Collaborative Multi-Camera Tracking of Athletes in Team Sports

Wei Du, Jean-Bernard Hayet, Justus Piater, Jacques Verly

Department of Electrical Engineering and Computer Science
Montefiore Institute, University of Liège, Belgium

CVBASE 2006
Player tracking is a basic task in sports video analysis.

Single-camera tracking: noise-corrupted observations
- occlusion
- motion blur, appearance change, ...
Multi-camera tracking: we are able to

- deal with **occlusions**

- reduce **uncertainties** by fusing multi-camera information
Multiple cameras are often available in team-sports scenarios

A commonly-used framework

- local trackers in each camera
- a fusion module for integrating local information
Inconsistent Observations

Problem: cameras do not agree with each other!

Solution: assess tracking performance by
- computing the quality of the match
- estimating the variance of the target distribution

Question:
- does a high matching score mean a good performance?
Approximation of Similar Objects

In sports video analysis, the approximation of similar objects renders the assessment of tracking performance infeasible.

Wrong estimates can have a high matching score!
What is wrong with this fusion framework?

- The correlation between multi-camera data is not modeled
- No collaboration between local trackers

Additional information in other cameras is not taken into account!
Overview of Our Approach

Cameras talk to each other
- good observations compensate for poor observations
- no wrong information propagation
- no performance assessment required

The features of our approach
- Particle-Filter based local trackers
- Graphical Models for sharing geometric data
- Belief Propagation for collaboration
recursive inference of target state

\[ P(x_t|z_{1:t}) \propto P(z_t|x_t) \int P(x_t|x_{t-1})P(x_{t-1}|z_{1:t-1}) \, dx_{t-1} \]

- target state: position, size, and velocity
- temporal prior: constant-velocity motion model
- Likelihood: color-based observation model
Assume $L$ cameras are used

- $x_{t,0}$, target state on the ground plane at time $t$
- $x_{t,j}$, target state in camera $j$ at time $t$

$x_{t,j}, j = 1, \ldots, L$, are independent given $x_{t,0}$

undirected link between $x_{t,0}$ and $x_{t,j}$:

potential function $\psi_{0,j}(x_{t,0}, x_{t,j})$
BP can have exact inference on the tree-structured model

The tree-structured model is flexible
Belief Propagation in Tree-Structured graphical models

BP solves inference by passing messages.

belief of the central node:

$$p(x_{t,0}|z_{t,1:L}) \leftarrow \int p(x_{t,0}) \prod_{k} m_{0k}(x_{t,0}) \psi_{0,j}(x_{t,0}, x_{t,j}) \, dx_{t,0}$$

belief of a leaf node:

$$p(x_{t,j}|z_{t,1:L}) \leftarrow \int p(z_{t,j}|x_{t,j}) p(x_{t,j}) m_{j0}(x_{t,j}) \psi_{0,j}(x_{t,0}, x_{t,j}) \, dx_{t,0}$$
Belief Propagation in Tree-Structured graphical models

BP solves inference by passing messages.

message from a leaf node to central node:

\[ m_{0j}(x_{t,0}) \leftarrow \int p(z_{t,j} | x_{t,j}) p(x_{t,j}) \psi_{0,j}(x_{t,0}, x_{t,j}) \, d\, x_{t,j} \]

message from central node to a leaf node:

\[ m_{j0}(x_{t,j}) \leftarrow \int p(x_{t,0}) \prod_{k \neq j} m_{0k}(x_{0}) \psi_{0,j}(x_{t,0}, x_{t,j}) \, d\, x_{t,0} \]
Potential function $\psi_{i,j}^t(x_{t,i}, x_{t,j})$ models the spatial correlation between $x_{t,i}$ and $x_{t,j}$.

Assume that targets move on a calibrated ground plane.

Image-to-ground homography is used to compute the correlation.
message passing is asymmetric: desirable for dealing with inconsistent observations

- occlusion in camera 1: distribution with large variance
- no occlusion in camera 2: distribution with small variance
- after message passing, the variance in camera 1 becomes small, while the variance in cameras 2 won’t grow too much. Energy decreases!
Connecting tree-structured models at different time

Thus, multi-camera tracking is to infer

\[ p(x_{t,j}|z_{1:t}), \quad z_{1:t} = z_{1,1:t}, \ldots, z_{L,1:t}, \quad j = 0, \ldots, L \]

We use **Sequential Belief Propagation** to solve the inference problem
Hua et.al., "Multi-scale Visual Tracking by SBP", CVPR2004

Particle Filters + BP = SBP

- A particle filter infers $x_{t,j}$ at node $j$, $j = 0, \ldots, L$
- BP: passes messages to have the particle filters collaborate
Sequential Belief Propagation

belief of central node:

\[ p(x_{t,0}|Z_{1:t}) \propto \prod_k m_{0k}(x_{t,0}) \int_{x_{t-1,0}} p(x_{t,0}|x_{t-1,0})p(x_{t-1,0}|Z_{1:t-1}) \]

belief of a leaf node:

\[ p(x_{t,j}|Z_{1:t}) \propto p(z_{t,j}|x_{t,j})m_{j0}(x_{t,j}) \int_{x_{t-1,j}} p(x_{t,j}|x_{t-1,j})p(x_{t-1,j}|Z_{1:t-1}) \]
message from central node to a leaf node:

\[ m_{j0}(x_{t,j}) \leftarrow \int_{x_{t,j}} \psi^t_{0,j}(x_{t,0}, x_{t,j}) \prod_{k \neq j} m_{0k}(x_{t,0}) \int_{x_{t-1,j}} p(x_{t,j}|x_{t-1,j})p(x_{t-1,j}|Z_{1:t-1}) \]

message from a leaf node to central node:

\[ m_{0j}(x_{t,0}) \leftarrow \int_{x_{t,j}} \psi^t_{0,j}(x_{t,0}, x_{t,j})p(z_t|i|x_{t,i}) \int_{x_{t-1,0}} p(x_{t,0}|x_{t-1,0})p(x_{t-1,0}|Z_{1:t-1}) \]
SBP – message from a leaf node to central node:

\[ m_{0j}(x_{t,0}) \leftarrow \int x_{t,0} \psi_{0,j}^t(x_{t,0}, x_{t,j}) p(z_{t,i} | x_{t,i}) \int p(x_{t,0} | x_{t-1,0}) p(x_{t-1,0} | Z_{1:t-1}) \]

BP – message from a leaf node to central node:

\[ m_{0j}(x_{t,0}) \leftarrow \int p(z_{t,j} | x_{t,j}) p(x_{t,j}) \psi_{0,j} (x_{t,0}, x_{t,j}) \, dx_{t,j} \]

SBP – message from central node to a leaf node:

\[ m_{j0}(x_{t,j}) \leftarrow \int x_{t,j} \psi_{0,j}^t(x_{t,0}, x_{t,j}) \prod_{k \neq j} m_{0k}(x_{t,0}) \int p(x_{t,j} | x_{t-1,j}) p(x_{t-1,j} | Z_{1:t-1}) \]

BP – message from central node to a leaf node:

\[ m_{j0}(x_{t,j}) \leftarrow \int p(x_{t,0}) \prod_{k \neq j} m_{0k}(x_{0}) \psi_{0,j} (x_{t,0}, x_{t,j}) \, dx_{t,0} \]
Sequential Belief Propagation

- $p(x_{t,j} | Z_{1:t}) \sim \{s_{t,j}^n, \pi_{t,j}^n\}_{n=1}^{N}$
- $m_{ji} (x_{t,j}) \sim \{s_{t,j}^n, \omega_{t,j}^n\}_{n=1}^{N}$
- $\{s_{t,j}^n\}_{n=1}^{N}$ are particles sampled from $p(x_{t,j} | x_{t-1,j})$
Results

(a) Camera M (432)  (c) Camera B (432)

(e) Camera M (508)  (g) Camera B (508)

(i) Camera M (519)  (k) Camera B (519)

(a) Camera M (419)  (c) Camera M (432)

(e) Camera M (508)  (g) Camera M (519)
Conclusions and Future Work

A collaborative multi-camera tracking scheme

- rigorous formulation for integrating multi-camera information
- tree-structured graphical models that allows bidirectional message passing
- systematic treatment of inconsistent observations

Future work: multi-target, multi-camera tracking

- modeling target interactions