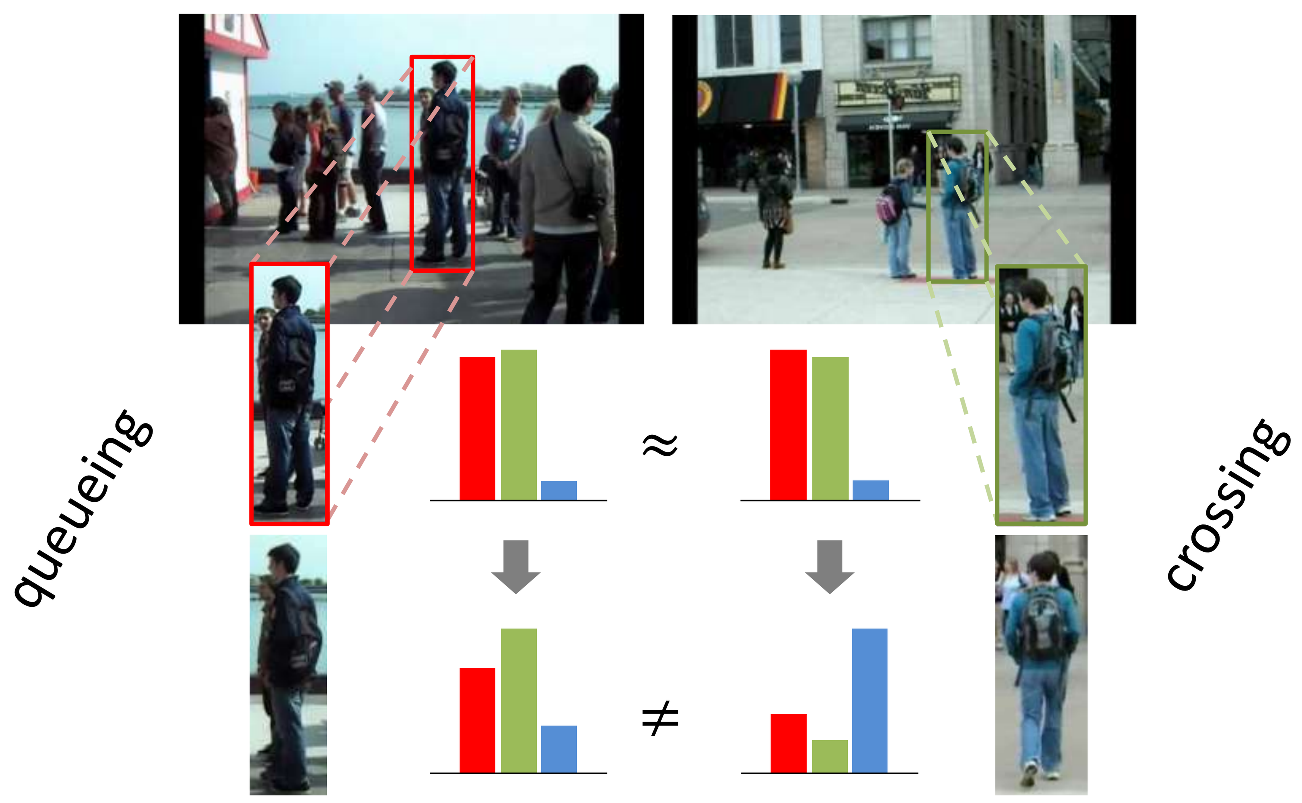
A Flow Model for Joint Action Recognition and Identity Maintenance

Sameh Khamis *Vlad I. Morariu* *Larry S. Davis*

CVPR 2012
Providence, Rhode Island

Motivation

- Different actions might have matching feature responses if they have visually similar bounding boxes.



- Action recognition can benefit from tracking, but tracking can also benefit from action recognition!
- We seek to improve action recognition performance by simultaneously solving both problems.

Proposed Approach

- Our goal is to formulate the problem as a tractable optimization function.
- The function should minimize
  - The action classification costs.
  - The identity association costs.
- The action classification cost is based on the Action-Context (AC) descriptor [3] using HOG as the underlying representation.
- The identity association cost penalizes appearance and action transition inconsistencies.
  - Appearances are modeled by a distance matrix learned using LMNN [4] between the blurred downsampled detection boxes as raw features.
  - Action transitions are modeled by a transition matrix learned by counting action pairs on the same track.
- We can leverage recently proposed formulations of tracking as network flow [5].

Model Formulation

- Our formulation can be represented as an integer linear program of a constrained minimum cost flow problem.

$$\min_{\{\mathbf{e},\mathbf{t},\mathbf{x}\}} \sum_i \sum_a \Big[(u_a(i) + \lambda_0) e_a(i) + \sum_{j \in \mathcal{P}(i)} \sum_b (u_a(i) + \lambda_d d(i,j) - \lambda_c \log(p_{ab})) t_{ab}(i,j) \Big]$$

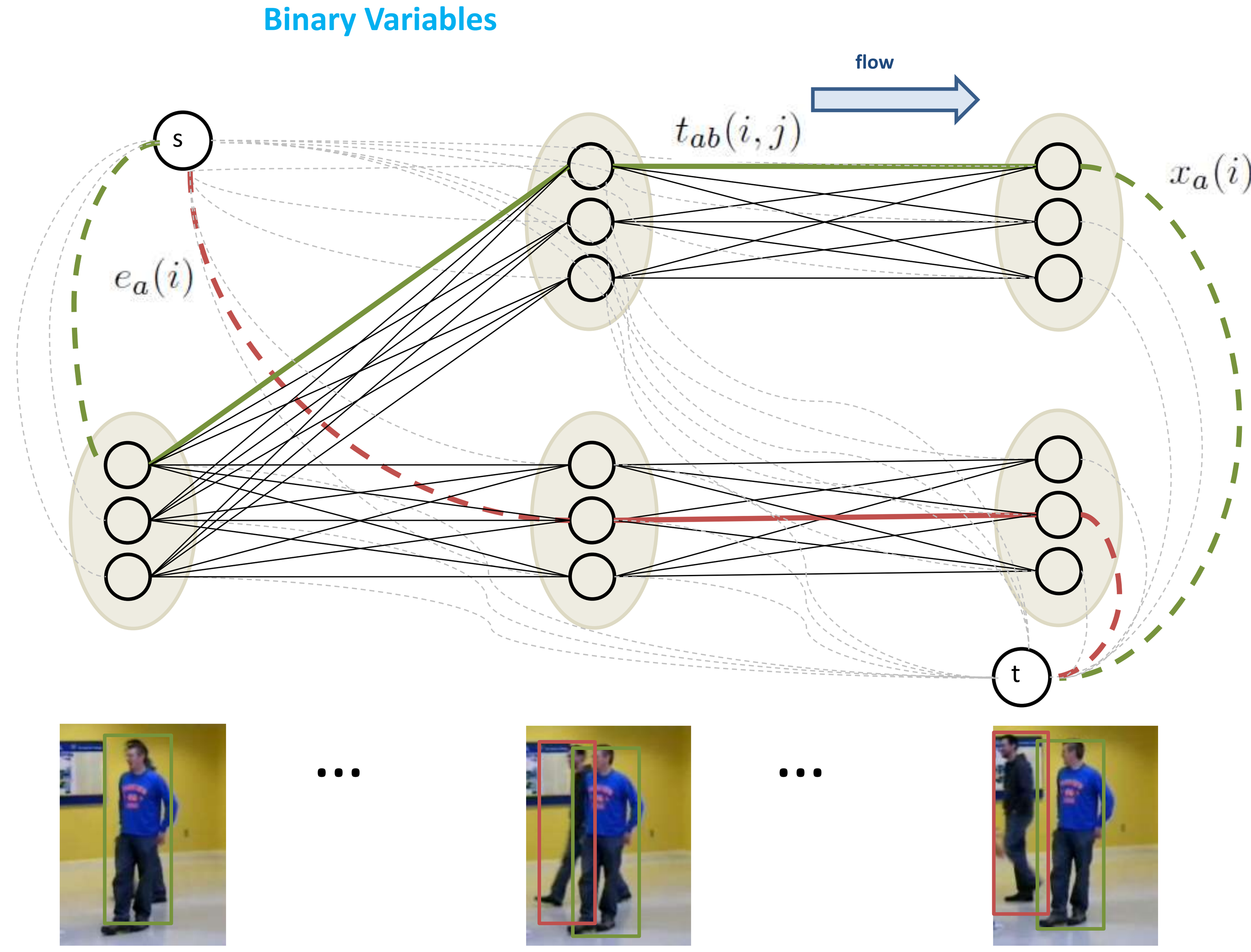
Appearance Consistency Weight, No Match Penalty, Action Transition Weight, Classification Cost, Appearance Consistency Cost, Action Transition Cost

$$\text{s.t.} \quad e_a(i) + \sum_{j \in \mathcal{P}(i)} \sum_b t_{ab}(i,j) = x_a(i) + \sum_{k \in \mathcal{S}(i)} \sum_c t_{ca}(k,i) \quad \forall i, a$$

$$\sum_a \Big[e_a(i) + \sum_{j \in \mathcal{P}(i)} \sum_b t_{ab}(i,j) \Big] = 1 \quad \forall i$$

$$\{\mathbf{e},\mathbf{t},\mathbf{x}\} \in \mathbb{B}^n$$

Enter Event, Transition Event, Exit Event, Flow Conservation Constraint, Explanation Constraint, Binary Variables



- Our ILP is constrained to the submodular polyhedron, therefore the constraint matrix is totally unimodular [2].
- Relax and solve!

Experimental Results

- We experimented on two public multi-person action recognition datasets [1]
- Our results improve on using unary potentials only and achieve state-of-the-art performance on both datasets

Approach	5-class	6-class
Classifier (AC Only)	68.8%	81.5%
Ours (AC + Tracking)	**70.9%**	**83.7%**

	crossing	waiting	queueing	walking	talking
crossing	67.9%	4.9%	2.0%	19.3%	1.7%
waiting	2.7%	58.1%	14.1%	10.4%	0.9%
queueing	4.2%	28.5%	78.5%	2.9%	5.9%
walking	24.6%	5.7%	1.4%	61.9%	3.6%
talking	0.5%	2.8%	4.1%	5.4%	87.9%

	crossing	waiting	queueing	talking	dancing	jogging
crossing	87.8%	5.8%	1.6%	4.0%	0.8%	0.5%
waiting	7.3%	57.5%	16.6%	1.5%	0.0%	0.1%
queueing	3.0%	30.3%	77.4%	4.6%	0.3%	4.1%
talking	0.4%	6.4%	4.1%	89.2%	1.4%	1.8%
dancing	0.3%	0.0%	0.2%	0.4%	97.0%	0.1%
jogging	1.2%	0.0%	0.0%	0.3%	0.4%	93.4%

References

1. W. Choi, K. Shahid, and S. Savarese. What are they doing?: Collective activity classification using spatio-temporal relationship among people. In VS, 2009.
2. S. Fujishige. Submodular Functions and Optimization. Elsevier Science, 2005.
3. T. Lan, Y. Wang, G. Mori, and S. N. Robinovitch. Retrieving actions in group contexts. In SGA, 2010.
4. K. Q. Weinberger and L. K. Saul. Fast solvers and efficient implementations for distance metric learning. In ICML, 2008.
5. L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In CVPR, 2008.