Video-Based People Tracking

P. Fua
EPFL IC-CVLab
Video-Based Tracking Challenges

- Thousands of frames
- Frequent occlusions
- Poor quality of input images
- Sudden illumination changes
Multi-Step Algorithm

Given cameras with overlapping fields of view and a discretized ground plane.

- Estimate ground occupancy probabilities in individual temporal time frames.
- Enforce temporal consistency under very weak assumptions.
- Assign identity and/or behavior.
Ground Occupancy

• **Input:** Binary images
• **Output:** Probability of Occupancy Map
Bayesian Formulation

Provide probabilistic estimates of

\[ P(X_1, \ldots, X_N \mid B_1, \ldots, B_C) \]

with

- \( X_i \) occupancy of location \( i \)
- \( B_j \) binary images from camera \( j \)

- For individual multi-view time frames.
- Given very noisy binary images.
- Using blob sizes to estimate distance.
- Consistent occlusion handling.
There are exactly three people at three different locations.

$A_c(...,1,....,1,....,1,...)$
Interpretation

Finds $P(X \mid B)$ such that the average synthetic image matches the binary images.

→ Solution of a fixed point problem
Four Cameras Looking at the Court
Enforcing Temporal Consistency

t=0
t=1
t=2
t=3
• People can transition from one location at one time to a neighboring one at the next instant.

• They can only enter and leave through virtual locations that correspond to exit or entrances.

• They are more likely to pass through high-probability locations.
Linear Program

Maximize \[ \sum_{t, i} \log \left( \frac{\rho^t_i}{1 - \rho^t_i} \right) \sum_{j \in \mathcal{N}(i)} f^t_{i,j} \]

subject to \[ \forall t, i, j, \quad f^t_{i,j} \geq 0 \]
\[ \forall t, i, \quad \sum_{j \in \mathcal{N}(i)} f^t_{i,j} \leq 1 \]
\[ \forall t, i, \quad \sum_{j \in \mathcal{N}(i)} f^t_{i,j} - \sum_{k : i \in \mathcal{N}(k)} f^{t-1}_{k,i} \leq 0 \]
\[ \sum_{j \in \mathcal{N}(v_{source})} f_{v_{source},j} - \sum_{k : v_{sink} \in \mathcal{N}(k)} f_{k,v_{sink}} \leq 0. \]

\[\rightarrow\] Can be solved in real-time using the K-Shortest Path Algorithm (KSP)!
Four Cameras Looking at the Court
Soccer

- Color-based appearance models.
- 2 goal keepers, 2 sets of 10 players each, 3 referees.
Handball
Monocular Result
Preserving Identity

- Read the numbers on the jersey whenever possible.
- In practice, not very often.
Linear Program

Appearance information used when available:

\[
\text{maximize } \sum_{t,i,l} \log \left( \frac{\rho_i(t) \varphi^l_i(t)}{1 - \rho_i(t)} \right) \sum_{j \in \mathcal{N}(i)} f^l_{i,j}(t)
\]

subject to \( \forall t, i, \sum_{j \in \mathcal{N}(i)} \sum_{l=1}^{L} f^l_{i,j}(t) \leq 1 \)

\[
\forall t, l, i, \sum_{j \in \mathcal{N}(i)} f^l_{i,j}(t) - \sum_{k : i \in \mathcal{N}(k)} f^l_{k,i}(t-1) \leq 0
\]

\[
\sum_{j \in \mathcal{N}(v_{\text{source}})} f^l_{v_{\text{source}},j} - \sum_{k : v_{\text{sink}} \in \mathcal{N}(k)} f^l_{k,v_{\text{sink}}} \leq 0
\]

\[
\forall t, l, \sum_{i=1}^{K} \sum_{j \in \mathcal{N}(i)} f^l_{i,j}(t) \leq N_l
\]

\[
\forall t, l, i, j, f^l_{i,j}(t) \geq 0
\]
Solving the LP in Real Time

- Run the KSP algorithm to select grid cells that are occupied.
- Segment the resulting trajectories into a set of tracklets that form an even more reduced graph.
- Solve the LP on the reduced graph.
Basketball

- Read the numbers on the jersey whenever possible.
- In practice, not very often.
Facial Identification
Facial Identification
Tracking the Ball

- Ambiguity
- Motion Blur
- Occlusion
Using Additional Knowledge

Model both the

- physics of the ball’s motion,
- interactions between ball and players.
Volleyball
Other Sports
Tech Transfer

- 2016: Player tracking deal announced by the NBA.

and it is SHALLOW!
Deep Background Subtraction

... and it could go Deep!
Conclusion

- Robust approach that can track arbitrary number of people over long periods of time.
- Does not require appearance information but can use it when available.
- Can handle the interaction between people and other moving objects.
- Real-time performance when using tracklets.

—> In use at the NBA and being extended to return 3D pose.
References and Code

• F. Fleuret, J. Berclaz, R. Lengagne and P. Fua, Multi-Camera People Tracking with a Probabilistic Occupancy Map, PAMI 2008.
• H. Ben Shitrit, J. Berclaz, F. Fleuret, and P. Fua, Multi-Commodity Network Flow for Tracking Multiple People, PAMI 2014.

• Code can be downloaded from