MINIMIZING HUMAN EFFORT IN INTERACTIVE TRACKING BY INCREMENTAL LEARNING OF MODEL PARAMETERS

Georgia Tech College of Computing School of Interactive Computing

Arridhana Ciptadi and James M. Rehg

College of Computing, Georgia Institute of Technology

Robotics@GT & Intelligent Machines

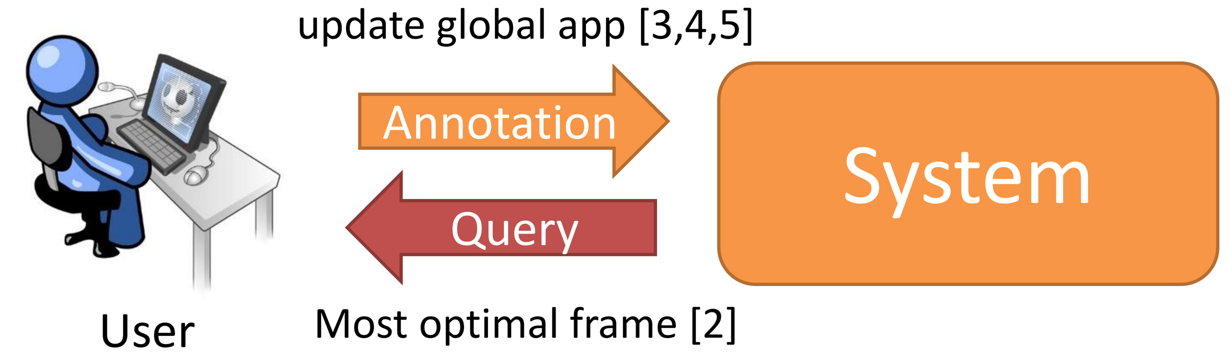
Project page: http://rehg.org/interactive-tracking/

Motivation

Minimizing human annotation effort (# of annotations per frame) is extremely important in interactive tracking.

More annotations = wasted resources!

Previous work



Intuition

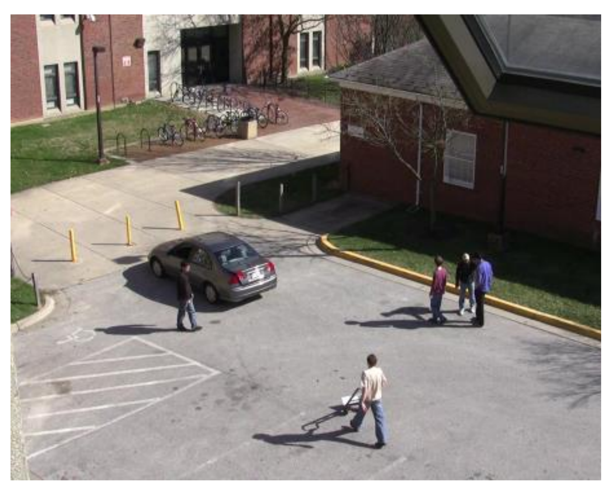
Suboptimal cost function parameters.

→ More tracking error!

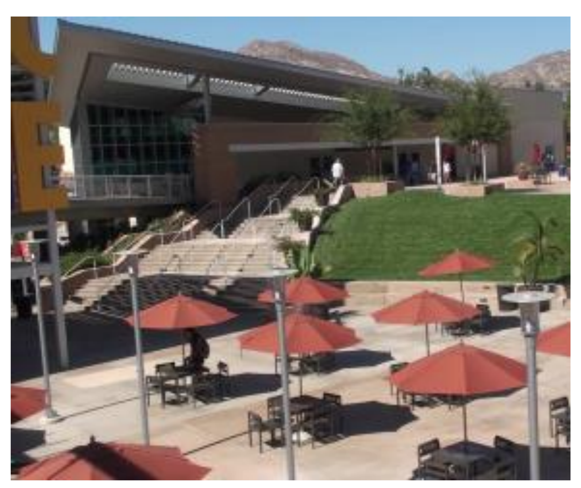
(requires more user annotation to fix)

Each tracking instance has different optimal parameters value.

→ Hand-tuning the parameters on a training set will not yield optimal results.



Instance 1



Instance 2

During the annotation process, incrementally learn ***instance specific*** model parameters for the tracking cost function.

Contributions

Novel annotation-driven maximum margin framework for efficiently learning instance-specific model parameters.

Problem

How to set the weight parameters of the tracking cost function?

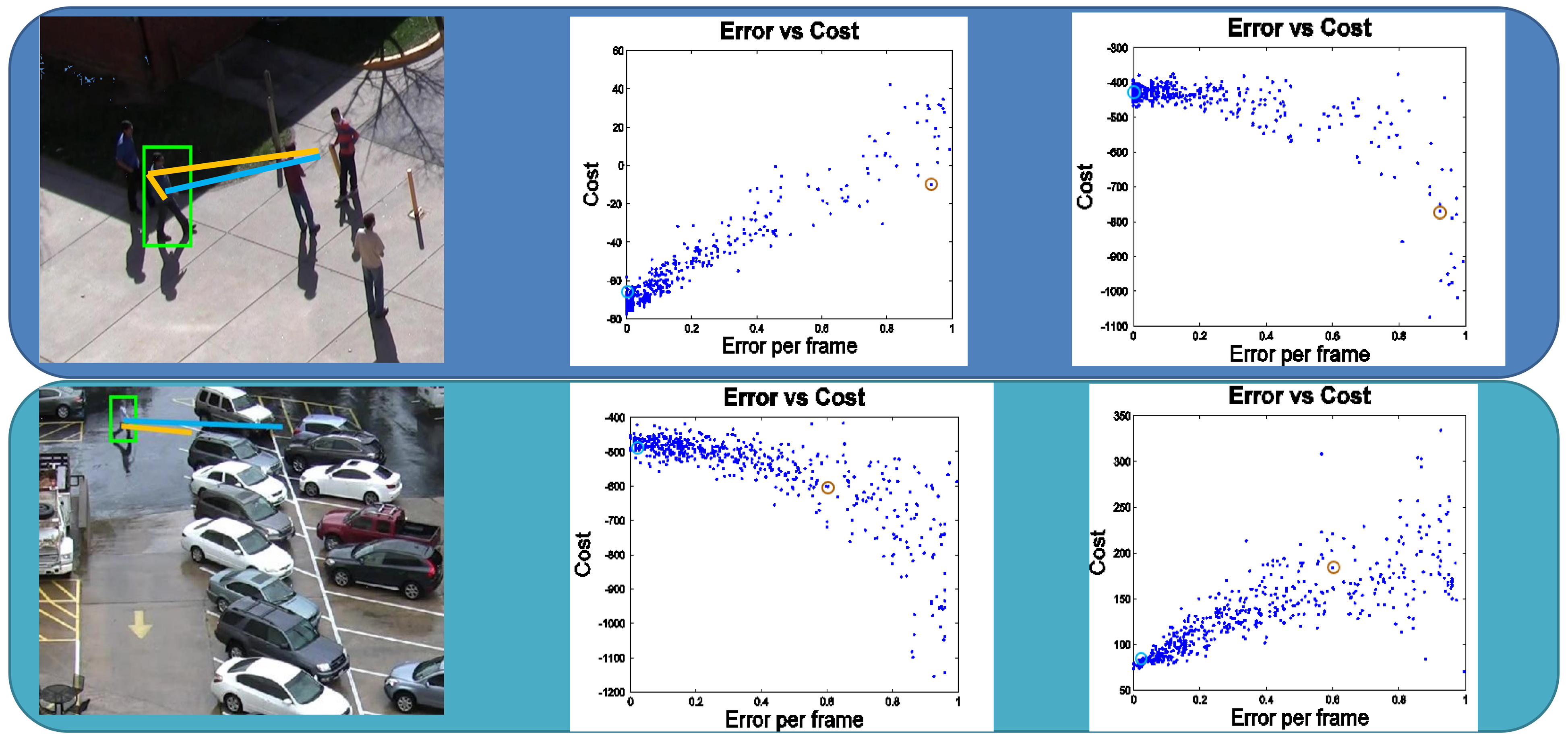
Tracking by detection: $E(Y;w) = \sum_t e(y_t;w)$

(Trajectory; Appearance & location at time t)

$$e(y_t;w) = w_1 d(y_t) + w_2 s_{app}(y_t, y_{t-1}) + w_3 s_{mot}(y_t, y_{t-1})$$

Global appearance cost; Local appearance similarity cost; Motion cost

Weights should be *instance specific*!



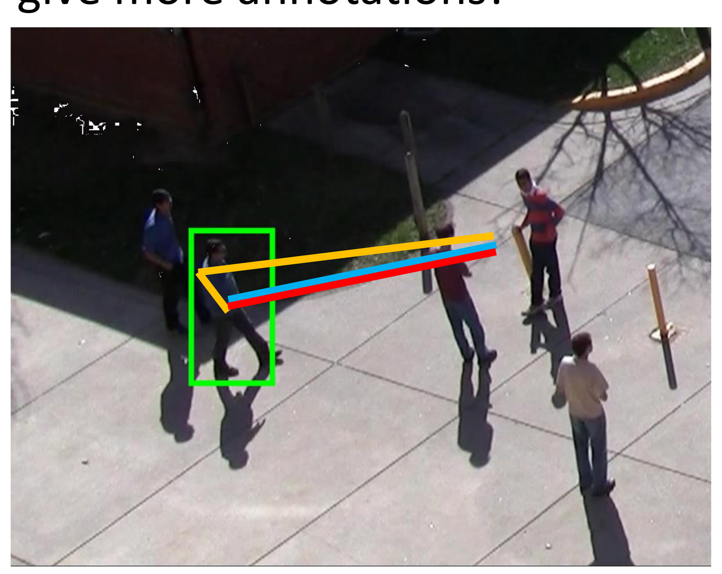
Low w_1, high w_2 and w_3 — High w_1, low w_2 and w_3

Solution

Exploit the incremental nature of interactive tracking

Each annotation results in a better track estimate.

→ Incrementally update w as the user give more annotations!



— Groundtruth trajectory

— Estimated trajectory after 2 annotations

— Estimated trajectory after 3 annotations

Max-margin formulation

$$\min_{w,\xi} \frac{1}{2}\|w\|^2 + \frac{C_1}{N}\sum_n \xi_n + C_2\left(E(Y^N;w) - E(Y^{N-1};w)\right)$$

$$E(Y^i;w) - E(Y^N;w) \ge \Delta(Y^i, Y^N) - \xi_n \qquad i = 1 \ldots N-1$$

$$w_j \ge 0 \qquad \forall w_j \in w$$

Trajectory estimate after i annotations

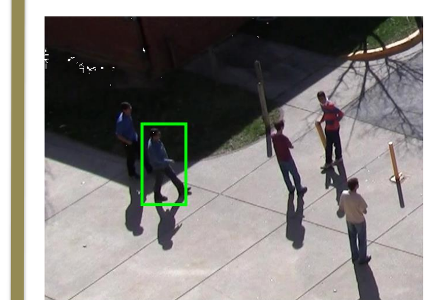
Current best trajectory estimate (after N annotations)

Search for the solution that maximizes separation between data points that are closest to the decision boundary

Results

Illustrative Example

Tracking an object (person) in a 300-frame sequence where there are many similar looking objects. Our approach quickly learn to put very little weight on the global appearance cost.

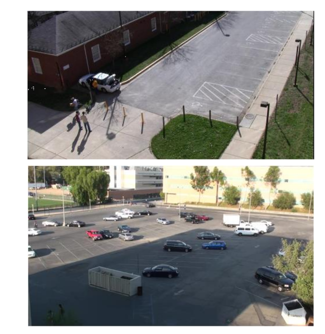


% annotations (compared to fixed)	w_1	w_2	w_3	Error/frame
10%	0.33	0.33	0.33	0.510
20%	0	0.49	0.51	0.15
30%	0	0.31	0.69	0

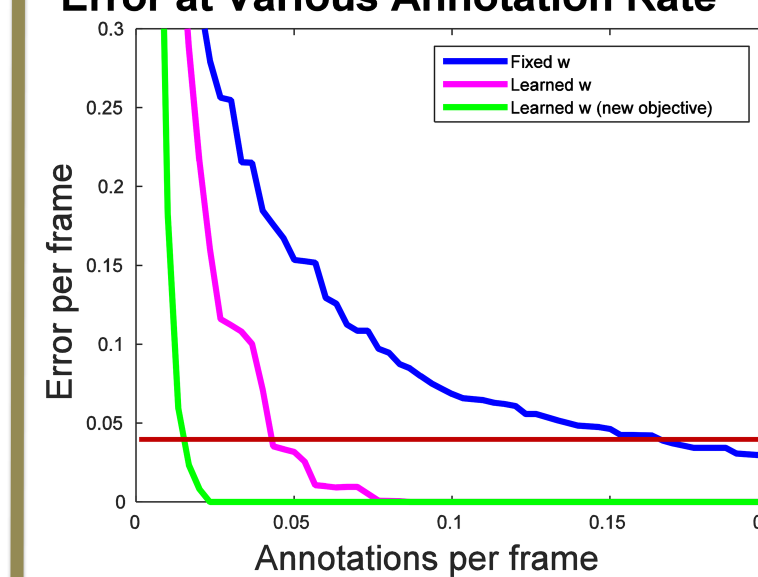
VIRAT Dataset [1]

300 surveillance video.

Task: track moving people and cars.



Error at Various Annotation Rate

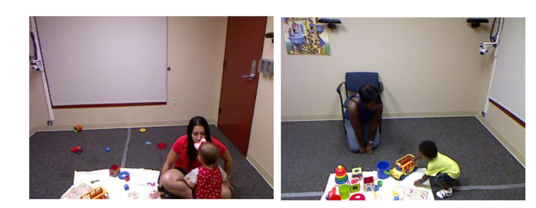


0.04 error-per-frame using only 0.017 annotations-per-frame compared to 0.17 using the fixed-weight approach. (90% savings)

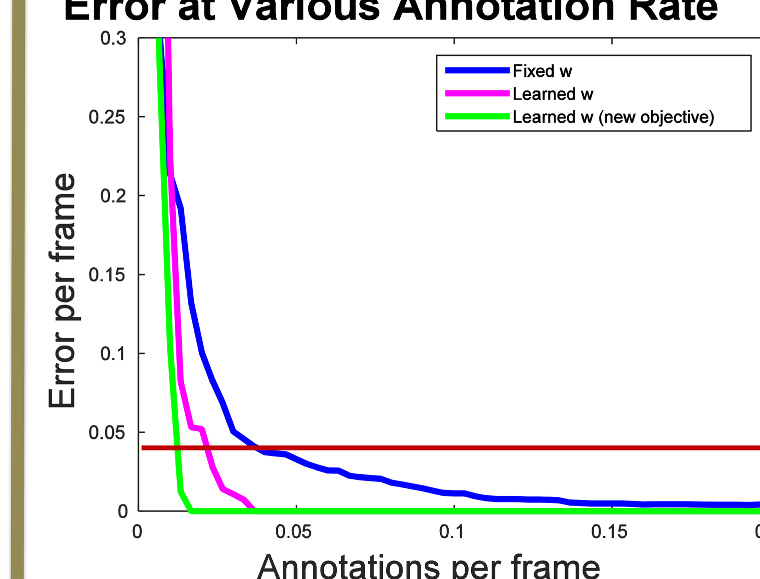
Infant-Mother Interaction Dataset

15 videos of infant-mother dyadic interaction.

Task: track the head of the people.



Error at Various Annotation Rate



0.04 error-per-frame using only 0.013 annotations-per-frame compared to 0.04 using the fixed-weight approach. (67.5% savings)

Conclusion

Easy-to-implement method for leveraging user annotations to set the cost function weight parameters. We have demonstrated on 2 real-world dataset that this approach saves a significant amount of annotation effort.

References:

[1] S. Oh, et. al. A large-scale benchmark dataset for event recognition in surveillance video. CVPR (2011).

[2] C. Vondrick and D. Ramanan. Video annotation and tracking with active learning. NIPS (2011).

[3] A. Buchanan and A. Fitzgibbon. Interactive feature tracking using kd trees and dynamic programming. CVPR (2006).

[4] Y. Wei, J. Sun, X. Tang, and H.-Y. Shum. Interactive offline tracking for color objects. ICCV (2007).

[5] C. Vondrick, D. Patterson, and D. Ramanan. Efficiently scaling up crowdsourced video annotation. IJCV (2013).