Efficient Tracking of Many Objects in Structured Environments

Nathan Jacobs¹, Michael Dixon¹, Scott Satkin², and Robert Pless¹

¹Computer Science and Engineering
Washington University in St. Louis
{jacobsn,msd2,pless}@cse.wustl.edu

²Robotics Institute
Carnegie Mellon University
satkin@cmu.edu

Abstract

We address the problem of tracking objects in scenes with highly-structured motion patterns.

System overview:
- Find the lanes of travel using a novel active contour energy function that is based only on accumulated pixel-level statistics.
- Simultaneously track dozens of objects using only pixels at the center of the lanes.

We present quantitative results in the context of vehicle tracking in urban environments.

Road Finding using Average Local-Motion and Active Contours

Two-stage road finding algorithm:
1. Collect image derivative statistics from video (no tracking required).
2. Fit active contours using the accumulated data.

Iteratively add active contours:
1. Initialize a snake at a region with a high road-score.
2. Optimize individual snake energy function (see below).
3. Set road-score to zero in region covered by the optimized snake.
4. Repeat if regions with high road-score remain.

Objective function of a single active contour:

\[ E_{total} = E_{int} + E_{ext} \]

\[ E_{int}(\phi) = -\alpha \int |\nabla \phi| \, dx + \beta \int |P(\phi)| \, dh \]

\[ E_{ext}(\phi) = -\gamma \int R(\phi(x))^4 \, dx + \int \|V(\phi(x)) - \nabla \phi\| \, dh \]

Tracking Objects in Space-Time Sheets

Given space-time sheets from every lane in the camera, our tracking system extracts a set of partial object tracks, or tracklets.

To do so, it first identifies an initial set of object detections using a color background model and local-motion cues to find moving objects.

It then generates a set of tracklets by temporally extending the detections forward and backward in time using a template-based appearance model coupled with an acceleration penalty. The final stage of our system combines the tracklets detected from different sheets throughout the camera network into a set of complete object tracks.

This data association problem is expressed as a MAP estimation over the posterior probability of the set of tracks, \( T \), given the tracklets, \( S \). This can be formulated as a min-cost flow problem and solved optimally [Zhang 2008].

\[ T^* = \arg\max_T P(S | T) \]

\[ = \arg\max_T P(S | T) P(T) \]

\[ = \arg\max_T \prod_i P(S_i | T_i) \prod_i P(T_i) \]

\[ = \arg\min_T \sum_i - \log P(S_i | T_i) + \sum_i - \log P(T_i) \]

Tracking Evaluation

The evaluation dataset, NGSIM Peachtree, was captured through a collaboration of researchers interested in modeling vehicle behavior. Cameras were set up in an urban environment and 15 minutes of data were recorded in each.

Conclusions

An automatic end-to-end tracking system:
- Finds roads without tracking using image derivatives.
- Uses roads to reduce 2D tracking problem to 1D.
- Tracks dozens of cars in real-time using very little image data (only 1% in our experiments).

Future work
- Improve tracklet generation by using longer time frames.
- Extend to additional domains, such as cell tracking.