Advances in video tracking:
adaptation, self-evaluation & context learning

Andrea Cavallaro
A simple example to start with ...
Multi-target tracking on confidence maps: an application to people tracking

F. Poiesi, R. Mazzon, A. Cavallaro

Multi-target tracking on confidence maps: an application to people tracking
F. Poiesi, R. Mazzon, A. Cavallaro
... and how about tracking these targets?
Reference

Book

Video Tracking
Theory and Practice

Wiley and Sons, 2011
ISBN: 978-0-470-74964-7

Website

www.videotracking.org
Advances in video tracking

adaptation  self-evaluation  context learning
Acknowledgements

adaptation  self-evaluation  context learning

Samuele Salti  Juan Carlos SanMiguel  Emilio Maggio
Introduction
Tracking: information flow

Feature extraction

$E_I$

$I_k$: input (image frame)

$E_o$

$z_k$: observation

$E_s$

$x_k$: output (target state)
Target state

- The state $x_k$ is a vector of object parameters
  - Shape and position
  - Appearance
  - Higher order moments (e.g. speed, acceleration)
Observation

• The observation vector $Z_k$ represents object information extracted from the image
  
  – Low-level information
    • e.g. pixel colour, gradient
  
  – Mid-level features
    • e.g. location of edges
  
  – High-level features
    • e.g. object detection
  
  – Combinations of the above
Challenges: appearance variations

- Pose variations
- Occlusions
Challenges: clutter

- Background objects with similar appearance
The video tracking pipeline

- Feature extraction
- Localisation
- Track management
- Target representation
- Meta-data extraction
- Trajectories (states)
- Meta-data

CIS centre for intelligent sensing

Queen Mary University of London
Particle Filter

- State
  \[ x_k = f_k(x_{k-1}, u_k) \]

- Observation
  \[ z_k = h_k(x_k, n_k) \]

- Objective
  - to estimate unknown state \( x_k \) based on a sequence of observations \( z_k, k = 0, 1, \ldots \)
  - to approx. the posterior distribution
    \[ p(x_k | z_{1:k}) \approx \sum_{i=1}^{L} w_k^i \delta(x_k - x_k^i) \]

- Solution (Bayesian)
  - Prediction step
    - Based on state equation
  - Update step
    - Based on likelihood function
Sequential Monte Carlo (Particle Filter)

• Monte Carlo approximation of the Bayes recursion
  – Approximate the posterior pdf

\[ p(x_k \mid z_{1:k}) \approx \sum_{i=1}^{L} \omega_k^{(i)} \delta(x_k - x_k^{(i)}) \]
  – Samples propagated using state and observation equations

\[
\omega_k^{(i)} \propto \omega_{k-1}^{(i)} \frac{g(z_k \mid x_k^{(i)}) f(x_k^{(i)} \mid x_{k-1}^{(i)})}{q(x_k^{(i)} \mid x_{k-1}^{(i)}, z_k)}
\]

• Sequential Importance Sampling (SIS)
  – Sample according to dynamics

\[ q(x_k \mid x_{k-1}, z_k) = f(x_k \mid x_{k-1}) \]
Resampling

• **SIS problem**
  – The sampling *does not depend on the measurement*
  – After some steps particles have *small weights*

• **Solution**
  – To eliminate particles with *small weight*
  – To propagate multiple times particles with *large weight*

• **Sequential Importance Resampling (SIR)**
  – Sample proportionally to \( \omega_{k-1}^{(i)} \)
  – New weight \( \omega_{k}^{(i)} \propto g(z_k \mid x_k^{(i)}) \) *likelihood*
CONDENSATION

\[ f(x_{k+1}^{(i)} | x_k^{(i)}) \]

resampling

\[ g(z_k | x_k^{(i)}) \]

\[ f(x_k^{(i)} | x_{k-1}^{(i)}) \]

\[ p(x_{k-1} | z_{1:k-1}) \]

CIS centre for intelligent sensing

Queen Mary University of London
The (simplified) video tracking pipeline

- Feature extraction
- Localisation
- Trajectories (states)
The (simplified) video tracking pipeline

1. Feature extraction
2. Localisation
3. Target representation
4. Trajectories (states)
Generative vs discriminative trackers

- **Generative tracker**
  - (trained and) updated based on object appearance only **without** considering the appearance of the background or other targets

- **Discriminative tracker**
  - (trained and) updated to learn a **decision** boundary that can separate target & background and/or one target against other targets
Fixed target model

invariant features
Fixed target model: another example
Coping with appearance changes

\[ \begin{align*}
Z & \rightarrow \text{Tracker} \\
\text{Tracker} & \rightarrow \text{Target model} \\
\text{Target model} & \rightarrow \text{Model update} \\
\text{Model update} & \rightarrow \text{Tracker} \\
\text{Tracker} & \rightarrow \chi
\end{align*} \]
Adaptive target model?
Plasticity vs stability dilemma

(model update)
'extreme' plasticity
always forget

last model only
sliding window
ranking
blending
subspace/manifold

'A_t' overall target model
'a_{t+1}' model estimated in I_{t+1}

'extreme' stability
full history (potentially)
Adaptive target models

Adaptive appearance modeling for video tracking: survey and evaluation
S. Salti, A. Cavallaro, L. Di Stefano
<table>
<thead>
<tr>
<th><strong>algorithm</strong></th>
<th><strong>reference</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boost</td>
<td>Grabner and Bischof, <em>On-line boosting and vision</em>, CVPR06.</td>
</tr>
<tr>
<td>SemiBoost</td>
<td>Grabner <em>et al.</em>, <em>Semi-supervised on-line boosting for robust tracking</em>, ECCV08</td>
</tr>
<tr>
<td>BeyondSemiBoost</td>
<td>Stalder <em>et al.</em>, <em>Beyond semi-supervised tracking</em>, ICCVW09</td>
</tr>
<tr>
<td>IVT</td>
<td>Ross <em>et al.</em>, <em>Incremental learning for robust visual tracking</em>, IJCV08</td>
</tr>
<tr>
<td>MILBoost</td>
<td>Babenko <em>et al.</em>, <em>Robust object tracking with online multiple instance learning</em>, PAMI11</td>
</tr>
<tr>
<td>TLD</td>
<td>Kalal <em>et al.</em>, <em>Online Learning of Robust Object Detectors During Unstable Tracking</em>, ICCV09</td>
</tr>
<tr>
<td>STRUCK</td>
<td>Hare <em>et al.</em>, <em>STRU CK: Structured Output Tracking with Kernels</em>, ICCV11</td>
</tr>
<tr>
<td>Mean-shift</td>
<td>Comaniciu <em>et al.</em>, <em>Kernel-based object tracking</em>, PAMI03.</td>
</tr>
<tr>
<td>Color-based PF</td>
<td>Pérez <em>et al.</em>, <em>Color-based probabilistic tracking</em>, ECCV02</td>
</tr>
<tr>
<td>FragTrack</td>
<td>Adam <em>et al.</em>, <em>Robust Fragments-based Tracking Using the Integral Histogram</em>, CVPR06</td>
</tr>
</tbody>
</table>
Adaptive target modelling: examples

Adaptive appearance modeling for video tracking: survey and evaluation
S. Salti, A. Cavallaro, L. Di Stefano
Adaptive target modelling: examples

Adaptive appearance modeling for video tracking: survey and evaluation
S. Salti, A. Cavallaro, L. Di Stefano
How well am I tracking?

Performance evaluation
Tracking quality estimation

tracking result

ideal result
So, what happened?
Performance evaluation

- Analytical methods
- Empirical methods
  - Standalone methods
  - Discrepancy methods

- based on quality criteria about the trajectory
- Quantify the deviation of the results from reference results
- Need a ground-truth
What quality criteria?

Observation likelihood

✓ Accurate if tracker is correct
✗ Affected by distractors

Covariance (uncertainty)

✓ Indicates target searching
✗ Affected by distractors
Analysis of the particle distribution

- Tracker status estimation
  - based on spatial uncertainty
  - multi-hypothesis tracker (e.g. Particle Filter → spread of particles)

Score: based on normalised determinant of the covariance matrix of the particle distribution

---

**Graph:**

- **Ground-truth error**
- **Uncertainty**

**Axes:**
- **Score**
- **Frame**

**Data Points:**

- Frame 15, 58, 69, 88
- Ground-truth error and uncertainty scores increasing with frame number.
Tracker condition estimation

- **Goal:** to determine transitions of tracker condition
  - **Locked-on:** tracking the target or a distractor (!)
  - **Locking-in:** re-focus on the target or a distractor (!) after a failure
  - **Scanning:** searching the target after a failure

![Diagram showing transitions between Tracker Condition states: Locked-on, Locking-in, Scanning with conditions for changes in uncertainty signal.](diagram.png)
Tracker condition estimation
Final criterion: reverse tracking

Reverse-based evaluation

✓ Checks trajectory consistency
× Exponential complexity

• Recovery from error
  – Reverse tracking for matching: recovery ↔ reference
How well am I tracking?

- Temporal segmentation of tracker operation
  - determines correct tracking
    - successful
    - unsuccessful
  - based on
    - tracker condition and
    - correct recovery from error

**H1:** tracker remains or move to **scanning**

**H2:** tracker moves from **locking-in** to **locked-on** & there is a correct recovery from error
To summarise ...
Adaptive on-line performance evaluation of video trackers
J.C. SanMiguel, A. Cavallaro and J.M. Martinez
On-line performance evaluation: results

Adaptive on-line performance evaluation of video trackers
J.C. SanMiguel, A. Cavallaro and J.M. Martinez
Can scene modelling help tracking?

(multiple objects)
Learning scene context for multiple object tracking
E. Maggio and A. Cavallaro
Multi-target observation and state

• Observations
  \[ Z_k = \{z_{k,1}, \ldots, z_{k,N(k)} \} \]
  \[ z = [y_1, y_2, w, h] \]

• State
  \[ X_k = \{x_{k,1}, \ldots, x_{k,M(k)} \} \]
  \[ x = [y_1, y_2, \dot{y}_1, \dot{y}_2, w, h] \]
Multi-target Bayes filter

**Single target**
Random vectors

\[
\begin{align*}
    x_k &= f(x_{k-1}, u_k) \\
    z_k &= g(x_k, n_k)
\end{align*}
\]

**Multiple targets**
Random Finite Sets (RFS)

\[
\begin{align*}
    \Xi_k &= S_k(X_{k-1}) \cup \Gamma_k \\
    \Omega_k &= \Theta_k(X_k) \cup K_k
\end{align*}
\]

- Recursively propagates the multi-target state
- Prediction

\[
p(X_k | Z_{1:k-1}) = \int f(X_k | X_{k-1}) p(X_{k-1} | Z_{1:k-1}) \mu_s(dX_{k-1})
\]

- Update

\[
p(X_k | Z_{1:k}) = \frac{g(Z_k | X_k) p(X | Z_{1:k-1})}{\int p(Z_k | X_k) p(X_k | Z_{1:k-1}) \mu_s(dX_k)}
\]
Multi-target Bayes filter (2)

- Model target \textit{birth}, \textit{death}, \textit{clutter} and \textit{missing detections}

- Monte Carlo approximation
  - Dimensionality of the state grows with the number of targets
  - Number of particles is exponential with the number of targets
  - Not feasible in real-world applications!

- Solution
  - Probability Hypothesis Density (PHD) filter
Probability Hypothesis Density (PHD)

- The first order moment of a RFS $\rightarrow$ PHD
  - The integral on any $R$ gives the expected number of targets in $R$

\[
\int_{R} D_{\Xi}(x) \, dx = E\left[|\Xi \cap R|\right]
\]

$|X \cap R| = 3$
Recursive PHD filter

- **Prediction**
  \[
  D_{k|k-1}(x_k) = \gamma(x) + \int \phi(x_k, x_{k-1}) D_{k-1|k-1}(x_{k-1}) \, dx_{k-1}
  \]
  - PHD state transition
    \[
    \phi(x_k, x_{k-1}) = e(x_k) f(x_k | x_{k-1})
    \]

- **Update**
  \[
  D_{k|k}(x_k) = \begin{bmatrix}
  p_M(x_k) + \sum_{z \in Z_k} \kappa(z) + \psi_z(x_k) \\
  \end{bmatrix} \underbrace{D_{k|k-1}(x_k)}_{\text{Normalization}}
  \]
  - Survival probability
  - Single target transition
  - Clutter intensity
  - Birth intensity
  - Missing likelihood
  - Normalization

\[
\psi_z(x) = (1 - p_M(x)) g(z | x)
\]

- Single target likelihood
Tracking with the PHD filter

- **Advantage**: Linear complexity with the target number
- **Drawback**: Data association is not integrated in the Bayesian recursion
Results – CLEAR/i-LIDS dataset

- Detections
- PHD output
Context information: examples

- Sidewalks and roads → likely entry/exit points
- Vegetation → likely source of clutter
Using scene context in the PHD recursion

- **Goal:** context adaptive filtering
  - Birth intensity -- to model entry areas
  - Clutter intensity -- to model detection errors

- **Data collection**
  - Birth events from the output of the tracker
  - Clutter events via user feedback

- **Density estimation**
Example: incremental learning birth intensity
Example: learning birth intensity
Birth and clutter intensities

\[ \gamma(x) \] (purple)

\[ \kappa(z) \] (cyan)

Queen Mary
University of London
Comparison: w. and w/o context learning

- Performance improvement on CLEAR-2007 dataset
- Reference: PHD tracker with uniform birth and clutter intensities

Learning scene context for multiple object tracking
E. Maggio and A. Cavallaro
Advances in video tracking: summary

adaptation  self-evaluation  context learning
Outlook – projects

EU FP7 project CENTAUR
Crowded environments monitoring for activity understanding and recognition
Marie Curie Industry-Academia Partnerships 2013 - 2016

UK EPSRC project MAVIP
Multisource audio-visual production from user-generated content 2013 - 2015

EU Erasmus Mundus Joint Doctorate ICE
Interactive & Cognitive Environments now - 2017

EU Artemis JU project COPCAMS
Cognitive & perceptive Cameras
Public-private partnership 2013 – 2016

CIS centre for intelligent sensing

Queen Mary University of London
www.eecs.qmul.ac.uk/~andrea

www.twitter.com/smartcameras

www.youtube.com/smartcameras
Multi-target tracking on confidence maps: an application to people tracking
F. Poiesi, R. Mazzon, A. Cavallaro

Adaptive appearance modeling for video tracking: survey and evaluation
S. Salti, A. Cavallaro, L. Di Stefano

Adaptive on-line performance evaluation of video trackers
J.C. SanMiguel, A. Cavallaro and J.M. Martinez

Learning scene context for multiple object tracking
E. Maggio and A. Cavallaro