



A Unified Framework for Multi-Target Tracking and Collective Activity Recognition

Wongun Choi and Silvio Savarese



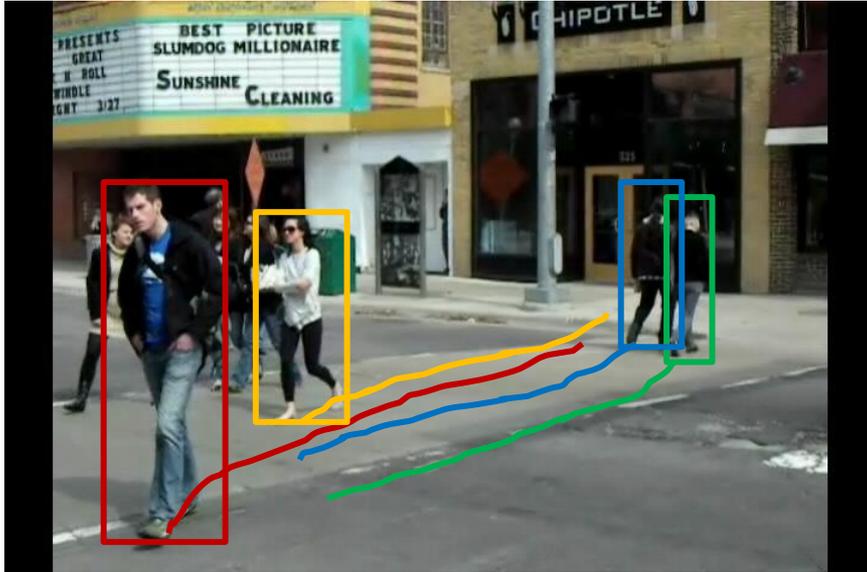
University of Michigan, Ann Arbor



Our Goal

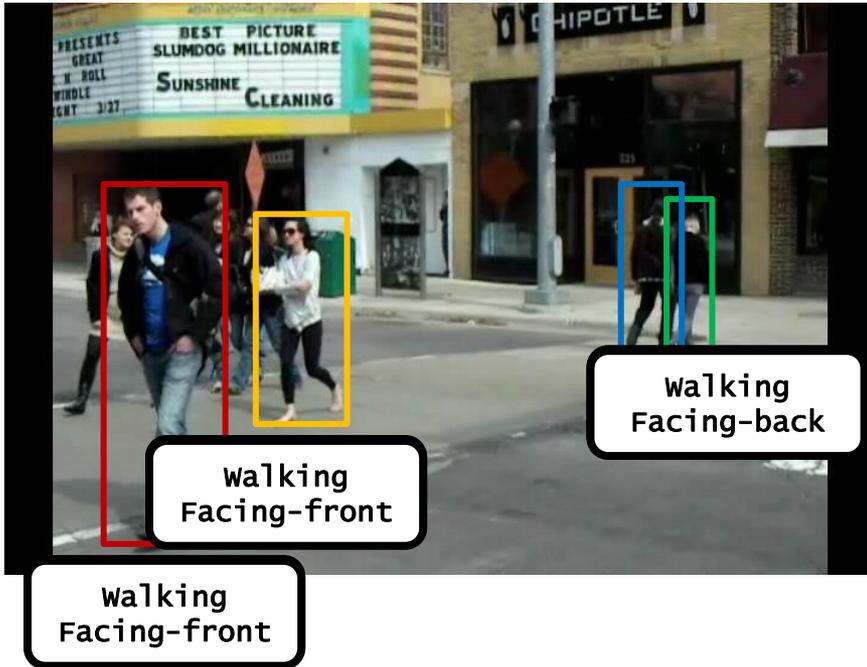


Our Goal



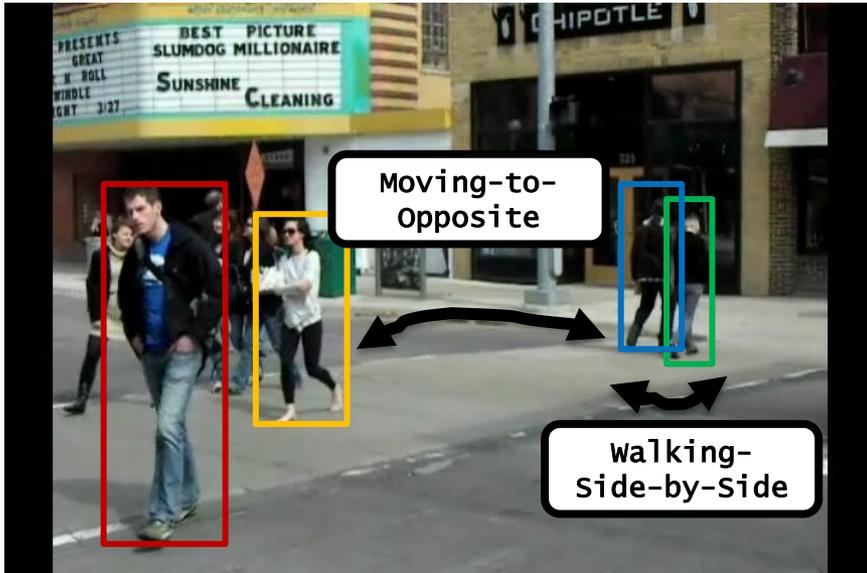
- Multiple target tracking

Our Goal



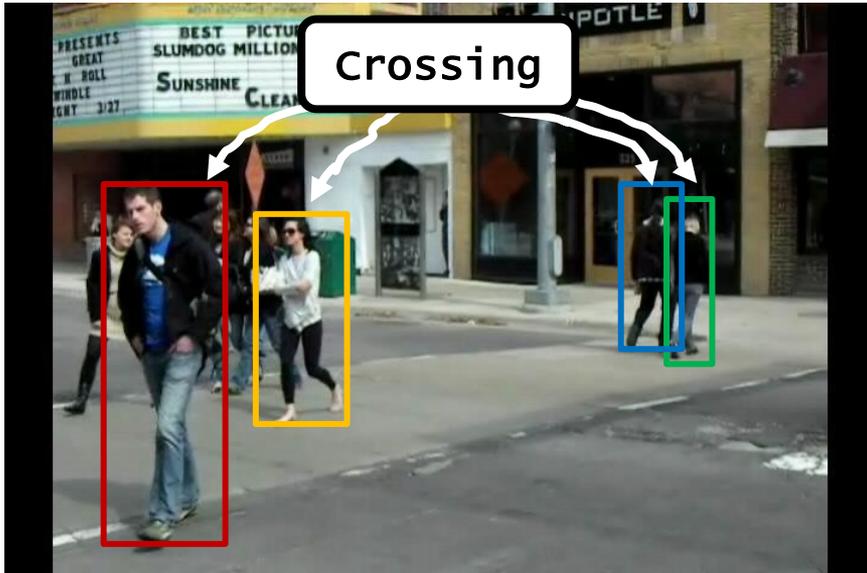
- Multiple target tracking
- Recognize activities at different level of granularity
 - Atomic activity & pose

Our Goal



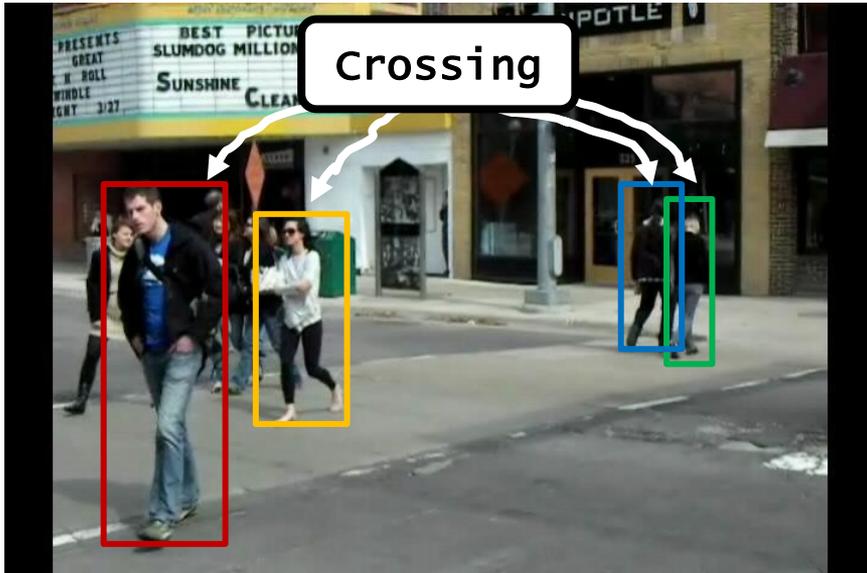
- Multiple target tracking
- Recognize activities at different level of granularity
 - Atomic activity & pose
 - Pairwise interaction

Our Goal



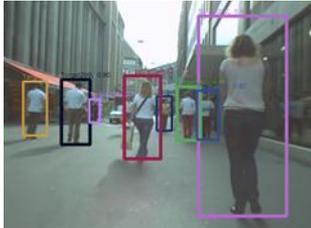
- Multiple target tracking
- Recognize activities at different level of granularity
 - Atomic activity & pose
 - Pairwise interaction
 - Collective Activity

Our Goal



- Multiple target tracking
- Recognize activities at different level of granularity
 - Atomic activity & pose
 - Pairwise interaction
 - Collective Activity
- solve all problems jointly!!

Background



Target Tracking

Atomic Activity

Pairwise Interaction

Collective Activity

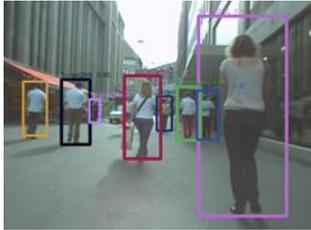
- Wu et al, 2007
- Avidan, 2007
- Zhang et al, 2008
- Breistein et al, 2009
- Ess et al, 2009
- Wojek et al, 2009
- Geiger et al, 2011
- Brendel et al, 2011
- Pirsivash et al, 2011

- Bobick & Davis, 2001
- Efros et al, 2003
- Schuldt et al, 2004
- Dollar et al, 2005
- Niebles et al, 2006
- Laptev et al, 2008
- Rodriguez et al, 2008
- Wang & Mori, 2009
- Gupta et al, 2009
- Liu et al, 2009
- Marszalek et al, 2009
- Liu et al, 2011

- Zhou et al, 2008
- Ryoo & Aggarwal, 2009
- Yao et al, 2010
- Choi et al, 2010
- Patron-perez et al, 2010

- Choi et al, 2009
- Li et al, 2009
- Lan et al, 2010
- Ryoo & Aggarwal, 2010
- Choi et al, 2011
- Khamis et al, 2011
- Lan et al, 2012
- Khamis et al, 2012
- Amer et al, 2012

Background



+

Target Tracking

+

Atomic Activity

+

Pairwise Interaction

+

Collective Activity

- Wu et al, 2007
- Avidan, 2007
- Zhang et al, 2008
- Breistein et al, 2009
- Ess et al, 2009
- Wojek et al, 2009
- Geiger et al, 2011
- Brendel et al, 2011
- Pirsivavash et al, 2011

- Bobick & Davis

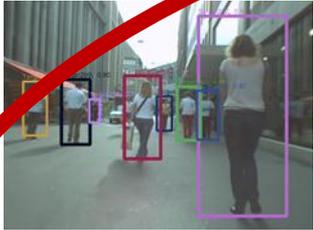
Investigated
in isolation

- Choi et al, 2009
- Li et al, 2009
- Lan et al, 2010
- Ryoo & Aggarwal, 2010
- Choi et al, 2011
- Khamis et al, 2011
- Lan et al, 2012
- Khamis et al, 2012
- Amer et al, 2012

Hierarchy of activities

- Lan et al, 2010
- Khamis et al, 2012
- Amer et al 2012

Background



Target
Tracking



Atomic
Activity



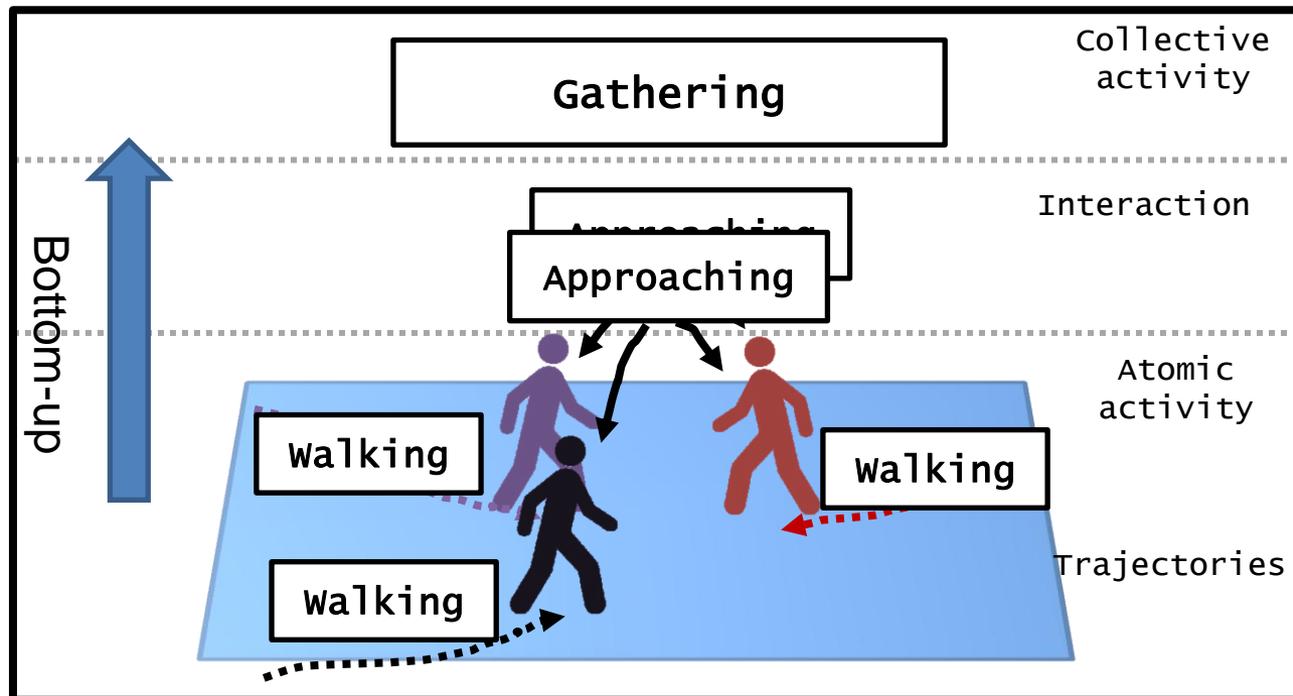
Pairwise
Interaction



Collective
Activity

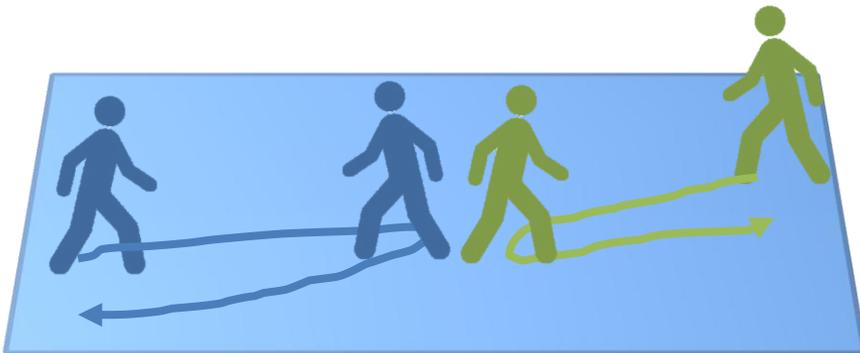
Contributions

- Bottom-up activity understanding.

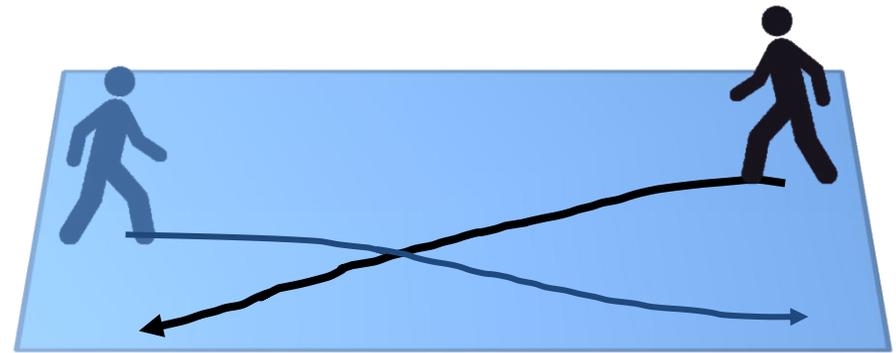


Contributions

- Bottom-up activity understanding.



Meeting and
Leaving

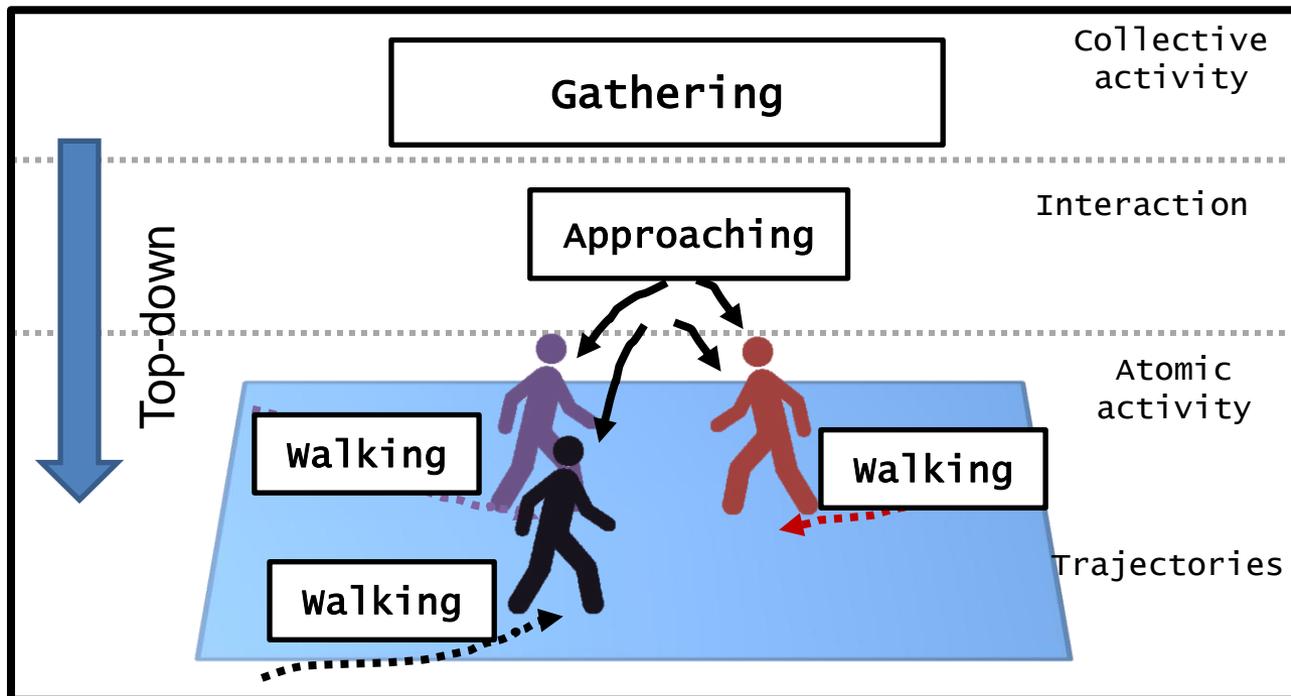


Crossing



Contributions

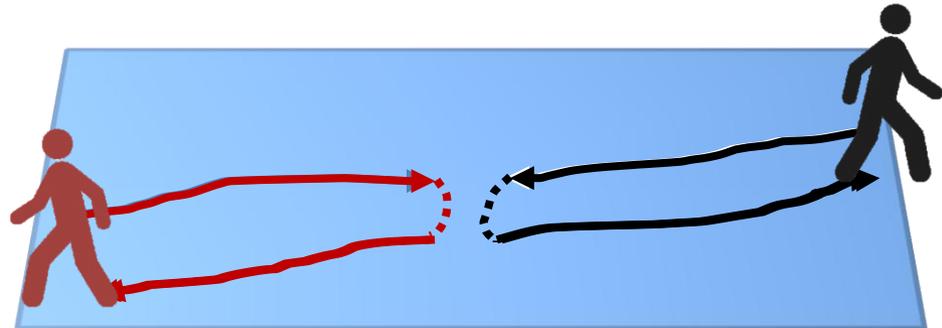
- Bottom-up activity understanding.
- Contextual information propagates top-down.



Contributions

- Bottom-up activity understanding.
- Contextual information propagates top-down.

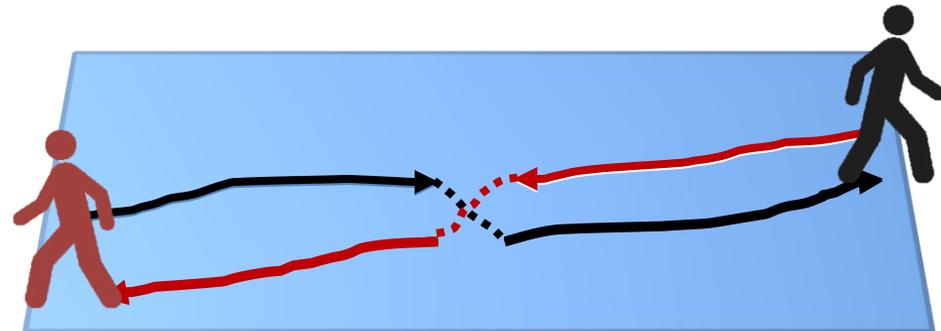
Meeting and Leaving



Contributions

- Bottom-up activity understanding.
- Contextual information propagates top-down.

Crossing



Simple Social Force Model
Pellegrini et al 2009
Choi et al 2010
Leal-Taixe et al 2012
etc

Repulsion &
Attraction

Outline

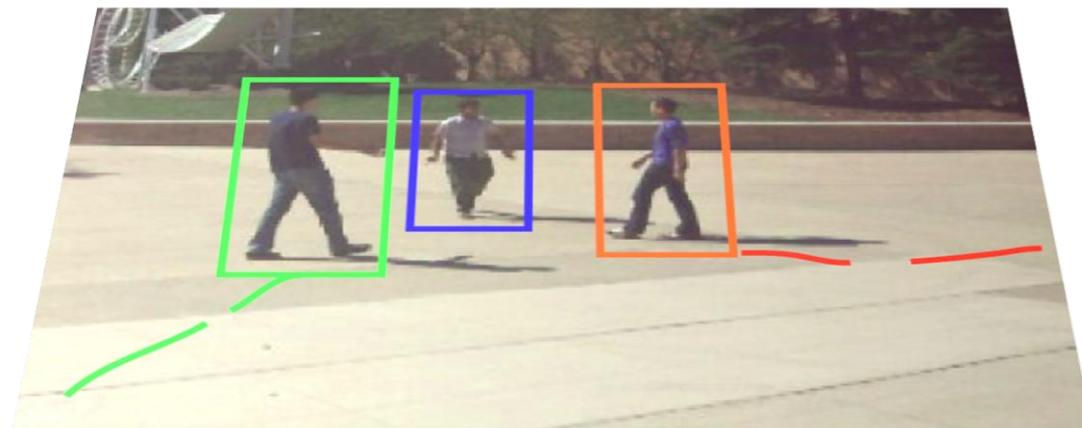
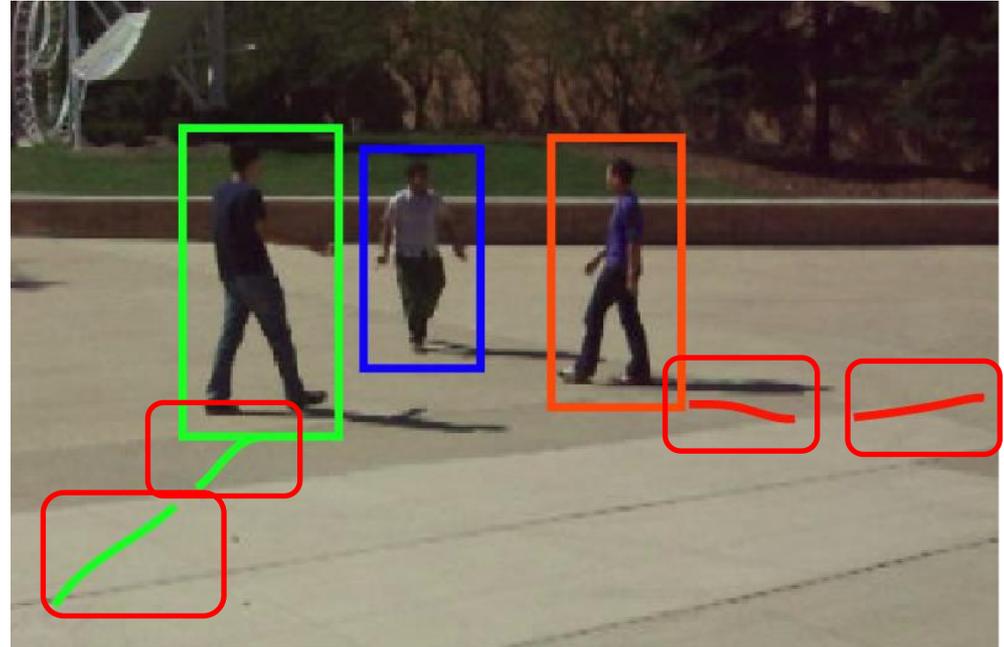
- Joint Model
- Inference/Training method
- Experimental evaluation
- Conclusion

Outline

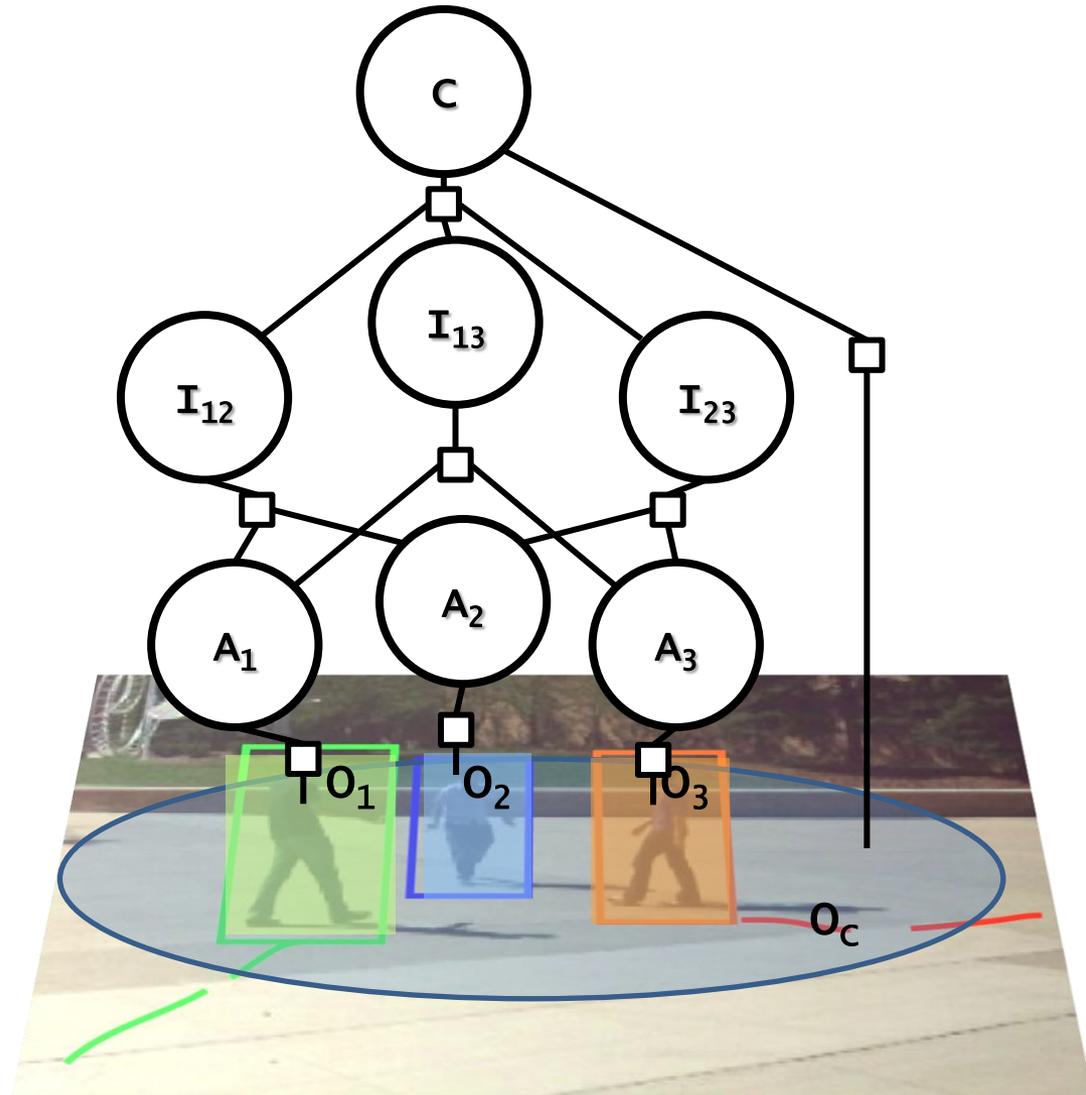
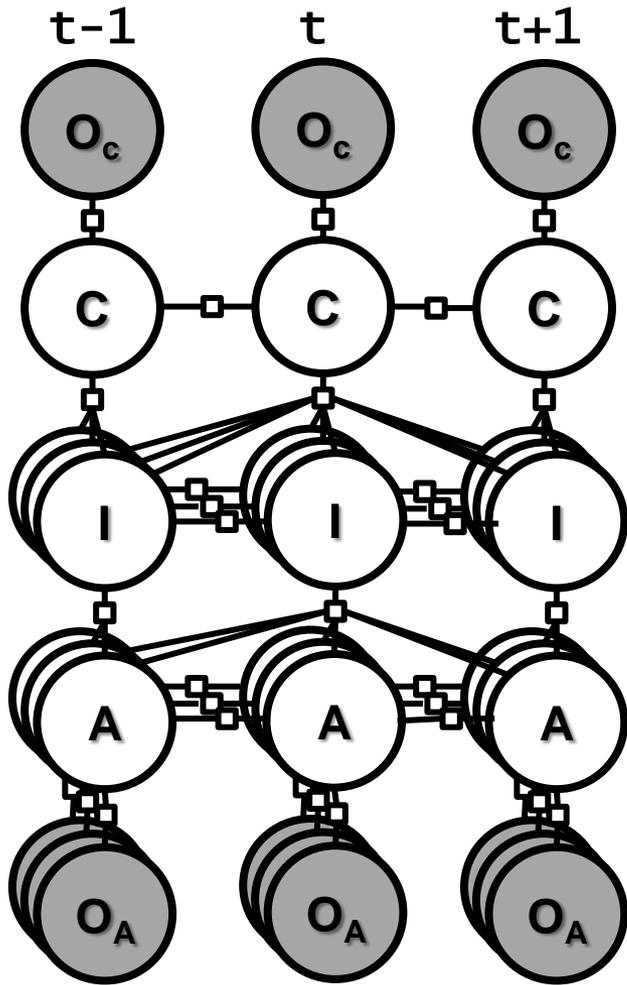
- Joint Model
- Inference/Training method
- Experimental evaluation
- Conclusion

Hierarchical Activity Model

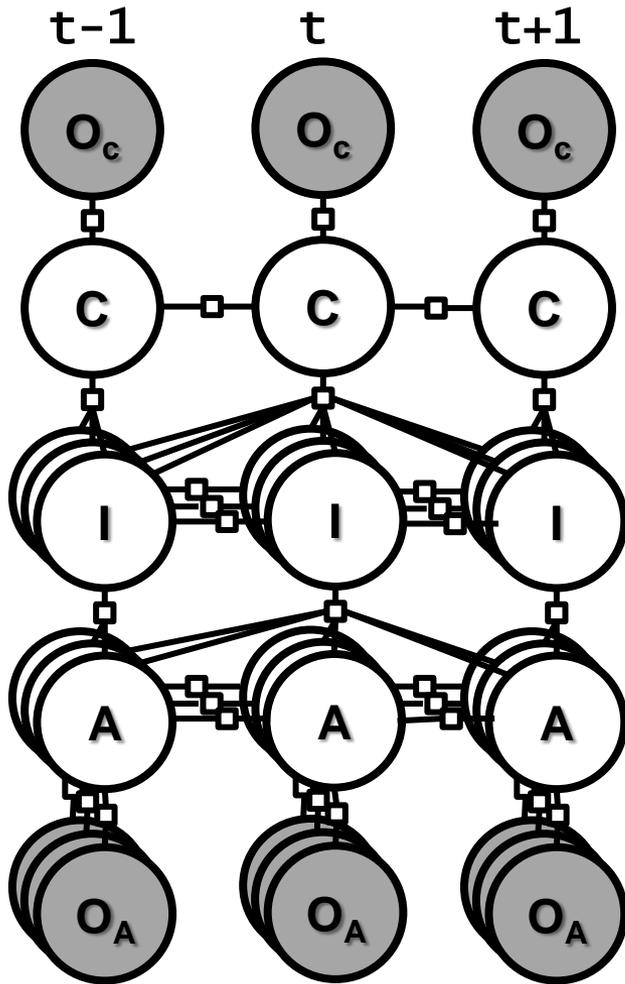
Input: video with tracklets



Hierarchical Activity Model

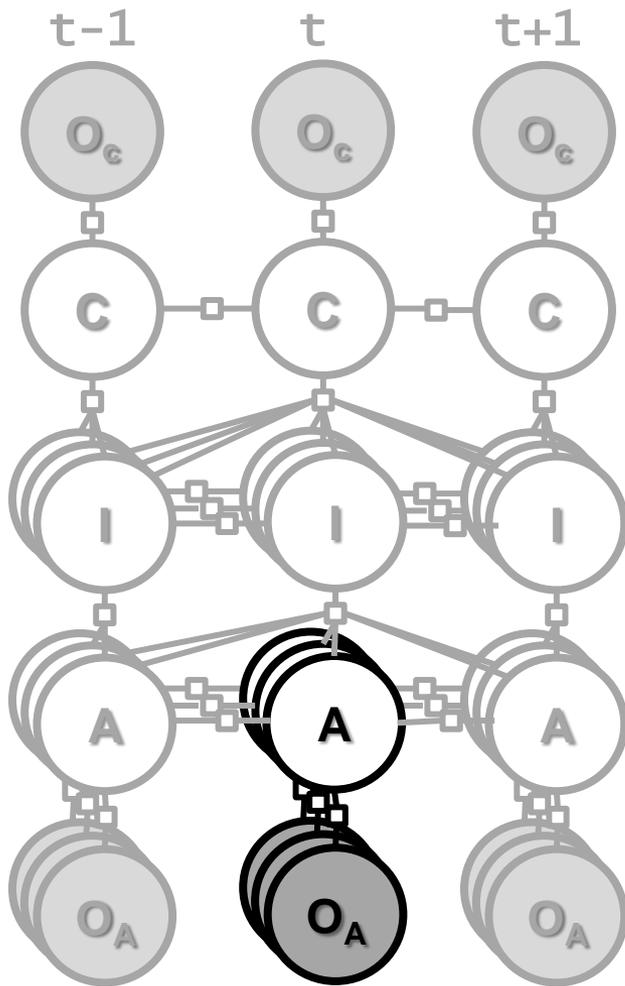


Hierarchical Activity Model



$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in \mathcal{S}$$

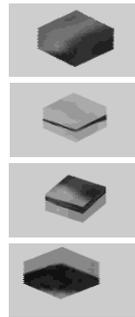
Atomic-Observation Potential



$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in S$$

Atomic Activity Models

- Action: BoW with STIP
- Pose: HoG

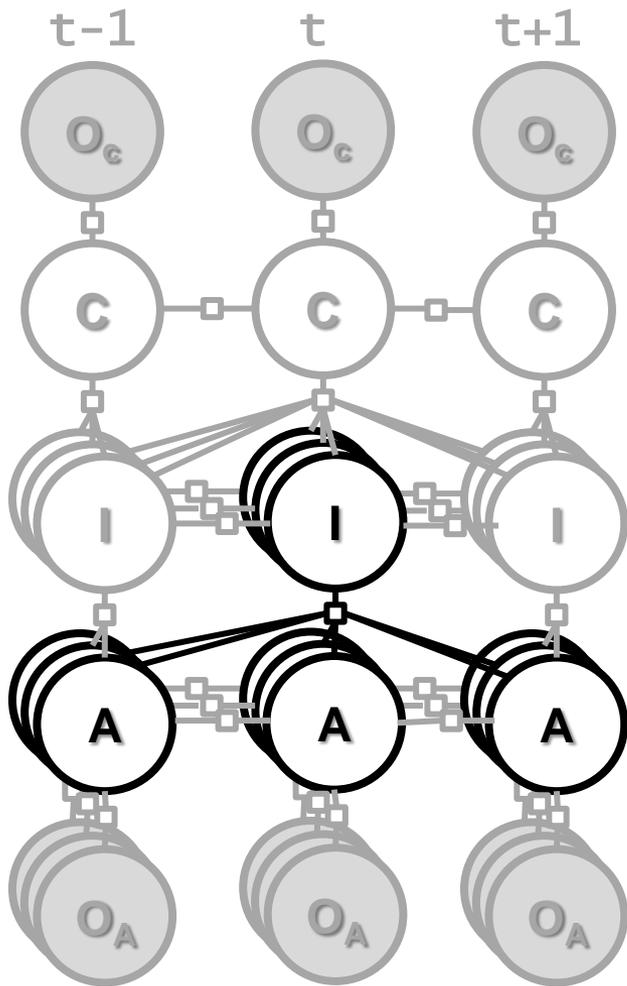


dollar et al, 06; Niebles et al, 07



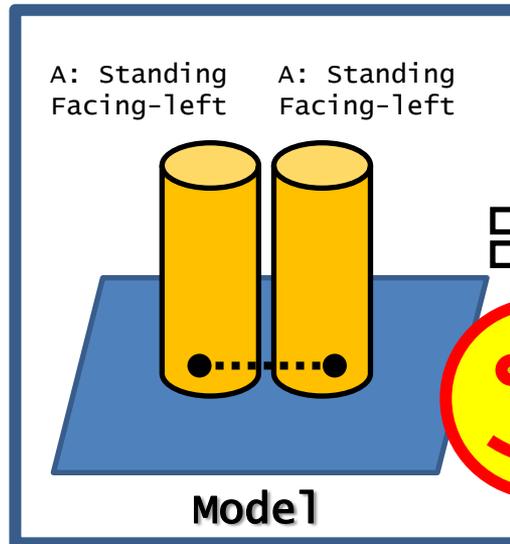
Dalal and Triggs, 05

Interaction-Atomic Potential

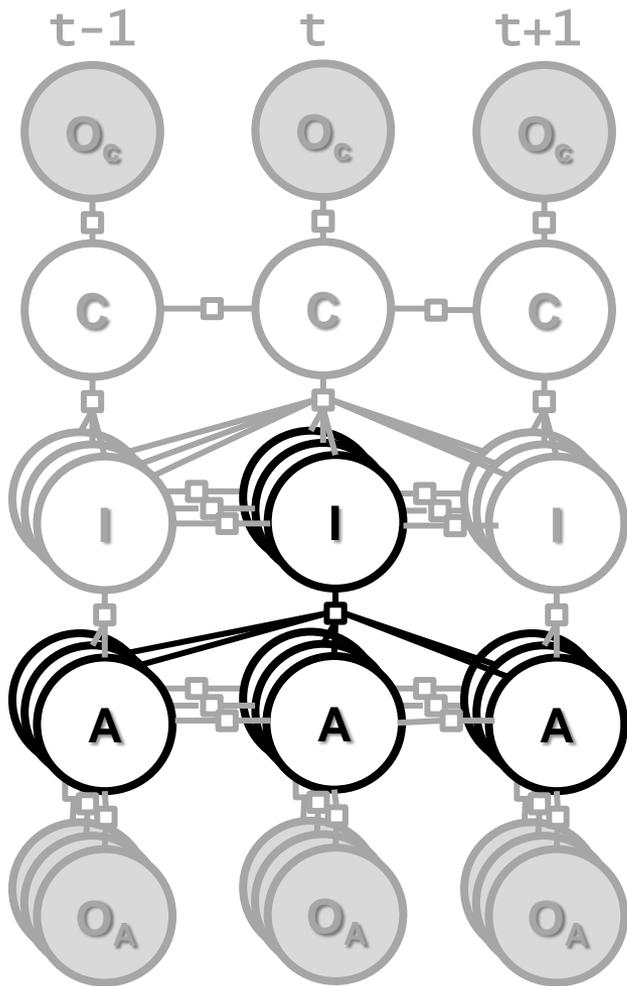


$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in S$$

I: Standing-in-a-line



Interaction-Atomic Potential



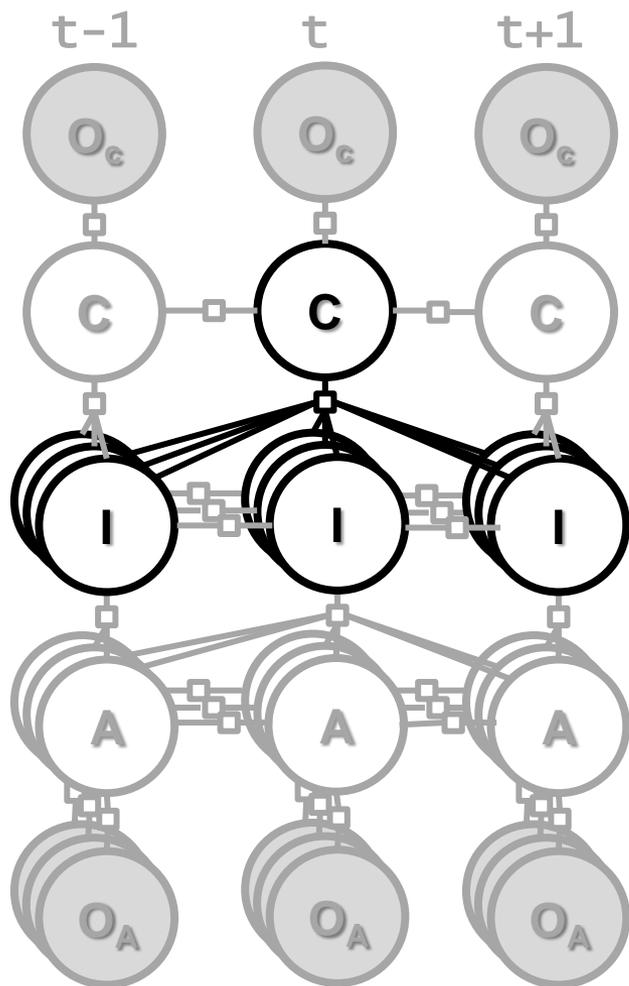
$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in S$$

I: Standing-in-a-line



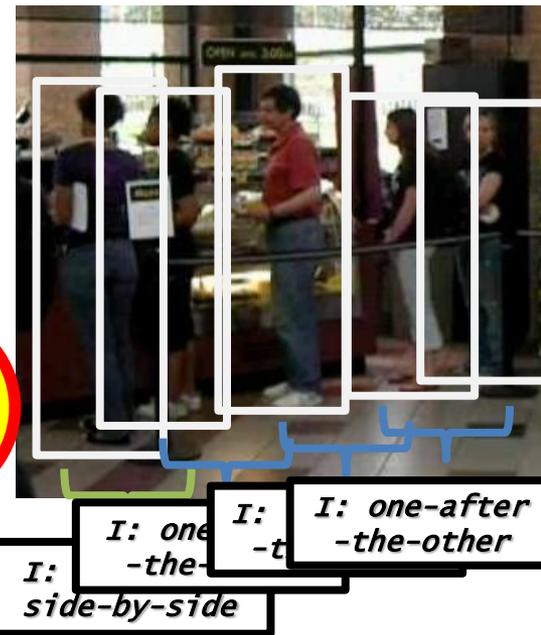
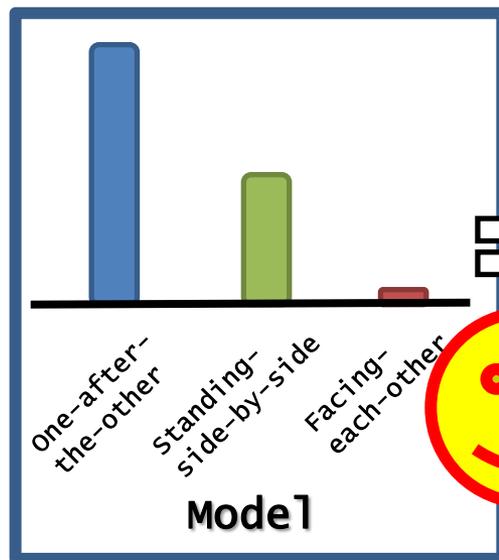
$$\Psi(A_i, A_j, I_{ij}, f) = \sum_{t \in T_V} \sum_{a \in \mathcal{I}} w^a \cdot \psi(A_i, A_j, f, t) \mathbb{I}(a, I_{ij}(t))$$

Collective-Interaction Potential

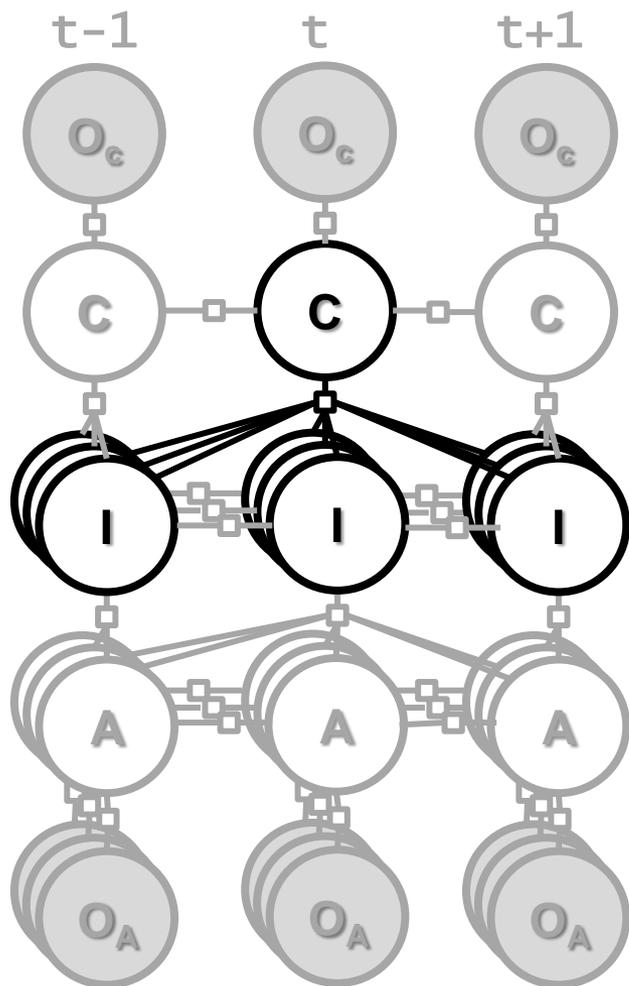


$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in S$$

C: Queuing

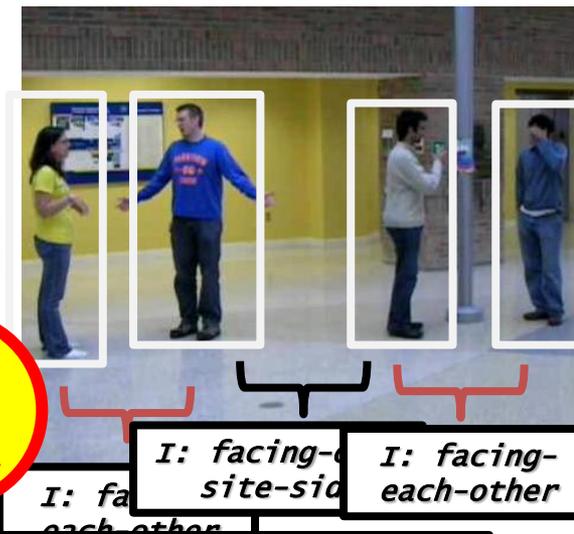
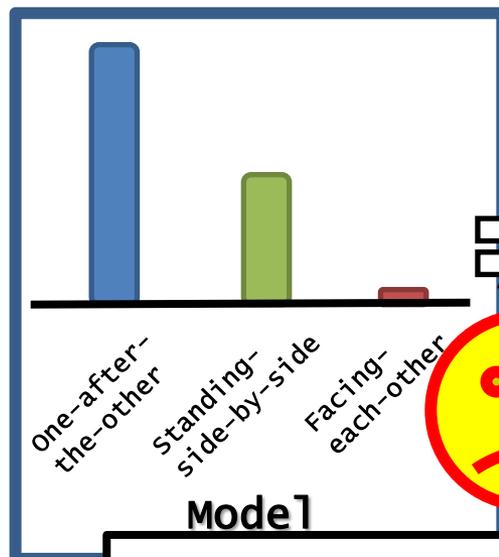


Collective-Interaction Potential



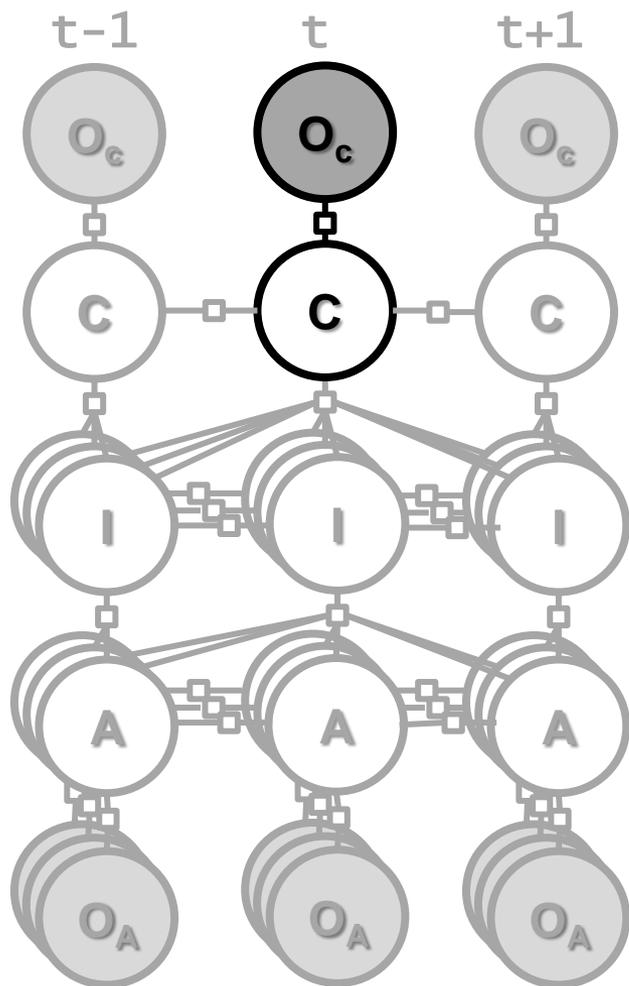
$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in S$$

C: Queuing



$$\Psi(C, I) = \sum_{t \in T_V} \sum_{c \in \mathcal{C}} w^c \cdot h(I, t) \mathbb{I}(c, C(t))$$

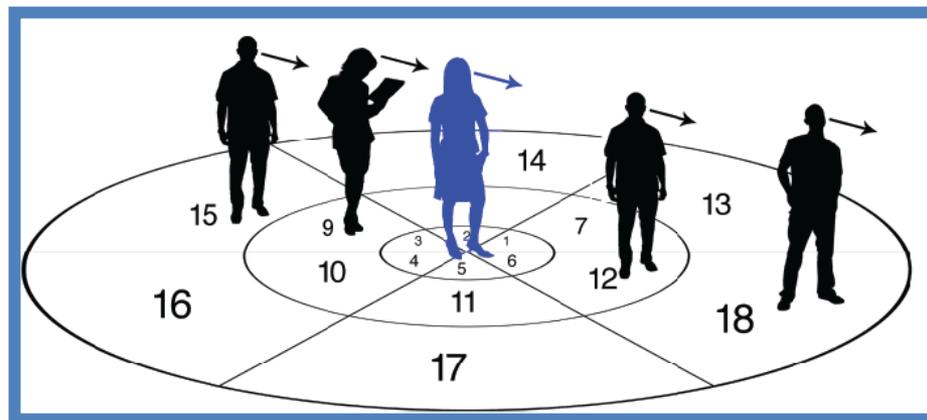
Collective-Observation Potential



$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in S$$

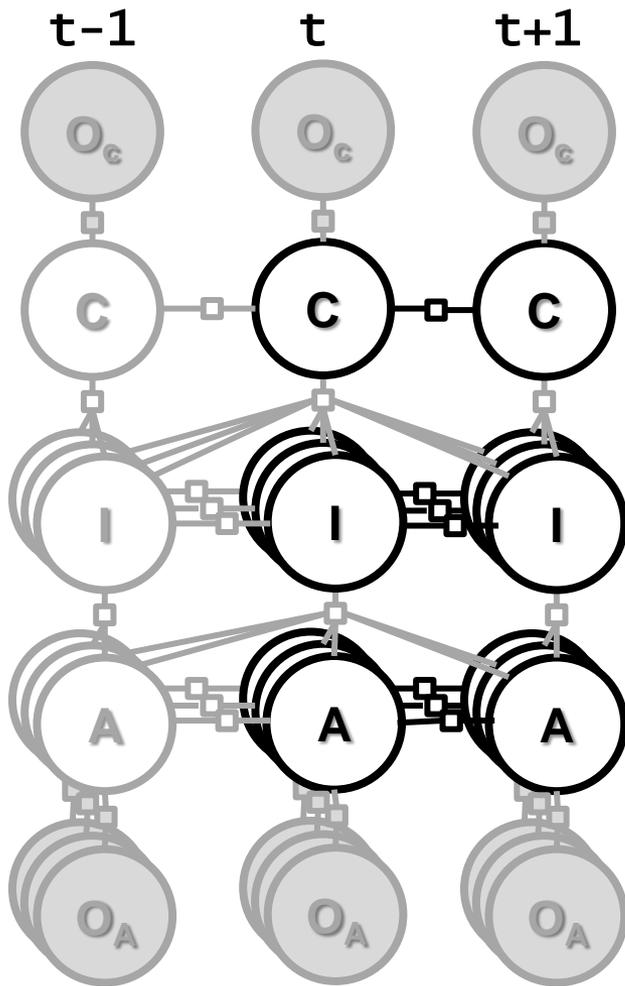
collective Activity

- STL of all targets



choi et al, 09

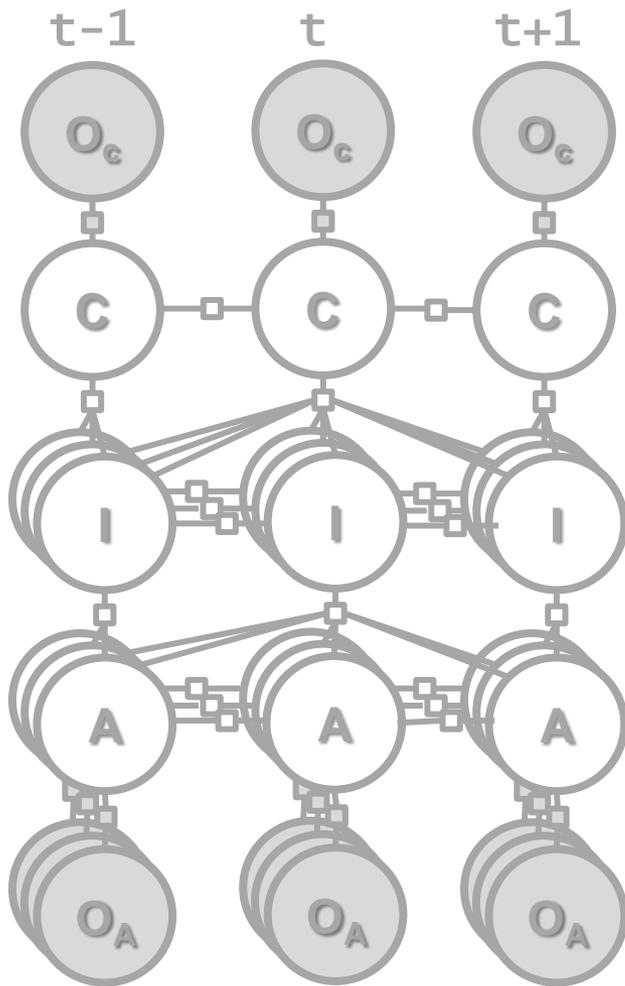
Activity Transition Potential



$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in S$$

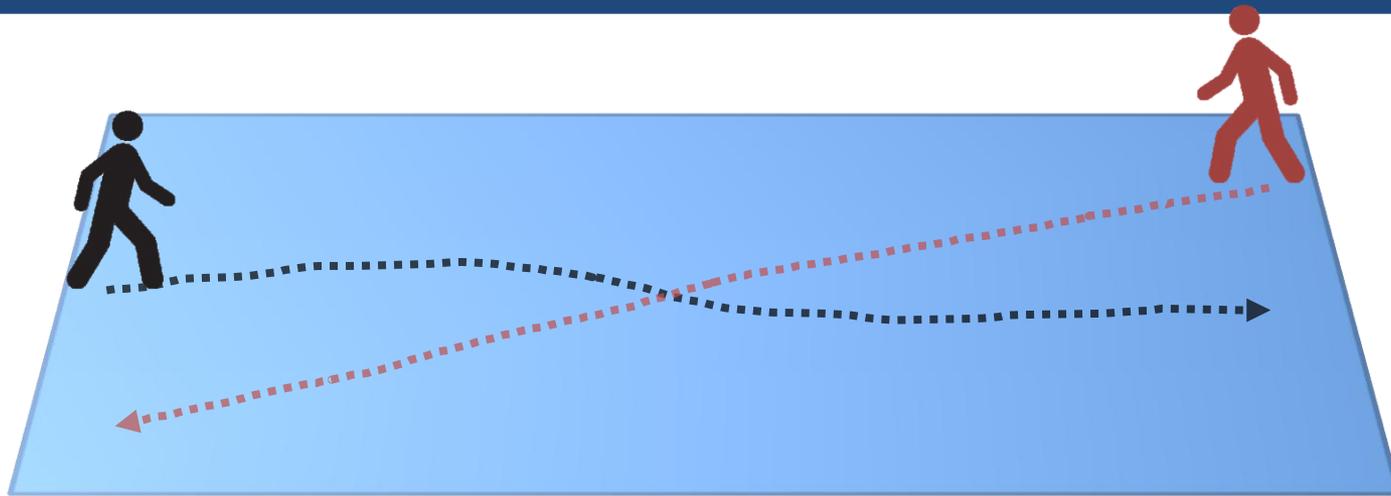
Smooth activity transition

Trajectory Estimation

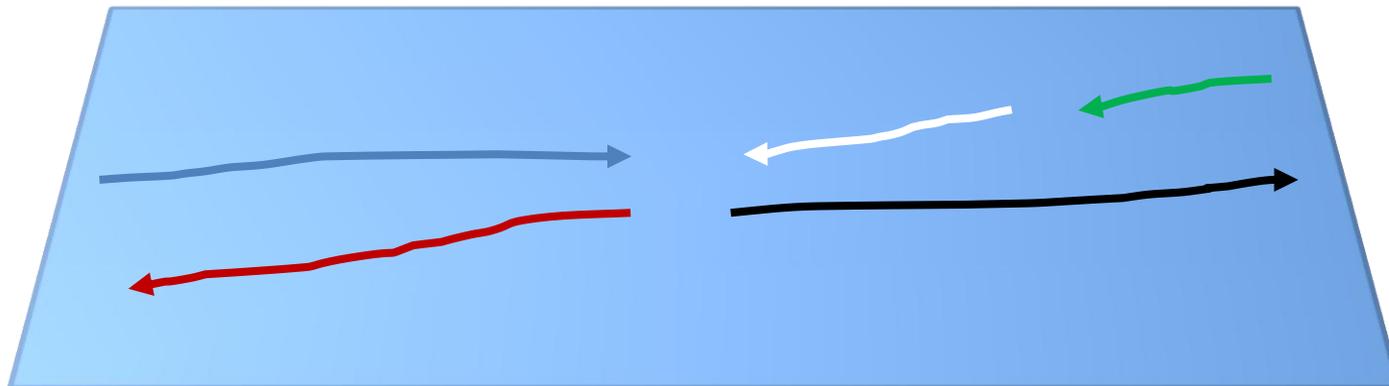


$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in \mathcal{S}$$

Tracklet Association Problem

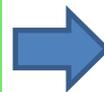


Tracklet Association Problem



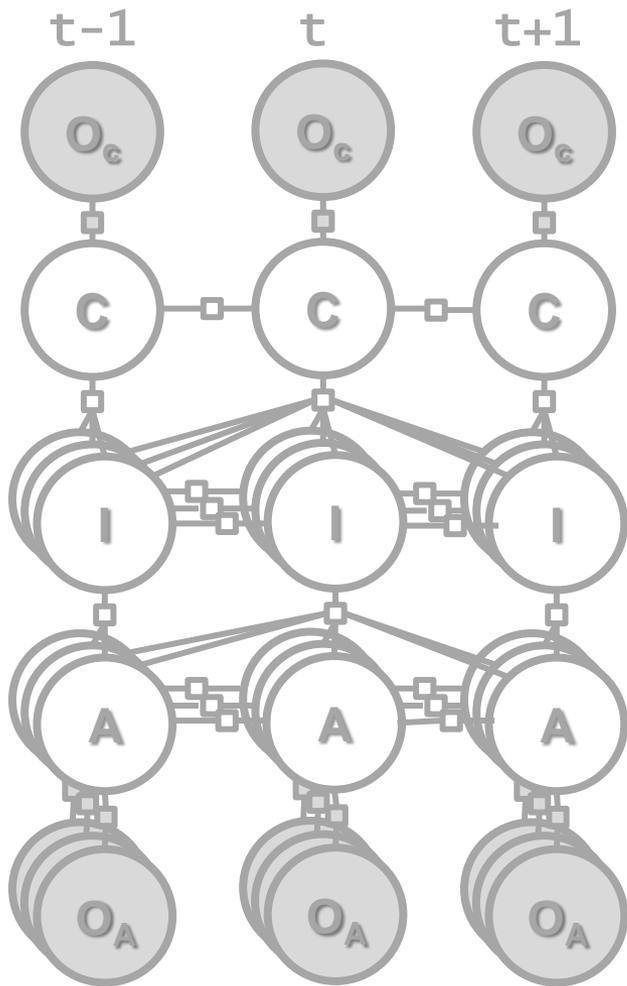
Input: Fragmented Trajectories (tracklets)

- Detector failures
- Occlusion between targets
- Scene clutter
- etc..

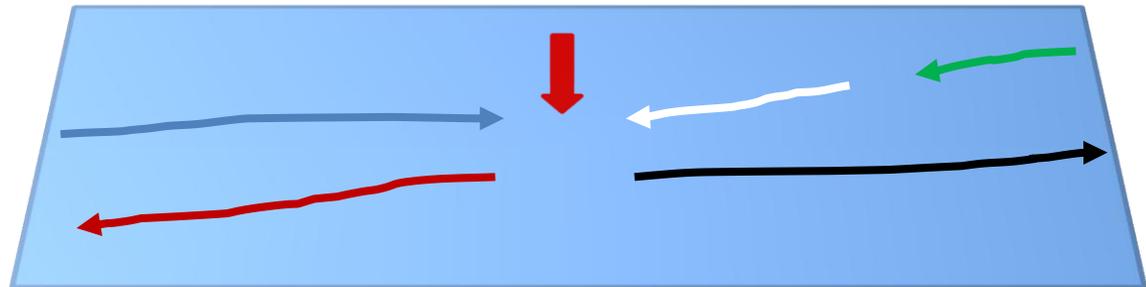


Output: set of trajectories with correct IDs (color)

Tracklet Association Model



$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in \mathcal{S}$$



Simple match costs, c

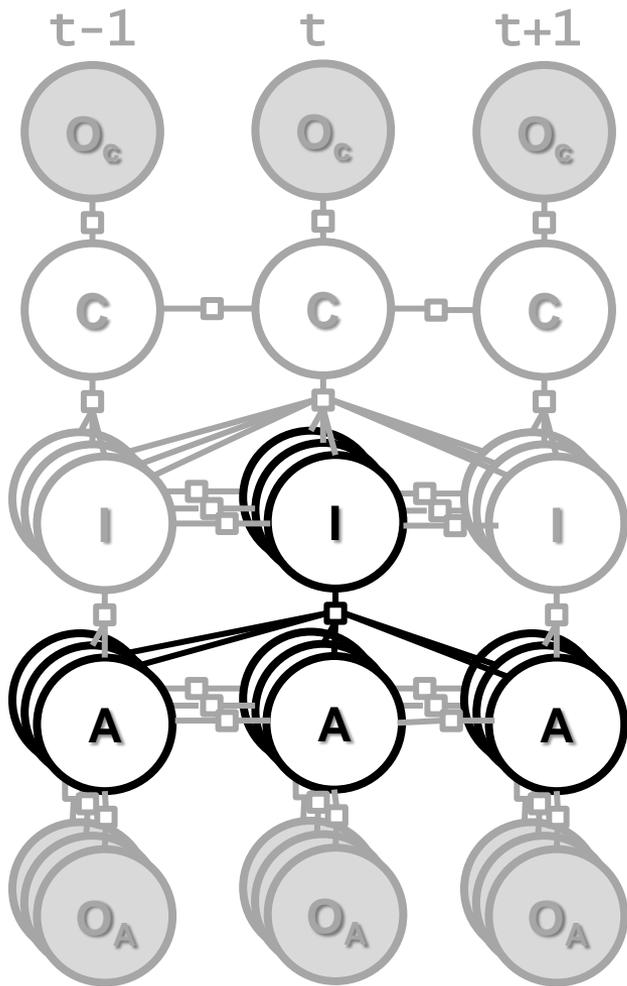
Location affinity ?

Appearance/Color ?

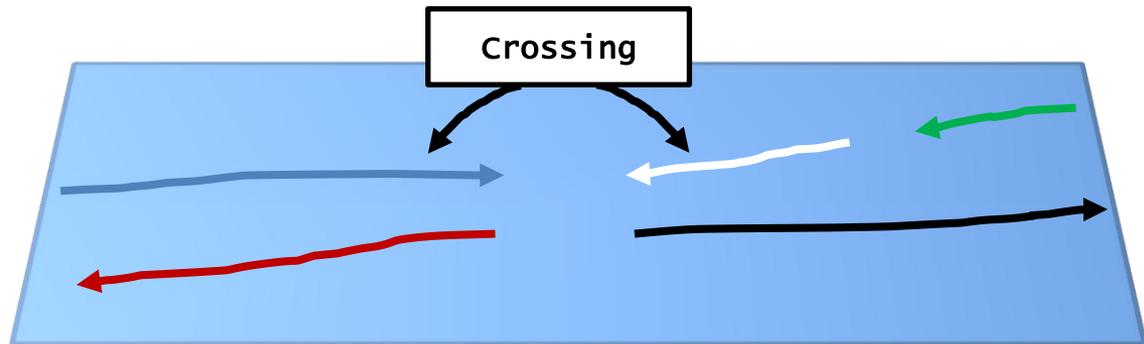
...



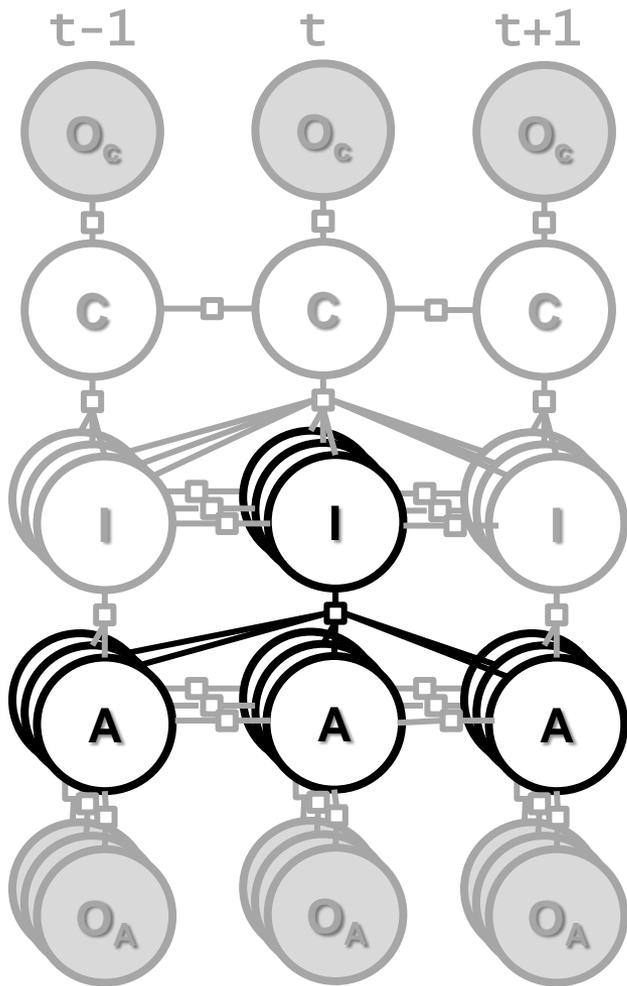
Tracklet Association Model



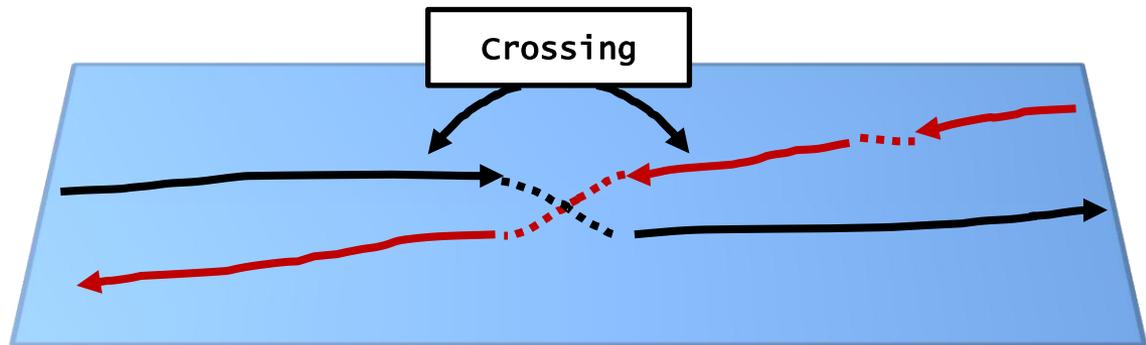
$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in \mathcal{S}$$



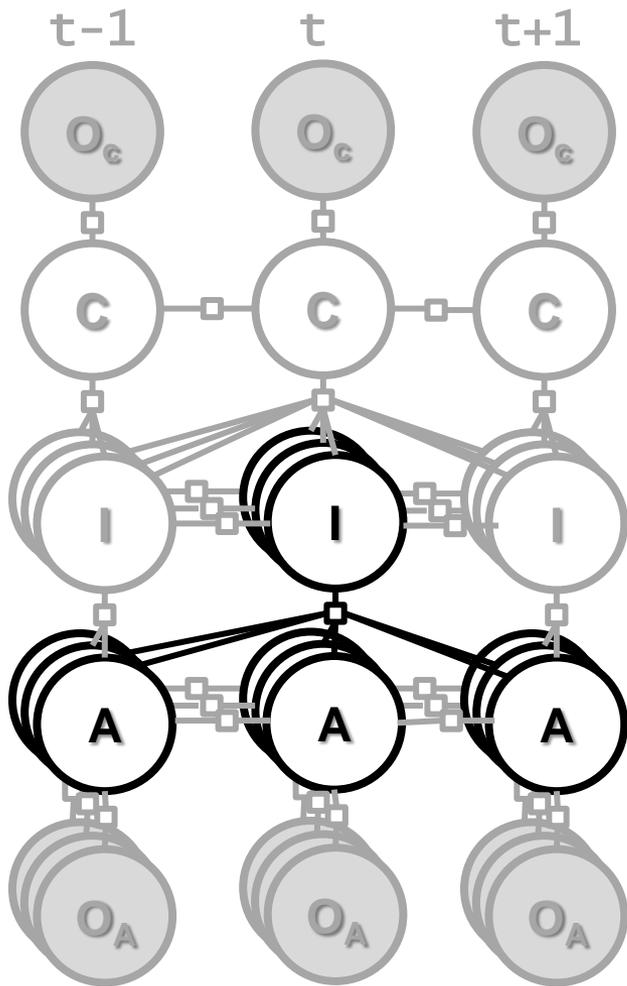
Tracklet Association Model



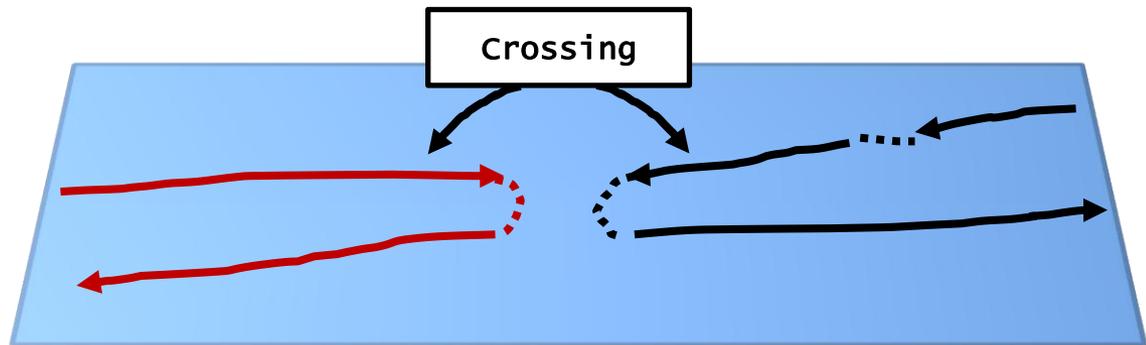
$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in \mathcal{S}$$



Tracklet Association Model



$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in \mathcal{S}$$



Outline

- Joint Model
- Inference/Training method
- Experimental evaluation
- Conclusion

Inference

$$\hat{C}, \hat{I}, \hat{A}, \hat{f} = \arg \max_{C, I, A, f} \left\{ \begin{aligned} \Psi(C, I, A, O, f) = \\ \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \\ \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in \mathcal{S} \end{aligned} \right.$$

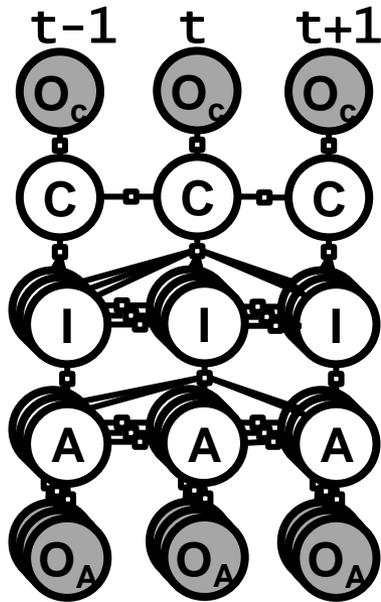
Non-convex Problem!!

Inference

$$\hat{C}, \hat{I}, \hat{A}, \hat{f} = \arg \max_{C, I, A, f}$$

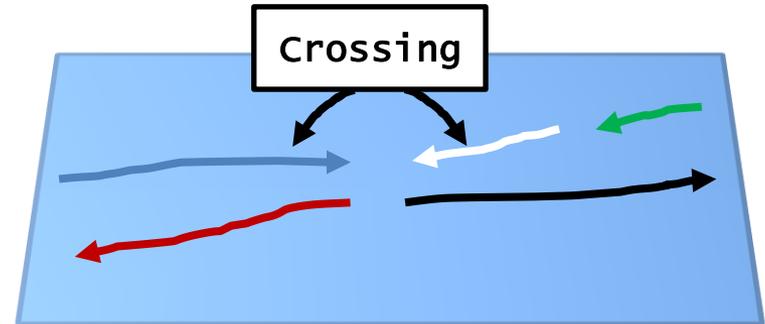
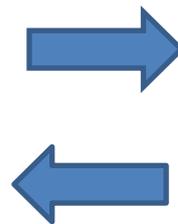
$$\Psi(C, I, A, O, f) = \Psi(A, O) + \Psi(I, A, f) + \Psi(C, I) + \Psi(C, O) + \Psi(C) + \Psi(I) + \Psi(A) - c^T f, f \in S$$

Activity Recognition Given f



Iterative Belief Propagation

Tracklet Association Given C, I, A



Novel Branch and Bound

Training

- Model weights are learned in a Max-Margin framework using Structural SVM.

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{n} \sum_{i=1}^n \xi_i, \text{ s.t. } \forall i, \xi_i \geq 0$$

$$\forall i, \forall \mathbf{y} \in \mathcal{Y} \setminus \mathbf{y}_i : \langle \mathbf{w}, \delta \Psi_i(\mathbf{y}) \rangle \geq \Delta(\mathbf{y}_i, \mathbf{y}) - \xi_i$$

Tsochantaridis et al, 2004

Outline

- Joint Model
- Inference/Training method
- **Experimental evaluation**
- Conclusion

Experiments

- Collective Activity Dataset choi et al, 2009
 - 44 videos with multiple people
 - Crossing, waiting, Queuing, walking, Talking



Crossing



waiting



Queuing



walking



Talking

Target identities

Interaction

- Approaching
- Leaving
- Passing-by
- Facing-each-other
- etc..

Atomic Activity

- facing-right
- facing-left
- ...
- walking
- standing

Experiments

- New Dataset
 - 32 videos with multiple people
 - Gathering, Talking, Dismissal, walking together, Chasing, Queuing



Gathering



Talking



Dismissal



walking-together



Chasing



Queuing

Target identities

Interaction

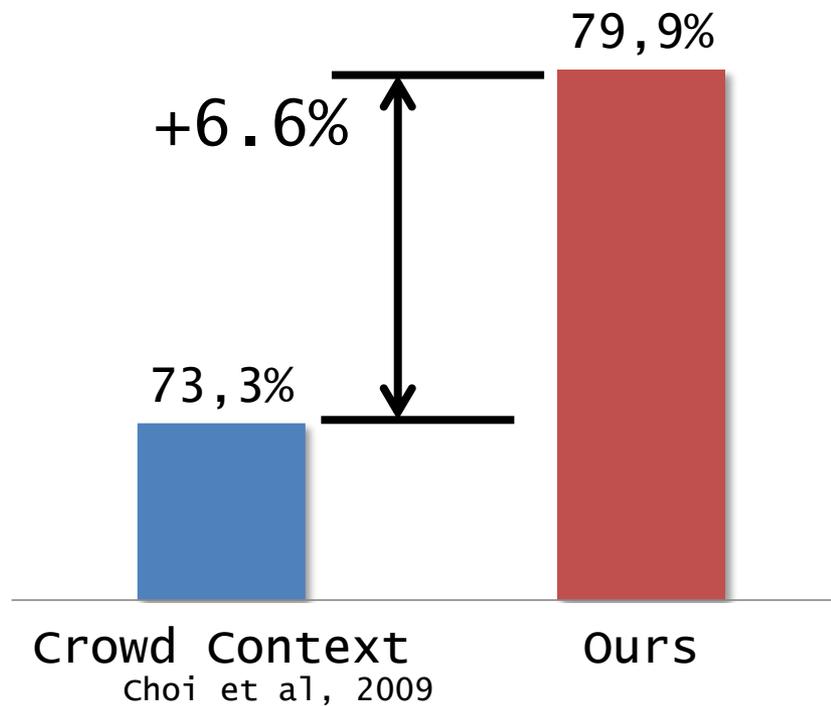
- Approaching
- walking-in-oppos.
- Facing-each-other
- Standing-in-a-row
- etc..

Atomic Activity

- facing-right
- facing-left
- ...
- walking
- Standing
- running

Classification Results

Overall Collective Activity Classification Accuracy



Collective Activity
Dataset, 2009

New Dataset

Target Association

	Tracklet			
# of error	1556			
Improvement over tracklet	0%			

Result of Dataset VSWS09

Target Association

	Tracklet	No Interaction		
# of error	1556	1109		
Improvement over tracklet	0%	28.73%		

Result of Dataset VSWS09

Target Association

	Tracklet	No Interaction	with Interaction	
# of error	1556	1109	894	
Improvement over tracklet	0%	28.73%	42.54%	

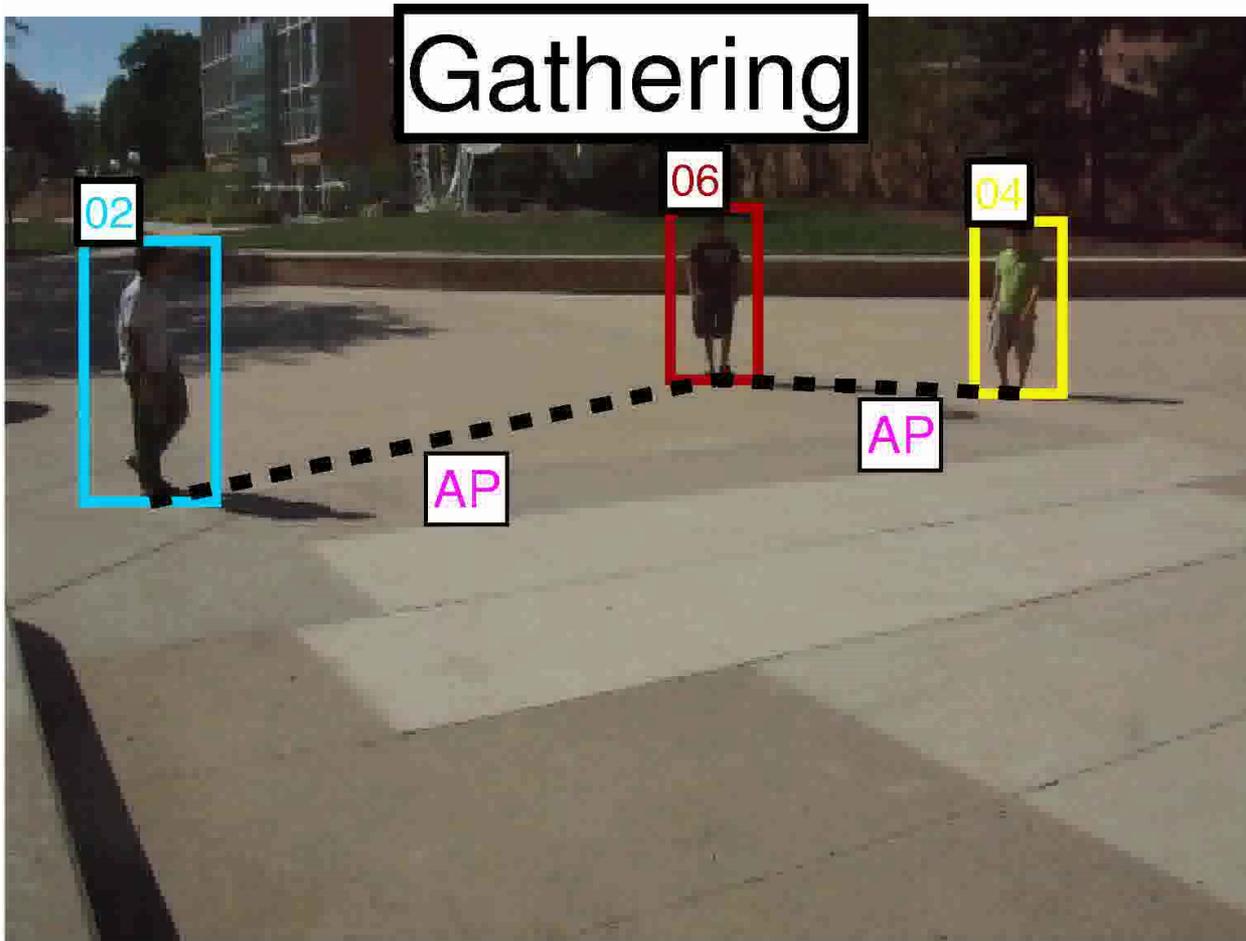
Result of Dataset VSWS09

Target Association

	Tracklet	No Interaction	with Interaction	with GT Activities
# of error	1556	1109	894	736
Improvement over tracklet	0%	28.73%	42.54%	52.76%

Result of Dataset VSWS09

Example Classification Result



Interaction Labels

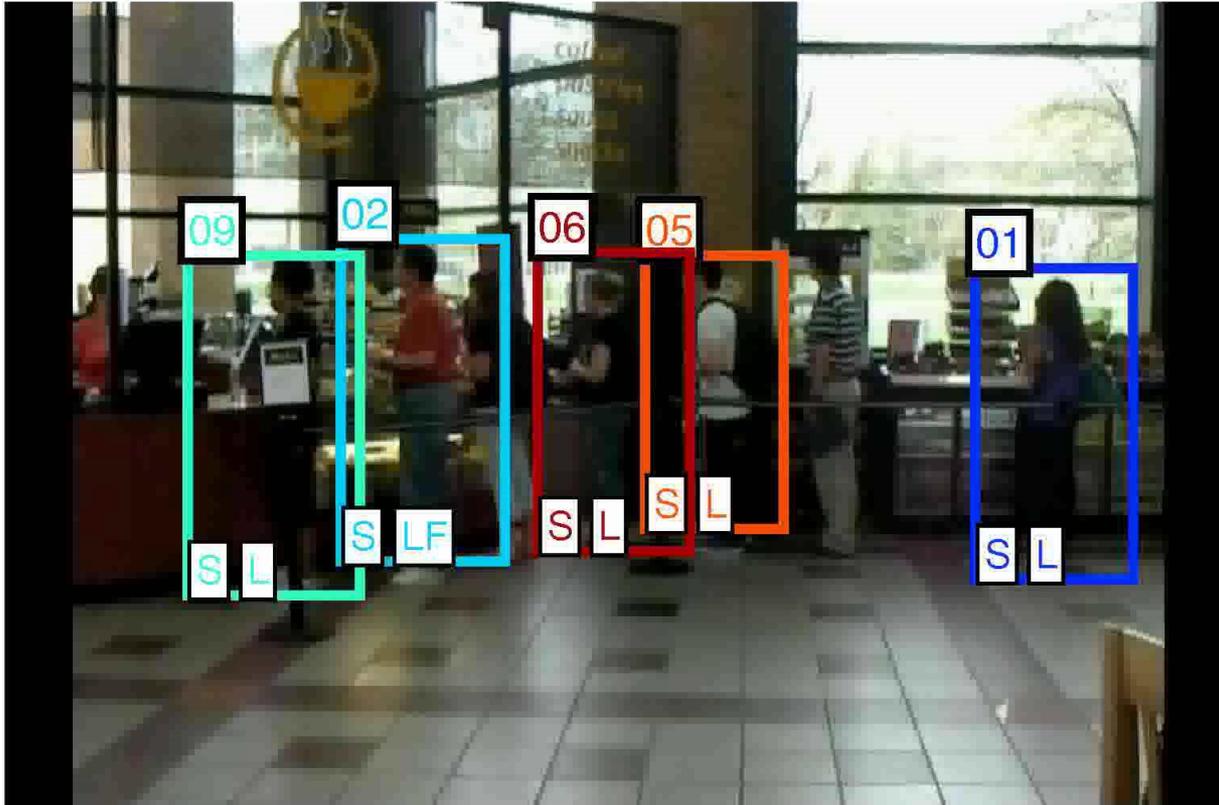
AP: approaching

FE: facing-each-other

SR: standing-in-a-row

...

Example Classification Result



Atomic Activities

Action:

W - walking

S - standing

Pose (8 directions)

L - left

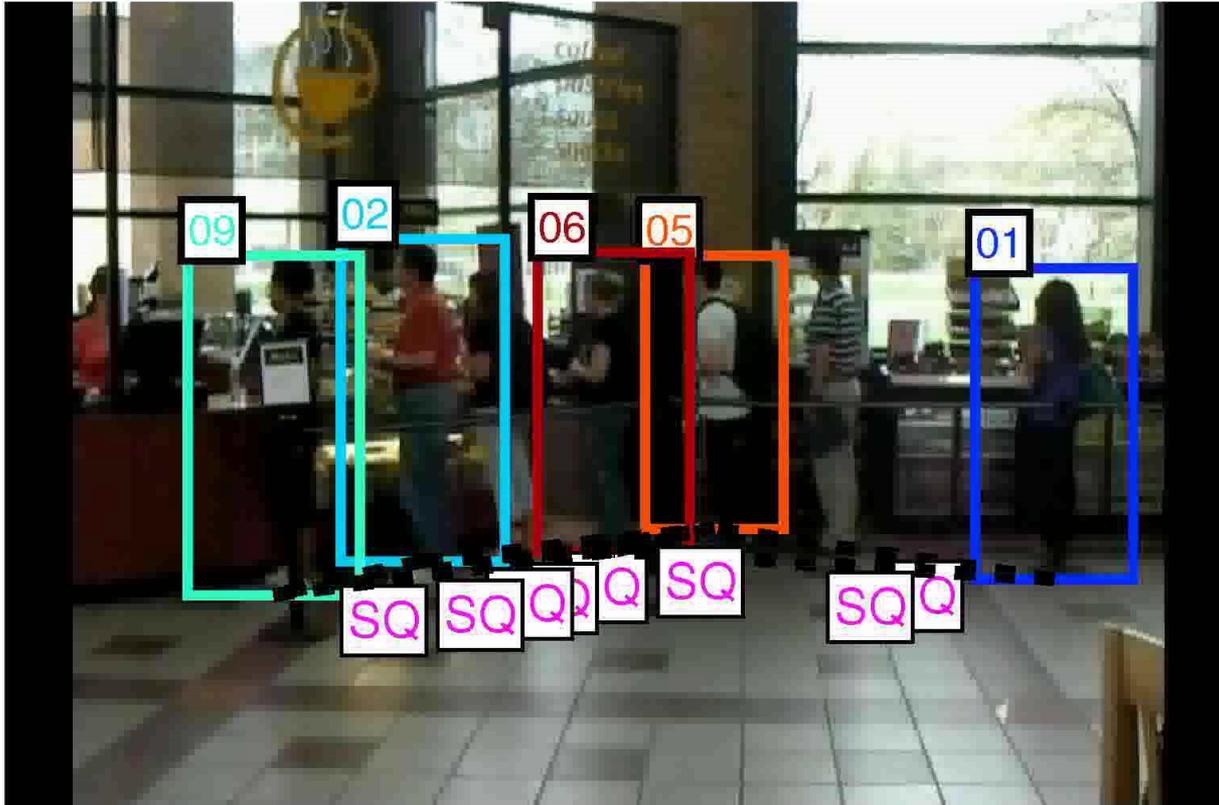
LF- left/front

F - front

RF- right/front

etc.

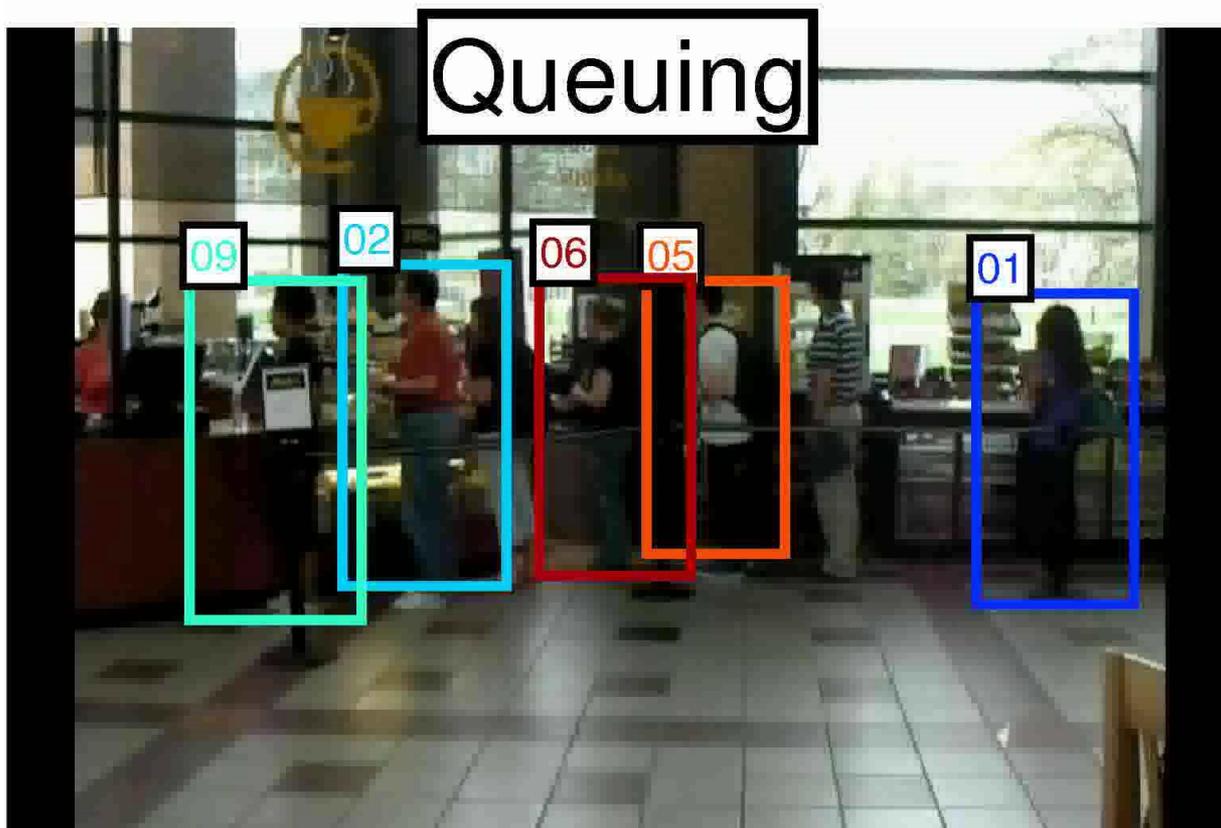
Example Classification Result



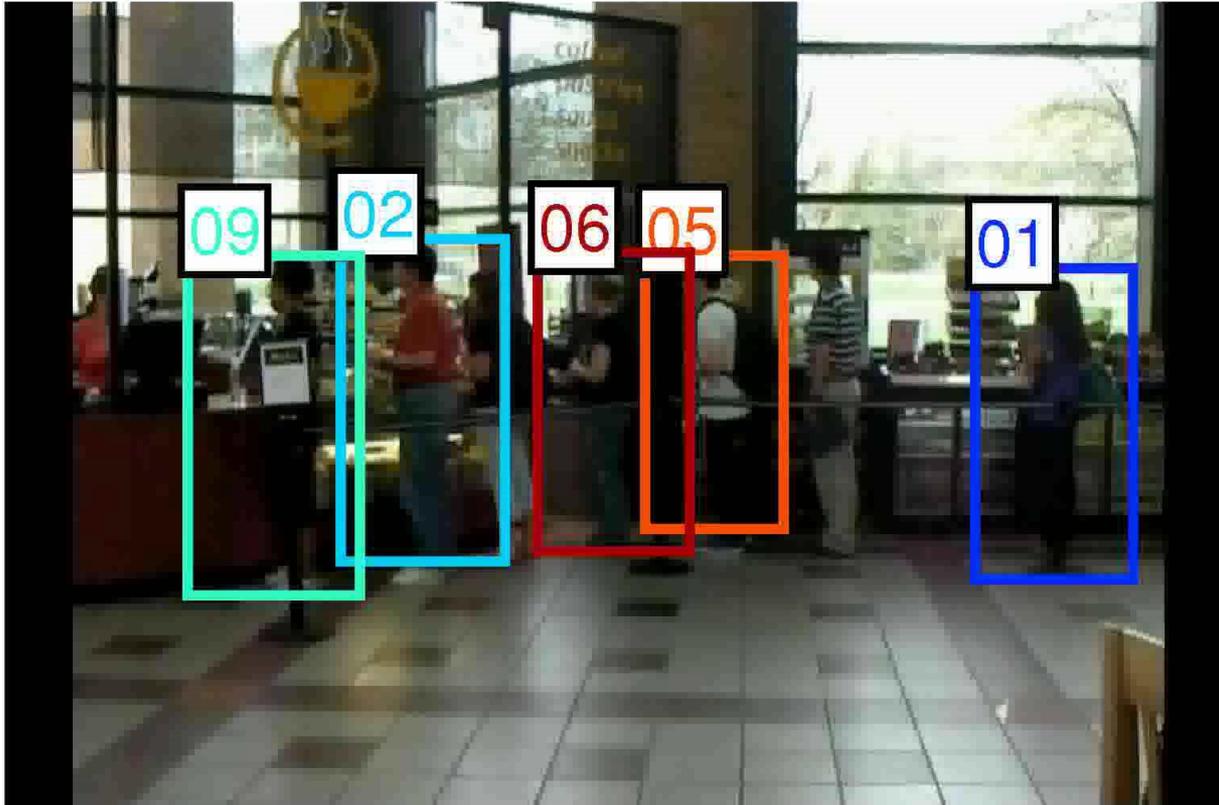
Pair-Interactions

- AP: approaching
-
- FE: facing-each-other
- SS: standing-side-by-side
- SQ: standing-in-a-queue

Example Classification Result



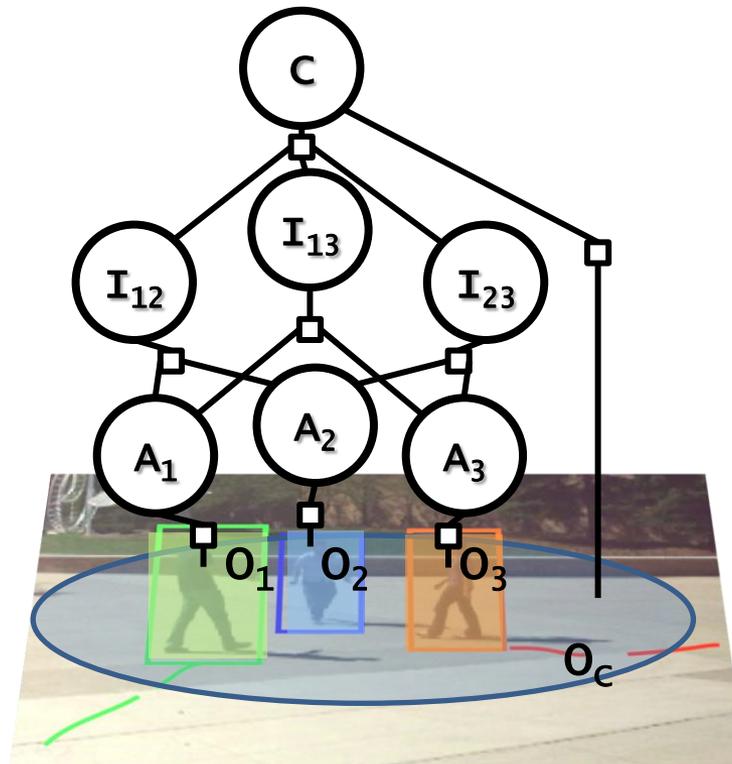
Example Classification Result



Tracklet Association
Color/nNumber: ID
Solid boxes: tracklets
Dashed boxes:
match hypothesis

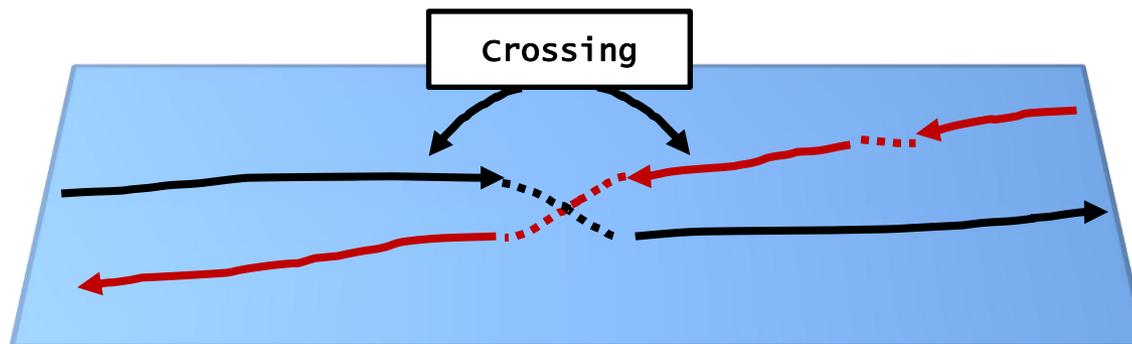
Conclusion

- Propose novel model for joint activity recognition and target association.



Conclusion

- Propose novel model for joint activity recognition and target association.
- High level contextual information help improve target association accuracy significantly.



Conclusion

- Propose novel model for joint activity recognition and target association.
- High level contextual information help improve target association accuracy significantly.
- Best classification results on collective activity up to date.

Thanks to

Yingze Bao, Byungsoo Kim, Min Sun, Yu Xiang,
ONR and anonymous reviewers

