate the location of basketball players at almost 500 frames per second. Consequently, we expect this work to be of strong interest to both the non-rigid structure from motion community, and those working in surveillance, or real time multi-target tracking. Code is available for download from
`http://www.eecs.qmul.ac.uk/~chrisr/tracking.tar.gz`.

†CR and FS assert joint first authorship. This work was funded by the European Research Council under the ERC Starting Grant agreement 204871-HUMANIS. FS was supported by the European Commission and Provincia Autonoma di Trento under Marie Curie Action - Cofund project ABILE.

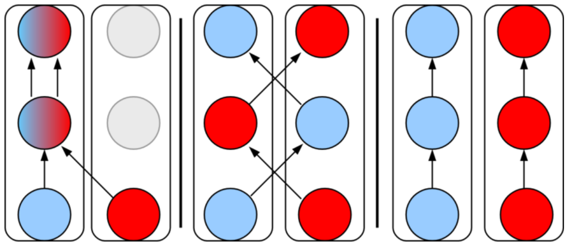

Figure 1: Diagram showing common failures of MTT, as two tracks closely pass by each other. The boxes show the true tracks of objects, while the colour of the balls and the arrows indicate their object labelling. **Left:** Without the use of exclusion constraints (forcing only one object to occur in a particular location at any time) object tracks may merge. **Centre:** Even with the use of exclusion constraints, if the distance between two object tracks is smaller than the distance moved by a single object in two frames flickering may occur. **Right** The use of second order terms to penalise objects sharply changing direction eliminates flicker.

Basketball	MOTP	MOTA
K-shortest path [1]	1.09	0.586
Track-before [3]	7.1	0.614
Pairwise only	0.6879	-1.064
Pairwise and unary	0.6021	-1.070
Pairwise and occlusion	0.8227	0.718
Pairwise, unary, and occlusion	0.8214	0.725
Complete method	0.7198	0.735

Figure 2: Quantitative evaluation of our new potentials. For MOTA higher is better, and 1 is optimal. For MOTP lower is better, and 0 is optimal.

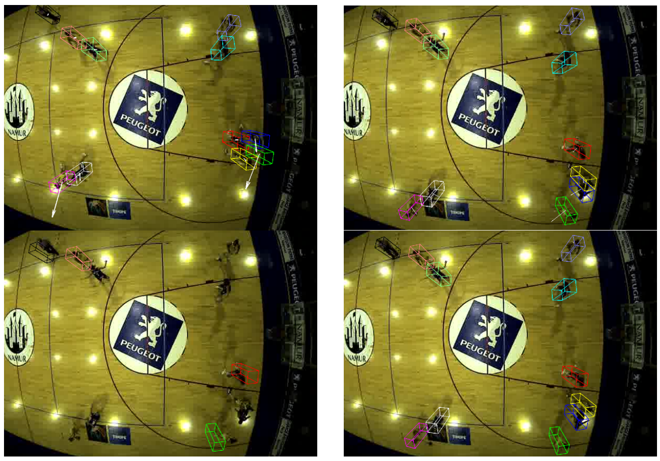

Figure 3: Qualitative evaluation of our new potentials **Top:** Velocity cues prevent the confusion of four players as they move (bottom right of image). **Bottom** These same second-order cues help recover from temporary system failure. See Figure 3 of main paper for comparison with other method and discussion.

[1] Jerome Berclaz, Francois Fleuret, Engin Turetken, and Pascal Fua. Multiple object tracking using k-shortest paths optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[2] Amlan Kundu, Yang He, and Paramvir Bahl. Recognition of handwritten word: first and second order hidden markov model based approach. *Pattern Recogn.*, 22:283–297, February 1989.

[3] M. Taj and A. Cavallaro. Multi-camera track-before-detect. In *Proc. of ACM/IEEE Int. Conf. on Distributed Smart Cameras*, 2009.