Improved Tracking by Decoupling Camera and Target Motion

Shawn Lankton Allen Tannenbaum
Georgia Institute of Technology, Atlanta, GA

Summary

Challenges

Moving Camera: Camera motion due to pans, tilts, or a moving platform can introduce large motion not associated with the object.

Disappearing Objects: Especially when camera motion is present, objects may come and go from the field of view. These objects should be uniquely reacquired when they return.

Frames 20, 30, 40, 50, 60, 70, 80, and 90 of the AERIAL sequence.

Our Contribution

Decoupling Motion: We estimate object and camera motion separatelyz

Completing Kalman Filtering: Estimating object motion with a Kalman filter and using camera motion as a control input completes the Kalman filter and creates a more principled system.

Inspiration

Mean Shift

Fast & Simple: Mean shift is a popular localization method which provides an efficient way to find a local image region whose histogram best matches some target histogram.

Basin of Attraction Dependence: This local method must be initialized near the object of interest in order to converge to the correct place. This is a major limitation of mean shift.

Global Motion Estimation

Equivalence: The global motion is equivalent to the inverse of camera motion when the object motion is relatively small.

Efficient Computation: We employ an efficient frame registration technique that uses only a fraction of image pixels to increase speed with almost no loss in accuracy.

Combining in a Kalman Framework

The Kalman filter is often used in tracking to add dynamics to a system. In the simplest form, the discrete Kalman filter describes the state $\mathbf{x} \in \mathbb{R}^n$, at iteration k, in terms of the state at the previous iteration, $k-1$, and a control signal $\mathbf{u}$. This relationship

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_{k-1}$$

is supplemented by a measurement $\mathbf{z} \in \mathbb{R}^m$

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k$$

In most tracking computer vision tracking applications, the $\mathbf{B}$ term is omitted by assuming that $\mathbf{u}$ is always 0. In this work, we use the control signal, $\mathbf{u}$ to incorporate the global motion estimate into the filter.

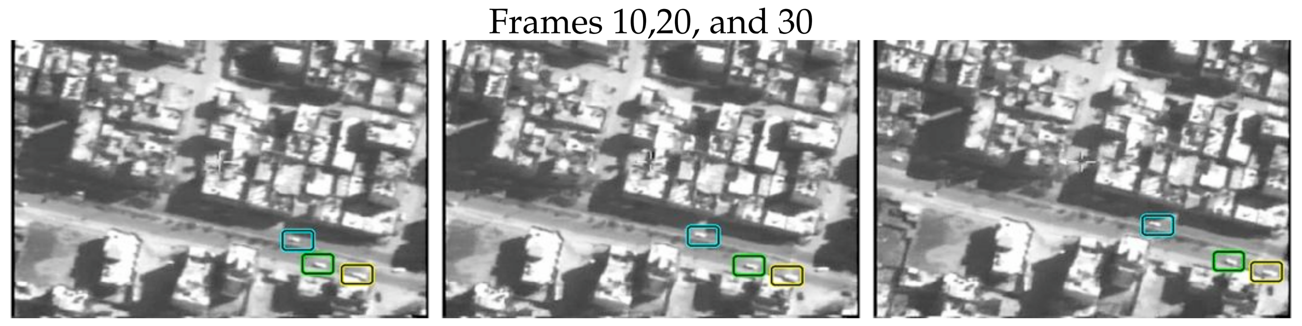
Result

The predicted target position at each frame is the previous location plus a decoupled combination of target motion and camera motion estimates.

Experiment: Disappearing Objects

In the DISAPPEAR sequence two of the tracked cars leave view. While out of sight, their position estimate is updated based on camera motion and the objects' motion model. Based on these two motion estimates, the system predicts where and when the target will come back into view, and reacquires track.

Frames 10,20, and 30



Frames 40,50, and 60

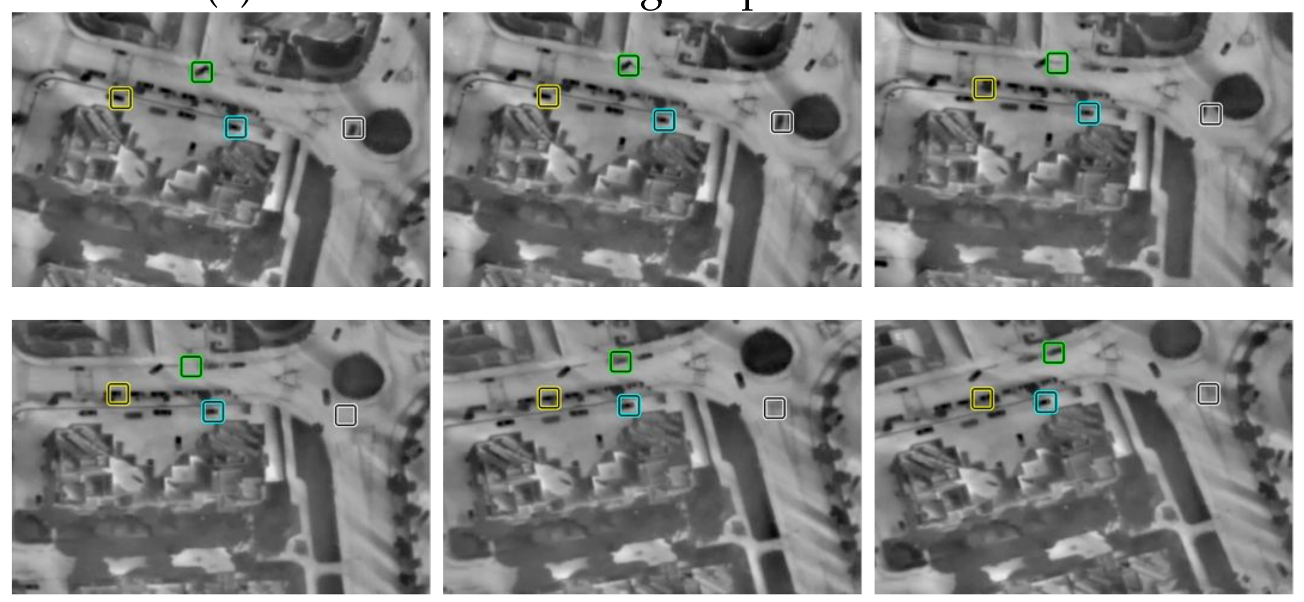
Frames 70,80, and 90

The cars tracked in green and yellow disappear for approx. 50 frames

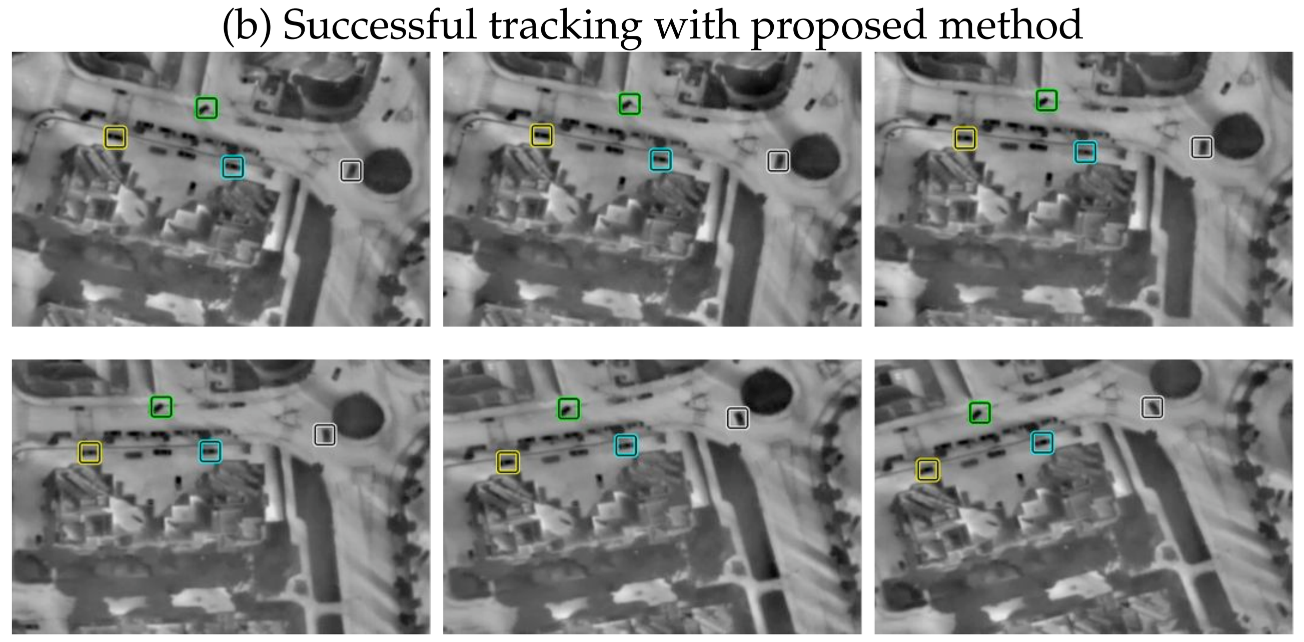
Experiment: Frame Rate

The FRAMERATE sequence demonstrates the technique's ability to maintain track despite large camera motion. Notice the shifting position of the background throughout these six consecutive frames.

(a) Failure to track using coupled motion model



(b) Successful tracking with proposed method



(six consecutive frames shown in each experiment)

Robustness against these large jumps allows fewer frames to be processed freeing resources for other computations.

Efficiency

The chosen frame registration algorithm and mean shift are both efficient techniques. Thus the overall system is capable of tracking in real time. Our prototype Matlab implementation achieved speeds of at least 15 fps.

Sequence	Frame Rate
FRAMERATE	15.15fps
DISAPPEAR	16.87fps

slankton@ece.gatech.edu