F3: Tracking in Large Public Spaces

Mustafa Ayazoglu, Caglayan Dicle, Binlong Li, Fei Xiong, Octavia Camps, Mario Sznaier
ayazoglu.m@neu.edu, dicle.c@neu.edu, li.b@neu.edu, xiong.f@neu.edu, camps@coe.neu.edu, msznaier@coe.neu.edu

Northeastern University

Abstract

This research proposes a comprehensive "robust tracking" framework that can substantially enhance automatic surveillance systems. The proposed approach exploits the dynamics of the target and associated "contextual" objects across multiple cameras. The use of multiview contextual dynamics allows for persistent tracking through long occlusions, even if involving targets with similar appearances. The proposed techniques require neither rearrangement nor calibration of already deployed cameras.

Relevance

Tracking is a core module for automatic surveillance systems.

•Current surveillance systems have difficulties handing occlusion/ similar appearance / calibration.

• Our method provides an unified solution for addressing these three critical problems.

• Transferring this technology to industry may have an immediate impact on surveillance systems.

• End users include law enforcement agencies, security service companies, major transit services & airports, high-volume venues.

Technical Approach

Multicamera Information ← **Tracking in Large Spaces** → **Context Information**

Background:

n^{th} order autoregressive model

$y_k = a_1 y_{k-1} + a_2 y_{k-2} + \ldots + a_n y_{k-n}$

It's associated Hankel matrix

$$H_y^{s,r} \doteq \begin{bmatrix} y_0 & y_1 & \cdots & y_r \\ y_1 & y_2 & \cdots & y_{r+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_s & y_{s+1} & \cdots & y_{r+s} \end{bmatrix}$$

holds

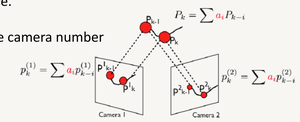
$$H_y^{s,n} \begin{bmatrix} a \\ -1 \end{bmatrix} = 0$$

Fusing Information from Multiple Cameras:

All the 2D affine projections of the 3D trajectory of a target, captured simultaneously by a set of affine cameras, lie on a single subspace.

$$H_{y^{(i)}}^{s,n} \begin{bmatrix} a \\ -1 \end{bmatrix} = 0 \qquad i = 1, \ldots, M$$

i denotes the camera number



How to do it?:

When the target is occluded in a view, it is possible to Information from other cameras as dynamic and epipolar geometry constraints to estimate the current location of the target in the occluded view and use this estimate to predict the location of the target in the next frame in all views

$$A = \begin{bmatrix} H_{y^{(1)}}^{s,n} & 0 & 0 \\ H_{y^{(2)}}^{t,n} & 0 & 0 \\ \hline Y_{k-1-n:k-1}^{(2)} & -I_{2\times 2} & 0_{2\times 1} \\ \hline 0_{1\times n} & \ell_1 & \ell_2 & -\ell_3 \end{bmatrix} \quad b = \begin{bmatrix} \text{vect}\left(Y_{k-n+1:k}^{(1)}\right) \\ \text{vect}\left(Y_{k-n:k-1}^{(2)}\right) \\ \hline \\ \hline \end{bmatrix}$$

DYNAMIC CONSTRAINTS

Estimation

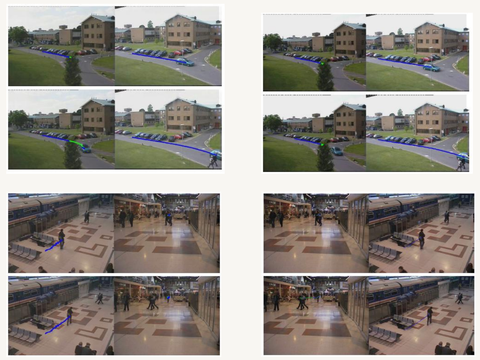
Epipolar constraints

$$x = \begin{bmatrix} a \\ y_k^{(2)} \end{bmatrix}; \; Y_{j:k}^{(i)} = \begin{bmatrix} y_j^{(i)T} & y_{j+1}^{(i)T} & \cdots & y_k^{(i)T} \end{bmatrix}^T$$

$$\begin{bmatrix} \ell_1 & \ell_2 & \ell_3 \end{bmatrix} = \begin{bmatrix} y_k^{(1)T} & 1 \end{bmatrix} F \qquad Ax = b$$

Multiview Tracking

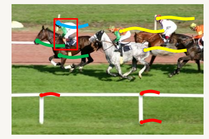
OUR METHOD | COMPETING METHOD



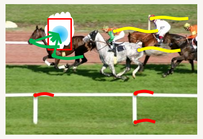
Contextual Tracking

Legends: Ground Truth; ST[3] ; RT[1]; Proposed RST.

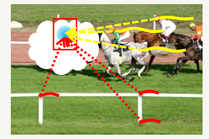
Fusing information from Context Features



A frame with the target marked by a red rectangle and feature trajectories.



The rigidly coupled supporters are helpful to estimate the position of the occluded target.

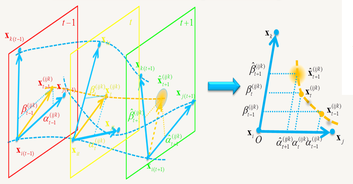


When there are no rigidly correlated features, the target motion can still be modeled and estimated.

How to do it?:
Local Autoregressive Dynamic Models

Local Affine Reference

$$x_t - x_{it} = \begin{bmatrix} \alpha_t & \beta_t \end{bmatrix} \begin{bmatrix} x_{jt} - x_{it} \\ x_{kt} - x_{it} \end{bmatrix}$$



Autoregressive Dynamic Models

$$\Delta\hat{x}_t^{(ijk)} = \underset{\Delta w, \Delta x_t^{(ijk)}}{\arg\min} H_{\Delta x^{(ijk)}}$$

$$\text{where} H_{\Delta x^{(ijk)}} \doteq \begin{bmatrix} \Delta x_1^{(ijk)} & \Delta x_2^{(ijk)} & \cdots & \Delta x_c^{(ijk)} \\ \Delta x_2^{(ijk)} & \Delta x_3^{(ijk)} & \cdots & \Delta x_{c+1}^{(ijk)} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta x_{t-c+1}^{(ijk)} & \Delta x_{t-c}^{(ijk)} & \cdots & \Delta x_t^{(ijk)} \end{bmatrix}$$

$$\hat{x}_t^{(ijk)} = \hat{x}_{t-1}^{(ijk)} + \Delta\hat{x}_t^{(ijk)}$$

Vote by Rank Minimization Estimates

$$p(x_t|(x_{it}, x_{jt}, x_{kt})) \sim \frac{1}{||H_{\Delta x^{(ijk)}}||_*} N(x_t|\hat{x}_t^{(ijk)}, \Sigma)$$

$$p(x_t|I_t) = \sum_{i \in S} p(x_t|(x_{it}, x_{jt}, x_{kt})) p((x_{it}, x_{jt}, x_{kt})|I_t)$$

$$\hat{x}_t = \underset{x_t}{\arg\max}\, p(x_t|I_t)$$

Accomplishments Through Current Year

•Algorithm Development
•Unified Theory
•Early testing of the algorithm and comparison with existing methods

Future Work

Research is currently underway seeking to extend these results to

•Perspective cameras.
•Multiple Targets.
•Optimized Implementation

Opportunities for Transition to Customer

Most large public spaces are already equipped with surveillance cameras. Our method can substantially improve tracking robustness in large, crowded environments where long occlusions are frequent. Further, this improvement is achieved without the need to add equipment or recalibrate and redeploy existing cameras.

Publications Acknowledging DHS Support

•Ayazoglu M., Li B., Dicle C., Camps O. and Sznaier M.: Dynamic Subspace-Based Coordinated Multicamera Tracking In IEEE ICCV, 2011.

Other References

1. Ding, T., Sznaier, M., Camps, O.: Receding horizon rank minimization based estimation with applications to visual tracking. In: CDC. (2008) 3446 –3451
2. Yang, M., Wu, Y., Hua, G.: Context-aware visual tracking. IEEE TRANS. ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 31 (2009) 1195–1209
3. Grabner, H., Matas, J., Van Gool, L., Cattin, P.: Tracking the invisible: Learning where the object might be. In: CVPR. (2010) 1285 –1292
4. Z. Wu, N. I. Hristov, T. L. Hedrick, T. H. Kunz, and M. Betke. Tracking a large number of objects from multiple views. In ICCV, 2009.

This material is based upon work supported by the U.S. Department of Homeland Security under Award Number 2008-ST-061-ED0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of the U.S. Department of Homeland Security.