Wide-Area Tracking

Amit K. Roy-Chowdhury
University of California, Riverside
Overview of the Talk

- Fundamentals of Tracking
- Challenges in Multi-Target Tracking
- Some Basic Tracking Approaches, their strengths and limitations
- Critical review of a few recent wide-area tracking methods
- Directions of future research
What is Visual Tracking

Generating an inference about the motion of an object given a sequence of images.
Challenges

- Illumination Change
- Occlusion
- Change of Appearance and Pose
- Clutter
Additional Challenges for Multiple Targets

- Data Association
- Interactions between Targets
  - Missed Detections/False Detections
  - Switching between Tracks
Challenges Specific to Multiple Cameras

- All of the previous.
- In addition – Handoff problem.
- Becomes extremely challenging without training data especially in non-overlapping views.
Dynamic Bayesian Estimate

Model:
\[ X_k = f_k(X_{k-1} + v_{k-1}) \]  
System/Transition model
\[ Y_k = h_k(X_k + n_k) \]  
Measurement/Output model

\[ X = \text{State vector} \]
\[ Y = \text{Measurement vector (From Detection Stage)} \]
\[ v = \text{i.i.d. process noise sequence} \]
\[ n = \text{i.i.d. measurement noise sequence} \]
\[ K = \text{Time instant} \]
Intuition

Figure courtesy: Kristen Grauman - UT Austin
Solution Strategy

- Assume, Posterior density \( p(X_{k-1}/Y_{1:k-1}) \) is known.
- **Prediction Step:** Use following equation to compute the prior:
  \[
p(X_k|Y_{1:k-1}) = \int p(X_k/X_{k-1})p(X_{k-1}/Y_{k-1})dX_{k-1}
\]
- **Update step:** Compute the Posterior distribution using Bayes rule.
  \[
p(X_k|Y_{1:k}) = \frac{p(Y_k|X_k)p(X_k|Y_{1:k-1})}{Z_k}
\]
  \[
  Z_k = \int p(Y_k|X_k)p(X_k|Y_{1:k-1})dX_k
  \]

[Normalizing Constant]
Kalman Filter

- Optimal with Linear Gaussian models.

But for Non-Gaussian and/or Non-Linear model?
Known: $p(X_{t-1}|Y_{1:t-1})$ i.e. $\{x_{t-1}^{(i)}, w_{t-1}^{(i)}\}$

To be known: $p(X_t|Y_{1:t}) = \{x_t^{(i)}, w_t^{(i)}\}$

Easy to get $x_t^{(i)} \rightarrow$ from transition model $p(X_t/X_{t-1})$

New weight

$w_t^i = w_{t-1}^i p(y_t|x_t^i)$

- Challenge is to get correct likelihood in visual tracking.
- Computationally Intensive
DATA ASSOCIATION IN MULTIPLE TARGETS

- Involves a Data Association Stage

JPDAF
- Single Hypothesis Tracker
- Computationally efficient

Assumptions:
- All past is summarized in a single hypothesis.
- Number of targets are fixed.

MHT
- Multiple Hypothesis Tracker
- Computation Heavy

Assumptions:
- Multiple hypotheses are kept running.
- Variable number of targets.
Overview of Some Wide Area Tracking Algorithms

- Features:
  - Color Histogram, Harris Corner Features, SIFT, HoG, SURF, CHoG
  - Optical Flow, STIP (Laptev 2005)

- Categorization according to methods:

<table>
<thead>
<tr>
<th>Training</th>
<th>Batch Process</th>
<th>Sequential Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Training</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Offline</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Online</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
HybridBoosted Tracker

- Overall approach:
  - Ranking
  - Classification

- Contributions: automatically select among various features and corresponding non-parametric models.

- Limitations: Batch process.

Handoff at FOV Lines

- Estimate FOV line of one in others.
- Draw the correspondence

- Works for overlapping FOVs.
- Experiments in controlled environment with 3 overlapping cameras.

Multi-Camera Track Association

- MAP estimation framework using
  - Appearance cues
  - Intercamera space-time probability.
- Tested on controlled environment
- Heuristically updates parameters
- No attempt to correct for wrong associations.

Handling Change of Appearance

- Change of appearance – modeled by Brightness Transfer Function (BTF).

- Experiments – indoor and/or outdoor with 2 or 3 cameras

No training, Batch process

- Overall approach:

- Contributions: Search for optimal associations between tracklets using a stochastic sampling method and consider long-term interdependencies between the target tracklet features.

- Limitations: Cannot be applied to on-line system.

Linear Trajectory Avoidance

- **Overall approach:** using linear trajectory avoidance (LTA) to predict the target’s expected point of closest approach and make it move in the optimal direction.

- **Contributions:** social behavior model in tracking

Multi-social-factor Behavior Model

- Overall method: energy function is a linear combination of six components: damping, speed, direction, attraction, grouping, collision => (Position prediction)

- Contributions: multi-social-factor (a complex behavior model)

- Experiments: ETH & HOTEL (the same as LTA), ZARA

Learn Incrementally

- Incremental learning of the cues

- Incremental region subdivision

A. Gilbert and R. Bowden, "Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity" ECCV 2006.
Learn Incrementally

- Transformation matrices of inter-camera color changes are learned
- No a priori information used and learning is unsupervised
- Improving performance as more data is accumulated

A. Gilbert and R. Bowden, "Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity" ECCV 2006.
Tracking as Maximum Weight Independent Set

$$y^* = \arg\max_y \sum_{i \in V} w_i \sigma(y_i) \prod_{j \in V} (1 - \sigma(y_j))^{B_{ij}}$$

Tracking as Maximum Weight Independent Set

- The proposed MWIS algorithm
  - polynomial in time.
  - guaranteed to converge to optimum.
- Similarity, constraints learned online.
- Learned pairwise motion correlation is used as contextual constraint.
- Long term occlusions handled by iterative application of MWIS.
- Favorable results on 5 challenging datasets.

Online Adaboost Learning

- Overall approach
  - Online sample selection
  - The appearance learning model

- Contributions: Collect training samples online
- Limitations: the positive samples lack diversity (because collected in a few neighboring frames)

Online Learned Appearance Model

- Scenarios: multiple non-overlapping cameras
- Overall Method:

- Contributions: the first work using online learning of a discriminative appearance affinity model across cameras

C. Kuo, C. Huang, and R. Nevatia. “Inter-camera association of multi-target tracks by on-line learned appearance affinity models”. In ECCV, 2010.
Online Learned Appearance Model

- Experiments: three-camera with disjoint FOVs
- 25 min videos and each frame has 2 to 10 people

Topology learning

- Scenarios: multiple cameras

  - Automatic calibration
  - Automatically learn entry/exit zones between two cameras (Transitional probability)
  - Provide a means for tracking targets across the “blind” areas of the network

- Overall method

- Contributions:
  - Automatically learns the network calibration over a wide area
  - Work for both non-overlapping and overlapping scenarios

Statistical Dependence between Cameras

- Scenarios: multiple non-overlapping cameras

  Multi-model transition distributions between two cameras

  Infer the topology of a camera network by measuring degree and nature of statistical dependence between observations in different cameras.

  Consider the correspondence problem and handle general types of statistical dependence by using mutual information and non-parametric density estimates.

- Overall method

- Experiments: simulated traffic network and real traffic network (5 cameras)

What are the “big” challenges?

- Discussions
Thank you

Contact: amitrc@ee.ucr.edu
Adaptive Learning

- Scenarios: multiple non-overlapping cameras
- Overall method:
  - Spatio-temporal relationships between targets
  - Prior knowledge of camera network topology
  - An adaptive method for learning spatio-temporal relationships
  - Brightness transfer functions

- Contributions: applied to long-term tracking, consider the environment changes.
- Limitations: on negative method is considered during learning process.