Video tracking

overview, applications and recent developments

Emilio Maggio

VICON
VISION + IMAGINATION

emilio@videotracking.org

Andrea Cavallaro

Queen Mary
University of London

andrea@videotracking.org

Resources

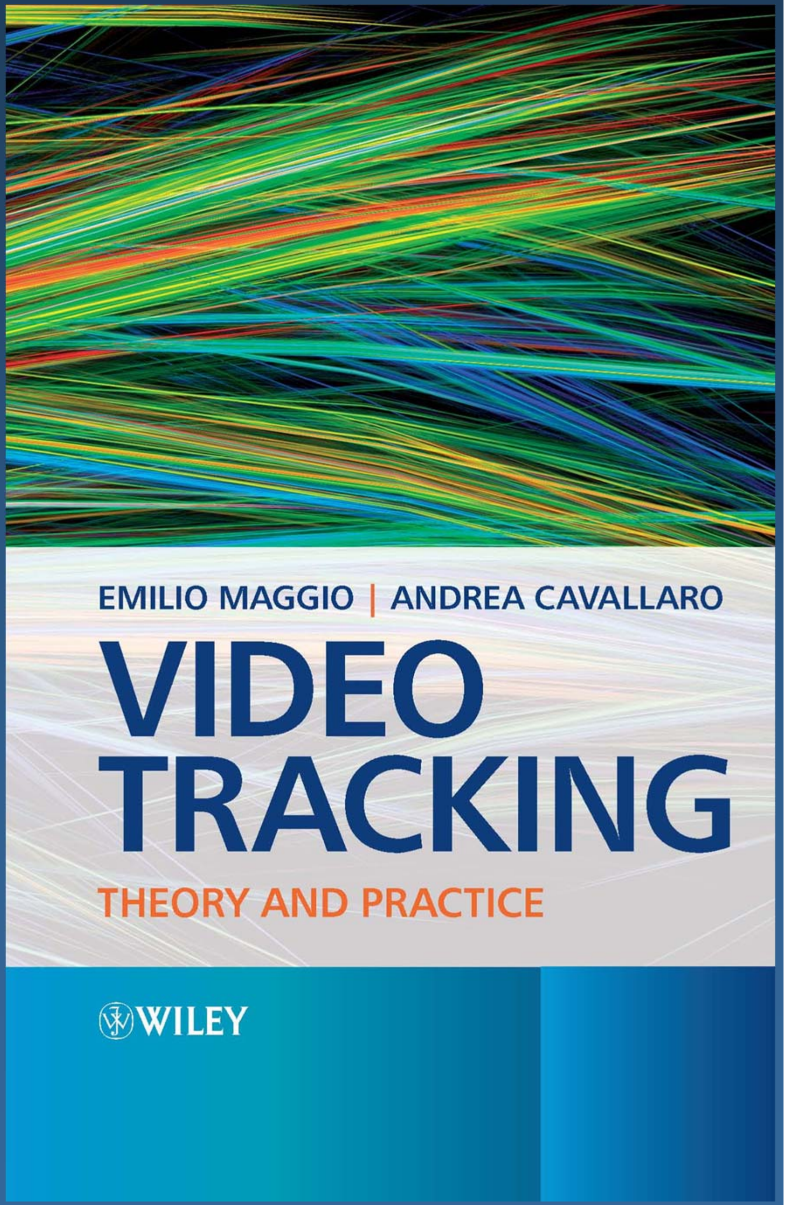
Book

Video Tracking: Theory and Practice

E. Maggio, A. Cavallaro

Wiley and Sons, Jan 2011

ISBN: 978-0-470-74964-7



Website

http://www.videotracking.org

Outline

- Video tracking: introduction
  - applications
  - problem statement
- Object representation
  - template and histograms
  - multiple features
- Single-target tracking
  - gradient-based trackers
  - Bayes' tracking
  - hybrid methods
- Multi-target tracking
  - data association
  - Random Finite Sets for tracking
- Outlook, links and references

applications

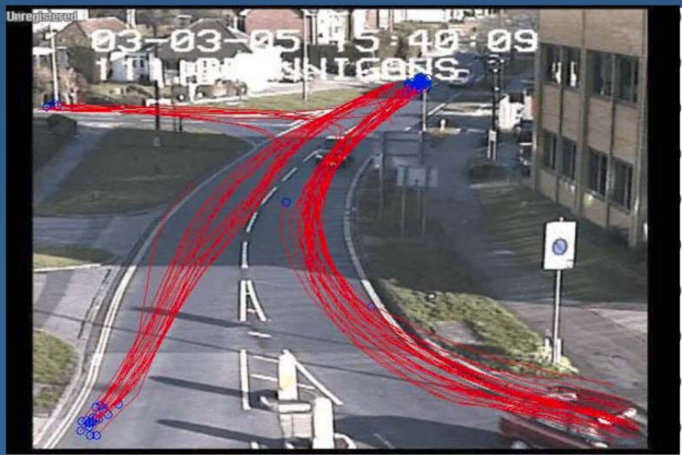

NPA
326
VICONPEAK
2D & 3D Motion Capture Systems

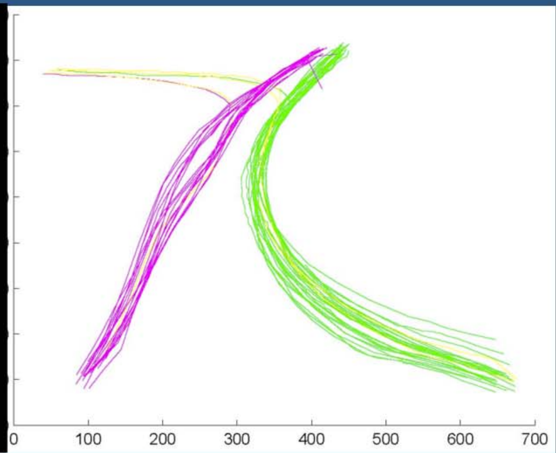
Spider-Man 3 (SPI)

Beowulf (Sony Pictures Imageworks)

Video tracking: applications

- Scene understanding
  - indexing events
  - clustering actions
    (e.g., retail intelligence)
  - semantic scene interpretation
  - mining video collections

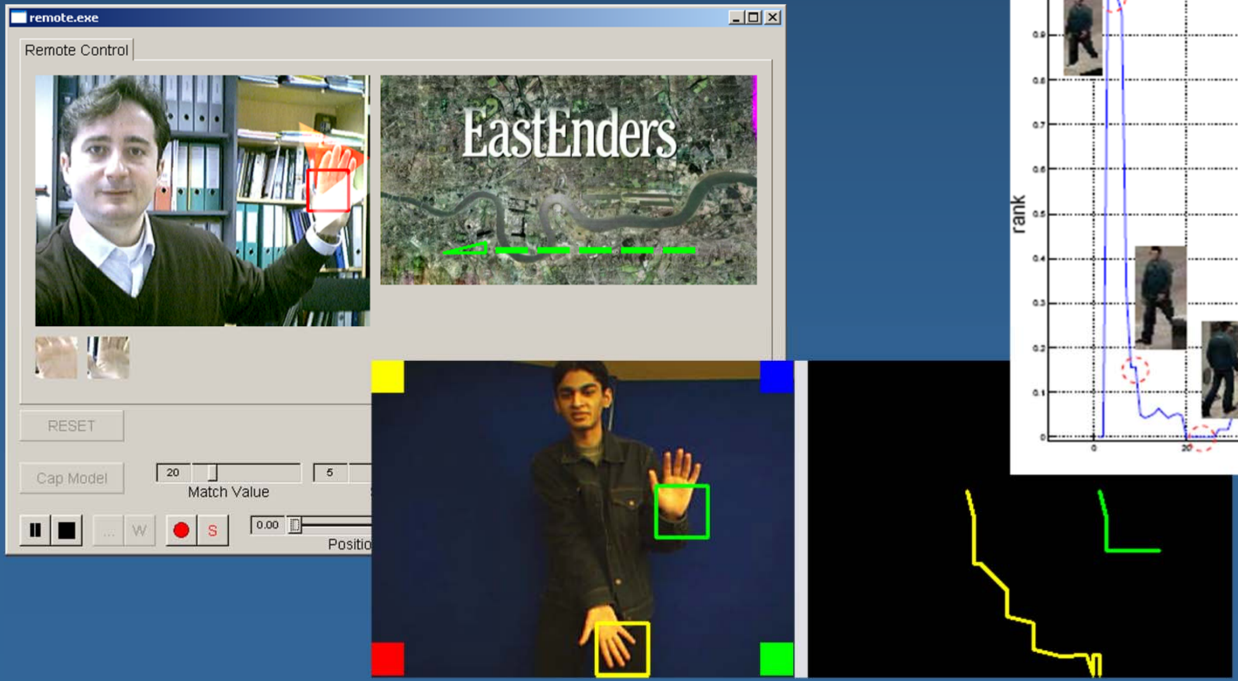


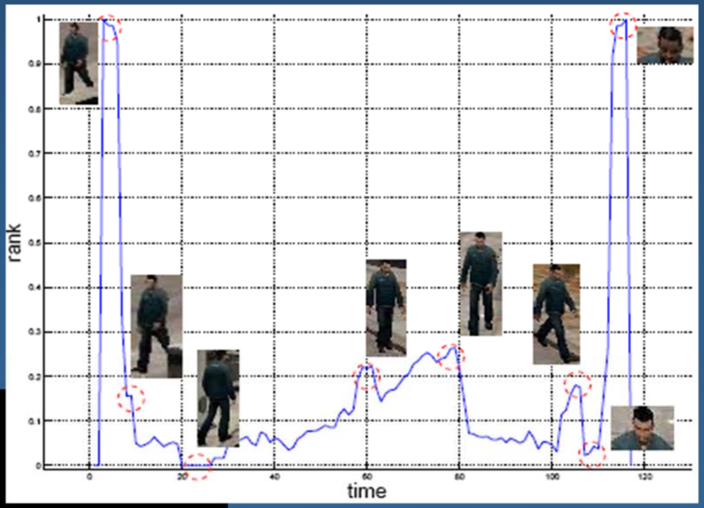


[Video]

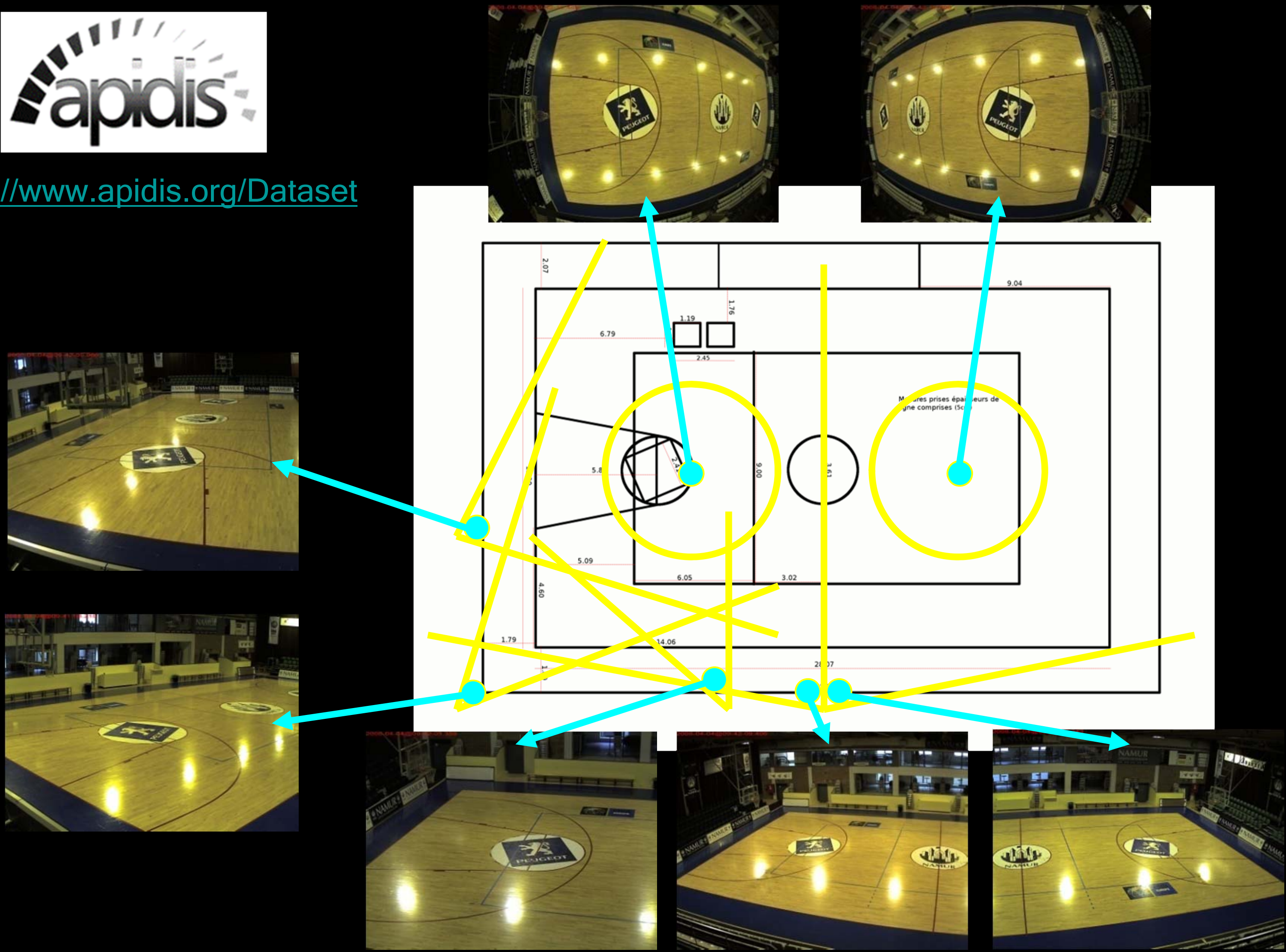
Video tracking: applications

- Camera control and natural interfaces
  - Prioritization/selection of multiple cameras
  - Smart rooms
  - Optimized scene sampling (PTZ cameras)
  - Multi-party immersive gaming (control)





http://www.apidis.org/Dataset



P1

P2

R1

R2

what’s behind all this?

Problem statement

- Objective
  - To estimate the target state over time (e.g. position, shape)

- Choices
  - object representation → target model
  - a searching process → generates candidate targets
  - a similarity measure → model-candidate distance (matching)

Note

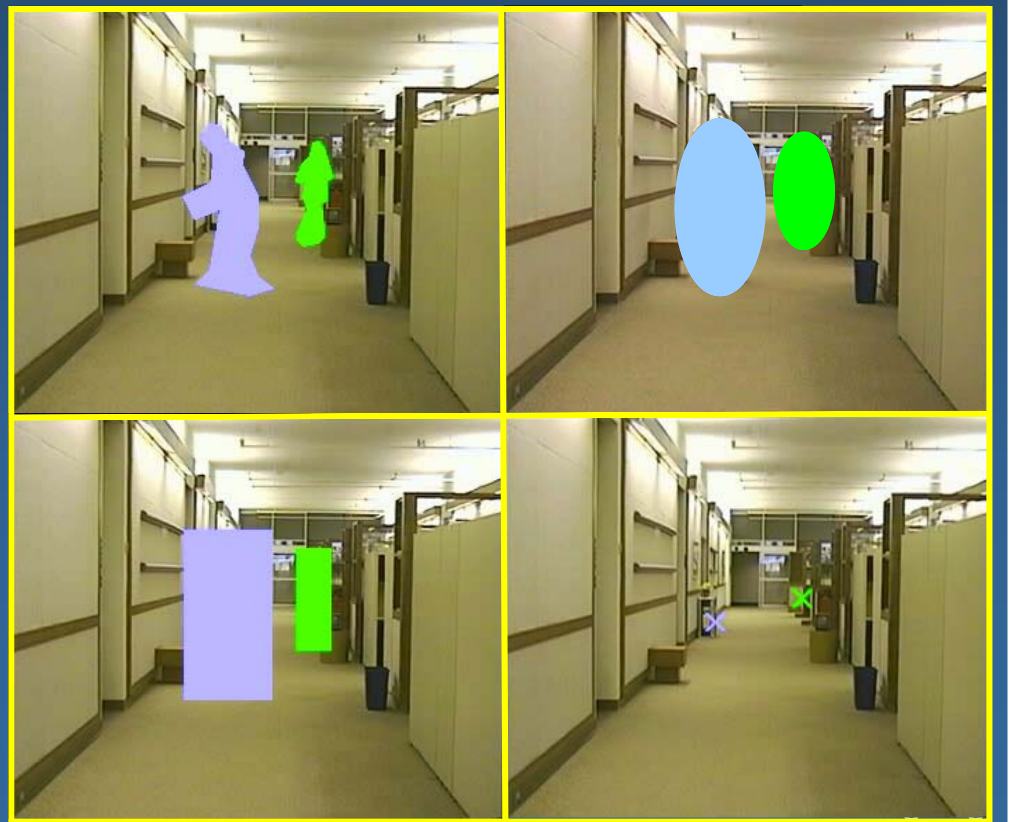
once you define the object of interest (target) → ~ recognition!

- but target resolution may be very small
  (compared to “traditional” recognition tasks), and
- objects do not dominate the image (large portions are background)

Video tracking: approximations

- Annotation of a video in terms of its component objects
  - <u>to localize</u> objects of interest
  - <u>to link</u> instances of the same object over time (tracking)

- Typical object approximations for tracking
  - Polygonal approximation
  - Bounding ellipse
  - Bounding box
  - Position only



Objective

- To estimate the state x_k from the observations $z_{1:k}$
  - x_k is a vector of object parameters (position, velocity, shape, etc.)
  - z_k is
    - the frame at time k -- appearance-based
    - the output of an object-detector -- detection-based
- State: examples
  - affine transformation parameters of a patch w.r.t. the first image
  - parameters of a shape or contour
  - parameters describing target appearance

Tracking: challenges

- Track management issues
  - Track initiation & termination
  - Occlusions handling
    - Partial / total occlusions
  - Spawning and splitting / merging
  - Target model update (drift problem)

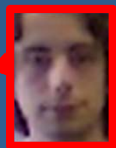
- What makes video tracking difficult?
  - Noisy observations
  - Changes in
    - object pose / scale
    - scene illumination
  - Other objects and background with similar appearance (clutter)

Outline

- Video tracking: introduction
  - applications
  - problem statement
- Object representation
  - template and histograms
  - multiple features
- Single-target tracking
  - gradient-based trackers
  - Bayes' tracking
  - hybrid methods
- Multi-target tracking
  - data association
  - Random Finite Sets for tracking
- Outlook, links and references

Target model: template

model I_T



Score function:

$$d(x) = \sum_{w} \left| I_T\left(A(x,w)\right) - I(w) \right|_2$$

Kanade-Lucas-Tomasi (KLT) tracker

- Gradient-based iterative approach [Lucas1981][Tomasi1991]
- Optimisation step
  - Objective: to minimise the score (assumes constant illumination)
  - Example: translation-only state

$$d(\Delta x) = \sum_w \left| I_T(w + \Delta x) - I(w) \right|_2$$

  - Template first-order approximation for small translations

$$I_T(w + \Delta x) \approx I_T(w) + \frac{\partial I_T(w)}{\partial w}' \Delta x$$

$$g(w)$$

Pre-computed gradient (of the template itself)

  - The optimal step is

$$\Delta x = \frac{\sum_w \left[I_T(w + \Delta x) - I(w) \right] g(w)}{\sum_w g(w)' g(w)}$$

Template: discussion

- Advantages
  - Simple → computationally fast
- Disadvantages
  - Over time → becomes non-representative of the object appearance (partial occlusions, out-of-plane rotations, noise)
  - Feature tracking: aperture problem
- Improved template
  - Temporal update
    - model each pixel in the template as a mixture of Gaussians that is updated over time [Zhou 2004]
    - include initial template to reduce drifting [Matthews 2004]
    - evolve based on a modified Kalman filter [Nguyen 2004]
  - Robustness to illumination variations
    - phase of wavelet coefficients (instead of color) [Jepson 2001]

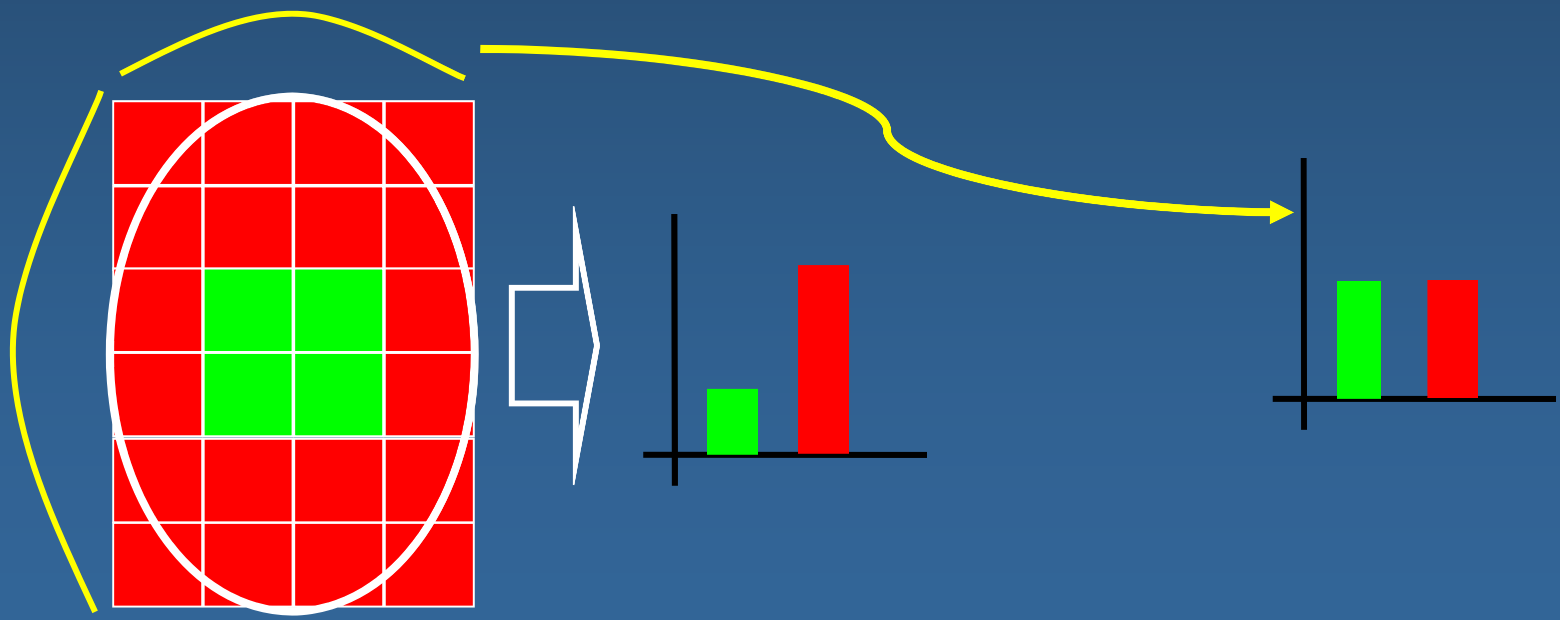
Target representation: can we do any better?

- Observation
  - The complete pixel information may be (is) redundant / misleading for the tracking task
- Target representation: desired properties
  - descriptive enough to disambiguate object vs. background
  - flexible enough to cope with changes of target
    - scale
    - pose
    - scene illumination
    - partial occlusions

→ *histograms*

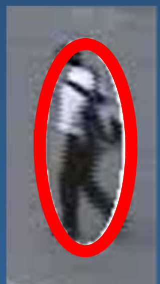
Color histograms for tracking

- Color histograms [Birchfield 1998] [Isard 2001] [Perez 2002] [Comaniciu 2003]
  - invariance to
    - scaling (normalized)
    - rotation
  - robustness to partial occlusions
  - data reduction and efficient computation

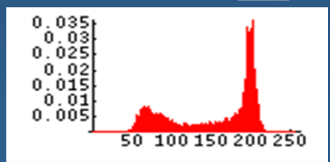


Likelihood of a candidate (score)

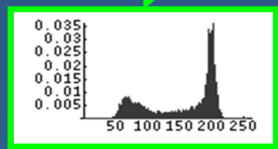
initialization

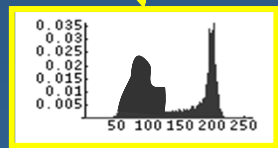


model $\hat{\mathbf{q}}$



k^{th} frame





candidates

$\hat{\mathbf{p}}(\mathbf{y})$

$$d\left[\hat{\mathbf{p}}(\mathbf{y}),\hat{\mathbf{q}}\right]=\sqrt{1-\sum_{u=1}^{m}\sqrt{\hat{p}_u(\mathbf{y})\hat{q}_u}}$$

⇨ *0.1* *0.7*

Target representation: template vs. color histogram

	Pixel-based representation (template)	*Statistical representation (normalized color distribution)*
data reduction	*No*	*Yes*
rotation & size invariance	*No*	*Yes*
flexibility	*No*	*Yes*
spatial information	*Yes (*)*	*No*

(*) If correctly updated is more descriptive

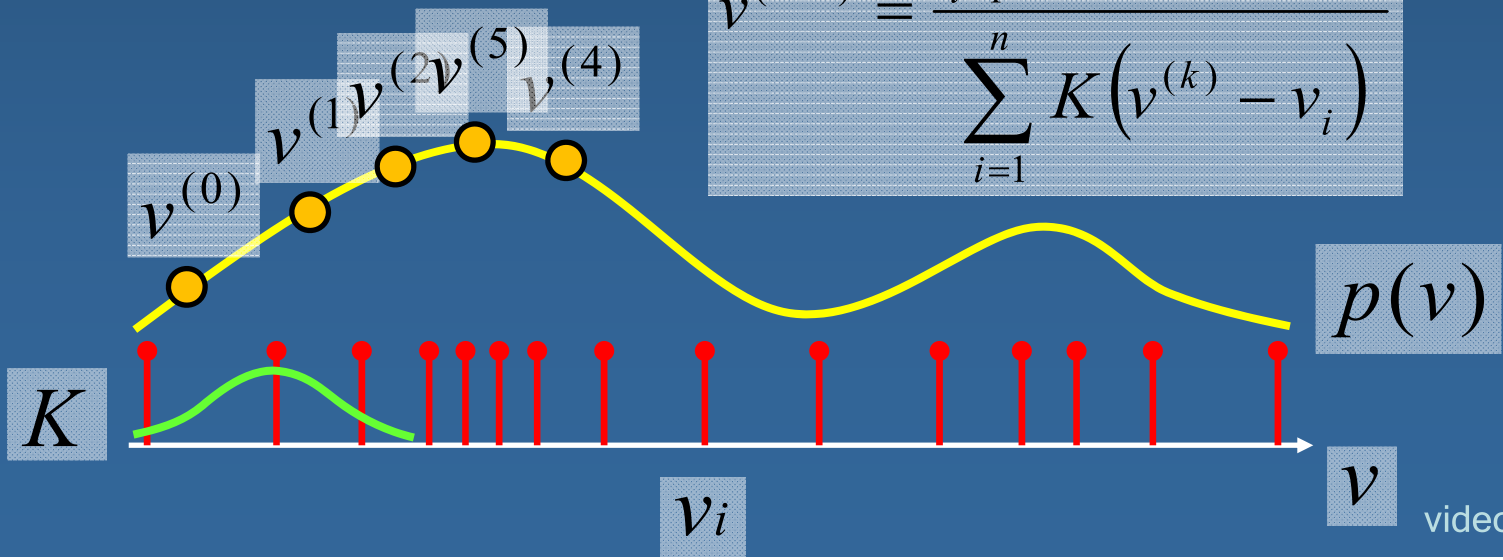
Mean shift (1)

- Sample Mean shift
  - Find a local maximum of a probability distribution *p(v)*
  - Input: set of samples drawn from *p(v)*
  - Kernel estimate

$$\hat{p}(v) = \frac{1}{n}\sum_{i=1}^{n} K\left(v - v_i\right)$$

  - Recursive optimization

$$v^{(k+1)} = \frac{\sum_{i=1}^{n} v_i K\left(v^{(k)} - v_i\right)}{\sum_{i=1}^{n} K\left(v^{(k)} - v_i\right)}$$



Mean shift (2)

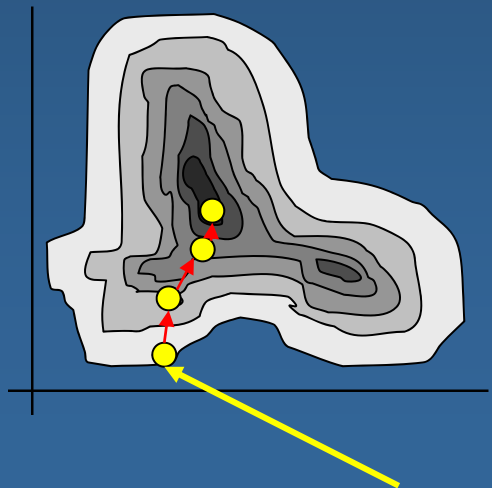
- Color Mean shift [Comaniciu 2001/3]
  - To minimize histogram distance
  - First-order Taylor approximation around $\hat{\mathbf{p}}(x^{(0)})$

$$d\left[\hat{\mathbf{p}}(x), \hat{\mathbf{q}}\right] \approx \frac{1}{2}\sum_{i=1}^{m}\sqrt{\hat{p}_i(x^{(0)})\hat{q}_i} + D\sum_{i=1}^{n} a_i K\left(\left\|\frac{x - w_i}{h}\right\|\right)$$

  - [First term]: constant with respect to the state x
  - [Second term]: weighted kernel density estimation
    - The samplers are the pixel locations w_i
    - The weights a_i depend on the two histograms
  - We can minimize the distance using a weighted Mean shift

Mean shift: summary

- Mean shift [Comaniciu 2001/3]
  - Deterministic non-parametric approach
  - Iterative procedure
  - Kernel-based
  - Gradient-based approach
    - If the distance function is smooth (kernel) → effective



Previous frame position

videotracking.org

Recursive Bayes tracking

- Propagates the posterior $p(x_k \mid z_{1:k})$
  - State eq. (dynamics) $\mathbf{x}_k = \mathbf{f}\left(x_{k-1}, \mathbf{u}_k\right) \Rightarrow f\left(x_k \mid x_{k-1}\right)$
  - Observation eq. (image) $\mathbf{z}_k = \mathbf{g}\left(x_k, \mathbf{n}_k\right) \Rightarrow g\left(z_k \mid x_k\right)$
- Prediction step

$$p(x_k \mid z_{1:k-1}) = \int f(x_k \mid x_{k-1}) p(x_{k-1} \mid z_{1:k-1}) dx_{k-1}$$

- Update step

$$p(x_k \mid z_{1:k}) = \frac{g(z_k \mid x_k) p(x_k \mid z_{1:k-1})}{\int g(z_k \mid x_k) p(x_k \mid z_{1:k-1}) dx_k}$$

Solving the Bayes equations

- Gaussian and linear
  - Kalman filter [Kalman 1960]
- Gaussian non-linear
  - Extended Kalman filter
    - first order Taylor expansion
  - Unscented Kalman filter -- [Julier 1997]
    - non-linear propagation of the first two order moments
  - Problem
    - Gaussian assumption does not hold in real-world image sequences
      - posterior multimodality, non-Gaussian noise
- Non-Gaussian non-linear
  - Monte Carlo methods -- Color feature: [VanGool 2002], [Perez 2002]

Sequential Monte Carlo (Particle Filter)

- Monte Carlo approximation of the Bayes recursion
  - Approximate the posterior *pdf*

$$p(x_k \mid z_{1:k}) \approx \sum_{i=1}^{L} \omega_k^{(i)} \delta(x_k - x_k^{(i)})$$

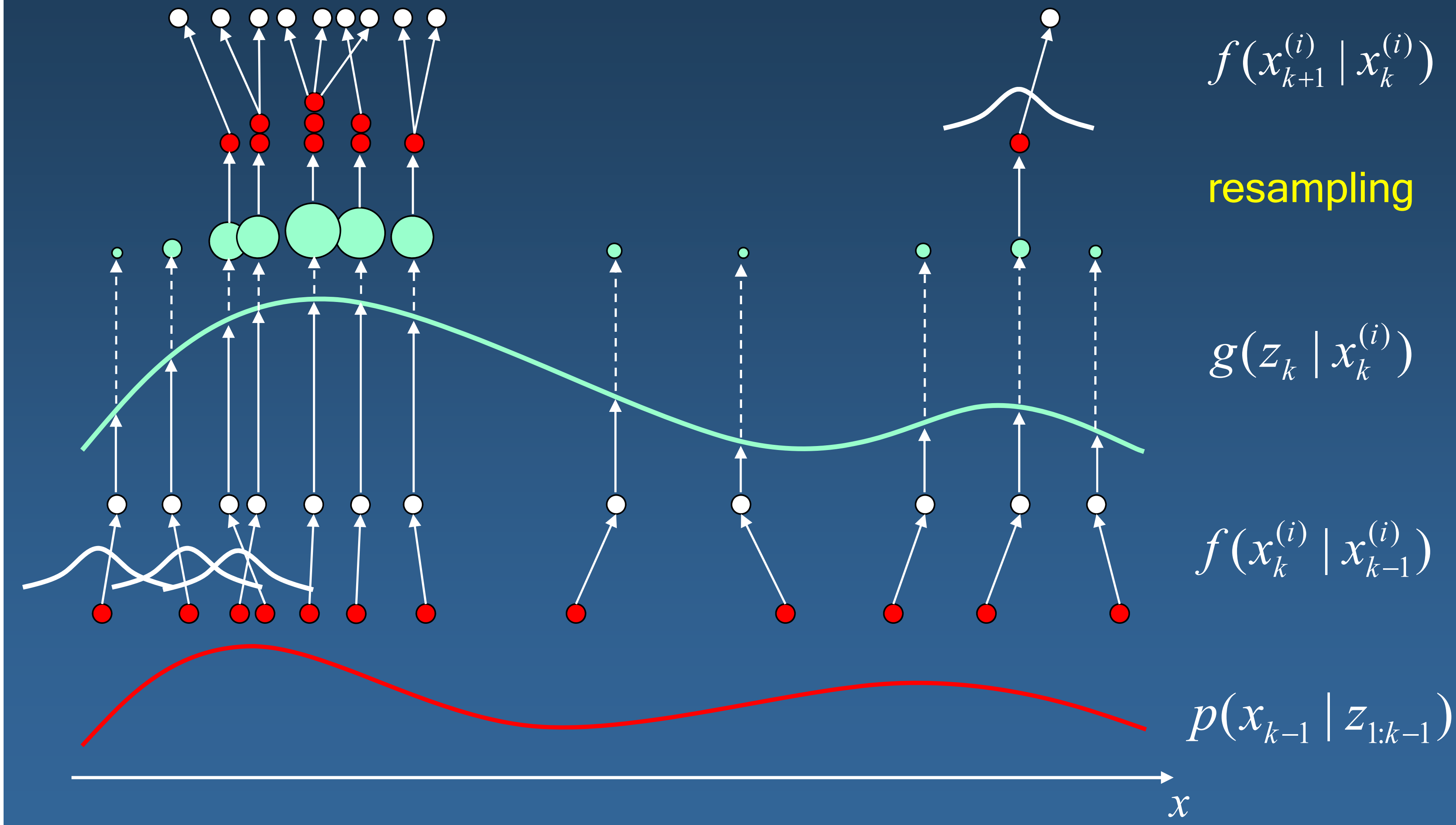
  - Samples propagated using state and observation equations

$$\omega_k^{(i)} \propto \omega_{k-1}^{(i)} \frac{g(z_k \mid x_k^{(i)}) \cancel{f(x_k^{(i)} \mid x_{k-1}^{(i)})}}{\cancel{q(x_k^{(i)} \mid x_{k-1}^{(i)}}, z_k)} \qquad E\left[x_k \middle| p\right] = \sum_{i=1}^{L} \omega_k^{(i)} x_k^{(i)}$$

- CONDENSATION [Isard 1996]
  - Sample according to dynamics $q(x_k \mid x_{k-1}, z_k) = f(x_k \mid x_{k-1})$
  - Systematic multinomial resampling $\omega_k^{(i)} \propto g(z_k \mid x_k^{(i)})$

likelihood

CONDENSATION



Particle filter: summary

- State $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{u}_k)$
- Observation $\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{n}_k)$
- Objective
  - to estimate unknown state $\mathbf{x}_k$ based on a sequence of observations $\mathbf{z}_k, k = 0,1,\ldots$
  - find the posterior distribution

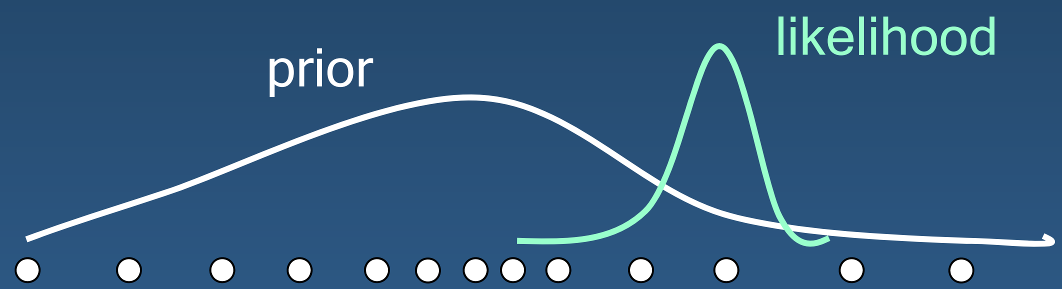
$$p(x_k \mid z_{1:k}) \approx \sum_{i=1}^{L} w_k^i \, \delta(x_k - x_k^i)$$

- Solution (Bayesian)
  - Prediction step
    - Based on state equation
  - Update step
    - Based on likelihood function

Particle filtering: resampling

- CONDENSATION sampling: no use of latest observation

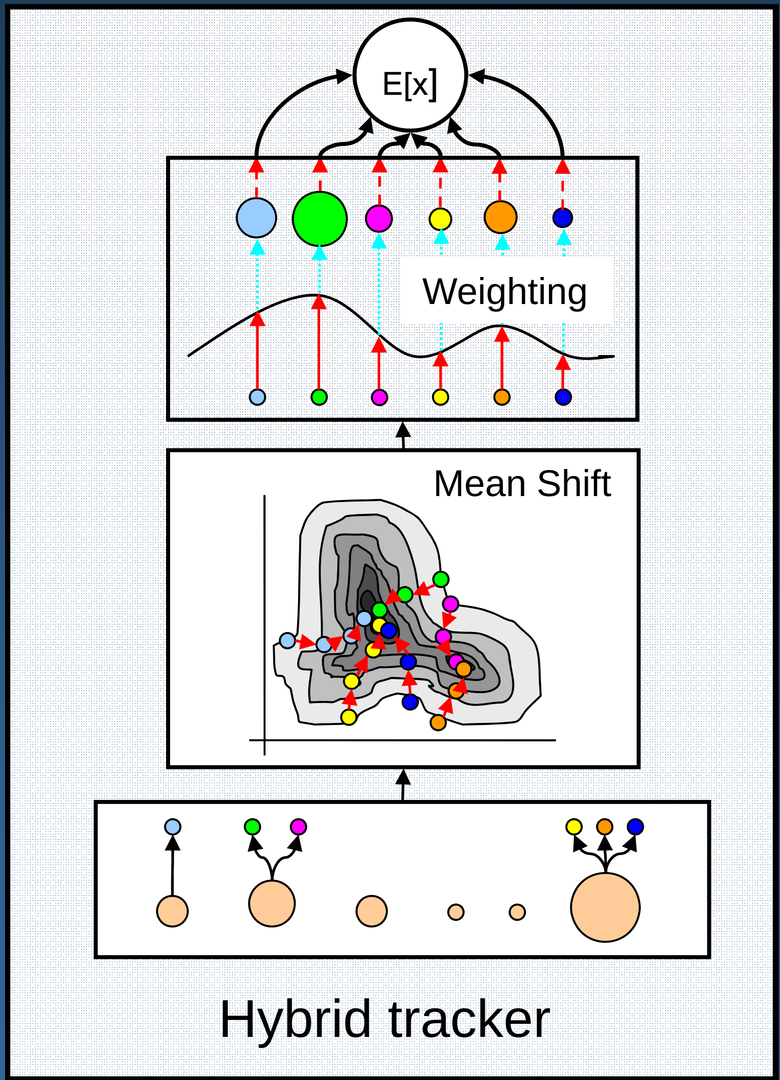
$$q(x_k \mid x_{k-1}, z_k) = f(x_k \mid x_{k-1})$$



- Just few samples near the peak of the posterior when:
  - Narrow likelihood ← discriminative target model
  - Likelihood lies on a tail of the prior ← manoeuvring target

- More particles sampled toward the peak of the likelihood

Hybrid (probabilistic-deterministic) sampling



- Advantages
  - After MS → each particle is near a local maximum of the likelihood
  - The efficiency of the particles is increased
  - Multi-modality of the posterior is maintained
  - Extra computation is compensated by fewer particles

Comparison: “Table tennis”

MS

PF-C

HY

Comparison: “Hand”

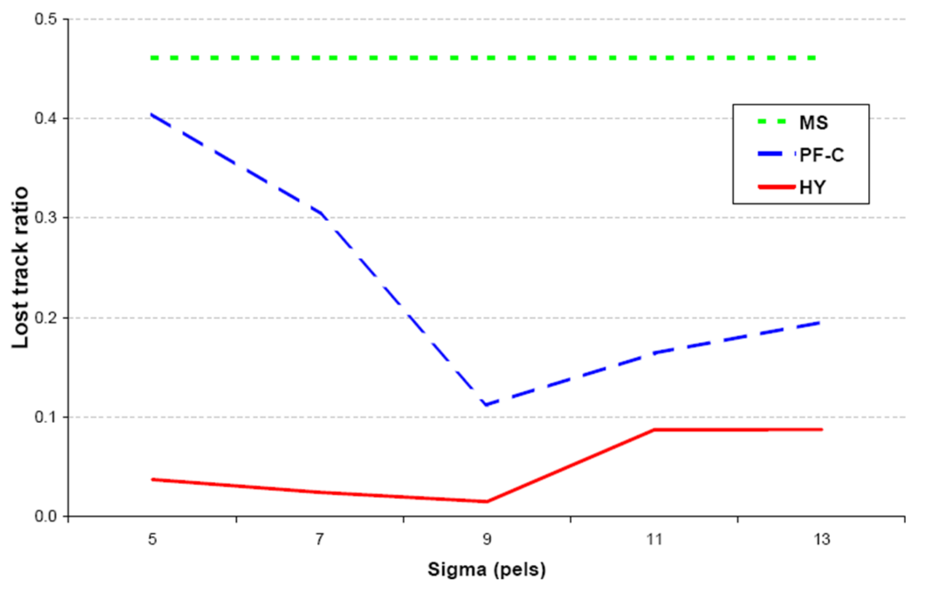
MS

PF-C

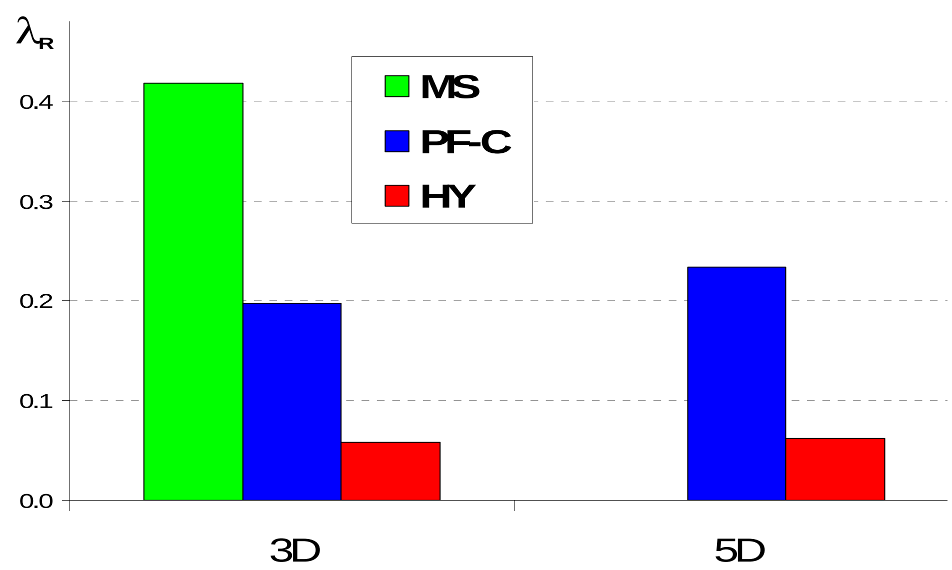
HY

Results

- What if we change the std of the dynamic noise (spread)?
  - Small std → PF-C cannot cope with fast motion
  - Large std → PF-C samples are not sufficiently dense
    → higher risk of attraction to false targets



Hand tracking



Dataset composed of 10 targets
Different motion behaviours
(hand, balls, faces, pedestrians)

Summary

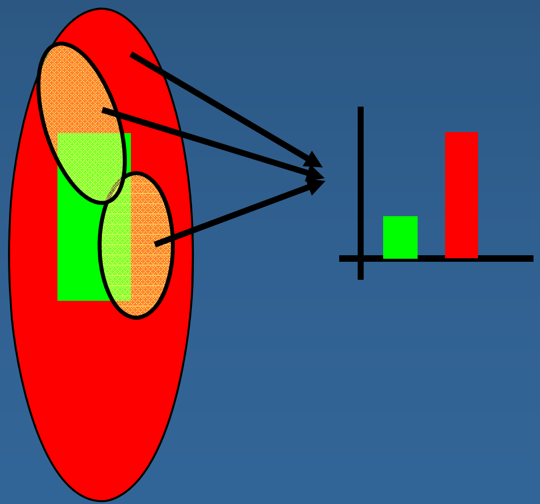
	Gradient-based search	Particle filter
advantages	Non parametric Low complexity (simple)	Multimodal *pdfs* Occlusion handling Flexibility (State)
disadvantages	Multimodal *pdfs* Occlusions Limits with fast targets	Parametric High complexity

end of part I

Color histogram: limitation

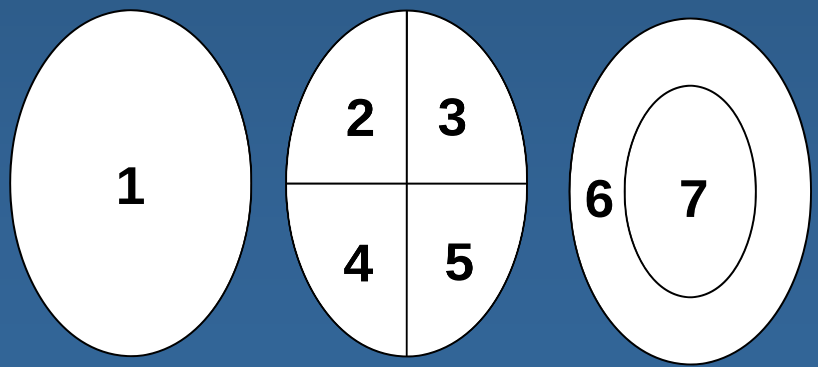
- What is the problem?

  lack of spatial information → difficult to discriminate targets with similar color properties



Improved color histogram representations

- Including spatial information
  - multiple histograms on different parts of the target [Perez 2002]
  - 2 non-overlapping areas (top and bottom parts) [Okuma 2004]
    - effective for the specific application (i.e., tracking ice-hockey players)
    - not necessarily effective on a generic target
  - More generic: 7 semi-overlapping areas [Maggio 2005]
    - global object information + spatial information
    - include rotation information
    - include size information

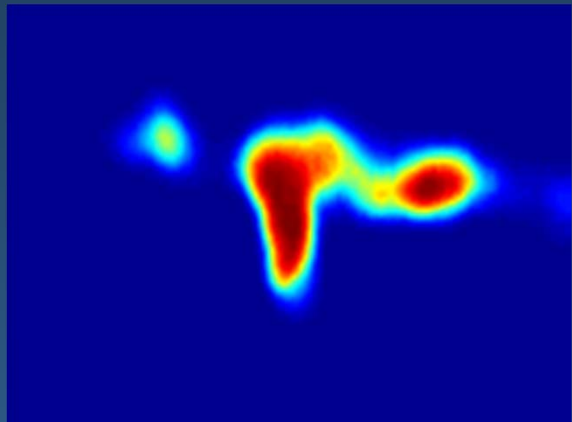


Multi-part histogram representation: example

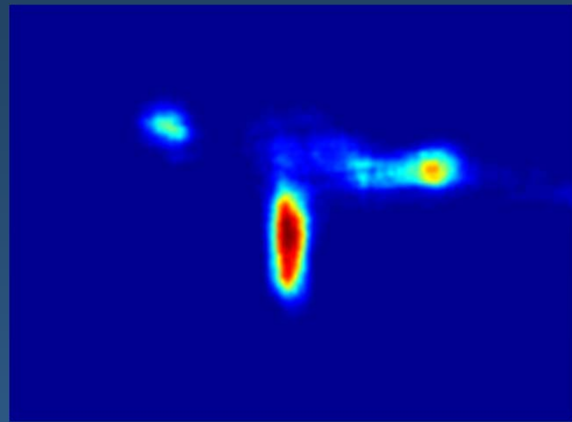
test image (crop)

model-candidate Bhattacharyya coefficients

(target: person)



with single
color histogram



with multi-part
color histogram

The videos with the results are available at
http://www.eecs.qmul.ac.uk/~andrea/MP.html

Comparison (using PF)

single color histogram

multi-part color histogram

time

Comparison (using MS)

single color histogram

multi-part color histogram

time

Comparison (using PF)

single color histogram

multi-part color histogram

time

To play with:

http://www.eecs.qmul.ac.uk/~andrea/dwnld/tsht/Setup.htm

Other ways to improve the target model

- Multiple features
  - to increase distinctiveness of the target model
  - e.g., gradient information to complement color information
    [Birchfield 1998] [Liu 2004] [Maggio 2005b/07]
  - Disadvantage: computationally more expensive

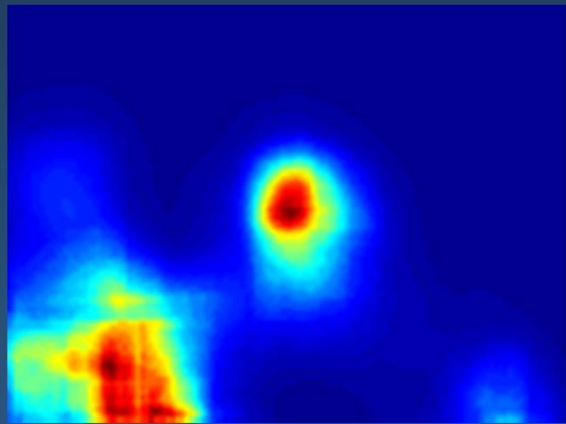
- Orientation histogram
  - the magnitude of the gradient
    at each pixel position is accumulated
    on the bin corresponding to its orientation
  - use estimate of the target orientation (provided by its state vector)
    to increase invariance to rotations

Color histogram vs. gradient histogram

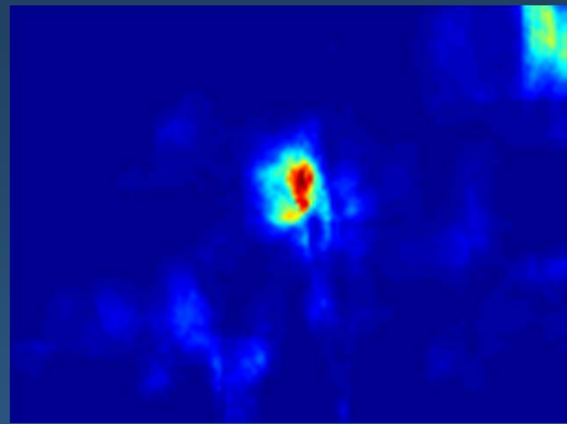
test image

(target: face)

color likelihood



gradient likelihood



- Likelihood-level fusion (mixture)

$$g(z_k \mid x_k) = \alpha_{k,c}\, g_c(z_k \mid x_k) + \alpha_{k,o} g_o(z_k \mid x_k)$$

- Adapt the weights $\alpha_{k,c}$ $\alpha_{k,o}$ according to feature reliability

Sample results

Color | Gradient | Color + Gradient

Fusion of multiple features

- Measurement level
  - color histogram +
    - edge map [Birchfield 1998]
    - edge density [Liu 2004]
    - shape [Shen 2003]
    - orientation histogram [Maggio 2005b/07]
      - feature reliability → spatial uncertainty → Particle spread weighted by the single-feature likelihood → adaptive

- Tracker level
  - Separate trackers
    - color, template, blob [Leichter 2004]
    - color, contour [Moreno-Noguer 2005]
  - Sequentially
    - color, blob, geometry [Veeraraghavan 2006]

Search process: summary

- KLT-template-based tracker
- Mean Shift (MS) tracker
  - Deterministic non-parametric approach, iterative procedure
  - Gradient-based approach
- Particle Filter (PF, PF-C)
  - Bayes' approach (prediction and update)
  - more reliable than Mean Shift with fast targets
- Hybrid approach(es) (HY)
  - Example: move the particles of PF toward the peaks of the likelihood

Outline

- Video tracking: introduction
  - applications
  - problem statement
- Object representation
  - template and histograms
  - multiple features
- Single-target tracking
  - gradient-based trackers
  - Bayes' tracking
  - hybrid methods
- Multi-target tracking
  - data association
  - Random Finite Sets for tracking
- Outlook, links and references

Multi-target tracking

- Detections -- input

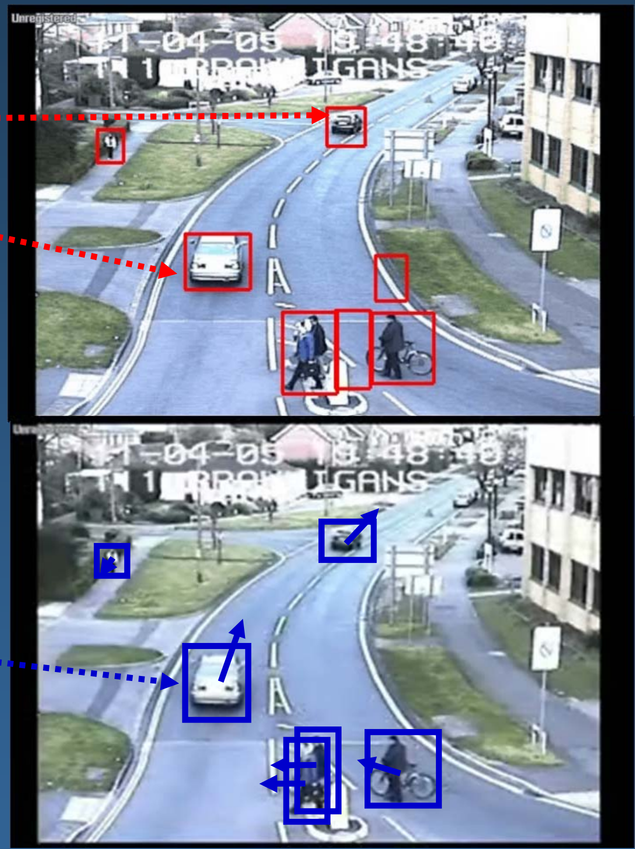
$$Z_k = \{z_{k,1}, \ldots, z_{k,N(k)}\}$$

$$z = [y_1, y_2, w, h]$$

- State -- output

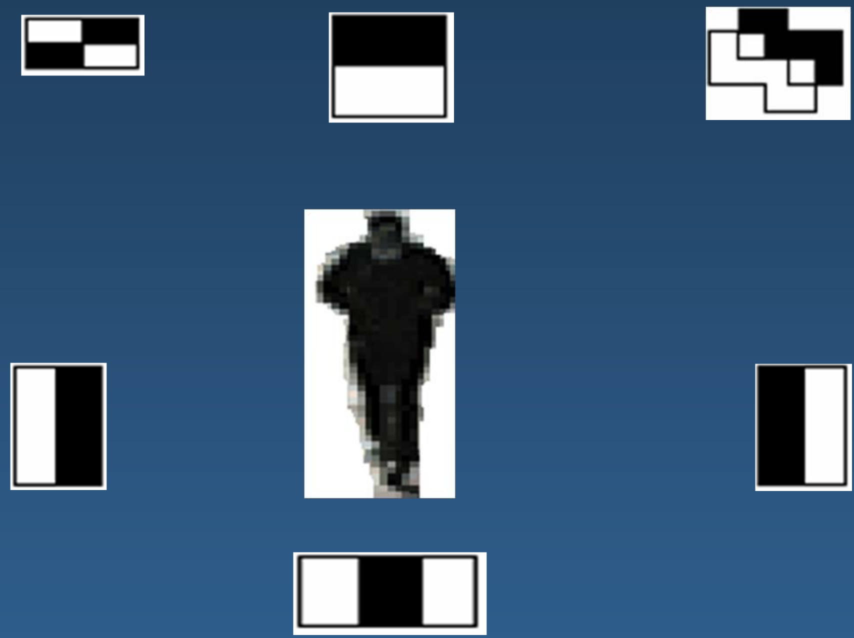
$$X_k = \{x_{k,1}, \ldots, x_{k,M(k)}\}$$

$$x = [y_1, y_2, \dot{y}_1, \dot{y}_2, w, h]$$



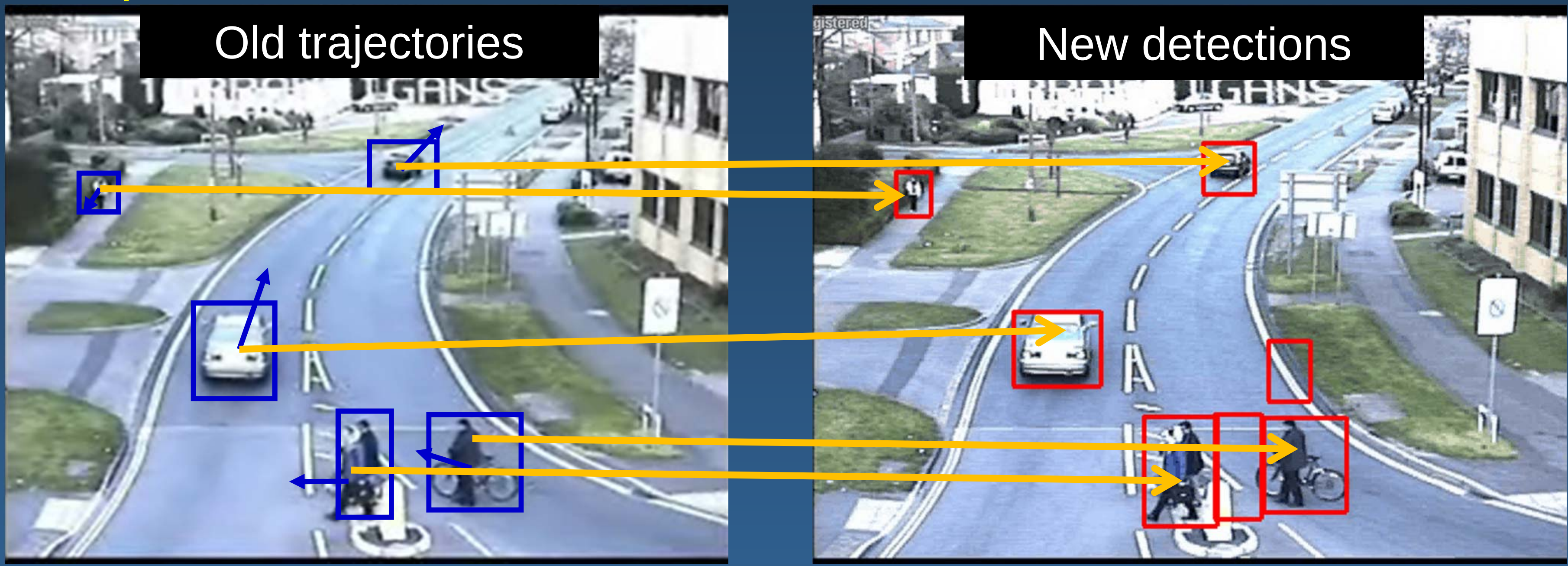
Target detection

- Detection examples
  - frame-based (usually)
    - can be improved with temporal features (e.g., pedestrians)
  - **trained classifier**
    - choice of training set
      - negative examples
      - poses
  - **change detector**
    - choice of features to be used
    - choice of background model
      - updating rate
    - global motion compensation (moving cameras)
    - global and local illumination variations



Data association (basics)

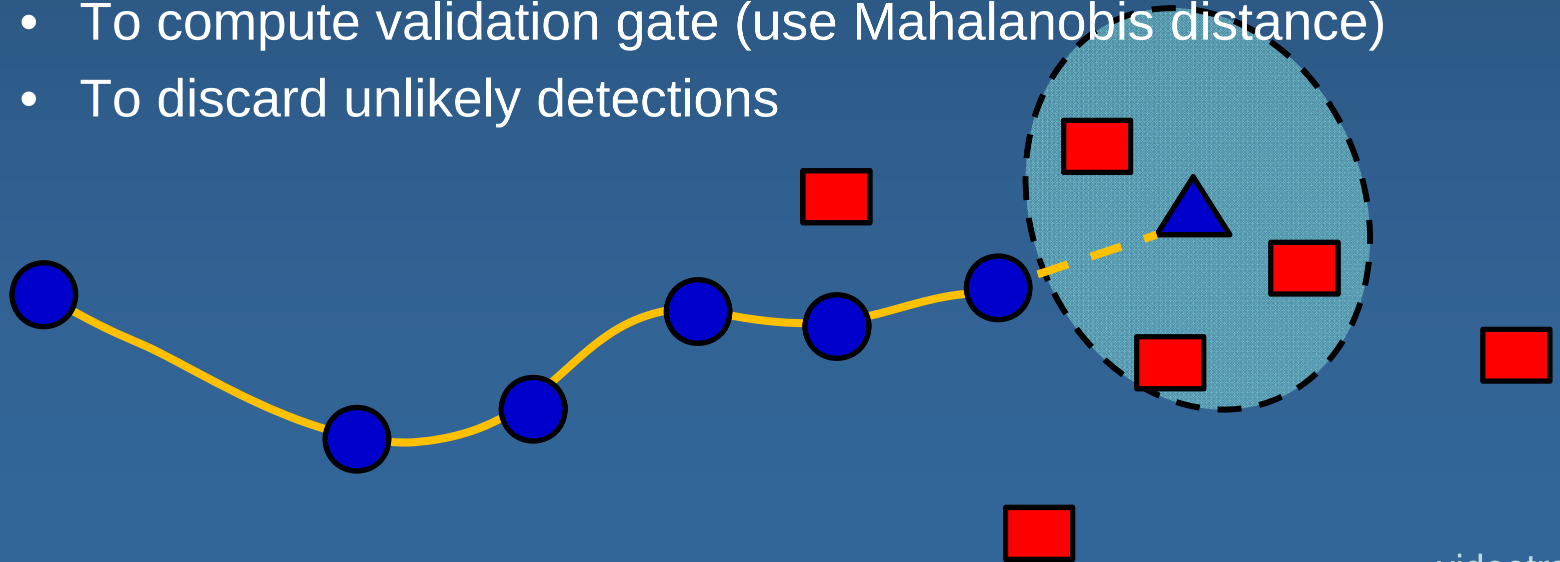
- Input



- Common assumption
  - Each target generates at most one detection
- Goal
  - To find the best assignment between trajectories and detections
  - To use the assignment to estimate the current state

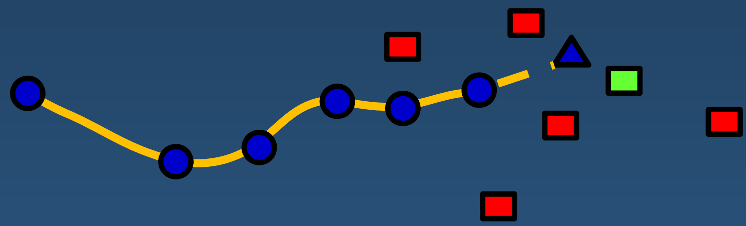
Data association (gating)

- Problem
  - The number of association hypotheses <u>grows quickly</u> with the number of targets
- Observation
  - Not all the detections are compatible with the trajectories
- Gating (to find the good candidates)
  - To predict the target position
  - To compute validation gate (use Mahalanobis distance)
  - To discard unlikely detections



Nearest neighbor (NN)

- Sub-optimal solution
  - Associate to each trajectory the closest detection

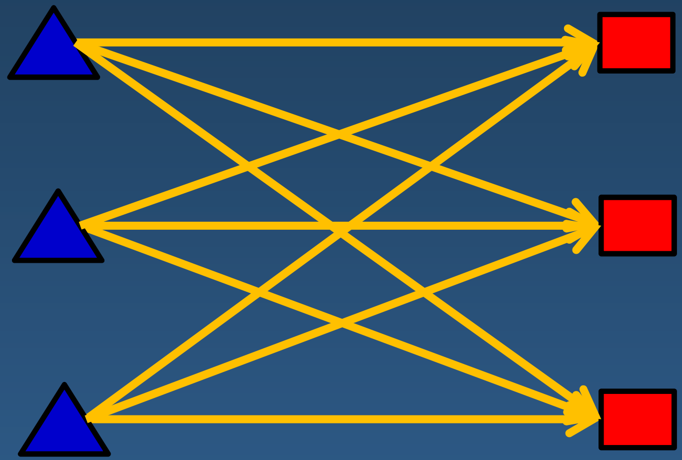


- Problem
  - Each detection may be associated to multiple trajectories
- Global solution
  - To solve the global association problem (global NN)
  - To find the best association ζ_{best} that maximise the score

$$\zeta_{best} = \arg\max_{\zeta} \sum_{z \in Z} c\left(z, \zeta(x)\right)$$

Linear assignment

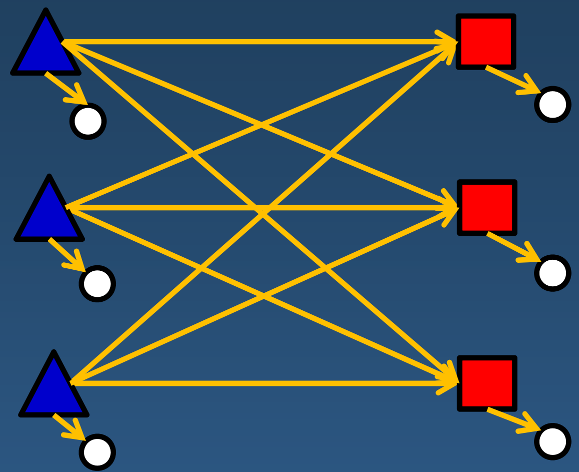
- Global solution (graph interpretation)
  - nodes
    - predictions
    - detections
  - edges
    - associations



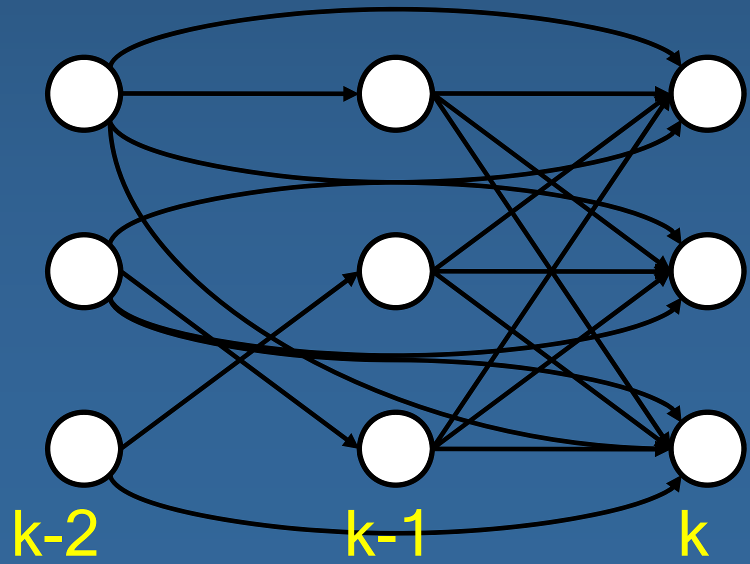
- Best association = maximum path-cover of the graph
  - algorithm: Hungarian search
  - complexity: cubic with the number of targets

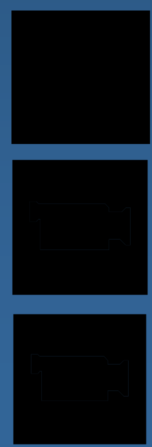
Linear assignment (2)

- Extra sink and source nodes can simulate
  - Target births and deaths
  - False and missing detections



- Can be done on multiple frames [Shafique 2005]





ex 1

ex 2

ex 3

Multi-target trackers

- Joint Probabilistic Data Association Filter (JPDAF)
  - shares detections between multiple trajectories (two frames)
  - can handle clutter
  - complexity: **quadratic** with number of targets
- Multiple Hypotheses Tracker (MHT)
  - evaluate hypotheses over multiple frames
  - complexity: **exponential** with time; **cubic** with number of targets
- PHD filter-based tracking
  - extend Bayesian recursion to multiple targets (w. Random Finite Sets)
  - can incorporate contextual information
  - complexity: **linear** with the number of targets

Multi-target Bayes filter

Single target

Random vectors

$$\mathbf{x}_k = \mathbf{f}\left(x_{k-1}, \mathbf{u}_k\right)$$
$$\mathbf{z}_k = \mathbf{g}\left(x_k, \mathbf{n}_k\right)$$

Multiple targets

Random Finite Sets (RFS)

$$\Xi_k = S_k\left(X_{k-1}\right) \cup \Gamma_k$$
$$\Omega_k = \Theta_k\left(X_k\right) \cup \mathrm{K}_k$$

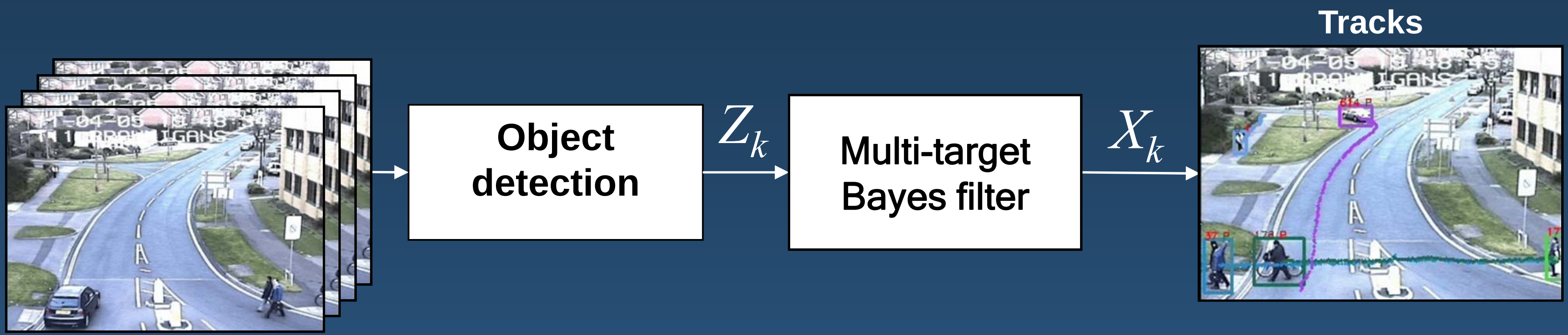
- Recursively propagates the multi-target state
  - Prediction

$$p(X_k \mid Z_{1:k-1}) = \int f(X_k \mid X_{k-1})\, p(X_{k-1} \mid Z_{1:k-1})\, \mu_s\left(dX_{k-1}\right)$$

  - Update

$$p(X_k \mid Z_{1:k}) = \frac{g(Z_k \mid X_k)\, p(X \mid Z_{1:k-1})}{\int p(Z_k \mid X_k) p(X_k \mid Z_{1:k-1}) \mu_s\left(dX_k\right)}$$

Multi-target Bayes filter (2)



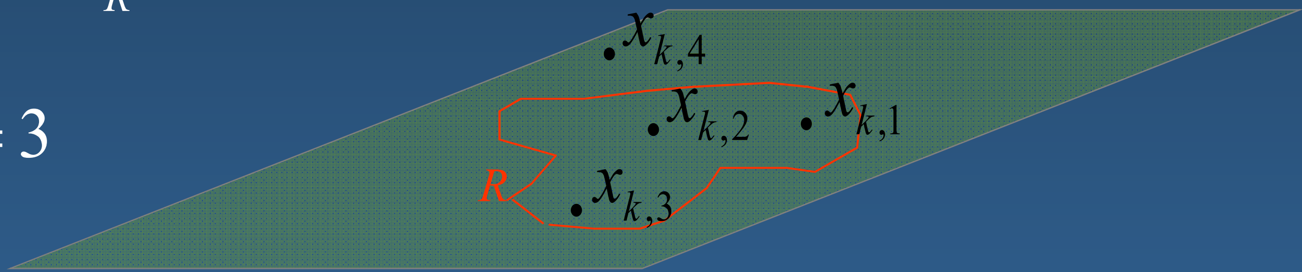
- Model target birth, death, clutter and missing detections
- Monte Carlo approximation:
  - Dimensionality of the state grows with the number of targets
  - Number of particles is exponential with the number of targets
  - Not feasible in real-world applications! ☹
- Solution: to propagate the first order moment of the RFS only

Probability Hypothesis Density (PHD)

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

$$\int_R D_\Xi(x)dx = E\left[\left|\Xi \cap R\right|\right]$$

$$\left|X \cap R\right| = 3$$



Multiple-target Bayes filter	$\longrightarrow$	$p(X_{k-1}\|Z_{1:k-1})$	$\longrightarrow$	$p(X_k\|Z_{1:k-1})$	$\longrightarrow$ ($\uparrow Z_k$)	$p(X_k\|Z_{1:k})$	$\longrightarrow$	…
		$\downarrow$		$\downarrow$	Z_k	$\downarrow$		
PHD Filter	$\longrightarrow$	$D_{k-1\|k-1}$	$\longrightarrow$	$D_{k\|k-1}$	$\longrightarrow$ ($\downarrow Z_k$)	$D_{k\|k}$	$\longrightarrow$	…

Recursive PHD filter

- Prediction

$$D_{k|k-1}(x_k) = \gamma(x) + \int \phi(x_k, x_{k-1}) D_{k-1|k-1}(x_{k-1}) dx_{k-1}$$

Birth intensity

  - PHD state transition

$$\phi(x_k, x_{k-1}) = e(x_k) f(x_k \mid x_{k-1})$$

Survival probability

Single target transition

- Update

$$D_{k|k}(x_k) = \left[p_M(x_k) + \sum_{z \in Z_k} \frac{\psi_z(x_k)}{\kappa(z) + \langle \psi_z, D_{k|k-1}(x_k) \rangle} \right] D_{k|k-1}(x_k)$$

Missing detection

Clutter intensity

Normalization

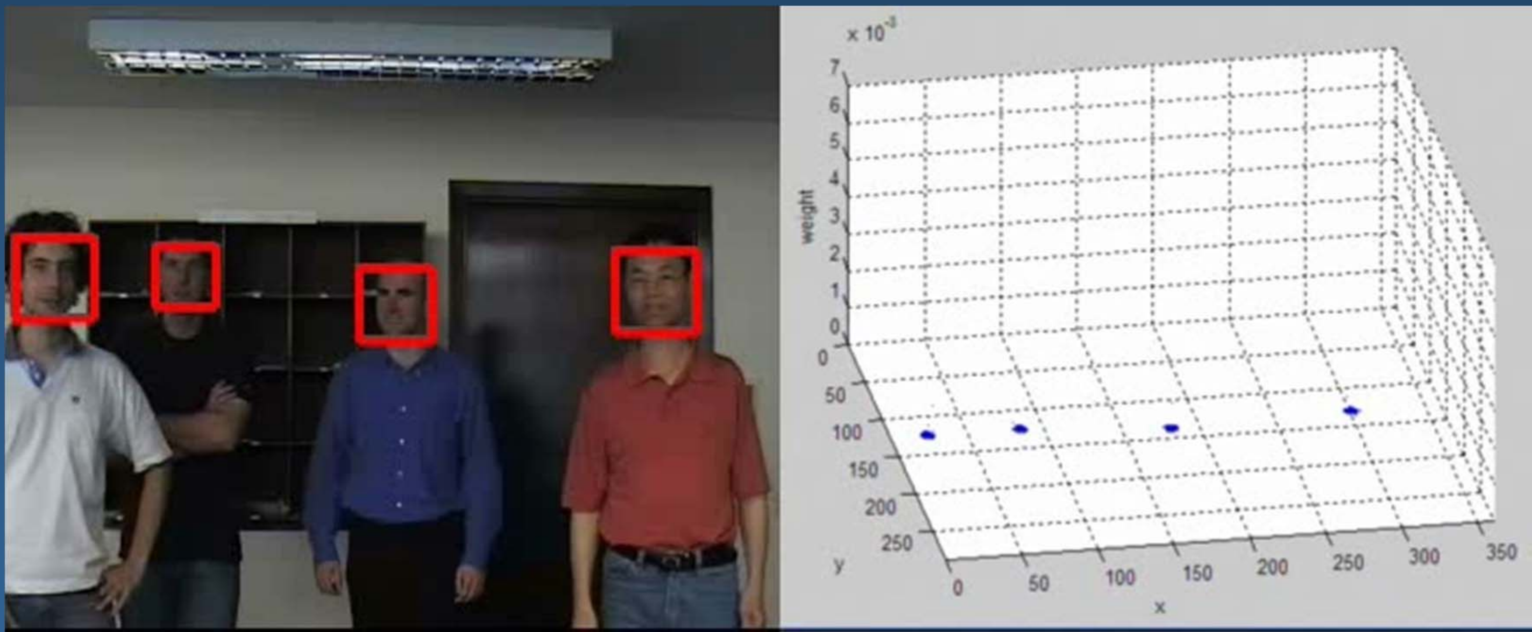
  - PHD likelihood

$$\psi_z(x) = \left(1 - p_M(x)\right) g(z \mid x)$$

Single target likelihood

Example: PHD propagation

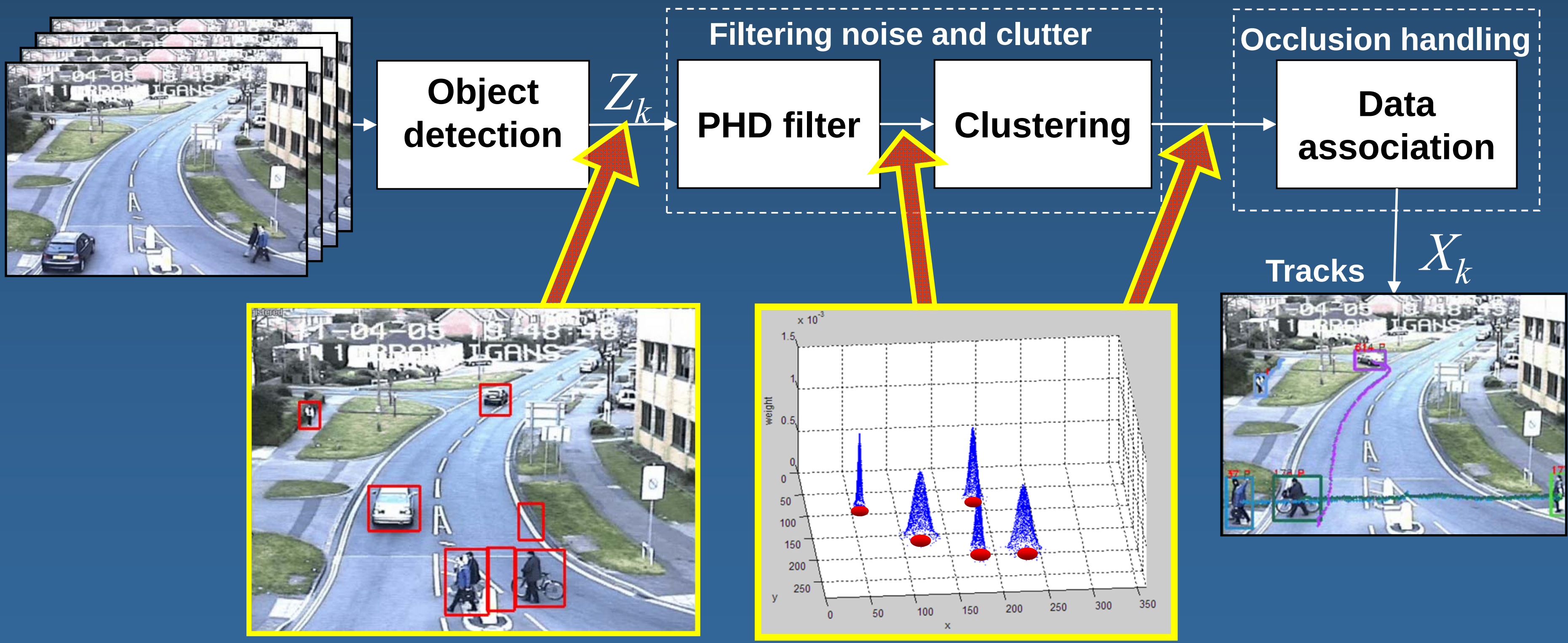
- Input: Face detector
- Output: Monte Carlo approximation of the PHD



When a new target appears the mass of the corresponding peak smoothly grows and converges to 1

Tracking with the PHD filter

- Advantage: Linear complexity with the target number
- Drawback: Data association not integrated, use multi-frame graph-based data association



Results – CLEAR/i-LIDS dataset

Detections PHD output

Object trajectories

Efficient multi-target visual tracking using Random Finite Sets

E. Maggio, M. Taj, A. Cavallaro

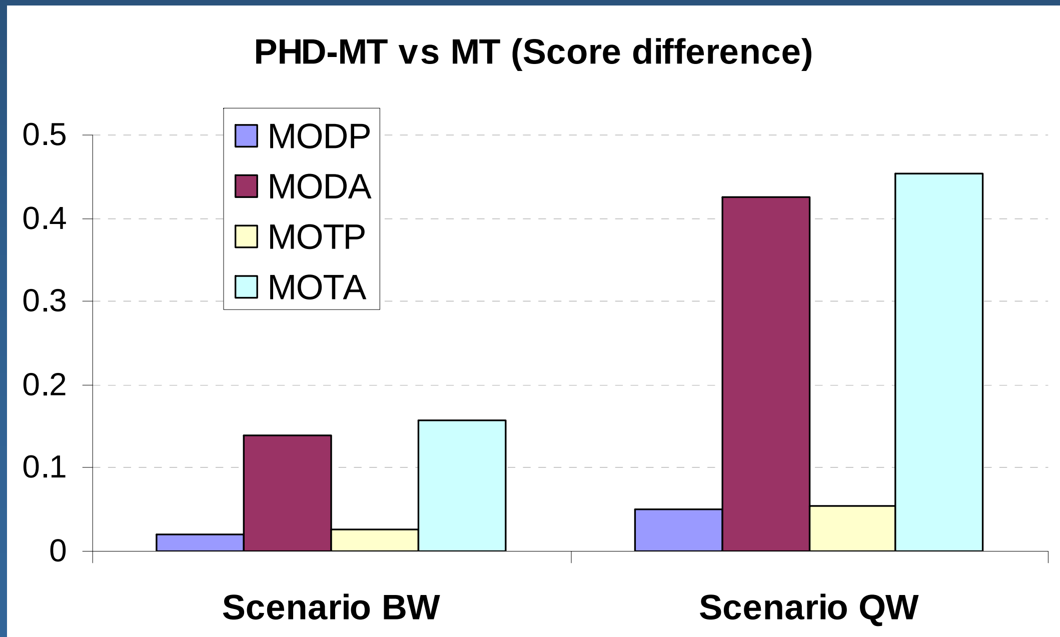
IEEE Trans. on Circuits and Systems for Video Technology, August 2008, pp.1016-1027

Video (with face detector and motion detector)

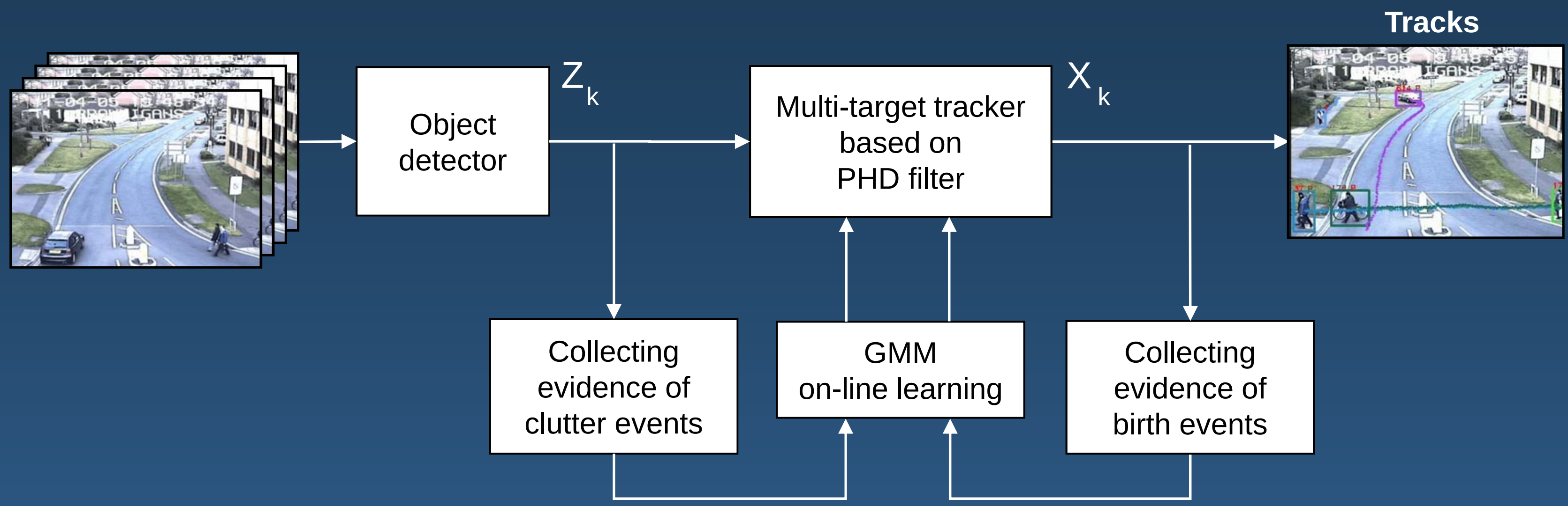
Evaluation: CLEAR-2007 dataset

- Dataset
  - 2 scenarios
  - 1 hour 20 minutes of annotated video
  - 50 evaluation segments

- Methodology: CLEAR-VACE (4 scores) [Kasturi '06]



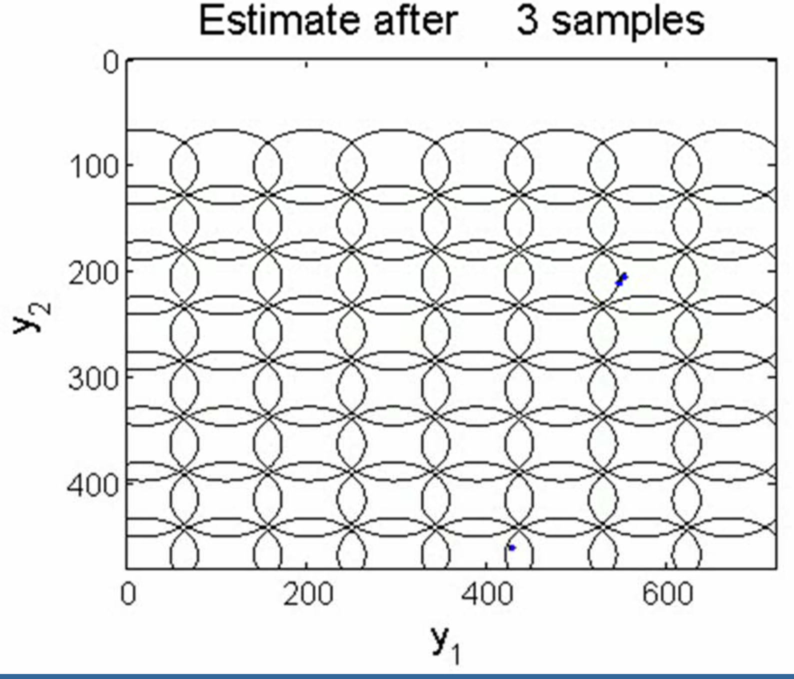
Using scene context in the PHD recursion

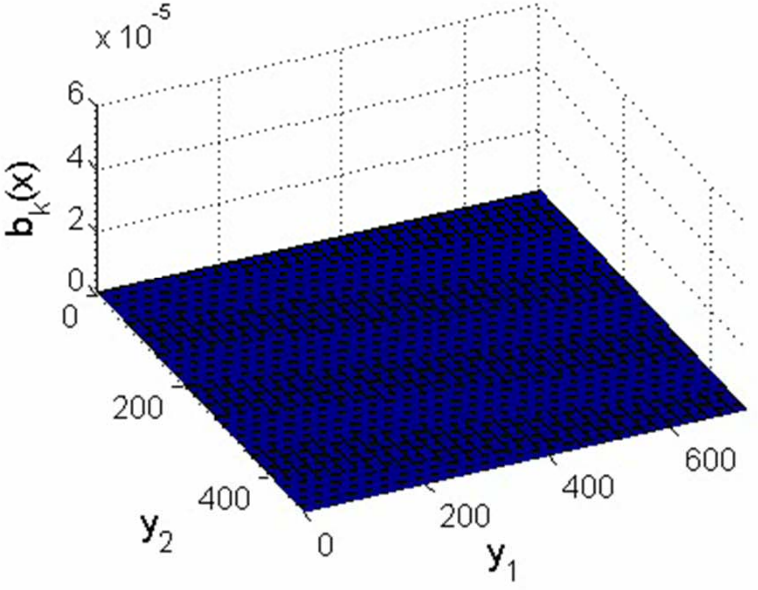


- Goal: context adaptive filtering
  - Birth intensity $\gamma(x)$ -- to model entry areas
  - Clutter intensity $\kappa(z)$ -- to model detection errors
- Data collection
  - Birth events from the output of the tracker
  - Clutter events via user feedback
- Density estimation -- GMM on-line learning

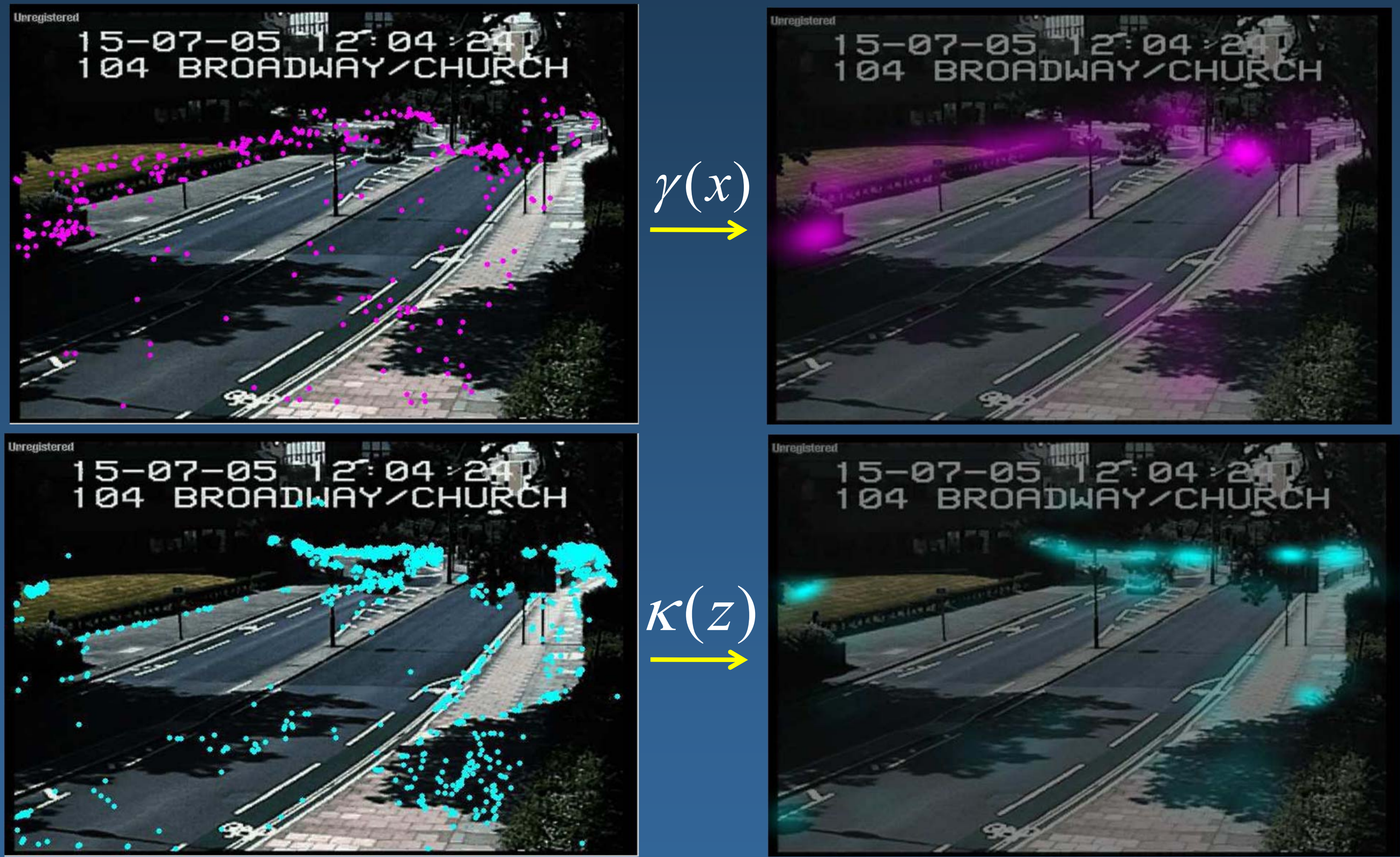
Example: learning birth intensity

Learning scene context for multiple object tracking
E. Maggio and A. Cavallaro
IEEE Trans. on Image Processing,
August 2009, pp. 1873-1884





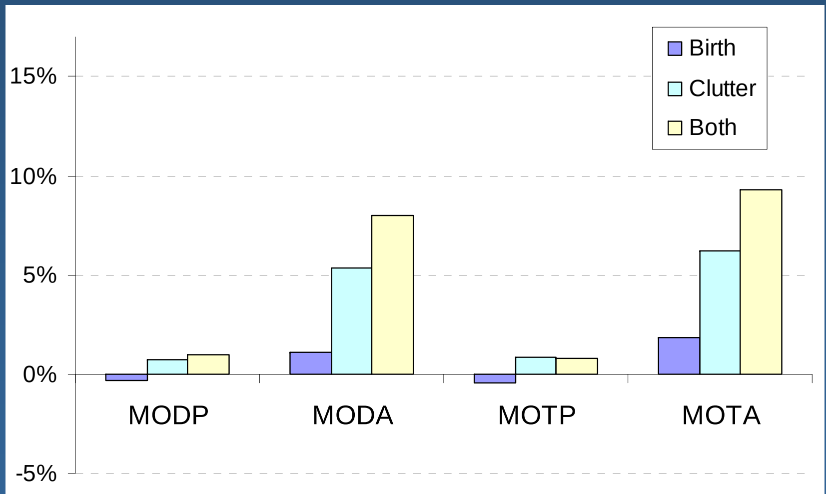
Birth and clutter intensities



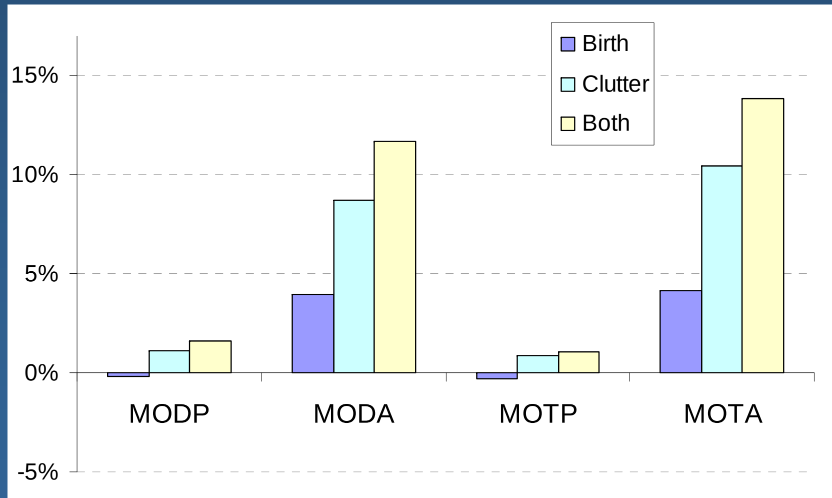
Comparison: w. and w.o. context learning

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

Scenario BW



Scenario QW



Where does the tracker still fail?

- Merging


- Persistent missing detections


Detections

PHD output

Outline

- Video tracking: introduction
    - applications
    - problem statement
- Object representation
    - template and histograms
    - multiple features
- Single-target tracking
    - gradient-based trackers
    - Bayes' tracking
    - hybrid methods
- Multi-target tracking
    - data association
    - Random Finite Sets for tracking
- Outlook, links and references

Research outlook / interesting topics

- What's next?
  - Algorithms
    - learning a model (on-line) / model update
  - Sensors
    - multi-camera: collaborative feature extraction
    - heterogeneous sensors, sensor fusion (video is good, but not the only modality!)
  - Applications
    - embedded systems, power consumption vs. accuracy
    - large-scale behavior recognition, detection of unusual events
  - Datasets
    - tracking on "real" sequences (crowds, UAVs, …)
  - Evaluation!

Evaluation

- UK i-LIDS: imagery Library for Intelligent Detection Systems
  - abandoned baggage, doorway surveillance, parked vehicle, sterile zone
  - multiple-camera tracking
- ETISEO: French Techno-Vision evaluation network
- PETS workshop series
- CAVIAR project: Context Aware Vision using Image-based Active Recognition
- US ARDA VACE program
- VERAAE project: Video Event Recognition Algorithm Assessment Evaluation
- CLEAR workshops: Classification of Events, Activities and Relationships
- CHIL project: Computers In the Human Interaction Loop
- AMI project: Augmented Multi-part Interaction
- Evaluation
  - distribute datasets
  - ground-truth
  - performance measure to compare results

Surveillance Performance EValuation Initiative

- SPEVI
  - Surveillance Performance EValuation Initiative – www.spevi.org
  - One-stop web site collecting existing datasets
    - Pointers to other evaluation programmes / datasets
    - Server hosting received datasets and ground-truth
    - Free distribution (for research)
    - Requested citation acknowledgment
  - If you want to contribute to this collection of datasets please contact info@spevi.org

To play with …

- OpenCV library
  http://sourceforge.net/projects/opencvlibrary/
- Image Processing & Computer Vision library Camellia
  http://camellia.sourceforge.net/
- Hybrid tracker
  http://www.eecs.qmul.ac.uk/~andrea/dwnld/tsht/Setup.htm
- 3D, pattern recognition, tracking, …
  http://peipa.essex.ac.uk/info/software.html
- Updated list of links …
  http://www.videotracking.org

Acknowledgements

Algorithms

Murtaza Taj

Nadeem Anjum

Face dataset:

Huiyu Zhou

Samuel Pachoud

Tim Popkin

EU FP7 project DEXMART www.dexmart.eu

EU FP7 project APIDIS www.apidis.org

UK EPSRC project MOTINAS www.spevi.org

References

[Arulampalam 2002] M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking, IEEE Trans. Signal Processing 50 (2) (2002) 174–188.

[Birchfield 1998] S. Birchfield, Elliptical head tracking using intensity gradients and color histograms, in: Proc. of IEEE Conf. on Comp. Vision and Pattern Recognition, Santa Barbara, USA, 1998, pp. 232–237.

[Birchfield 2005] S. Birchfield, S. Rangarajan, Spatiograms versus histograms for region-based tracking., in: Proc. of IEEE Conf. on Comp. Vision and Pattern Recognition, San Diego, USA, 2005, pp. 1158–1163.

[Black 1998] M. Black, A. Jepson, Eigen–tracking: Robust matching and tracking of articulated objects using a view–based representation, International Journalon Computer Vision 36 (2) (1998) 63–84.

[Comaniciu 2003] D. Comaniciu, V. Ramesh, P. Meer, Kernel-based object tracking, IEEE Trans. Pattern Analysis Machine Intell. 25 (5) (2003) 564–577.

[Cootes 1998] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In Proc. of the European Conf. on Computer Vision, pages 484–498. London, UK, 1998.

[Isard 1998] M. Isard and A. Blake. CONDENSATION – conditional density propagation for visual tracking. International Journal of Computer Vision, 29(1):5–28, 1998.

[Isard 2001] M. Isard, J. MacCormick, Bramble: A bayesian multiple-blob tracker., in: Proc. of International Conf. on Computer Vision, Vancouver, CAN, 2001, pp. 34–41.

References

[Liu 2004] T. Liu, H. Chen, Real-time tracking using trust-region methods, IEEE Trans. Pattern Analysis Machine Intell. 26 (3) (2004) 397–402.

[Lucas 1981] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proc. of International Joint Conf. on Artificial Intelligence, 1981, pp. 674–679.

[Matthews 2004] I. Matthews, T. Ishikawa, S. Baker, The template update problem, IEEE Trans. Pattern Analysis Machine Intell. 26 (6) (2004) 810–815.

[Nguyen 2004] H. T. Nguyen, A. W. M. Smeulders, Fast occluded object tracking by a robust appearance filter., IEEE Trans. Pattern Analysis Machine Intell. 26 (8) (2004) 1099–1104.

[Nummiaro 2002] K. Nummiaro, E. Koller-Meier, L. V. Gool, An adaptive color-based particle filter, Image and Vision Computing 21 (1) (2002) 99–110.

[Okuma 2004] K. Okuma, A. Taleghani, N. D. Freitas, J. Little, D. G. Lowe, A boosted particle filter: Multitarget detection and tracking, in: Proc. of the European Conf. on Computer Vision, Prague, CZ, 2004, pp. 28–39.

[Jepson 2001] A. Jepson, D. Fleet, T. El-Maraghi, Robust online appearance models for visual tracking, in: Proc. of IEEE Conf. on Comp. Vision and Pattern Recognition, Vol. 1, Kauai, USA, 2001, pp. 415–422.

References

[Perez 2002] P. Perez, C. Hue, J. Vermaak, M. Gangnet, Color-based probabilistic tracking in: Proc. of the European Conf. on Computer Vision, Vol. 1, Copenhagen, DK, 2002, pp. 661–675.

[Shi 1994] J. Shi and C. Tomasi. Good features to track. In Proc. of IEEE Conf. on Comp. Vis. and Pattern Recog., pages 593–600. Seattle, USA, 1994.

[Vo 2003] B. Vo, S. Singh, and A. Doucet. Sequential Monte Carlo implementation of the PHD filter for multi-target tracking. In Proc. of International Conf. on Information Fusion, volume 2, pages 792–799. Cairns, AU, 2003.

[Wu 2003] Y. Wu, G. Hua, and T. Yu. Switching observation models for contour tracking inclutter. In Proc. of IEEE Conf. on Comp. Vis. and Pattern Recog., pages 295–304. Madison, WI, USA, 2003.

[Zhou 2004] S. K. Zhou, R. Chellappa, B.Moghaddam, Visual tracking and recognition using appearance-adaptive models in particle filters, IEEE Trans. Image Processing 13 (11) (2004) 1491–1506.

--

[Maggio 2005] E. Maggio, A. Cavallaro, Multi-part target representation for colour tracking, in: Proc. of IEEE International Conf. on Image Processing, Genoa, IT, 2005, pp. 729–732.

[Maggio 2005b] E. Maggio, F. Smeraldi, A. Cavallaro, Combining colour and orientation for adaptive particle filter-based tracking, in: Proc. of British Machine Vision Conf., Oxford, UK, 2005, pp. 659–668.

[Karlsson 2008] S. Karlsson, M. Taj and A. Cavallaro, “Detection and tracking of humans and faces”, EURASIP Journal on Image and Video Processing, Special Issue on Anthropocentric Video Analysis: Tools and Applications, 2008

References

[Maggio 2008] E. Maggio, M. Taj, A. Cavallaro, “Efficient multi-target visual tracking using Random Finite Sets”, IEEE Transactions on Circuits and Systems for Video Technology, to appear, 2008

[Maggio 2007] E. Maggio, F. Smeraldi, A. Cavallaro, “Adaptive multi-feature tracking in a particle filtering framework”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, Issue 10, October 2007, pp.1348-1359

[Cavallaro 2005] A. Cavallaro, O. Steiger, T. Ebrahimi “Semantic video analysis for adaptive content delivery and automatic description”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, Issue 10, October, 2005, pp. 1200-1209

[Cavallaro 2005b] A. Cavallaro, E. Salvador, T. Ebrahimi “Shadow-aware object-based video processing”, IEE Vision, Image and Signal Processing, Vol. 152, Issue 4, August 2005, pp. 14- 22

[Cavallaro 2005c] A. Cavallaro, O. Steiger, T. Ebrahimi, "Tracking video objects in cluttered background", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, Issue 4, April 2005, pp. 575- 584

[Cavallaro 2004] A. Cavallaro, T. Ebrahimi, "Interaction between high-level and low-level image analysis for semantic video object extraction", Journal on Applied Signal Processing, Special Issue on: Object-based and semantic image and video analysis, vol. 2004, n. 6, June 2004, pp. 786-797

[Taj 2007] M. Taj and A. Cavallaro, "Multi-camera scene analysis using an object-centric continuous distribution Hidden Markov Model", Proc. of IEEE Int. Conf. on Image Processing (ICIP), San Antonio, Texas (USA), 2007

[Taj 2006] M. Taj, E. Maggio, A. Cavallaro, "Multi-feature graph-based object tracking", Proc. of Classification of Events, Activities and Relationships (CLEAR) Workshop, Southampton, UK, 6-7 April, 2006

Additional references available at videotracking.org