INRIA-VISTA
Activities in Human Analysis

Emilie Dexter, Ivan Laptev, Patrick Pérez, Nicolas Gengembre
IRISA/INRIA, Rennes, France

Workshop, Barcelona, January 22-23
Outline

- Introduction
- Person and object detection
- Tracking
- Periodic motion detection and segmentation
- Conclusion
- Future work
Introduction

- INRIA – VISTA research group
  http://www.irisa.fr/vista/Vista.english.html
  - Spatio-temporal images
  - Dynamic scene analysis
  - Motion analysis (Detection, estimation, segmentation, tracking, recognition, interpretation with learning)
Person and object detection in static images

Ivan Laptev
IRISA/INRIA, Rennes, France

[Laptev, 2006]
Detection

Training:

AdaBoost Learning with Local Histogram Features

Region features

- $f_1$
- $f_2$
- $f_3$
- $f_4$

Histograms of gradient orientation

positive samples

negative samples

selected features

$H(z) = \text{sgn} \left( \sum_{t=1}^{T} \alpha_t h_t(f_t) \right)$

weak classifier

[Freund and Schapire, 1997]
[Viola and Jones, 2001]
Detection

Search:

Classify windows at all image positions and scales

Results:

people

cars

bicycles

horses

cows
Detection: Comparison

PASCAL VOC 2005:

<table>
<thead>
<tr>
<th>Method</th>
<th>Motorbikes</th>
<th>Bicycles</th>
<th>People</th>
<th>Cars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boosted Hist.</td>
<td>0.896</td>
<td>0.370</td>
<td>0.250</td>
<td>0.663</td>
</tr>
<tr>
<td>TU-Darmstadt</td>
<td>0.886</td>
<td>–</td>
<td>–</td>
<td>0.489</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>0.453</td>
<td>0.119</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>INRIA-Dalal</td>
<td>0.490</td>
<td>–</td>
<td>0.013</td>
<td>0.613</td>
</tr>
</tbody>
</table>

Average precision for object detection in “test1”

![Precision-recall curve for object detection](image)
Robust visual tracking with background analysis

Nicolas Gengembre, Patrick Pérez
IRISA/INRIA, Rennes, France
Robust visual tracking with background analysis

- **Context:** *generic* visual tracking
  - No prior on object to track
  - No prior on video

- **Requirements**
  - Simple appearance modeling
  - Instantiated/learnt on-line
  - Discriminant enough
  - \( \Rightarrow \) *Color histograms* are appealing

- **For improved robustness**
  - Probabilistic modeling
  - Background analysis (local or not)
“Mean-shift” tracking [Comaniciu et al., 2000]
- Kernel-based global color modeling
- No (or slow) adaptation
- Search by gradient ascent on histogram similarity $\rho[h^*, h_t(x_t)]$

Pros and cons
- Robust to appearance changes
- Fast
- Scale and rotation invariant
- Local search only (occlusion problem)
Remove background contamination

One-step update

- Initial bkg/fg models with $B$ bins
  \[ h^f_u \propto \sum_{x \in R} 1_u[I_0(x)], \quad u = 1 \cdots B \]
  \[ h^b_u \propto \sum_{x \in \partial R} 1_u[I_0(x)] \]

- Empty weak fg bins
  \[ h^*_u \propto h^f_u \cdot 1(h^f_u \geq h^b_u), \quad u = 1 \cdots B \]

amounts to ML classification in $R$ and re-estimation
Background Motion

- Apparent background motion usually induced by camera motion
- Its sequential estimation permits
  - More robust object tracking
  - Easier definition of meaningful object dynamics
  - Definition of adaptation modules
  - Display of tracking results in fixed mosaic or with incrementally warped trajectories
- Approach
  - Robust fit of parametric motion on sparse KLT vectors
  - Kalman filtering for robustness to breakdowns (e.g., due to flash lights)

\[ \hat{\theta}_i = [\hat{t}_i^T, \hat{s}_i]^T \]
Selective Adaptation

- **Adaptation**: less necessary than with detailed models
- **Still necessary**: drastic zooms, illumination changes, appearance of new parts
- **Drift problem**: not during partial/total occlusions

\[(\text{if not(occlusion)} \& \hat{s}_t > 0), \quad h^* \leftarrow \alpha h^* + (1 - \alpha)h_t^*\]
More robust to occlusions, clutter, large displacements…

Kalman [Comaniciu et al. 00]: deterministic tracker provides a unique measure

Particle Filter [Pérez et al. 02]: bootstrap PF with likelihoods
\[ p(y_t|x_t) \propto \exp \lambda \rho[h^*, h^f(x_t)] \]

Tracking conditional to \( \theta \)
- “Conditional” dynamics
\[ p(x_t|x_{t-1}, \hat{\theta}_t) = W_{\hat{\theta}_t}(x_{t-1}) + w_t \]
- Conditional filter [Arnaud et al. 03]: compute or approximate
\[ p(x_t|I_{1:t}, \hat{\theta}_{1:t}) \]
Multiple Object tracking

- **Joint particle filter** in compound state space [Vermaak et al. 05]
  - Upper bound on object number and binary auxiliary existence variables
  $$x_t = (x_t^k, e_t^k)_{k=1}^K \in (\Lambda \times \{0, 1\})^K$$
  - Markov process on $e$ parameterized by entrance/exit probabilities
  - Interaction via observation model (exclusion principle)
  - Efficiency issue
    - Curse of dimension
    - Combinatorial treatment of interactions
Multiple Object tracking

- **Marginal particle filters** with approximate interactions
  - Given $K$ predicted particle sets $(x_t^k(n), w_t^k(n))_{n=1:N}$
  - Build pixel ownerships
    \[ \beta_k(x) \propto h_u(x) \sum_n w_t^k(n) \mathbf{1}_{R(x_t^k(n))}(x), \sum_k b_k(x) = 1 \]
  - Build association probabilities
    \[ \alpha(x_t^k(n)) = \frac{1}{|R(x_t^k(n))|} \sum_{x \in R(x_t^k(n))} \beta_k(x) \]
  - Update weights
    \[ w_t^k(n) \propto w_t^k(n) \alpha(x_t^k(n)) \]
Multiple Object tracking

independent trackers

interacting trackers

[Gengembre and Pérez, 2006]
Periodic Motion Detection and Segmentation via Approximate Sequence Alignment

Ivan Laptev*, Serge Belongie**, Patrick Pérez*
*IRISA/INRIA, Rennes, France
**Univ. of California, San Diego, USA
Periodic views can be approximately treated as stereopairs

\[ \{s_t, \ldots, s_m\} \]

\[ \{s_{t+p}, \ldots, s_{n+p}\} \quad (p: \text{period}) \]

\[ \{s_{t+np}, \ldots, s_{m+np}\} \]
Periodic views can be approximately treated as stereopairs.

\[ \{s_t, \ldots, s_m\} \]

Fundamental matrix \( F \) is generally time-dependent.

\[ \{s_{t+np}, \ldots, s_{m+np}\} \]

\( \Rightarrow \) Periodic motion estimation \( \sim \) sequence alignment.
1. Corresponding points have similar motion descriptors
   [Laptev and Lindeberg, 2003], [Laptev and Lindeberg, 2004]

2. Same period for all features $p = \Delta t$

3. Spatial arrangement of features across periods satisfy epipolar constraint:
   $[x^t]^t F x^{t+p} = 0$

⇒ Use RANSAC to estimate $F$ and $p$
Periodic motion detection

Original space-time features

RANSAC estimation of F,p

period p = 24.00
Periodic motion detection

Original space-time features

RANSAC estimation of F,p

period p=31.00
period p=33.00
Assume periodic objects are planar

⇒ Periodic points can be related by a dynamic homography:

\[ x_t = H x_{t+p} \text{ with} \]

\[ H(t) = I + p(vn^\top - n^\top vI)/d - tnm^\top vI/d \]

linear in time
Periodic motion segmentation

- Assume periodic objects are **planar**

  \( \Rightarrow \) Periodic points can be related by a **dynamic homography**:

  \[
  x_t = H x_{t+p} \quad \text{with} \quad H(t) = I + p(vn^\top - n^\top vI)/d - tnm^\top vI/d
  \]

  \( \Rightarrow \) RANSAC estimation of \( H \) and \( p \)
Object-centered stabilization

Periodic frame matching and alignment
Segmentation

Disparity estimation

Graph-cut segmentation
[Boykov and Kolmogorov, 2004]
[Kolmogorov and Zabih, 2002]
Segmentation
Conclusion

- Present three different methods in the human analysis domain:
  - People detection
  - People tracking
  - Periodic motion detection and segmentation
- Detection and segmentation could initiate a tracker
- Tracker results can be used as training data for a machine learning like in the presented detection method
Future work: space-time alignment

- Definition
  - Correspondences in time (synchronization) and in space

- Prior work addresses special cases
  - Caspi and Irani “Spatio-temporal alignment of sequences”, PAMI 2002
  - Tuytelaars and Van Gool “Synchronizing video sequences”, CVPR 2004

- Several constraints
  - Static video cameras
  - Field of view overlap
  - Use of static background information
  - Correspondences manually established
Future work: space-time alignment

- Generally hard problem
  - Unknown positions and motions of cameras
  - Unknown temporal offset
  - Possible time warping

- Useful in
  - Reconstruction of dynamic scenes
  - Recognition of dynamic scenes
Future work: space-time alignment

- Video example
Future work: space-time alignment

- Example of awaited result
References

Detection
- P. Viola and M. Jones “Rapid object detection using a boosted cascade of simple features”, CVPR 2001

Tracking
- E. Arnaud, E. Mémin, B. Cernushi Frias. Filtrage conditionnel pour le suivi de points dans des séquences d’images, Congrès Francophone de Vision par Ordinateur, ORASIS’03, 2003
- D. Comaniciu and V. Ramesh “Mean Shift and Optimal Prediction for Efficient Object Tracking”, IEEE ICIP, 2000

Periodic motion Detection and Segmentation
- V. Kolmogorov and R. Zabih “Multi-camera scene reconstruction via graph cuts”, ECCV 2002
- I. Laptev and T. Lindeberg “Space-time interest points”, ICCV 2003
- I. Laptev and T. Lindeberg “Local descriptors for spatio-temporal recognition”, First International Workshop on Spatial Coherence for Visual Motion Analysis 2004
- I. Laptev, S.J. Belongie, P. Pérez and J. Wills “Periodic Motion Detection and Segmentation via Approximate Sequence Alignment”, ICCV 2005