William Brendel

TRACKING

Motivations

- Sport videos: tracking players
- Video surveillance
- Digital camera: face tracking
- …

Tracking = Grouping

- Grouping of entities across time
- Local grouping Vs Global grouping

Window of few frames is considered for grouping
→ Fast
→ Online approach

Ex: particle filtering

Entire video is considered

→ Robust to occlusion and object ambiguities

Ex: Data Association based Tracking

Tracking = Grouping

- Grouping of entities across time

- Local grouping Vs Global grouping

Window of few frames is considered for grouping
→ Fast
→ Online approach

Ex: particle filtering

Entire video is considered

→ Robust to occlusion and object ambiguities

Ex: Data Association based Tracking

Tracking = Grouping

- Gestalt features for tracking:
  - Proximity
  - Similarity
  - Smooth motion

Challenges

- Occlusion
- Ambiguity between similar objects
- Crowded Scene with cluttered background
- Motion blur
- Fast motion

Challenges

- Occlusion
- Ambiguity between similar objects
- Crowded Scene with cluttered background
- Motion blur
- Fast motion

Papers

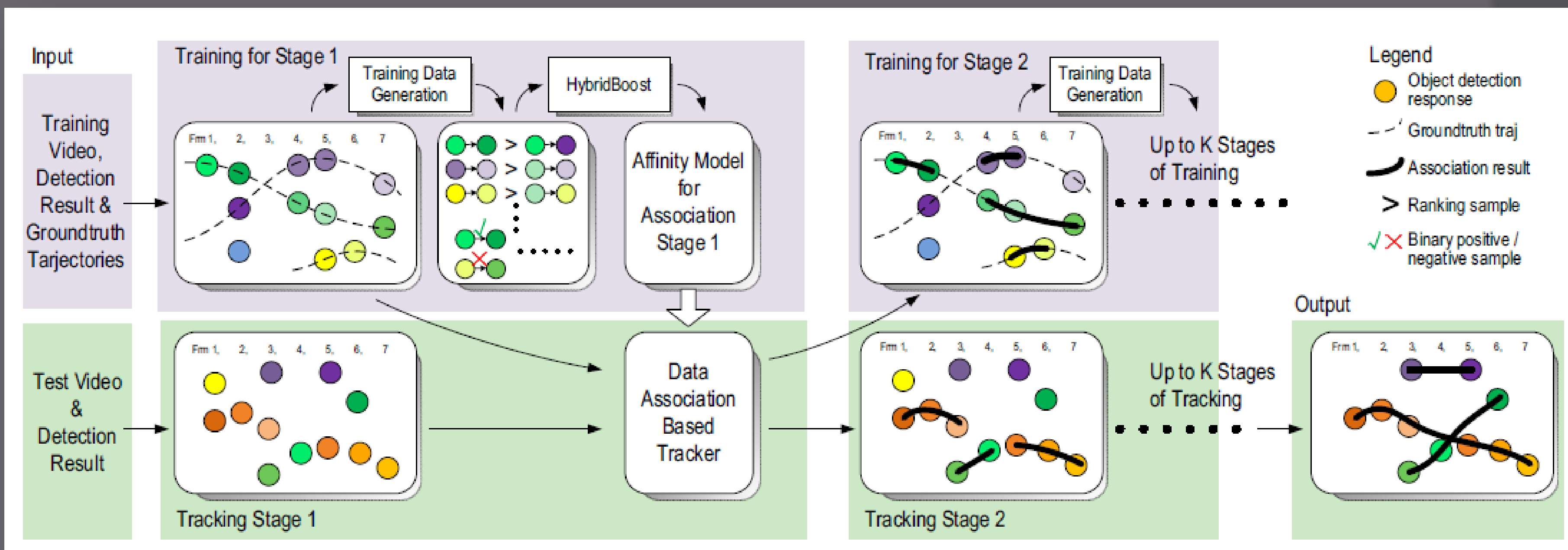
- "Learning to associate: HybridBoosted Multi-Target Tracker for Crowded Scene", Y. Li, C. Huang and R. Nevatia
- "Global Data Association for MultiObject Tracking Using Network Flows", L. Zhang, Y. Li and R. Nevatia

Learning to Associate: HybridBoosted Multi-Target Traker for Crowded Scene

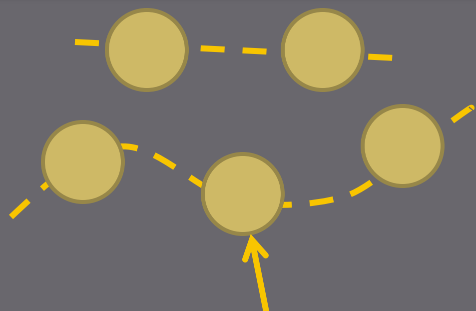
- Input:
  - Tracklets T_i
  - Pairwise tracklet properties
- Goal: group tracklets in longer tracks

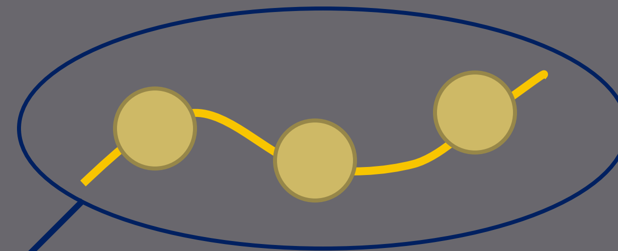
Overview

- N stages algorithm



MAP Formulation for Tracklet Association





$\mathcal{T}^{k-1} = \{T_i^{k-1}\}$

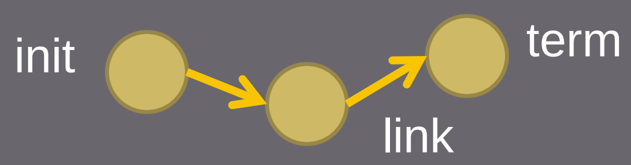
$\mathcal{T}^{k} = \{T_j^{k}\}$

$T_j^k = \{T_{i_0}^{k-1}, T_{i_1}^{k-1}, \cdots, T_{i_{l_j}}^{k-1}\}$

$$
\begin{aligned}
\mathcal{T}^{k*} &= \underset{\mathcal{T}^k}{\operatorname{argmax}}\, P(\mathcal{T}^{k-1} | \mathcal{T}^k) P(\mathcal{T}^k) \\
&= \underset{\mathcal{T}^k}{\operatorname{argmax}} \prod_{T_i^{k-1} \in \mathcal{T}^{k-1}} P(T_i^{k-1} | \mathcal{T}^k) \prod_{T_j^k \in \mathcal{T}^k} P(T_j^k)
\end{aligned}
$$

MAP Formulation for Tracklet Association

- Markov Chain formulation



$$\mathcal{T}^{k*} = \underset{\mathcal{T}^k}{\mathrm{argmax}} \prod_{T_i^{k-1}: \forall T_j^k \in \mathcal{T}^k, T_i^{k-1} \notin T_j^k} P_-(T_i^{k-1})$$

$$\prod_{T_j^k \in \mathcal{T}^k} \Big[P_{init}(T_{i_0}^{k-1}) P_+(T_{i_0}^{k-1}) P_{link}(T_{i_1}^{k-1} | T_{i_0}^{k-1}) \cdots$$

$$P_{link}(T_{i_{l_k}}^{k-1} | T_{i_{l_k-1}}^{k-1}) P_+(T_{l_k}^{k-1}) P_{term}(T_{i_{l_k}}^{k-1}) \Big], \qquad (2)$$

MAP Formulation for Tracklet Association

$$L_I(T_i^{k-1}) = \ln \frac{P_{init}(T_i^{k-1})P_+(T_i^{k-1})P_{term}(T_i^{k-1})}{P_-(T_i^{k-1})},$$

$$L_T(T_j^{k-1}|T_i^{k-1}) = \ln \frac{P_{link}(T_j^{k-1}|T_i^{k-1})}{P_{term}(T_i^{k-1})P_{init}(T_j^{k-1})}, \quad (3)$$

$$T^{k*} = \underset{T^k}{\mathrm{argmax}} \sum_{T_j^k \in T^k} \Big[L_I(T_{i_0}^{k-1}) + L_T(T_{i_1}^{k-1}|T_{i_0}^{k-1}) + L_I(T_{i_1}^{k-1}) \cdots + L_I(T_{i_{l_k}}^{k-1}) \Big]. \quad (4)$$

L_I = inner cost L_T = affinity measurement between 2 tracklets

Affinity Model Association

- Goal: estimate $L_T(T_j^{k-1}|T_i^{k-1})$

- Based on ranking and classification

→ HybridBoost algorithm

Affinity Model Association

- $x = \langle T_i, T_j \rangle$ = tracklet pair
- H(x) = ranking function

$$H(\langle T_1, T_2 \rangle) > H(\langle T_1, T_3 \rangle)$$

- H is also a binary classifier

$$H(\langle T, T' \rangle) < \zeta, \forall T' \in T$$

Affinity Model Association

- HybridBoost → minimize the hybrid loss Z:

$$Z = \beta \sum_{(x_{i,0}, x_{i,1}) \in \mathcal{R}} w_0(x_{i,0}, x_{i,1}) \exp\left(H(x_{i,0}) - H(x_{i,1})\right) + (1-\beta) \sum_{(x_j, y_j) \in \mathcal{B}} w_0(x_j, y_j) \exp\left(-y_j H(x_j)\right),$$

$$Z_t = \beta \sum_{(x_{i,0}, x_{i,1}) \in \mathcal{R}} w_t(x_{i,0}, x_{i,1}) \exp\left(\alpha_t (h_t(x_{i,0}) - h_t(x_{i,1}))\right) + (1-\beta) \sum_{(x_j, y_j) \in \mathcal{B}} w_t(x_j, y_j) \exp\left(-\alpha_t y_j h_t(x_j)\right), \quad (8)$$

h_t = best weak classifier for round t in the boosting process

Affinity Model Association

$f(x)$ = feature of tracklet pair

$$h(x) = \begin{cases} 1 & \text{if } f(x) > \eta, \\ -1 & \text{otherwise.} \end{cases} \qquad H(x) = \sum_{t=1}^{n} \alpha_t h_t(x)$$

$$L_T(T_2|T_1) = \begin{cases} H(x) & \text{if } H(x) > \zeta, \\ -\infty & \text{otherwise.} \end{cases}$$

Training

- Dataset generated for each stage k
- H_k learned for each stage k

Experiments - Results

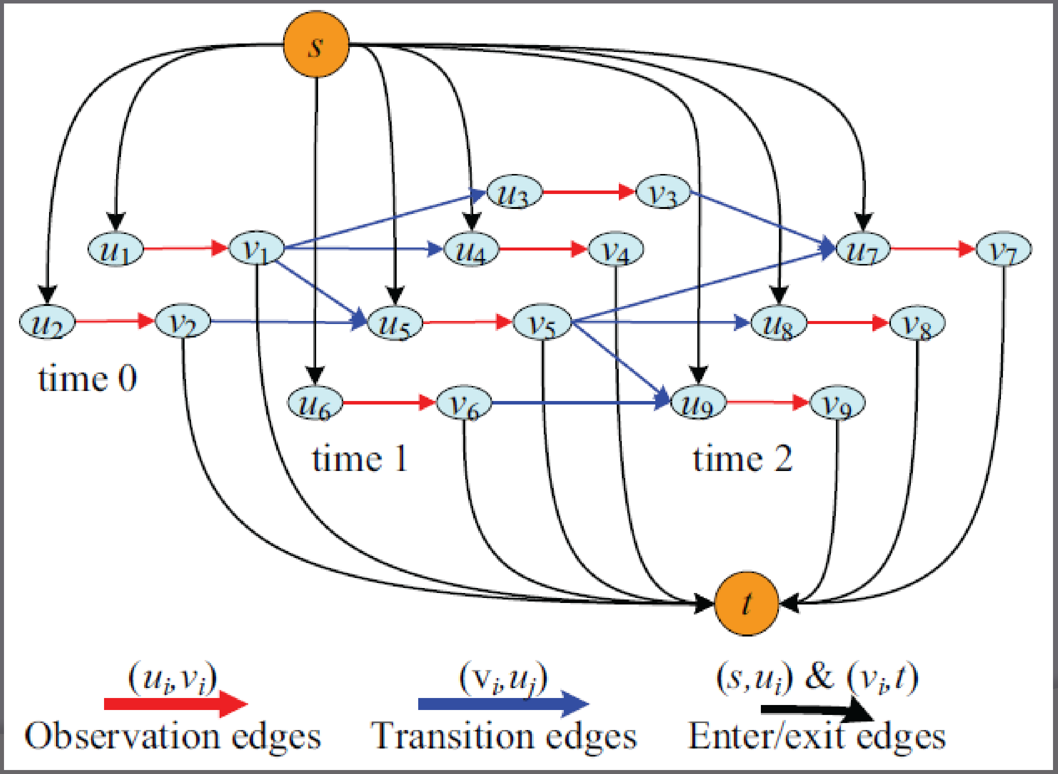
- Measure based on # track ID switches
- Hybrid approach ($0<\beta<1$) outperform state of the art

Global Data Association for Multi-Object Tracking Using Network Flows

- Input: object detections
- Goal: track objects with long-term inter-object occlusion

Overview

- Cost flow network with non-overlapping trajectories constraints



Detection input Tracking result

MAP under non-overlap constraint

$$\mathcal{X} = \{\mathrm{x}_i\} \qquad \mathrm{x}_i = (x_i, s_i, a_i, t_i) \qquad T_k = \{\mathrm{x}_{k_1}, \mathrm{x}_{k_2}, \ldots, \mathrm{x}_{k_{l_k}}\}$$

χ = set of objects

x_i = detected object

T_k = track = sequence of detected object through time

$$\mathcal{T} = \{T_k\}$$

$\mathcal{T}$ = set of tracks

$$\begin{aligned} \mathcal{T}^* &= \underset{\mathcal{T}}{\operatorname{argmax}}\, P(\mathcal{T}|\mathcal{X}) \\ &= \underset{\mathcal{T}}{\operatorname{argmax}}\, P(\mathcal{X}|\mathcal{T})P(\mathcal{T}) \\ &= \underset{\mathcal{T}}{\operatorname{argmax}} \prod_i P(\mathrm{x}_i|\mathcal{T})P(\mathcal{T}) \end{aligned}$$

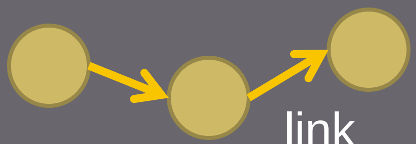
MAP under non-overlap constraint

- Non-overlap constraint: $\mathcal{T}_k \cap \mathcal{T}_l = \emptyset, \forall k \neq l$

$$\begin{aligned} \mathcal{T}^* &= \operatorname*{argmax}_{\mathcal{T}} \prod_i P(\mathrm{x}_i|\mathcal{T}) \prod_{\mathcal{T}_k \in \mathcal{T}} P(\mathcal{T}_k) \\ &\quad \text{s.t. } \mathcal{T}_k \cap \mathcal{T}_l = \emptyset, \forall k \neq l \end{aligned}$$

MAP under non-overlap constraint

- Markov Chain formulation



$$
\begin{aligned}
P(\mathbf{x}_i|\mathcal{T}) &= \begin{cases} 1-\beta_i & \exists \mathcal{T}_k \in \mathcal{T}, \mathbf{x}_i \in \mathcal{T}_k \\ \beta_i & \text{otherwise} \end{cases} && (4) \\
P(\mathcal{T}_k) &= P(\{\mathbf{x}_{k_0}, \mathbf{x}_{k_1}, \ldots, \mathbf{x}_{k_{l_k}}\}) \\
&= P_{entr}(\mathbf{x}_{k_0}) P_{link}(\mathbf{x}_{k_1}|\mathbf{x}_{k_0}) P_{link}(\mathbf{x}_{k_2}|\mathbf{x}_{k_1}) \\
&\quad \ldots P_{link}(\mathbf{x}_{k_{l_k}}|\mathbf{x}_{k_{l_k-1}}) P_{exit}(\mathbf{x}_{k_{l_k}}) && (5)
\end{aligned}
$$

Min-Cost flow Solution

- Need to map the MAP formulation into a flow graph

Min-Cost flow Solution

- Indicator function to incorporate non-overlap constraints

binary

$$f_{en,i} = \begin{cases} 1 & \exists \mathcal{T}_k \in \mathcal{T}, \mathcal{T}_k \text{ starts from } \mathbf{x}_i \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$f_{ex,i} = \begin{cases} 1 & \exists \mathcal{T}_k \in \mathcal{T}, \mathcal{T}_k \text{ ends at } \mathbf{x}_i \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$f_{i,j} = \begin{cases} 1 & \exists \mathcal{T}_k \in \mathcal{T}, \mathrm{x}_j \text{ is right after } x_i \text{ in } \mathcal{T}_k \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

unary

$$f_{i} = \begin{cases} 1 & \exists \mathcal{T}_k \in \mathcal{T}, \mathrm{x}_i \in \mathcal{T}_k \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Min-Cost flow Solution

$$
\begin{aligned}
T &= \arg\min_{\mathcal{T}} \sum_{\mathcal{T}_k \in \mathcal{T}} -\log P(\mathcal{T}_k) + \sum_i -\log P(\mathbf{x}_i|\mathcal{T}) \\
&= \arg\min_{\mathcal{T}} \sum_{\mathcal{T}_k \in \mathcal{T}} (C_{en,k_0} f_{en,k_0} \\
&\qquad + \sum_j C_{k_j,k_{j+1}} f_{k_j,k_{j+1}} + C_{ex,k_{l_k}} f_{ex,k_{l_k}}) \\
&\qquad + \sum_i (-\log(1-\beta_i) f_i - \log \beta_i (1-f_i)) \\
&= \arg\min_{\mathcal{T}} \sum_i C_{en,i} f_{en,i} + \sum_{i,j} C_{i,j} f_{i,j} \\
&\quad + \sum_i C_{ex,i} f_{ex,i} + \sum_i C_i f_i \qquad (11)
\end{aligned}
$$

Incorporate constraints and take the log of the MAP

Min-Cost flow Solution

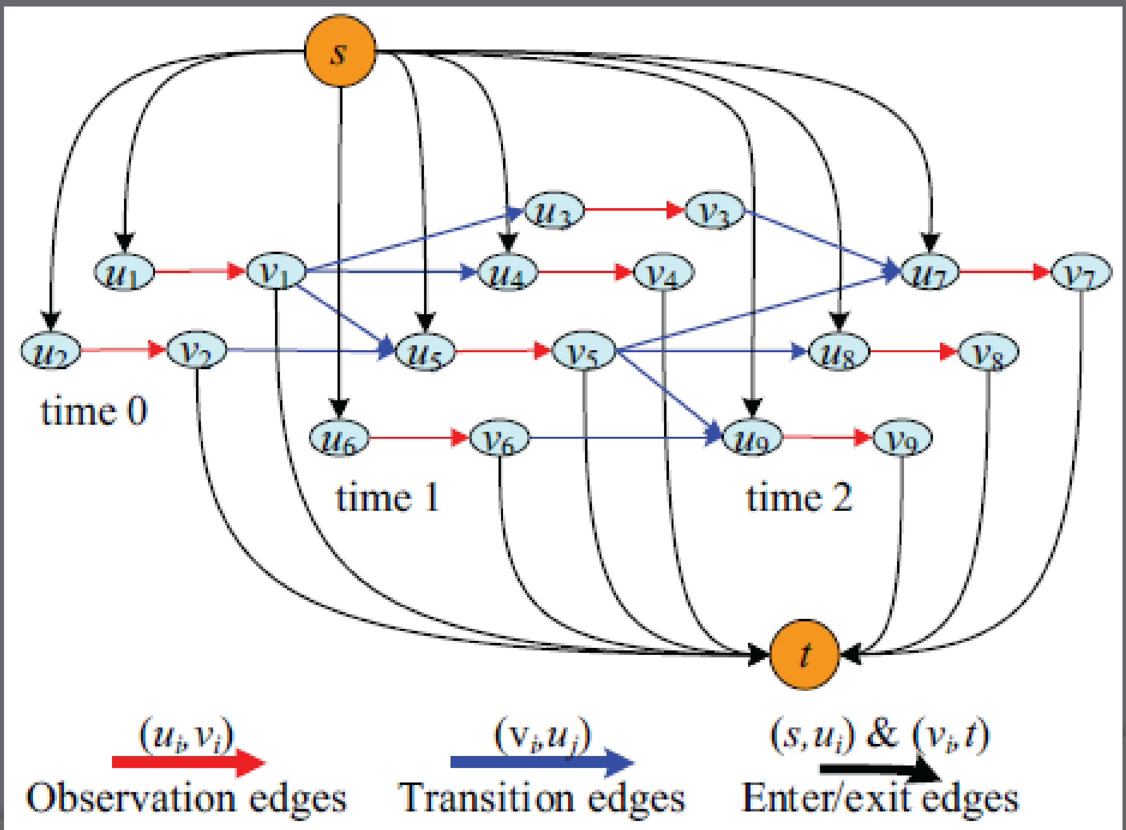
$$T = \underset{T}{\operatorname{argmin}} \sum_i C_{en,i} f_{en,i} + \sum_{i,j} C_{i,j} f_{i,j} + \sum_i C_{ex,i} f_{ex,i} + \sum_i C_i f_i \tag{11}$$

$$C_{en,i} = -\log P_{entr}(\mathbf{x}_i) \qquad C_{ex,i} = -\log P_{exit}(\mathbf{x}_i)$$

$$C_{i,j} = -\log P_{link}(\mathbf{x}_j|\mathbf{x}_i) \qquad C_i = \log \frac{\beta_i}{1-\beta_i}$$

Min-Cost flow Solution

- For each x_i, create 2 nodes u_i, v_i.



$$c(u_i, v_i) = C_i$$

$$f(u_i, v_i) = f_i$$

$$c(s, u_i) = C_{en,i}$$

$$f(s, u_i) = f_{en,i}$$

$$c(v_i, t) = C_{ex,i}$$

$$f(v_i, t) = f_{ex,i}$$

$$c(v_i, u_j) = C_{i,j}$$

$$f(v_i, u_j) = f_{i,j}$$

Explicit Occlusion Model

- Object parameters: $\mathrm{x}_i = (x_i, s_i, a_i, t_i)$

position scale appearance time

- Introduce occlusion nodes if

$|x_i - x_j|$

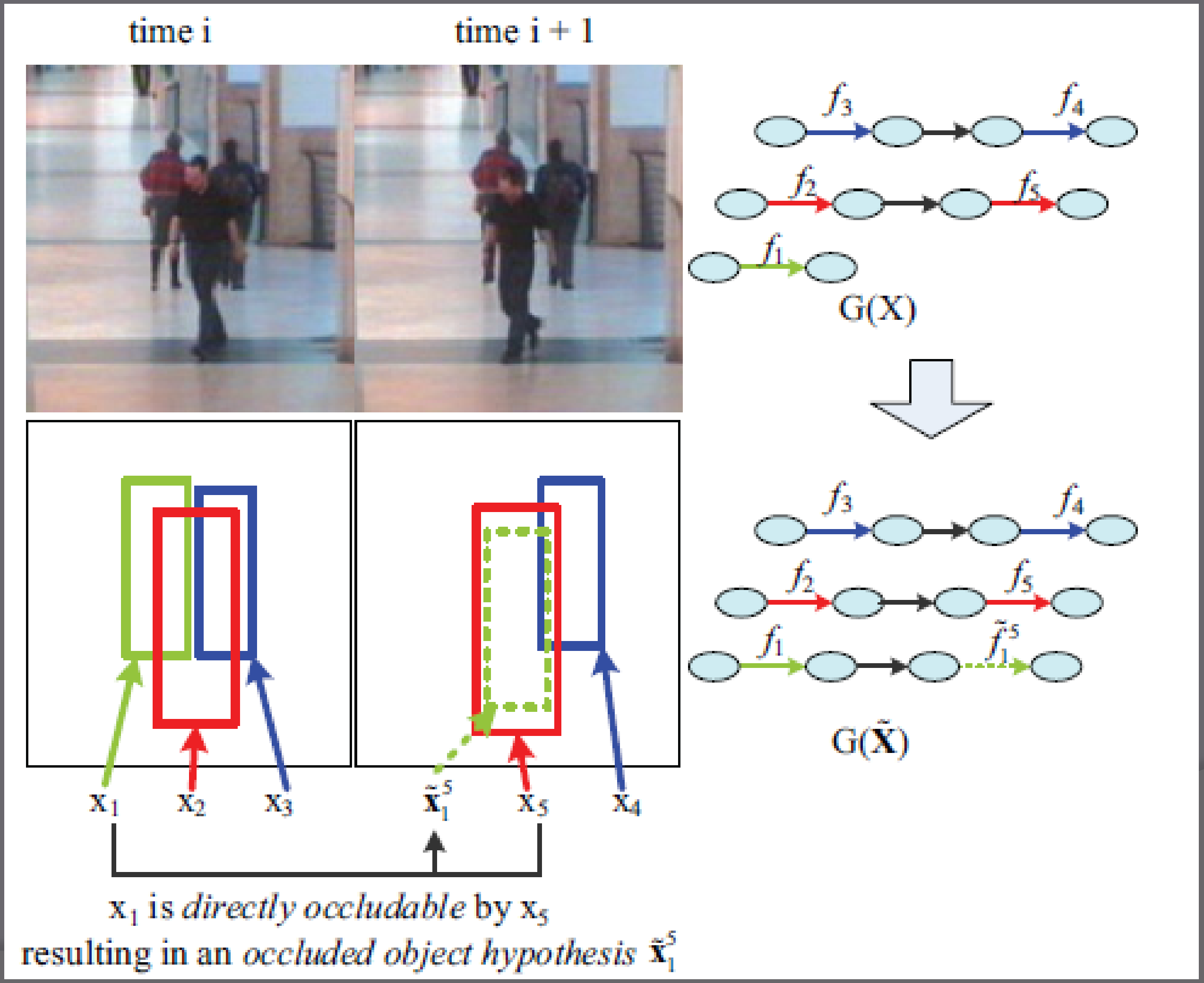
s_i / s_j

Less than a threshold

- Occlusion node parameters: $\tilde{\mathrm{x}}_i^j = (x_j, s_i, a_i, t_j)$

Where x_i is *directly occludable* by x_j

Explicit Occlusion Model



Explicit Occlusion Model

- Solve the min-cost flow problem using the augmented graph

Results

- Achieve stat-of-the-art
- Complexity of min-cost flow solver is polynomial in the number of node and edges

Conclusion

- Global analysis helps recovering long tracks under crowded scene and severe occlusion