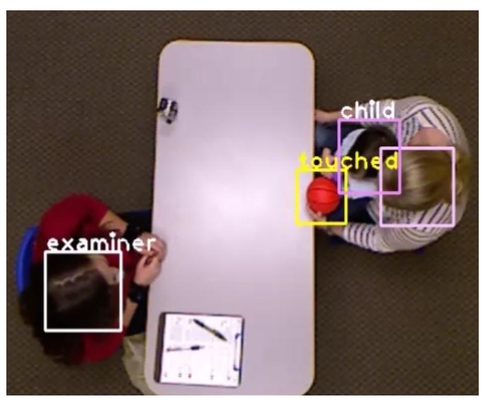
Tracking People and Objects with a Tracker Hierarchy and Template Ensembles for Engagement Estimation

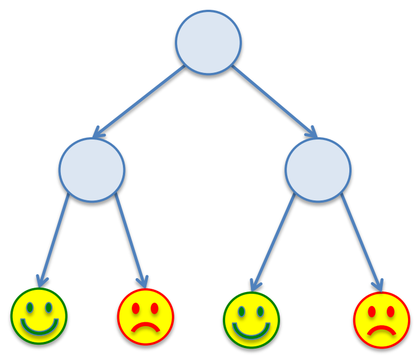
Jianming Zhang, Liliana Lo Presti, Stan Sclaroff



Detecting & tracking objects

Features	
Ball visibility	# frames the ball is detected
Ball touched	# frames the ball is touched
Ball touched by child	# frames the child touches…
…..	……

Feature extraction by trajectory analysis



Easy to engage vs. not easy to engage

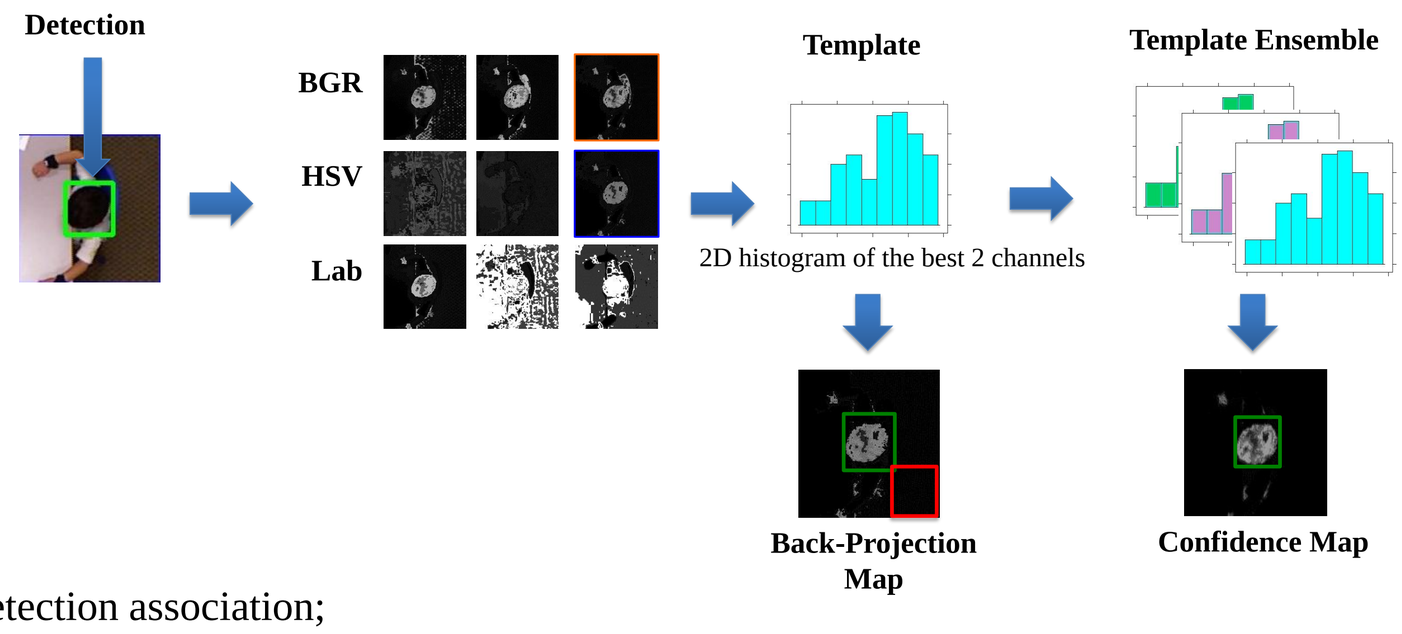
Template Ensemble and Tracker Hierarchy

Template Ensemble Management:

Based on the template scores, each tracker

- discards templates with negative scores;
- keeps at most the best N templates.

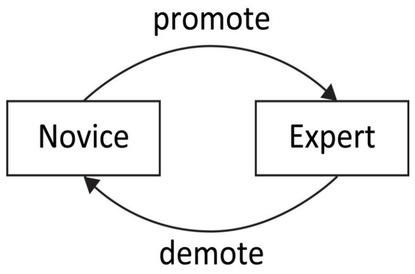
$$Score(f) = \underset{(i,j)\in f}{mean}(BP(i,j)) - \max_{w:w\cap f=\varnothing} \underset{(i,j)\in w}{mean}(BP(i,j))$$



Tracker Management:

- Tracker initialization based on frequency of the detection association;
- Online promotion/demotion of trackers to select the tracking strategy based on their demonstrated reliability.

Definition	Search Area	Tracking Strategy	Effect
Expert: a tracker with more than K templates	constant and small	mean-shift and Kalman filter	robust to false alarms
Novice: otherwise	increasing and large	links detection over time	fast recover of the tracking



Paper: J. Zhang, L. Lo Presti, S. Sclaroff, "*Online MultiPerson Tracking by Tracker Hierarchy*," IEEE AVSS, 2012.
Tracker source code: http://www.cs.bu.edu/groups/ivc/software/TrackerHierarchy/

"Easy to Engage" Score Prediction

Experiments on the RABC sessions to predict the "easy to engage" vs. "not easy to engage" for ball and book stages

- 16 training sequences (used in leave-1-out cross-validation);
- 15/14 test sequences.

Features	
✓Ball visibility	# frames the ball is detected
Ball touched by examiner	# frames the examiner touches the ball
Ball showed	# frames the examiner shows the ball
✓Ball on examiner side	# frames the ball is in the left side
✓Effort of the examiner	ball showed + ball touched by examiner

Raw Features → Mid-Level Features → Decision Tree

Confusion Matrix

Predicted \ Actual (On training set)	P	N
P	4	0
N	1	11

Predicted \ Actual (On test set)	P	N
P	2	1
N	0	11

Accuracy: in cross-validation: 93.75%, in testing: 92.85%

Features	
Book visibility	# frames the book is detected
✓Book touched	positive variations of skin in book area
✓Motion magnitude	of a sparse set of corners on the book
✓Histogram of Oriented Flow	of a sparse set of corners on the book

Raw Features → Linear SVM

Confusion Matrix

Predicted \ Actual (On training set)	P	N
P	4	3
N	5	3

Predicted \ Actual (On test set)	P	N
P	3	4
N	0	8

Accuracy: in cross-validation: 62.50%, in testing: 73.33%