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

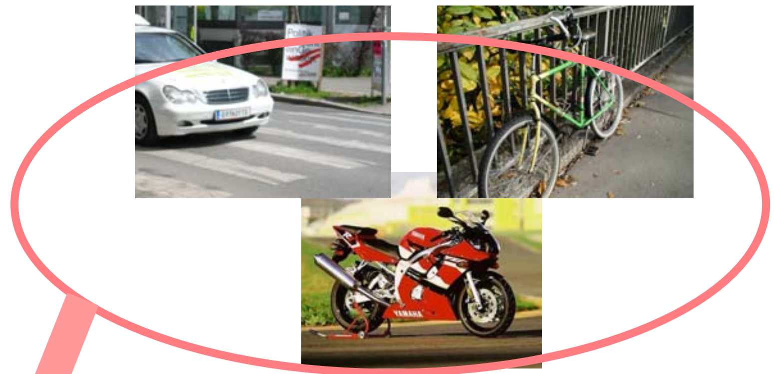
IRISA/INRIA, Rennes, France

[Laptev, 2006]

Detection

Training:

positive samples

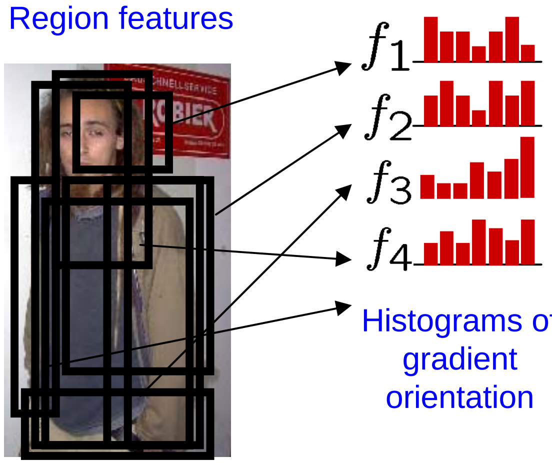


negative samples

AdaBoost Learning with Local Histogram Features

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

Region features



Histograms of gradient orientation

boosting

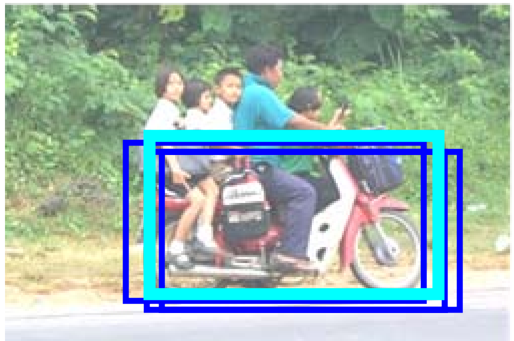
selected features

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

weak classifier

Detection

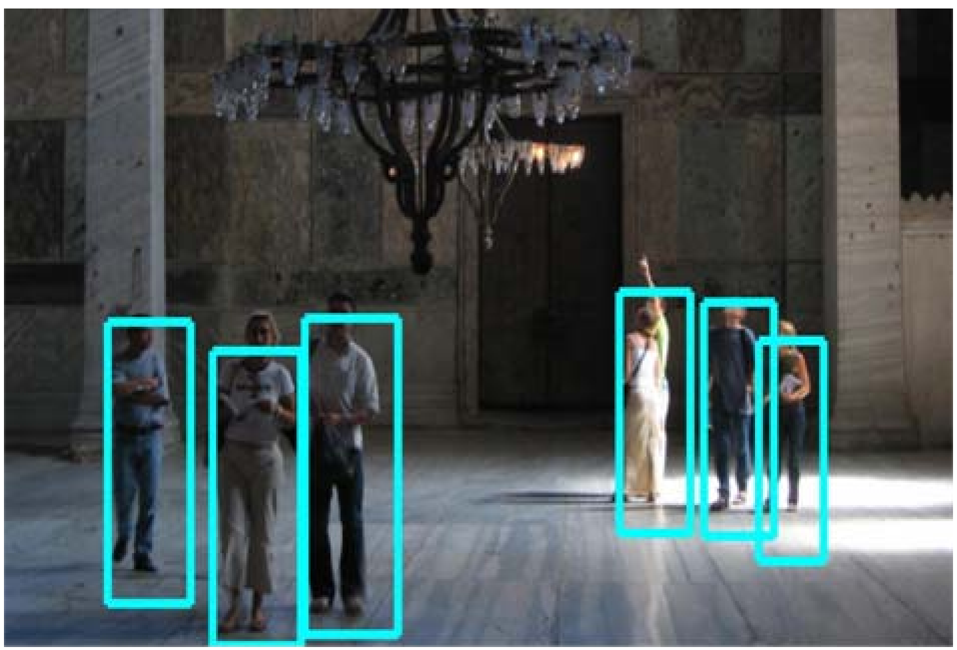
Search:



Classify windows at all image positions and scales

Results:

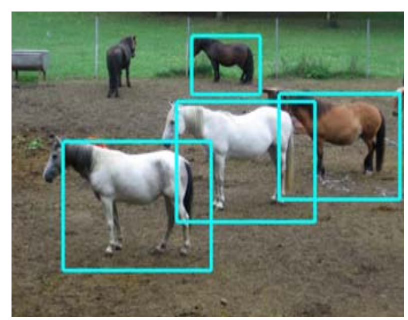
people



cars

bicycles

horses



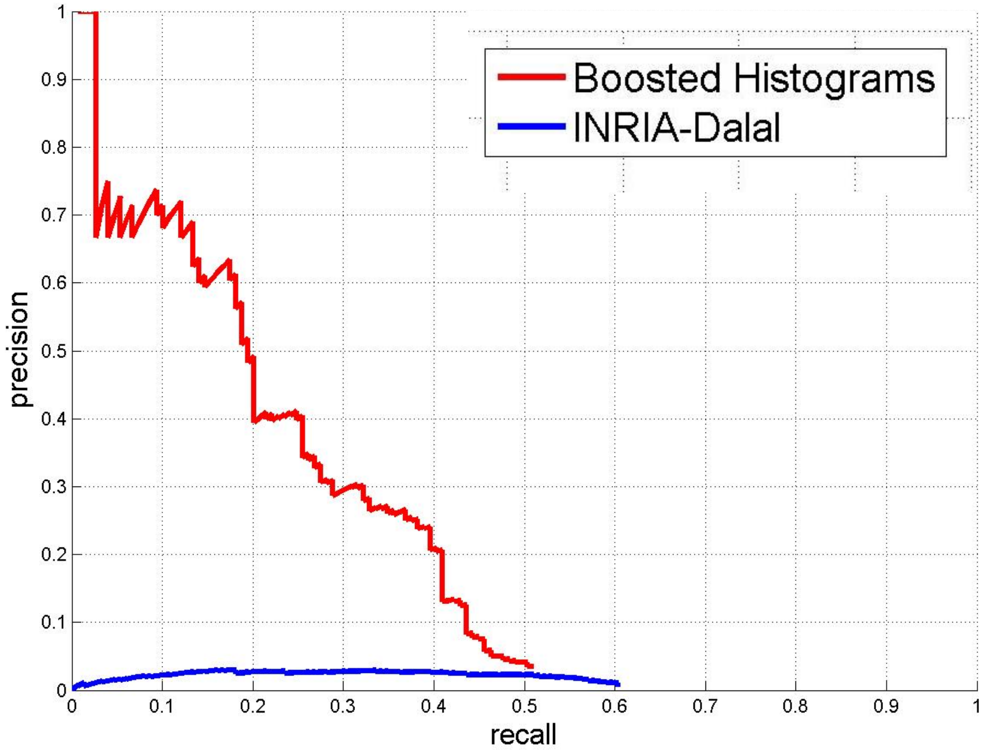
cows

Detection: Comparison

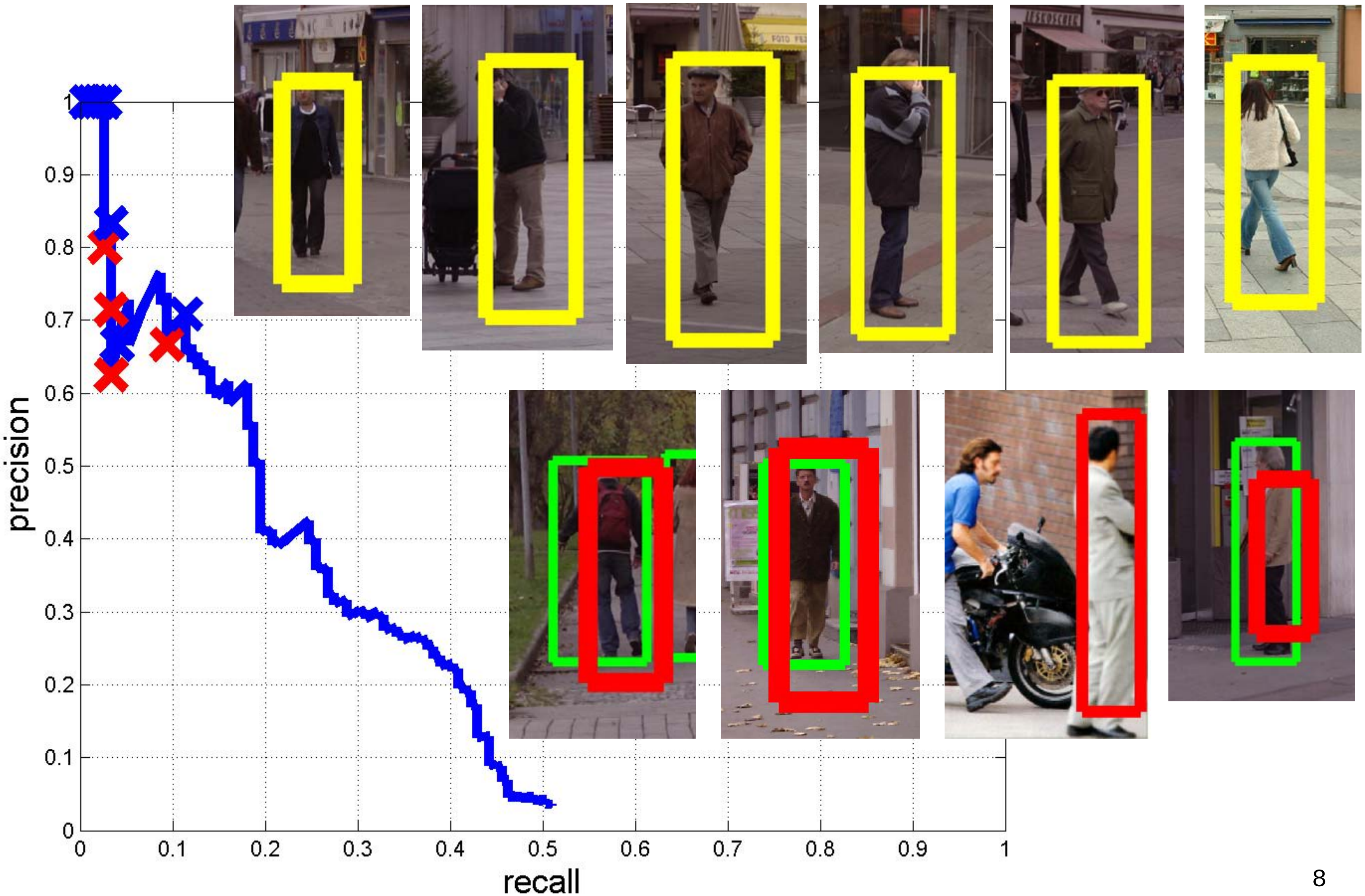
PASCAL VOC 2005:

Method	Motorbikes	Bicycles	People	Cars
Boosted Hist.	0.896	0.370	0.250	0.663
TU-Darmstadt	0.886	–	–	0.489
Edinburgh	0.453	0.119	0.002	0.000
INRIA-Dalal	0.490	–	0.013	0.613

Average precision for object detection in “test1”



Detection: Samples



Robust visual tracking with background analysis

Nicolas Gengembre, Patrick Pérez
IRISA/INRIA, Rennes, France

Robust visual tracking with background analysis

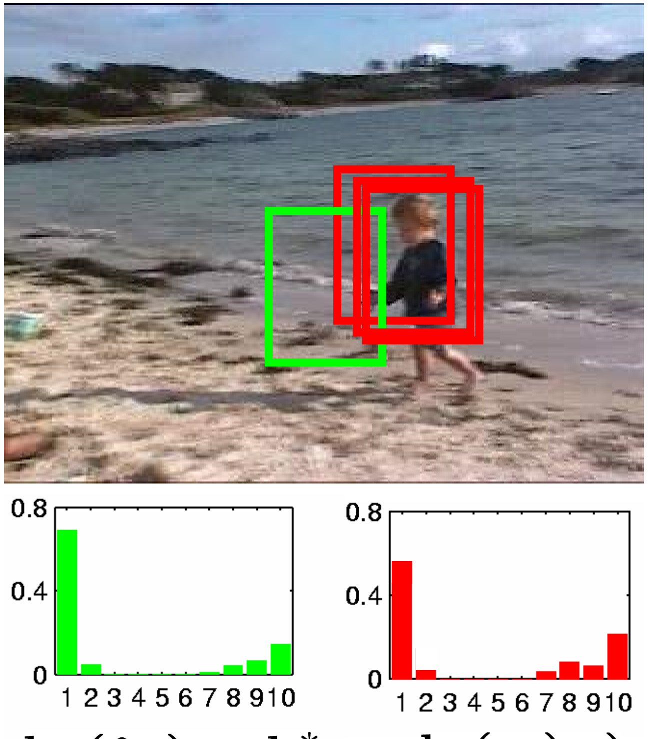
- Context: *generic* visual tracking
  - No prior on object to track
  - No prior on video
- Requirements
  - Simple appearance modeling
  - Instantiated/learnt on-line
  - Discriminant enough

  => *Color histograms* are appealing
- For improved robustness
  - Probabilistic modeling
  - Background analysis (local or not)

Deterministic Color-based Tracking

- “Mean-shift” tracking [Comaniciu *et al*.,2000]
  - Kernel-based global color modeling
  - No (or slow) adaptation
  - Search by gradient ascent on histogram similarity $\rho[\mathbf{h}^*, \mathbf{h}_t(\mathbf{x}_t)]$
- Pros and cons
  - Robust to appearance changes
  - Fast
  - Scale and rotation invariant
  - Local search only (occlusion problem)

$\mathbf{I}_t$



$\mathbf{h}_0(\hat{\mathbf{x}}_0) = \mathbf{h}^*$ $\mathbf{h}_t(\mathbf{x}_t)_{-1})$

Bkg / Fg Color Modeling

- Remove background contamination
- One-step update
  - Initial bkg/fg models with *B* bins

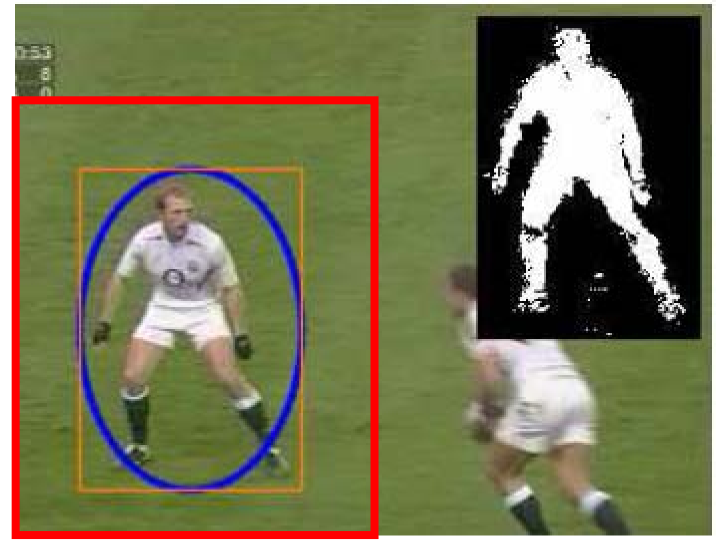
$$h_u^f \propto \sum_{\mathbf{x}\in R} \mathbf{1}_u[\mathbf{I}_0(\mathbf{x})],\ u = 1 \cdots B$$

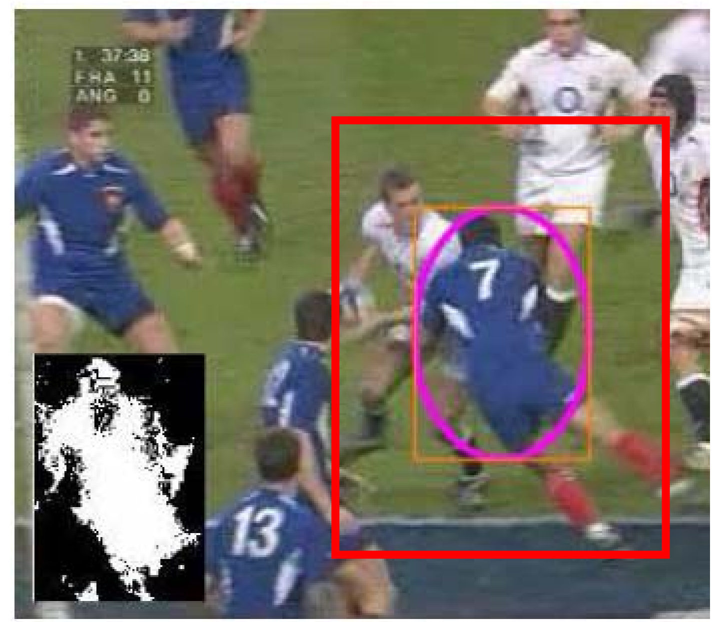
$$h_u^b \propto \sum_{\mathbf{x}\in \partial R} \mathbf{1}_u[\mathbf{I}_0(\mathbf{x})]$$

  - Empty weak fg bins

$$h_u^* \propto h_u^f \cdot \mathbf{1}(h_u^f \geq h_u^b),\ u = 1 \cdots B$$

amounts to ML classification in *R* and re-estimation





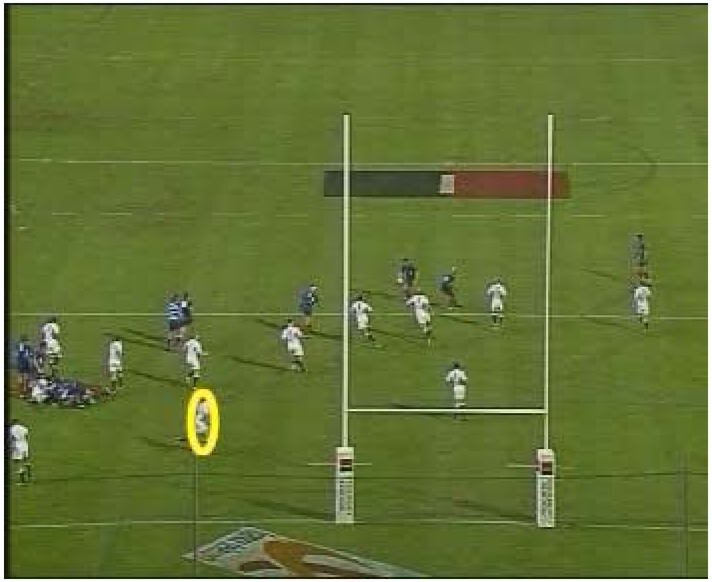
Background Motion

- 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

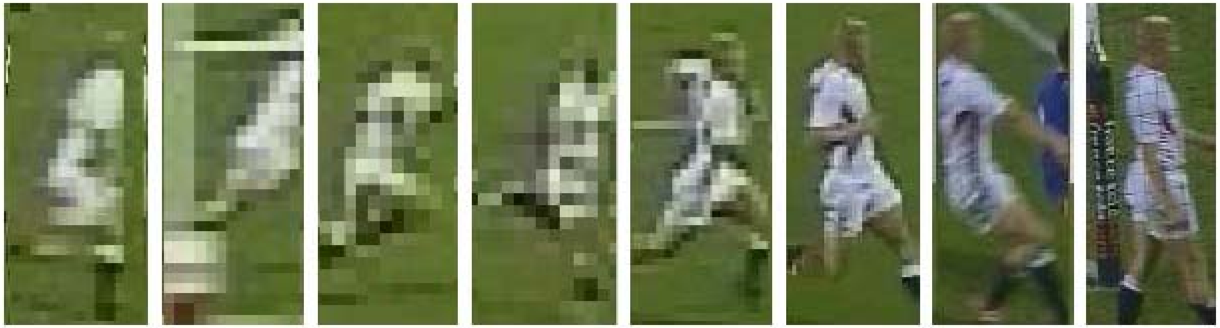
- Approach
  - Robust fit of parametric motion on sparse KLT vectors $\hat{\theta}_t = \left[\hat{t}_t^{\,T}, \quad \hat{s}_t \right]^T$
  - Kalman filtering for robustness to breakdowns (e.g., due to flash lights)

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



$$(\text{ifnot(occlusion)} \ \& \ \hat{s}_t > 0), \ \mathbf{h}^* \leftarrow \alpha \mathbf{h}^* + (1-\alpha)\mathbf{h}_t^*$$



Probabilistic Tracking

- 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

$$p(\mathbf{y}_t|\mathbf{x}_t) \propto \exp \lambda\rho[\mathbf{h}^*, \mathbf{h}^f(\mathbf{x}_t)]$$

- Tracking conditional to θ
  - "Conditional" dynamics

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}, \widehat{\boldsymbol{\theta}}_t) = W_{\widehat{\theta}_t}(\mathbf{x}_{t-1}) + \mathbf{w}_t$$

  - Conditional filter [Arnaud et al. 03]: compute or approximate

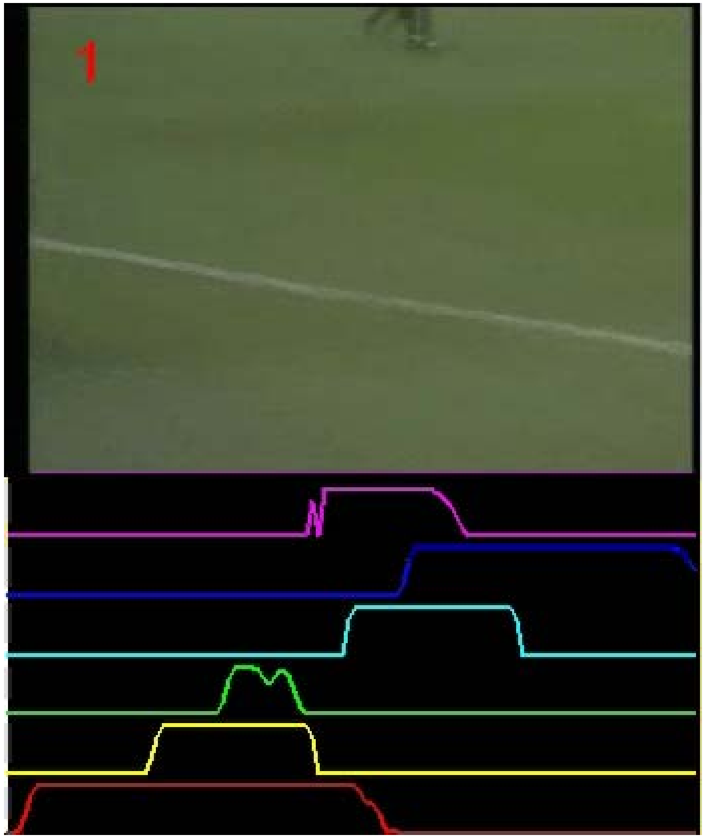
$$p(\mathbf{x}_t|\mathbf{I}_{1:t}, \widehat{\boldsymbol{\theta}}_{1:t})$$

Multiple Object tracking

- Joint particle filter in compound state space [Vermaak *et al*. 05]
  - Upper bound on object number and binary auxiliary existence variables

$$\mathbf{x}_t = (\mathbf{x}_t^k, e_t^k)_{k=1\cdots K} \in (\Lambda \times \{0,1\})^K$$

  - 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 $(\mathbf{x}_t^{k,(n)}, w_t^{k,(n)})_{n=1:N}$
  - Build pixel ownerships

$$\beta_k(\mathbf{x}) \propto h^*_{u(\mathbf{x})} \sum_n w_t^{k,(n)} \mathbf{1}_{R(\mathbf{x}_t^{k,(n)})}(\mathbf{x}), \quad \sum_k b_k(\mathbf{x}) = 1$$

  - Build association probabilities

$$\alpha(\mathbf{x}_t^{k,(n)}) = \frac{1}{|R(\mathbf{x}_t^{k,(n)})|} \sum_{\mathbf{x} \in R(\mathbf{x}_t^{k,(n)})} \beta_k(\mathbf{x})$$

  - Update weights

$$w_t^{k,(n)} \propto w_t^{k,(n)} \alpha(\mathbf{x}_t^{k,(n)})$$

Multiple Object tracking

independent trackers

interacting trackers

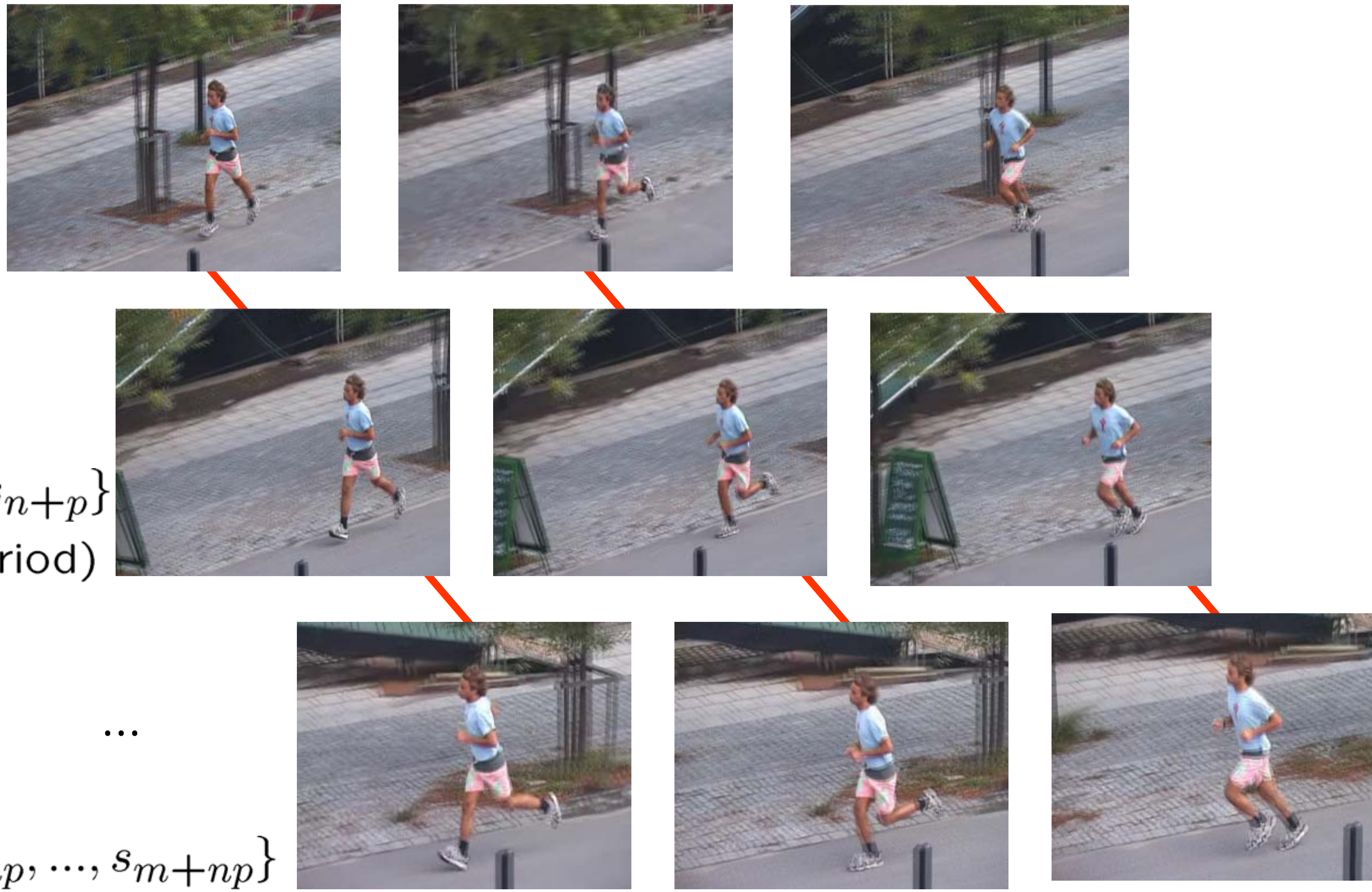
[Gengembre and Pérez, 2006]

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 motion

- **Periodic views can be approximately treated as** **stereopairs**



$\{s_t, \ldots, s_m\}$

$\{s_{t+p}, \ldots, s_{n+p}\}$

(p : period)

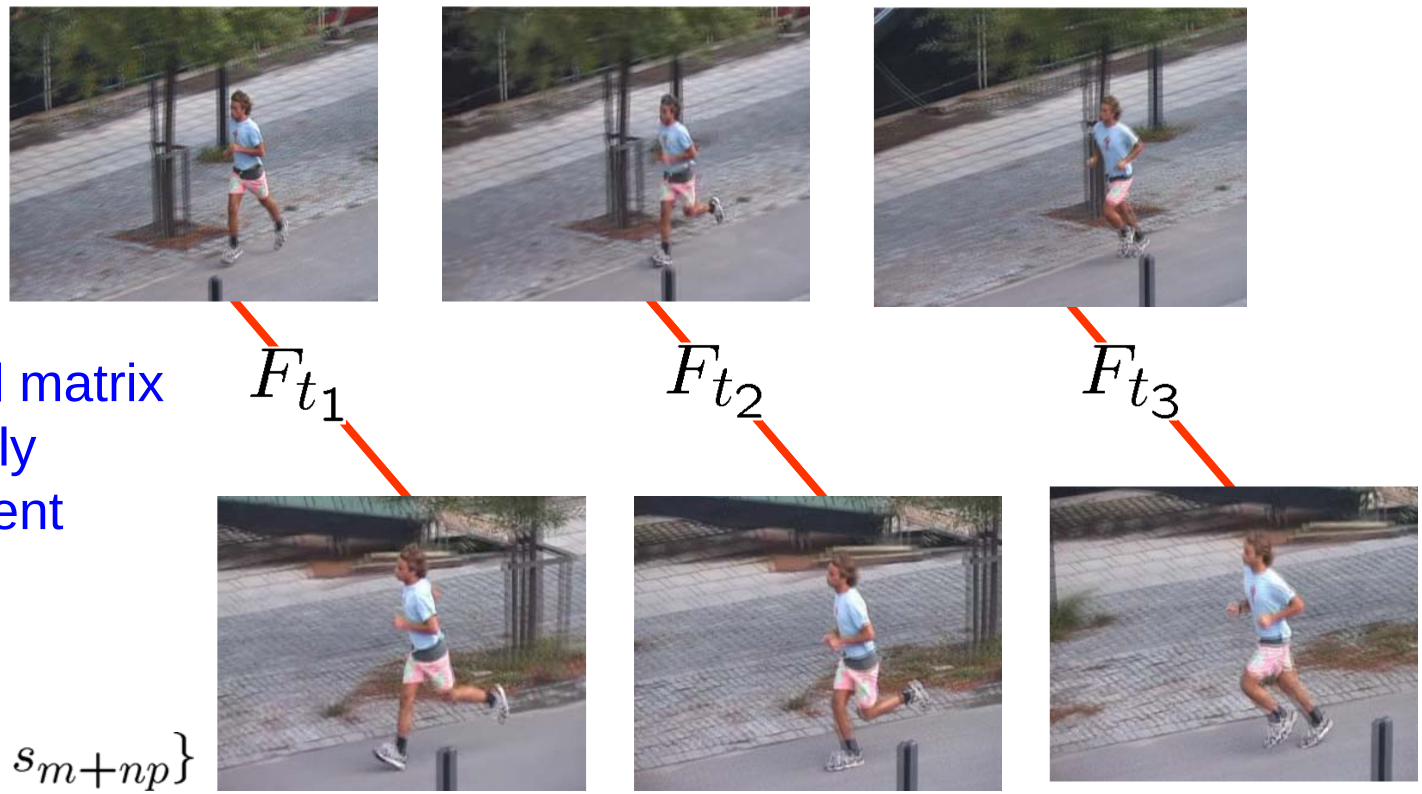
...

$\{s_{t+np}, \ldots, s_{m+np}\}$

Periodic motion

- **Periodic views can be approximately treated as** **stereopairs**

$\{s_t, \ldots, s_m\}$



Fundamental matrix F is generally time-dependent

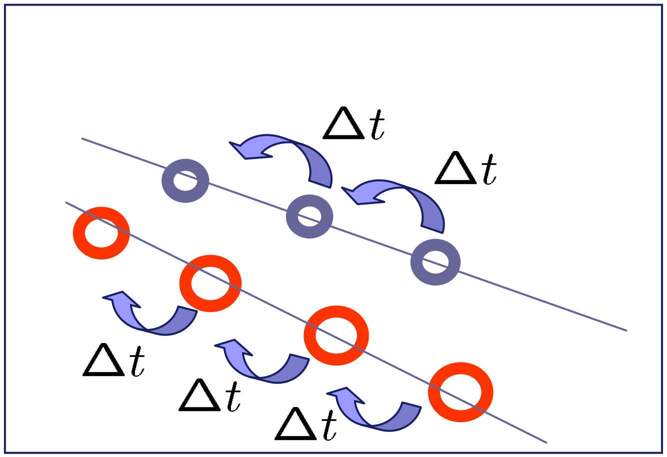
$\{s_{t+np}, \ldots, s_{m+np}\}$

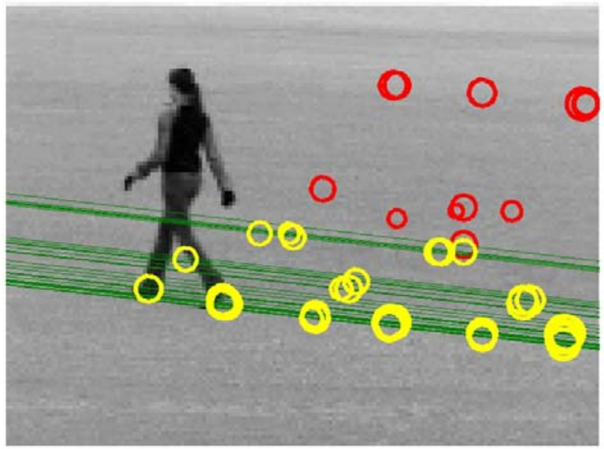
$\Rightarrow$ Periodic motion estimation ~ sequence alignment

Periodic motion detection

1. **Corresponding points have similar motion descriptors**

[Laptev and Lindeberg, 2003], [Laptev and Lindeberg, 2004]





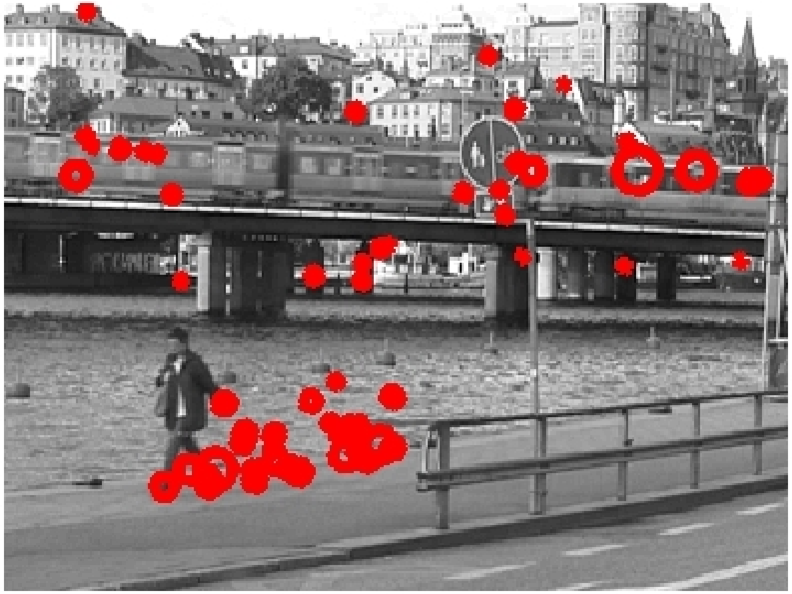
2. **Same period** $p = \Delta t$ **for all features**

3. **Spatial arrangement of features across periods satisfy epipolar constraint:** $[x^t]' F x^{t+p} = 0$

$\Rightarrow$ **Use RANSAC to estimate *F* and *p***

Periodic motion detection

Original space-time features

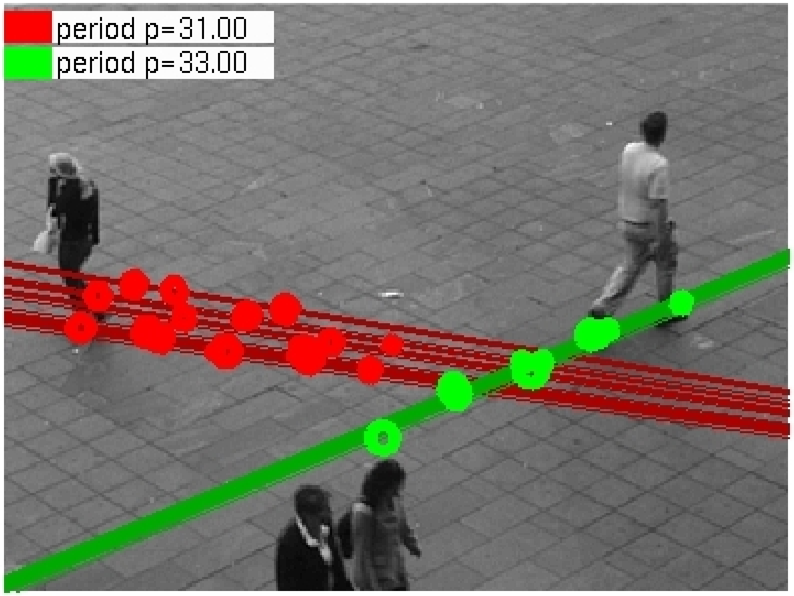


RANSAC estimation of F,p

Periodic motion detection

Original space-time features

RANSAC estimation of F,p



Periodic motion segmentation

- **Assume periodic objects are planar**

⇒ Periodic points can be related by a *dynamic homography:*

$x_t = H x_{t+p}$ with

$$H(t) = I + p(\mathbf{v}\mathbf{n}^\top - \mathbf{n}^\top\mathbf{v}I)/d - t\mathbf{n}^\top\mathbf{v}I/d$$

linear in time

Periodic motion segmentation

- **Assume periodic objects are planar**

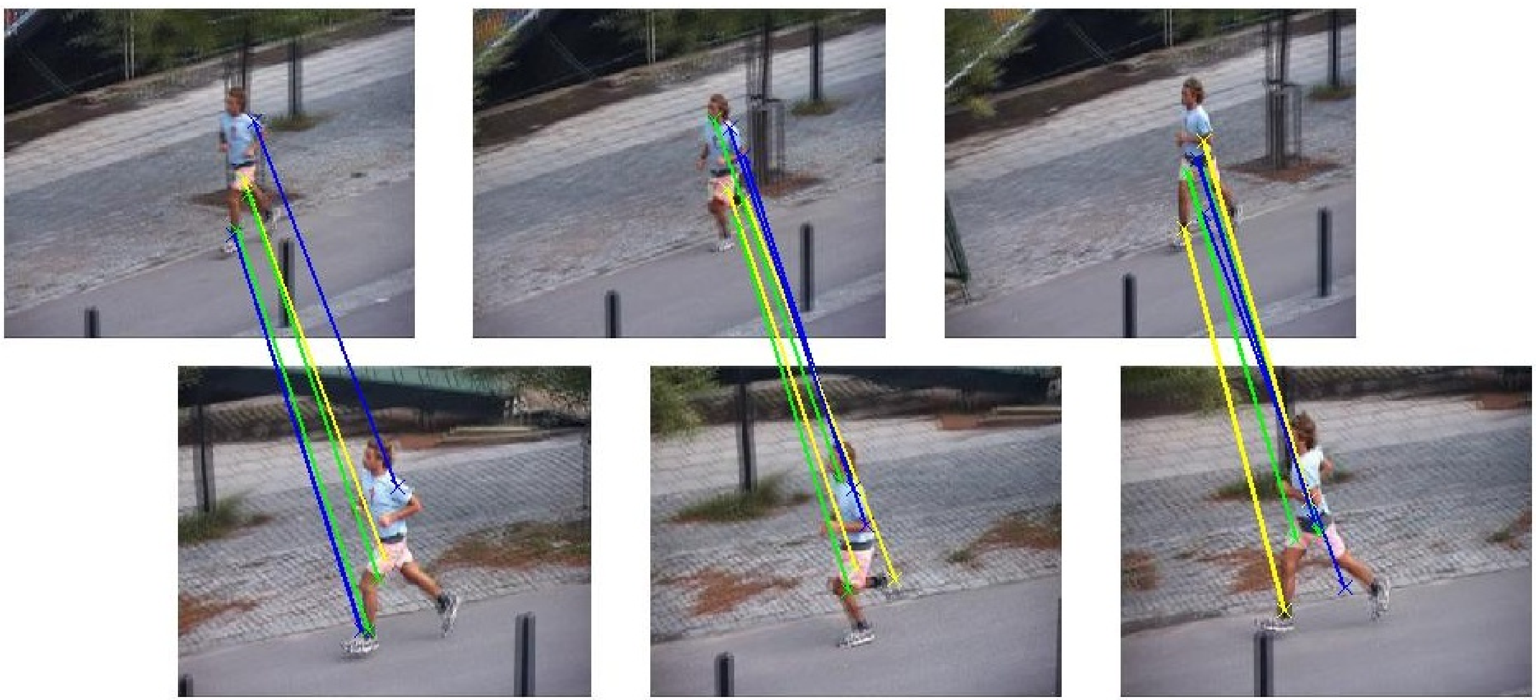
⇒ Periodic points can be related by a *dynamic homography:*

$$x_t = Hx_{t+p} \text{ with}$$

linear in time

$$H(t) = I + p(\mathbf{v}\mathbf{n}^\top - \mathbf{n}^\top\mathbf{v}I)/d - t\mathbf{n}^\top\mathbf{v}I/d$$

⇒ 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
  - Rao et.al. "*View-invariant alignment and matching of video sequences*", ICCV 2003
  - 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

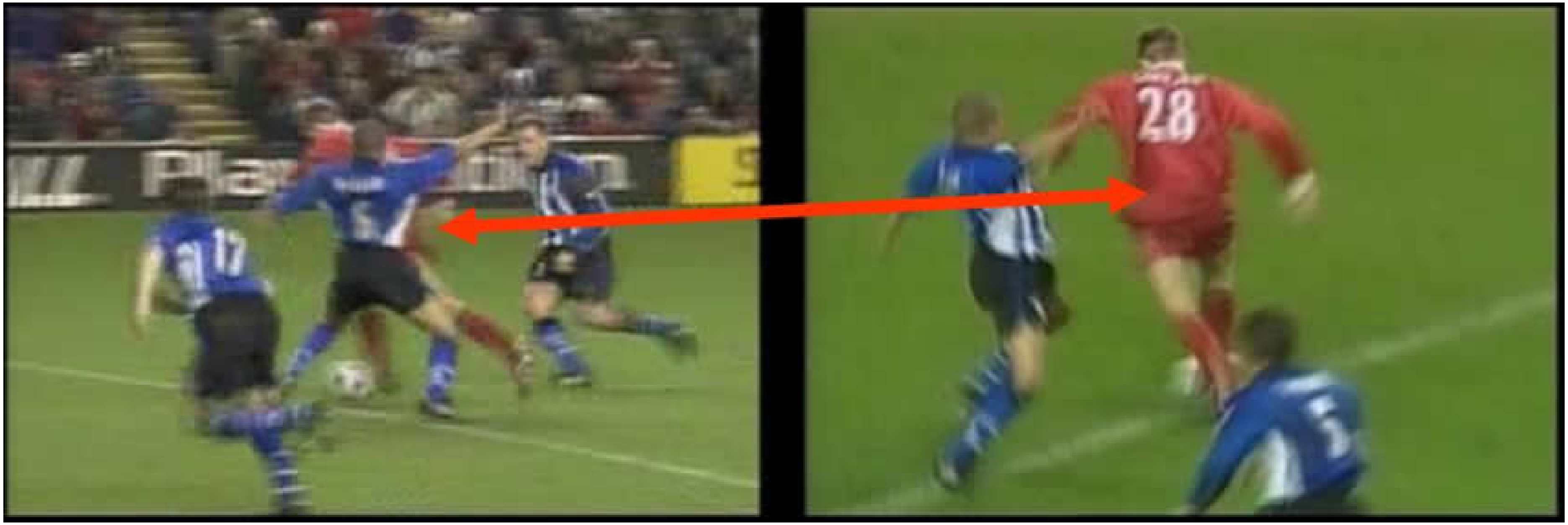
- Useful in
  - Reconstruction of dynamic scenes
  - *Recognition* of dynamic scenes

Future work: space-time alignment

- Video example

Future work: space-time alignment

- Example of awaited result



References

- Detection
  - Y. Freund and R. E. Schapire “*A decision-theoretic generalization of on-line learning and an application to boosting*”, J. of Comp. and Sys. Sc.1997.
  - I. Laptev “*Improvements of Object Detection Using Boosted Histograms*”, BMVC 2006
  - P. Viola and M. Jones “*Rapid object detection using a boosted cascade of simple features*”, CVPR 2001
  - http://www.pascal-network.org/challenges/VOC/voc2005/
  - http://www.pascal-network.org/challenges/VOC/voc2006/
- 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
  - P. Pérez, C. Hue, J. Vermaak, and M. Gangnet “*Color-based probabilistic tracking*”, ECCV 2002
  - J. Vermaak, S. Godsill, P. Pérez. “*Monte Carlo filtering for multi-target tracking and data association*” IEEE Trans. on Aerospace and Electronic Systems 2005
- Periodic motion Detection and Segmentation
  - Y. Boykov and V. Kolmogorov. “*An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision*”, IEEE-PAMI 2004.
  - 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