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

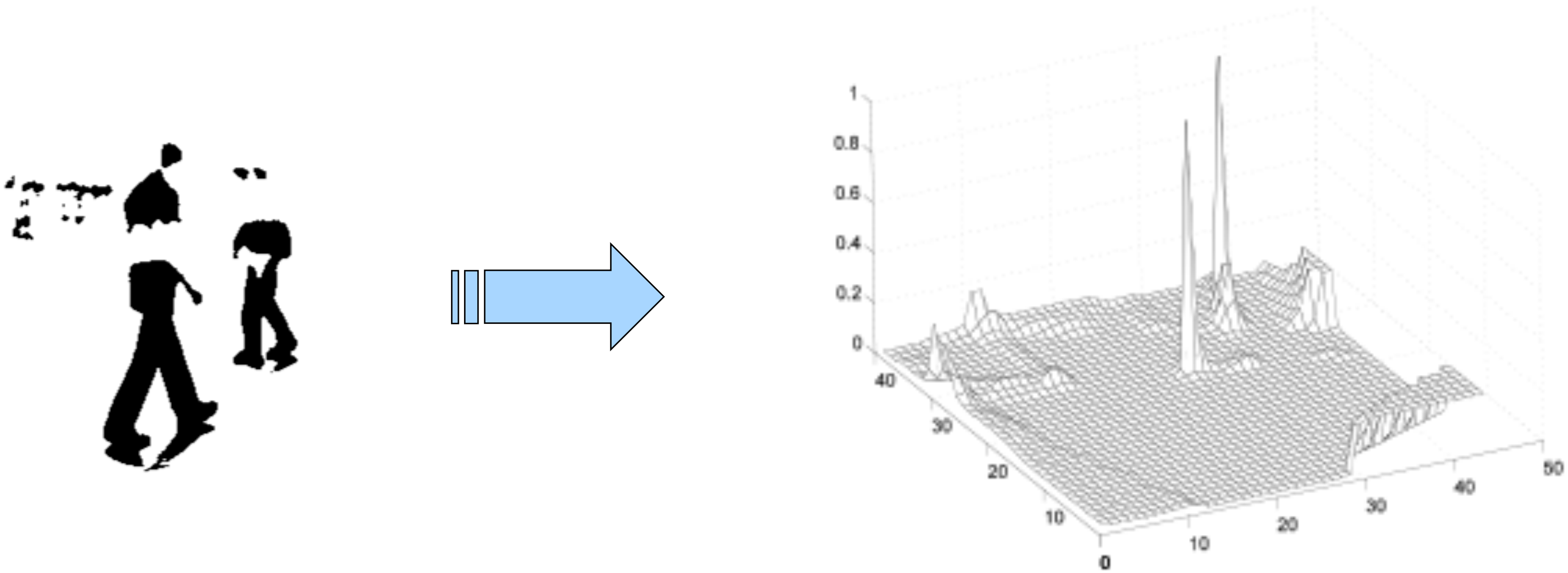
Multi-Step Algorithm

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

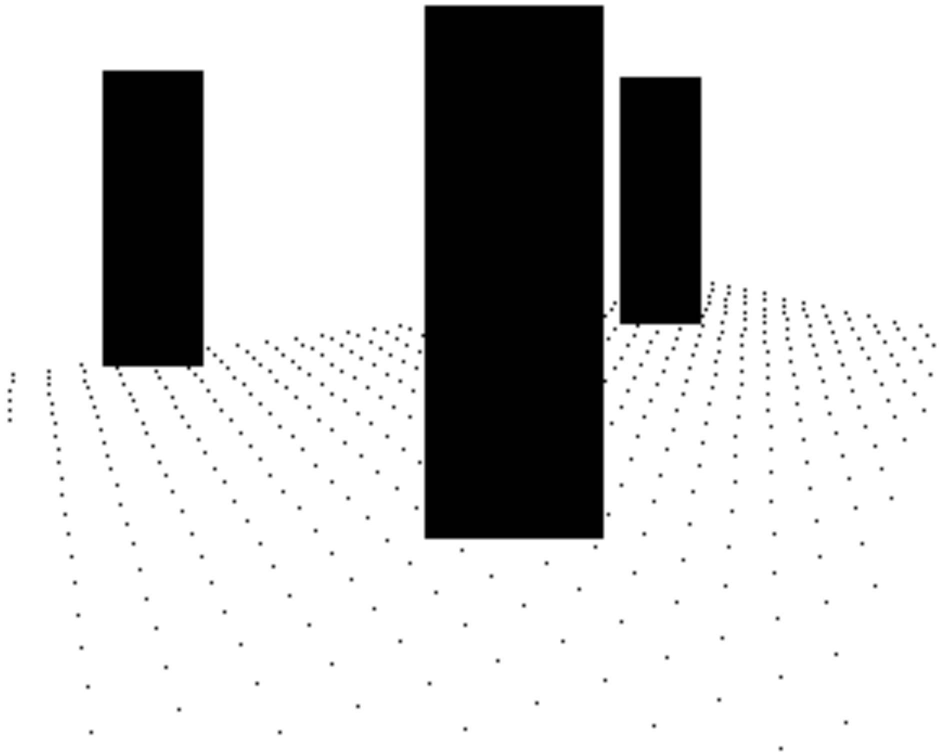
$$P\left(X_1,\ldots,X_N \mid B_1,\ldots,B_C\right) \text{ with } \begin{cases} X_i \text{ occupancy of location } i \\ B_j \text{ binary images from camera } j \end{cases}$$

- For individual multi-view time frames.
- Given very noisy binary images.
- Using blob sizes to estimate distance.
- Consistent occlusion handling.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

CvLab

Generative Model

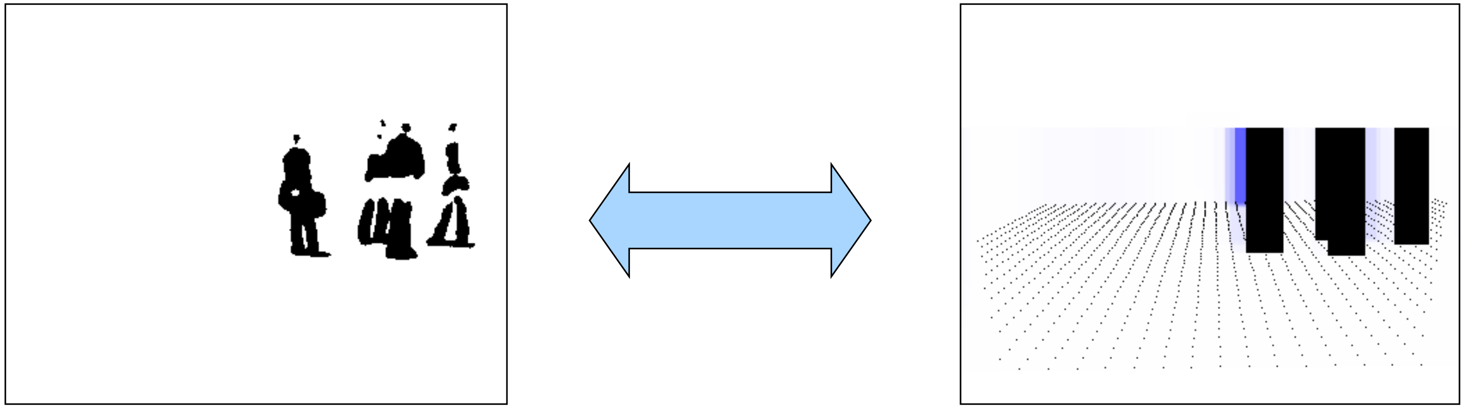


$$A_c(..,1,....,1,...,1,...)$$

There are exactly three people at three different locations.

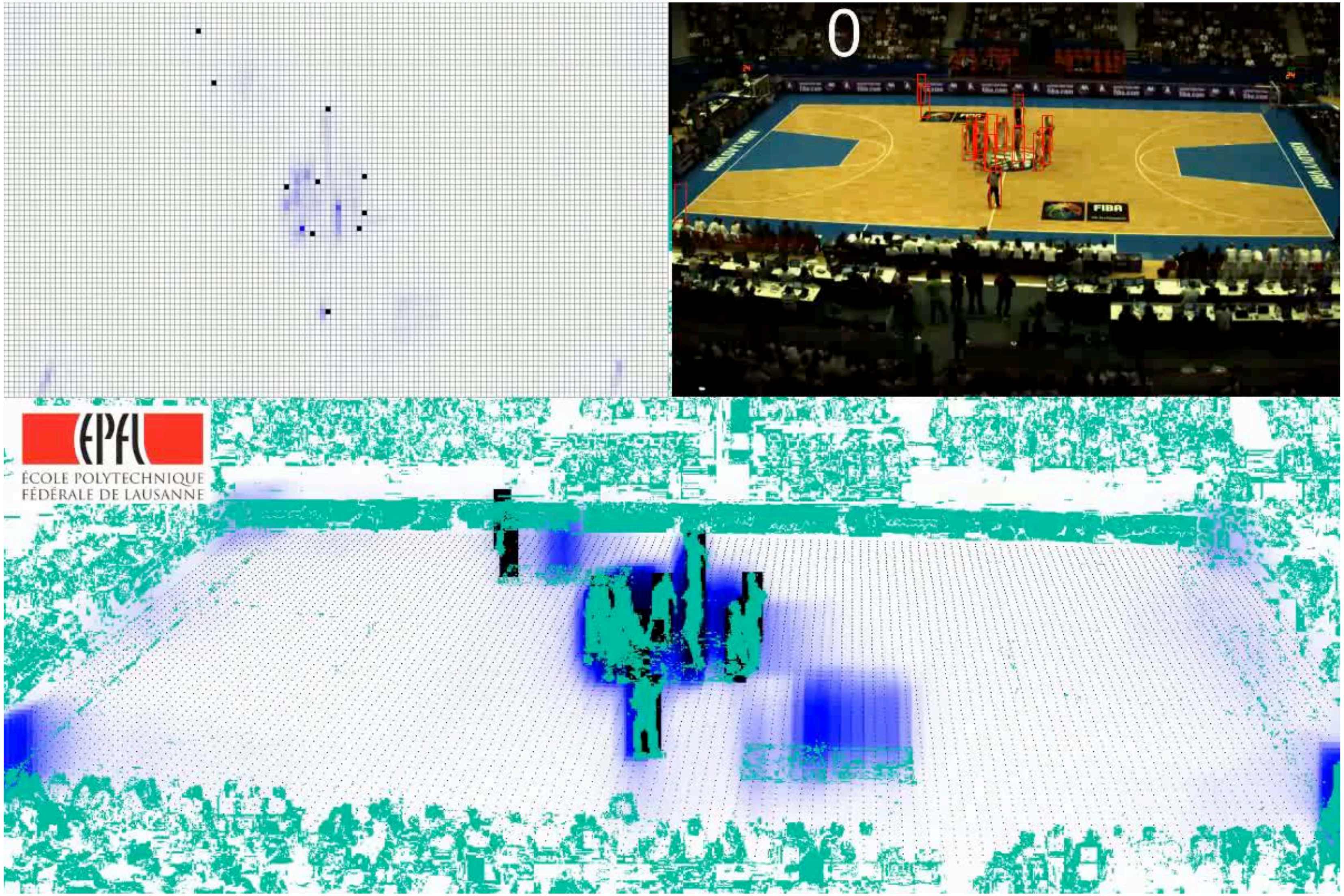
Interpretation

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

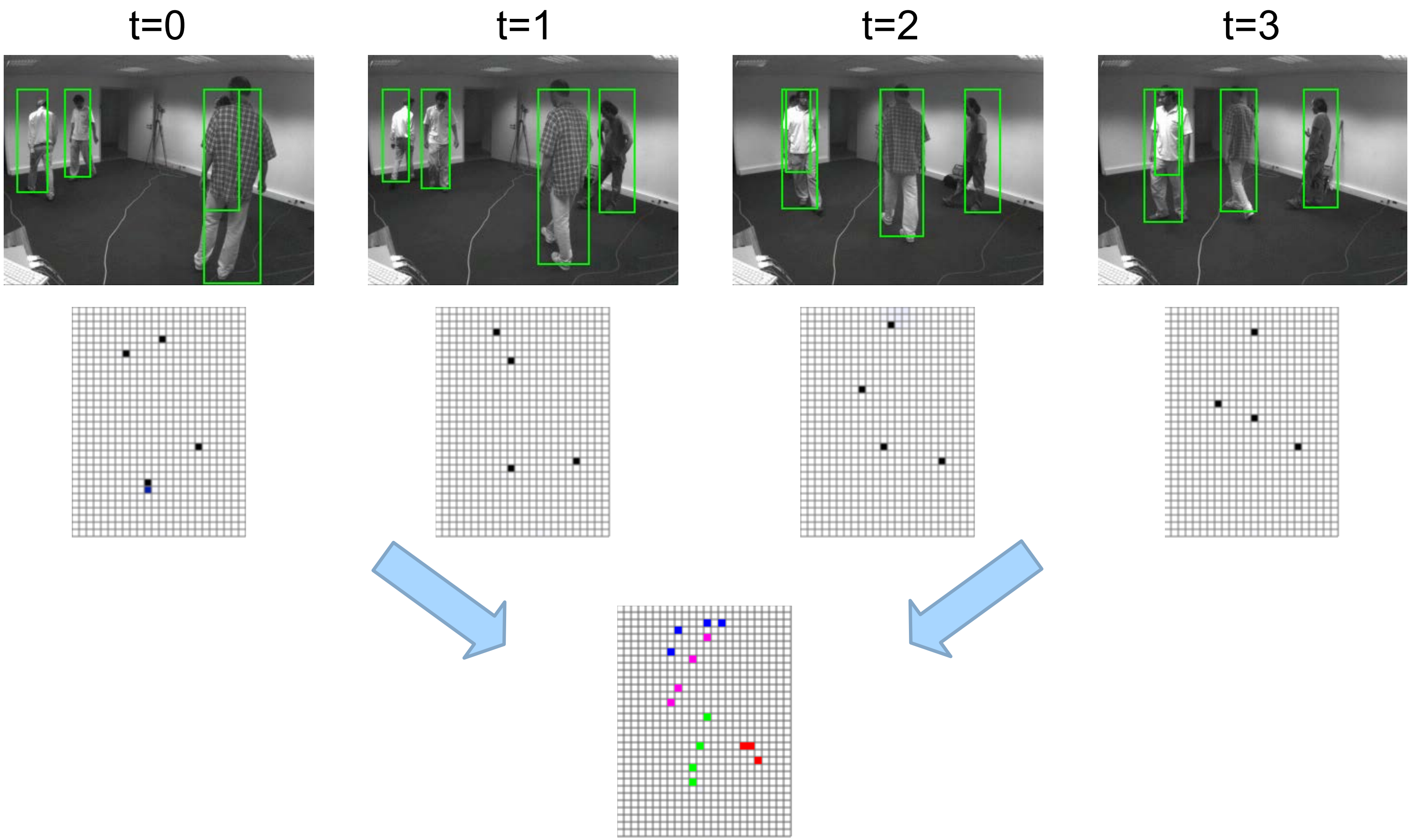


→ Solution of a fixed point problem

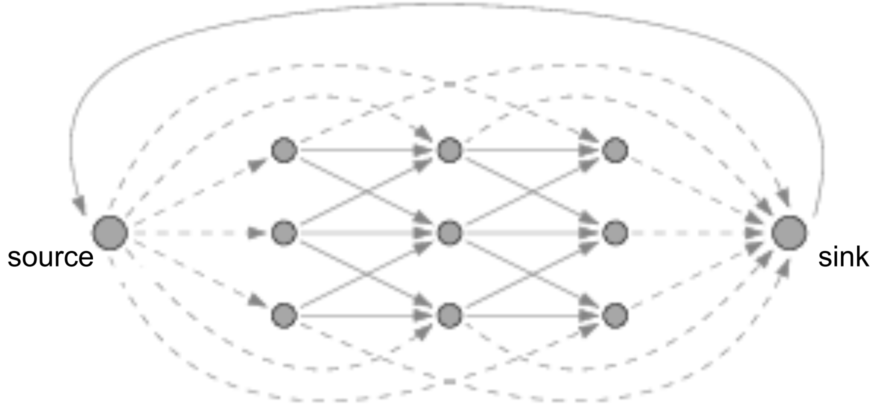
Four Cameras Looking at the Court



Enforcing Temporal Consistency



Graph-Based Formulation



- 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

$$\text{Maximize} \quad \sum_{t,i} \log\left(\frac{\rho_i^t}{1-\rho_i^t}\right) \sum_{j \in \mathcal{N}(i)} f_{i,j}^t$$

$$\text{subject to} \quad \forall t,i,j, \quad f_{i,j}^t \geq 0$$

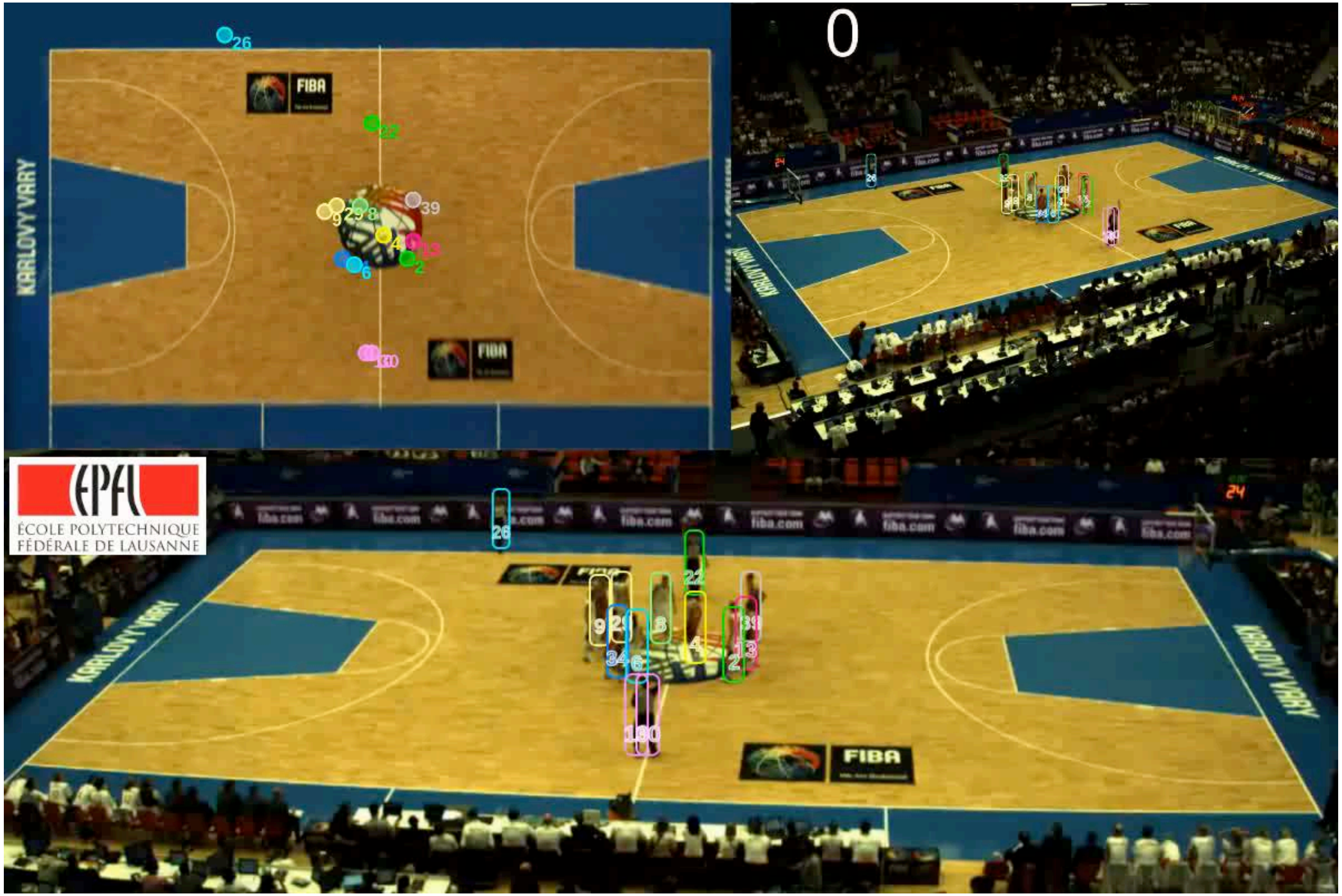
$$\forall t,i, \quad \sum_{j \in \mathcal{N}(i)} f_{i,j}^t \leq 1$$

$$\forall t,i, \quad \sum_{j \in \mathcal{N}(i)} f_{i,j}^t - \sum_{k: i \in \mathcal{N}(k)} f_{k,i}^{t-1} \leq 0$$

$$\sum_{j \in \mathcal{N}(\upsilon_{\text{source}})} f_{\upsilon_{\text{source}},j} - \sum_{k: \upsilon_{\text{sink}} \in \mathcal{N}(k)} f_{k,\upsilon_{\text{sink}}} \leq 0 \ .$$

—> 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.

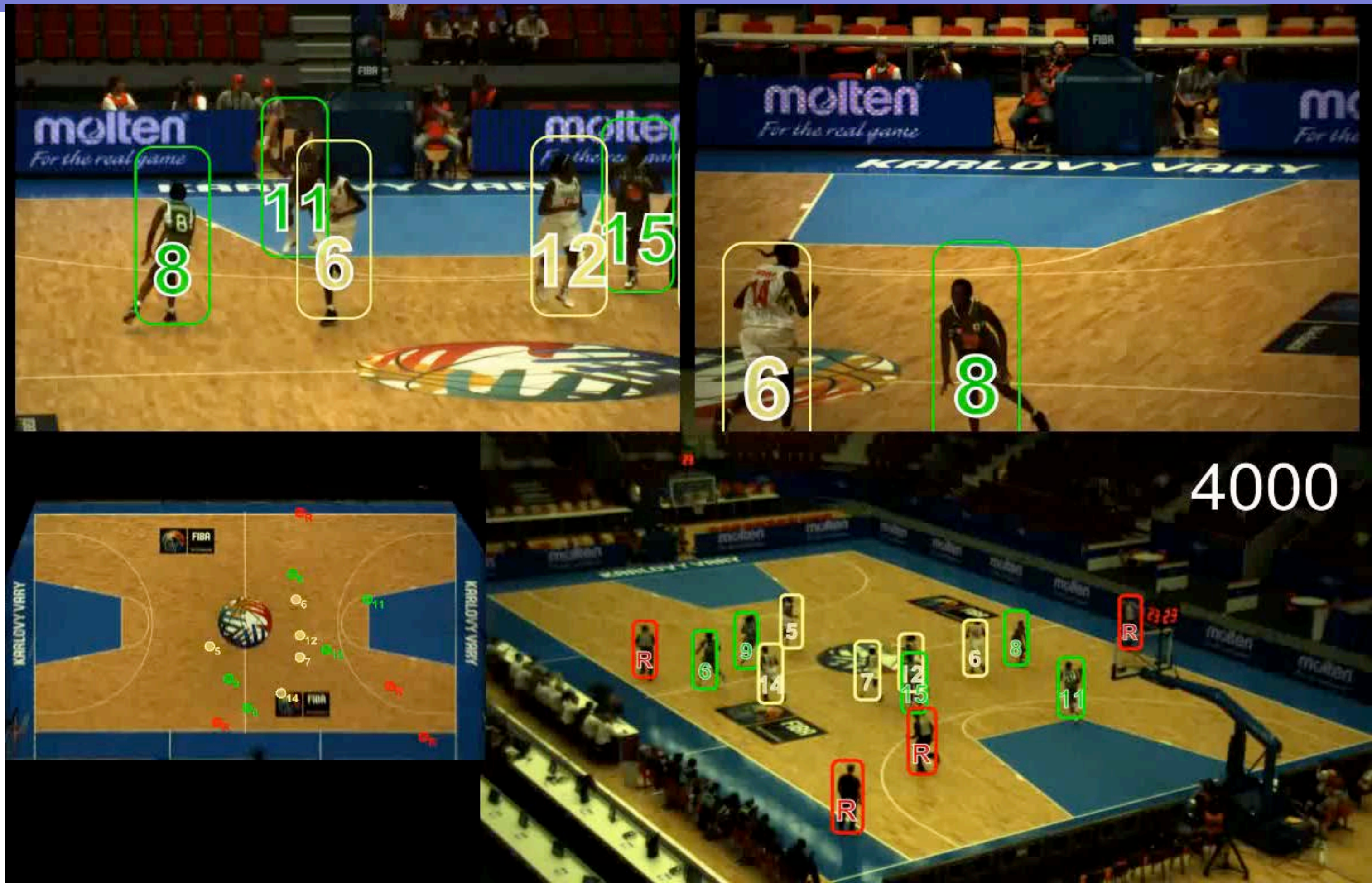
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Handball

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
22250
SPORTS PARTNER

Monocular Result

Preserving Identity



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

Linear Program

Appearance information used when available:

$$\text{maximize} \sum_{t,i,l} \log\left(\frac{\rho_i(t)\varphi_i^l(t)L}{1-\rho_i(t)}\right) \sum_{j\in\mathcal{N}(i)} f_{i,j}^l(t)$$

Probability of belonging to a specific group.

$$\text{subject to } \forall t, i, \quad \sum_{j\in\mathcal{N}(i)} \sum_{l=1}^{L} f_{i,j}^l(t) \leq 1$$

$$\forall t, l, i, \quad \sum_{j\in\mathcal{N}(i)} f_{i,j}^l(t) - \sum_{k:i\in\mathcal{N}(k)} f_{k,i}^l(t-1) \leq 0$$

$$\sum_{j\in\mathcal{N}(\upsilon_{\text{source}})} f_{\upsilon_{\text{source}},j} - \sum_{k:\upsilon_{\text{sink}}\in\mathcal{N}(k)} f_{k,\upsilon_{sink}} \leq 0$$

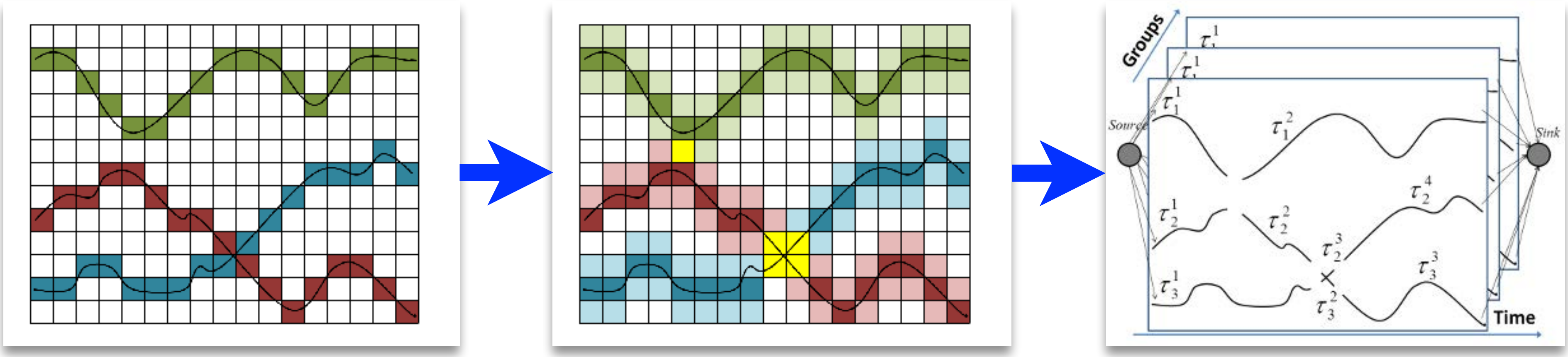
$$\forall t, l, \quad \sum_{i=1}^{K} \sum_{j\in\mathcal{N}(i)} f_{i,j}^l(t) \leq N_l$$

$$\forall t, l, i, j, \quad f_{i,j}^l(t) \geq 0 \ .$$

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

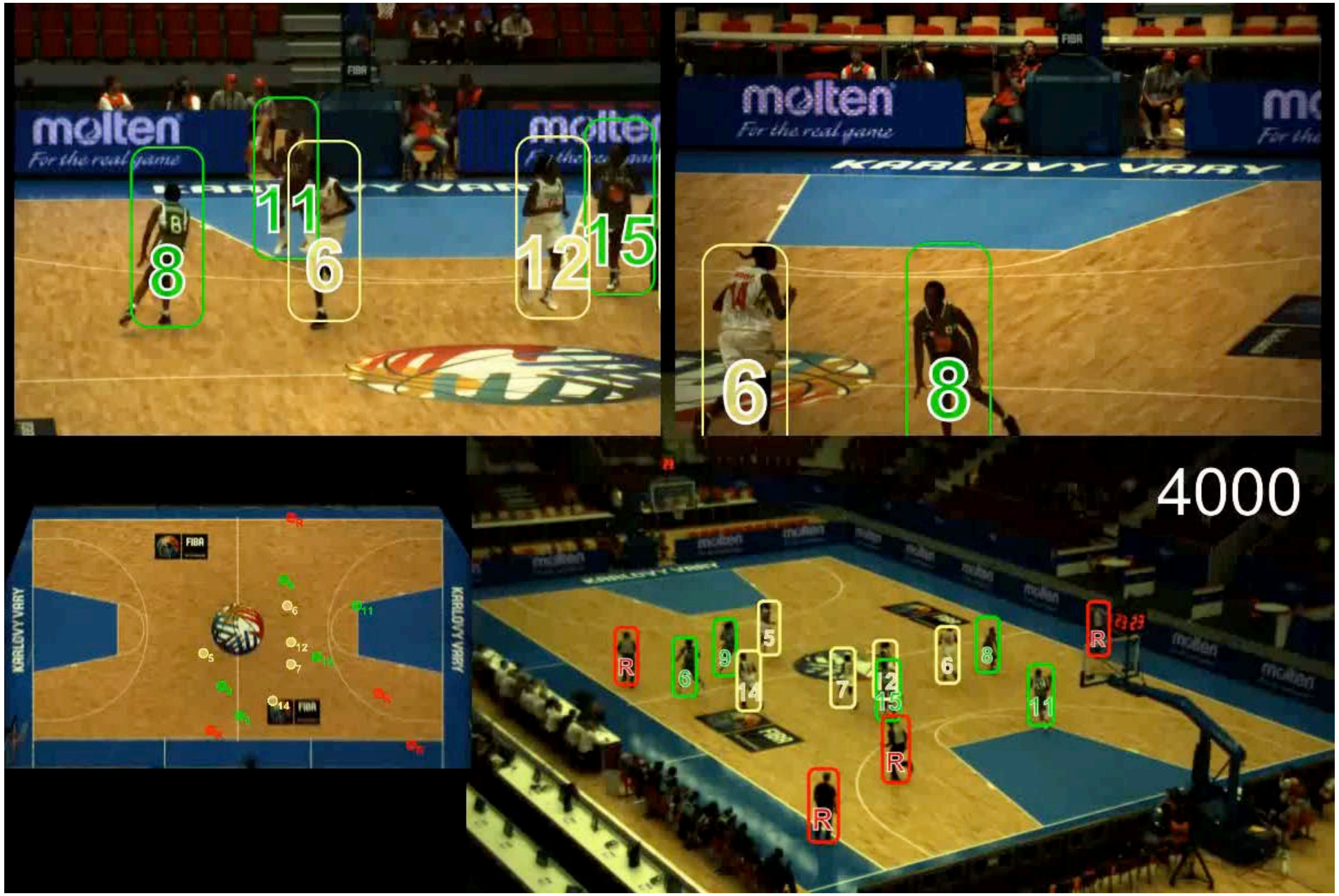
CVLab

Solving the LP in Real Time



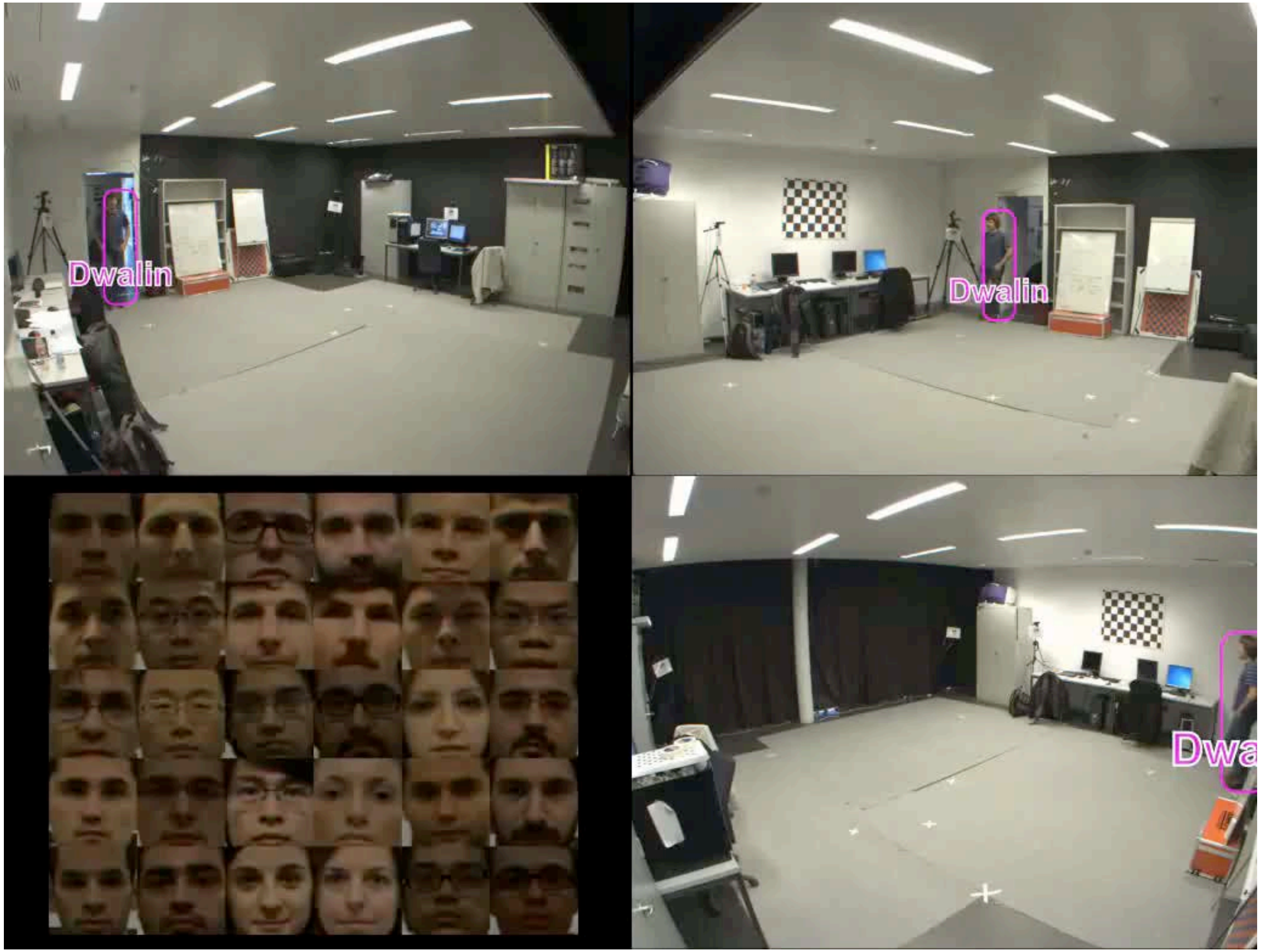
- Run the KSP algorithm to select grid cells that are occupied.
- Segment the resulting trajectories into a set of tracklets that form en even more reduced graph.
- Solve the LP on the reduced graph.

Basketball



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

Facial Identification



Facial Identification

Tracking the Ball

Ambiguity

Motion Blur

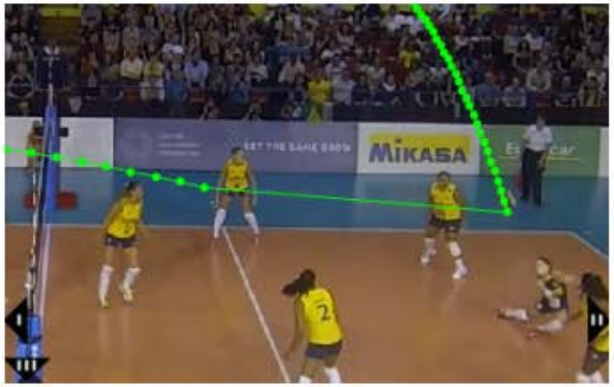
Occlusion

Using Additional Knowledge

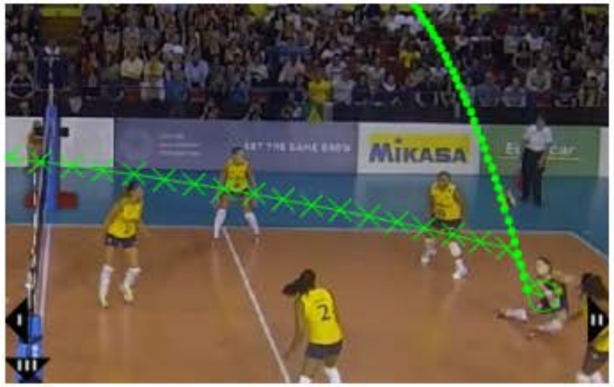
Model both the

- physics of the ball's motion,
- interactions between ball and players.

Interactions only



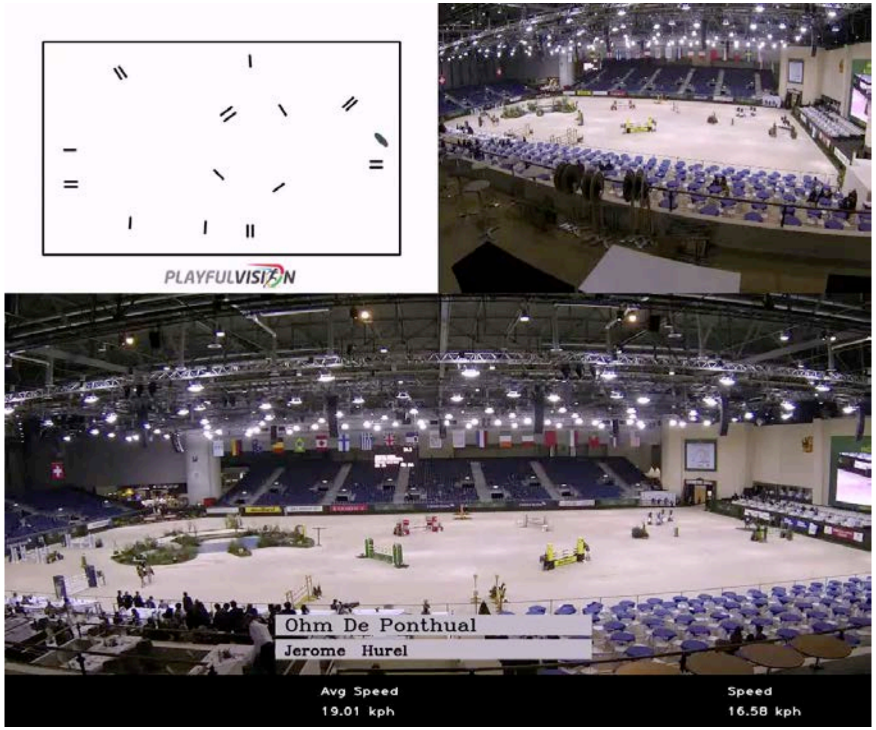
Ballistics only



Both

Volleyball

Other Sports



Tech Transfer

...and it is SHALLOW!

- 2014: PlayfulVision founded.
- 2015: Acquired by SecondSpectrum.
- 2016: Player tracking deal announced by the NBA.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

CVLab

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.
- J. Berclaz, F. Fleuret, E. Türetken and P. Fua, **Multiple Object Tracking using K-Shortest Paths**, PAMI 2011.
- H. Ben Shitrit, J. Berclaz, F. Fleuret, and P. Fua, **Multi-Commodity Network Flow for Tracking Multiple People**, PAMI 2014.
- X. Wang, E. Turetken, F. Fleuret, and P. Fua. **Tracking Interacting Objects Using Intertwined Flows,** *PAMI* 2016.

- Code can be downloaded from
  - http://cvlab.epfl.ch/software/pom/index.php
  - http://cvlab.epfl.ch/software/ksp/index.php