Coupled Detection and Trajectory Estimation for Multi-Object Tracking

Bastian Leibe[1], Konrad Schindler[1], Luc Van Gool[1,2]

[1]Computer Vision Laboratory, ETH Zurich, Switzerland

[2]VISICS, KU Leuven, Belgium

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Motivation

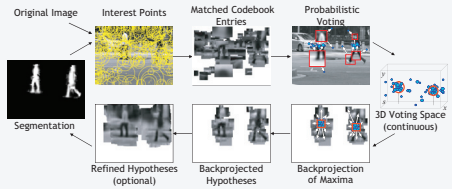
- Multi-object tracking-by-detection
- Improve robustness by coupling object detection and tracking
  - Strong object model + feedback from trajectory estimation to detection
- Global optimization to resolve trajectory interactions
  - Incorporate real-world physical constraints
  - Computationally feasible by formulation as hypothesis selection

Contributions

"We propose a novel approach for multi-object tracking that couples detection and trajectory estimation in a combined global optimization framework. At each time instant, our approach tries to find a globally optimal combined solution that provide the best explanation for the current image and all previous frames, while incorporating physical constraints such that no two objects may occupy the same space, nor explain the same image pixels at the same time."

Object Detection

ISM Recognition [CVPR'05]



- Hypothesis selection
  - Solve a Quadratic Boolean Optimization Problem
  - Constraint: each pixel may at most belong to a single detection.

$$\max_{n} \; n \; S \; n$$

support (Σ pixels)

interactions (Σ pixels)

Spacetime Trajectory Estimation

Spacetime Event-Cone Following [CVPR'07]

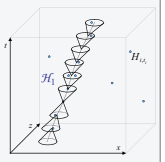
- Trajectory growing
  - Start from each detection (at each time step)
  - Collect detections in event cone
  - Evaluate under trajectory model (EKF)

$$p(H_{i,t+1}|\mathcal{H}_{t_0:t}) = p(H_{i,t+1}|\mathcal{A}_t)p(H_{i,t+1}|\mathcal{D}_t)$$

  - Adapt trajectory

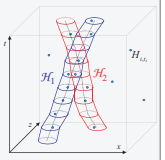
$$\mathbf{x}_{t+1} = \frac{1}{Z}\left(e^{-\lambda}\mathbf{x}^p_{t+1} + \sum_i p(H_{i,t+1}|\mathcal{H}_{t_0:t})\mathbf{x}_i\right)$$

  - Iterate



- Hypothesis selection
  - Solve a Quadratic Boolean Optimization Problem

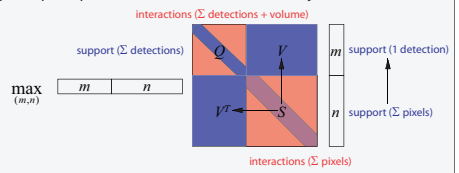
$$\max_{m} \; m \; Q \; m$$



support (Σ detections) interactions (Σ detections + volume)

Coupled Detection & Trajectory Estimation

- Basic idea:
  - Couple the two optimization problems into a single one.
  - Move support for current detections into coupling terms.
- Coupling terms:
  - Express support for certain trajectories from new detections.
  - Express spatial prior for detection locations from trajectories.



$$\max_{(m,n)} \; m \; n \qquad \begin{bmatrix} Q & V \\ V^T & S \end{bmatrix} \begin{matrix} m \\ n \end{matrix}$$

- Problem: Asymmetric relationship
  - Trajectories rely on continuing detections for support.
  - But detections can exist without supporting trajectories (e.g. when a new object enters the scene).

⇒ Introduce *virtual trajectories* v with interaction matrix R.

  - Enable detections to survive without contributing to an existing trajectory

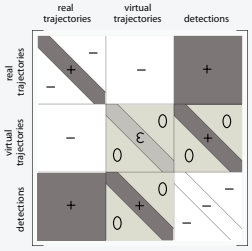
$$\max_{(m,v,n)} \; m \; v \; n \qquad \begin{bmatrix} \tilde{Q} & U & V \\ U^T & R & W \\ V^T & W^T & \tilde{S} \end{bmatrix} \begin{matrix} m \\ v \\ n \end{matrix}$$

- Interpretation
  - W: Support of each detection (for virtual trajectory)
  - V: Coupling between detections and real trajectories
  - U: Exclusion constraints

Problem: Matrix Structure

- Optimization problem no longer submodular

$$\begin{bmatrix} \tilde{Q} & U & V \\ U^T & R & W \\ V^T & W^T & \tilde{S} \end{bmatrix} = \begin{bmatrix} - & - & + \\ - & \varepsilon & + \\ + & + & - \end{bmatrix}$$



Iterative Solution

- Fix trajectories, solve

$$\max_{n} \left[n^T\left(R+S+2\,\mathrm{diag}(V^T m)+2\,\mathrm{diag}(W^T n^-)\right)n\right]$$

- Fix detections, solve

$$\max_{m} \left[m^T\left(Q+2\,\mathrm{diag}(Vn)+2\,\mathrm{diag}(Uv)\right)m\right]$$

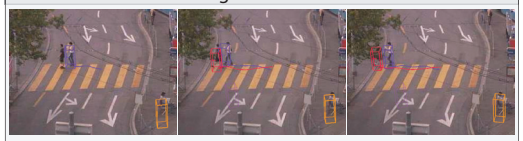
- (Repeat if necessary...)

Effects of Coupled Optimization

Spatial Detection Prior from Trajectories

Modeled as a Gaussian around projected object location.

Non-Markovian Tracking



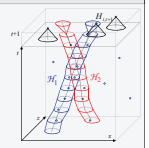
- Hypothesize-and-select paradigm
  - Tracks can adjust according to new evidence
  - The system can automatically recover from mismatches and errors.

⇒ ***Just need to make sure correct hypotheses are among the candidates, the model selection framework will (ideally) take care of the rest!***

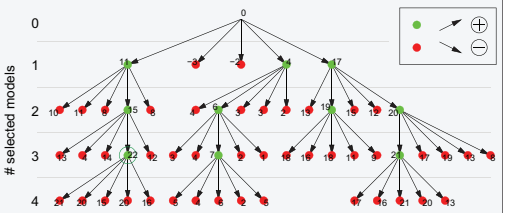
Efficient Implementation

Incremental Computation

- Trajectories can be constructed incrementally
  - Try to extend old trajectories by new evidence.
  - Only grow new trajectories from the last n frames.
- Interaction matrix S can be reused
  - Many entries don't need to be updated.
  - Just need to be weighted with temporal discount.



Multibranch Gradient Ascent



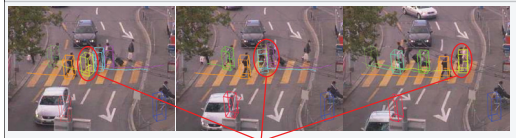
- Bounded breadth-first search for first steps
  - Guaranteed to be at least as good as greedy with constant overhead.
  - Empirically better performance than Taboo search or LP relaxation.

Limitations of the Approach

- Based on output of object detector
  - Will generate false tracks where detector produces consistent false alarms.
  - Will lose tracks if detector fails to return detections (e.g. in long occlusion).
- Track history may change over time as a result of model selection.
  - This is our strength! – but which track version should be evaluated?

Experimental Results

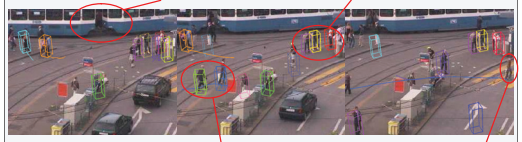
Qualitative Results (also see Videos)



Tracking through occlusion

Recovery from mismatches/lost tracks

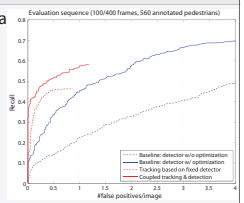
Large-scale background changes Static objects



Automatic track initialization Tracking over long sequences

Quantitative Results

- Challenging surveillance video data
  - Public webcam in Zurich
  - Low-resolution video (320x240)
  - Strong MPEG compression artifacts
- Results for PedXing sequence
  - Annotated every 4th frame
  - Evaluated bounding box overlap (not identities)



Conclusions

- Proposed framework for non-Markovian multi-object tracking.
- Formulation as coupled model selection problem.
- Approach has several interesting properties:
  - Tolerates large-scale background changes.
  - Can track a large number of static and moving objects.
  - Able to recover from errors and temporarily lost tracks.
  - Global optimization resolves trajectory interactions.

New Results

- Combination with moving camera system:

A. Ess, B. Leibe, L. Van Gool,
"Depth and Appearance for Mobile Scene Analysis"
in ICCV'07, Rio de Janeiro, Oct. 2007.

⇒ Ask to see the video!