Human Detection and Tracking in Crowded Scenes Using a Fast Mean Shift Procedure

C. Beleznai[1], T. Schlögl[1], B. Frühstück[2], H. Bischof[3]

- *[1]Advanced Computer Vision GmbH. – ACV, Vienna, Austria*
- *[2]Siemens AG Österreich, Programm- und Systementwicklung, Graz, Austria*
- *[3]ICG, Graz University of Technology, Austria*

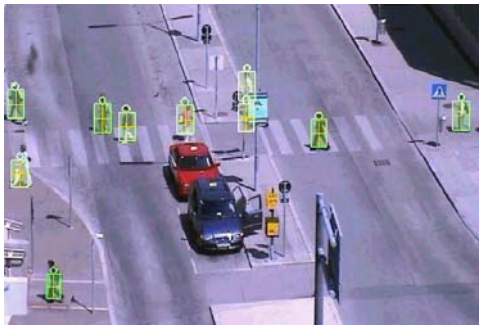
Contents

1. Introduction
2. Human detection and tracking in crowded scenarios
   - state-of-the art
   - human detection using mean shift
   - fast mean shift computation
   - results and evaluation
3. Mean shift mode tracking
4. Extensions
   - arbitrary camera geometry
   - fast data-driven clustering
5. Concluding remarks

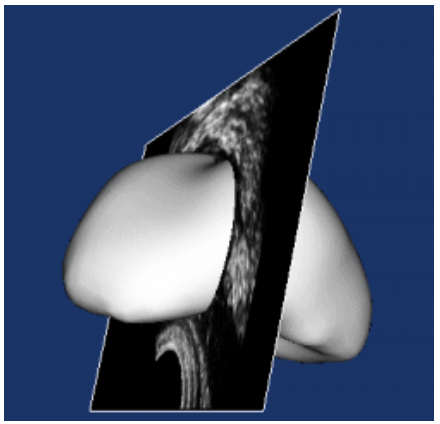
Introduction - the company

Kplus Research Areas and Industrial Applications :

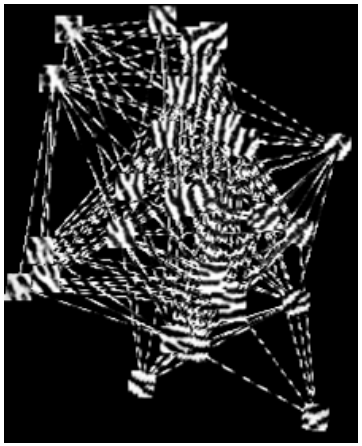
A Visual surveillance and tracking



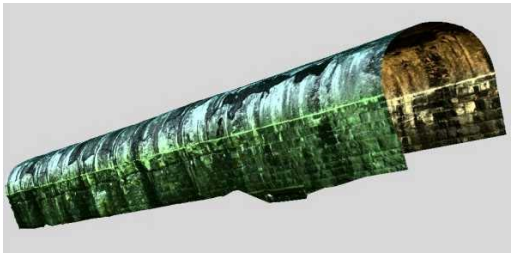
B 3-Dimensional Segmentation, Modelling and Reconstruction



C Statistical methods and learning



D Matching



Human detection and tracking in crowded scenarios

Introduction:

Application: Video surveillance

Objectives:

Robust detection of moving humans by a static camera.

- Detection in occluded situations
- Tracking (short duration of occlusions)

Human detection and tracking in crowded scenarios

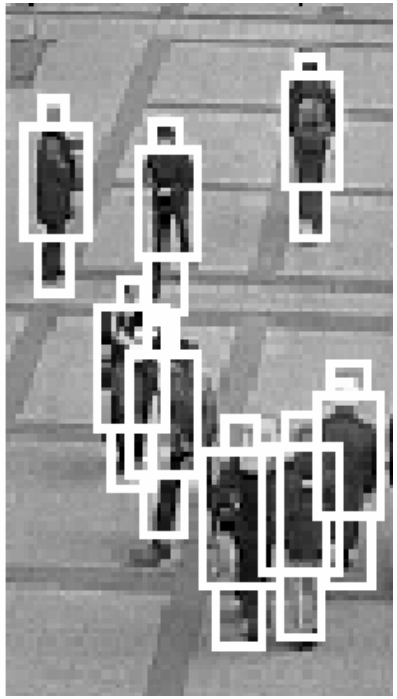
Desired output:

- how many objects,
- approximate location
- consistent motion path

original

“blob” image

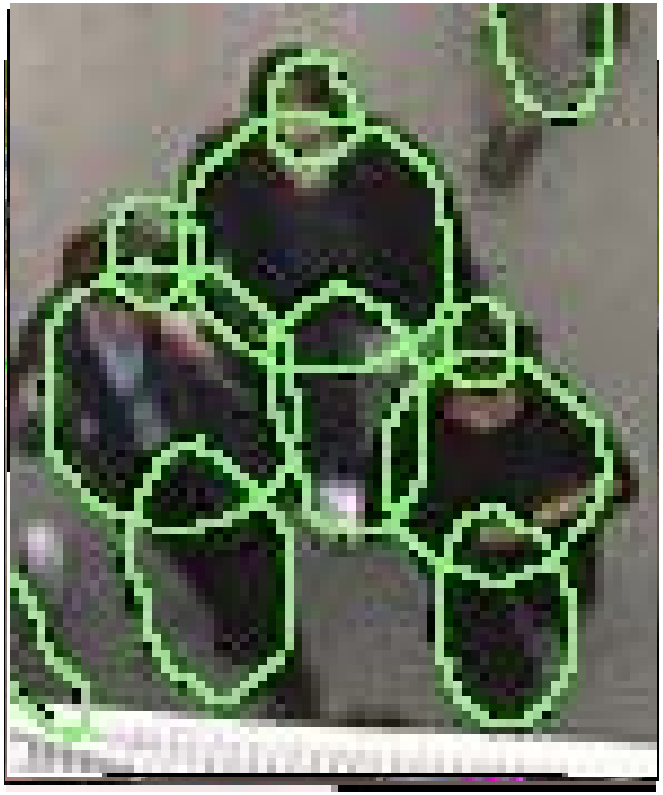
detected “blob” objects



proposed method

State-of-the-art approaches

- Silhouette analysis
  (Kuno 1996, Haritaoglu 2001)
- Color-based segmentation / tracking
  (Elgammal 2001, Comaniciu 2000)
- Particle filter
  (Kuno 1996, Isard 2001)
- EM clustering
  (Pece 2000)
- Appearance models
  (Senior 2001)
- Stochastic segmentation
  (Zhao 2003)



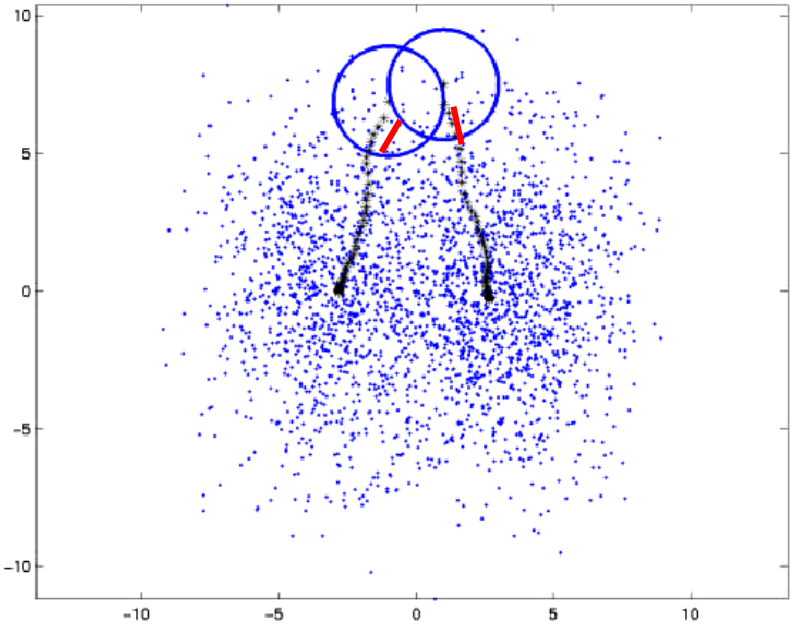
Real-time mean shift-based human localization

- change detection – difference between frame and a reference
- no thresholding – clustering
- fast mean shift algorithm
- model-based validation

Mean shift

Mean shift offset:

$$\Delta_x = \frac{\sum_a K(a-x)\ w(a)\ a}{\sum_a K(a-x)\ w(a)} - x$$



K is a kernel function

w(a) is the weight (intensity) at data point (pixel) *a*

Concept introduced by:

Fukunaga and Hostetler (1975)

Cheng (1995)

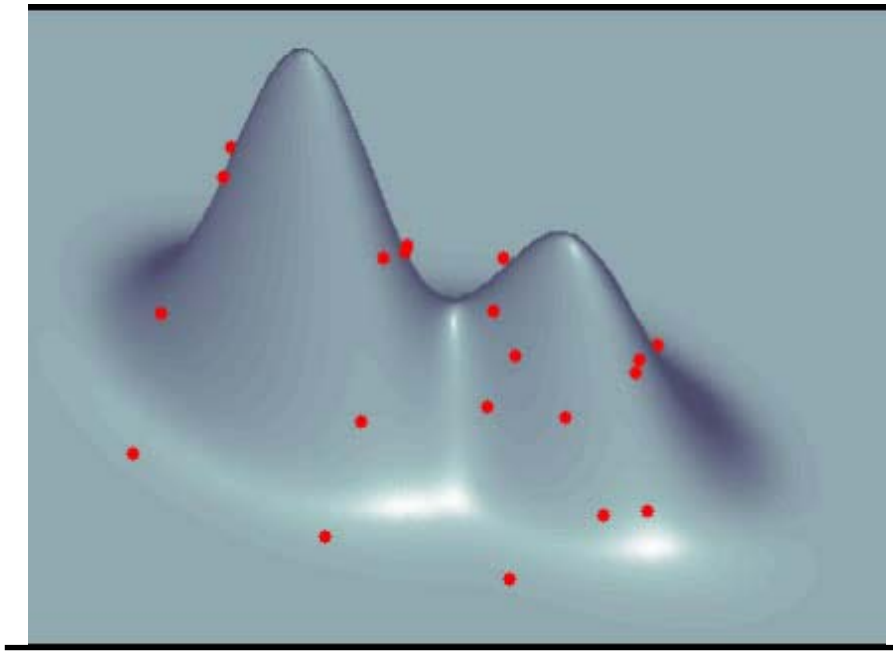
Comaniciu and Meer (1998)

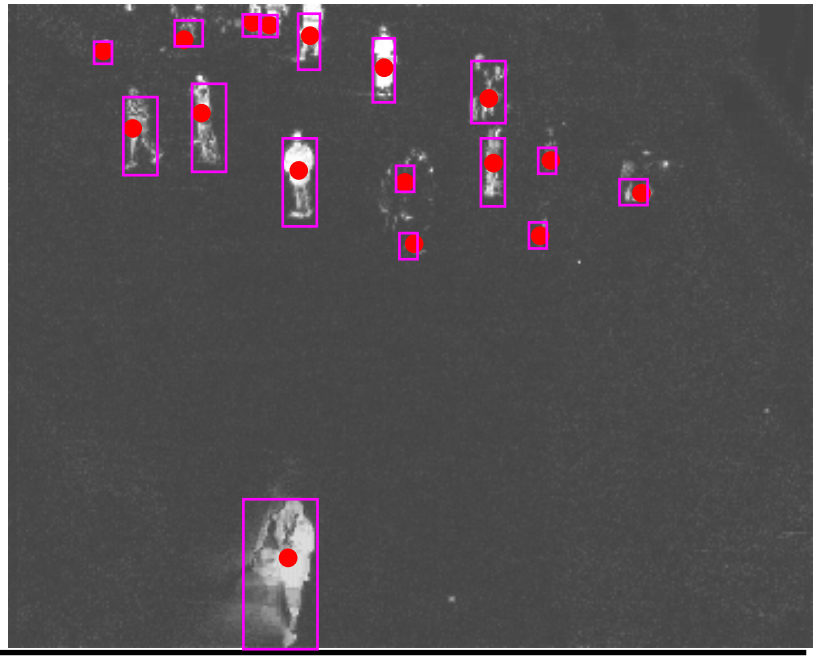
Mean shift clustering

Assumption: Difference image : high intensity ~ high probability of motion

1. locating initial points (sample set),
2. mean shift procedure until convergence,
3. mode grouping

Output: mode, basin of attraction, attraction path, points along attraction path





Fast Mean Shift Computation

Boxlets - P. Simard et al. (1999)

A fundamental property of convolution operation:

$$(f * g)^n = f^n * g = f * g^n$$

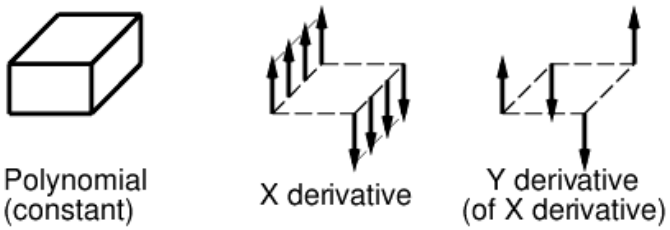
g – signal, *f* – filter

Thus:

$$(f'') * \left(\int\int g\right) = f * g.$$

integral image

Convolution can be significantly accelerated, if the 2nd derivative of *f* is sparse.

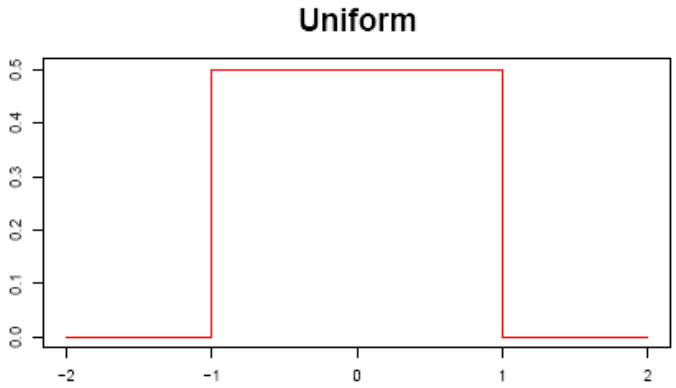


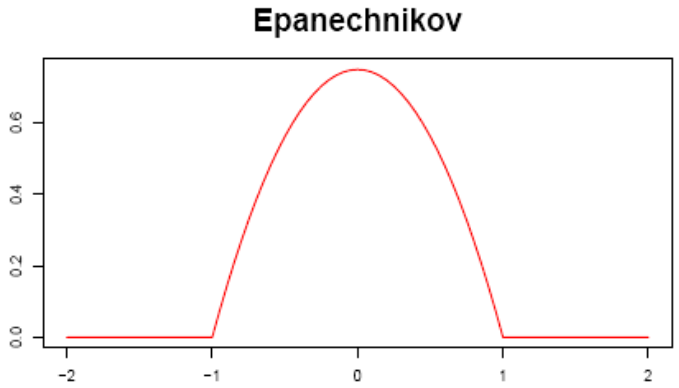
$$\Delta_x = \frac{\sum_a K(a-x)\ w(a)\ a}{\sum_a K(a-x)\ w(a)} - x$$

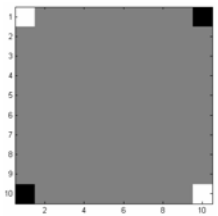
Fast Mean Shift Computation using Integral Images

using: $$K * (w(a)\ a) = K'' * \iint w(a)\ a$$

Mean shift offset: $$\Delta_x = \frac{\sum_a \left[K''(a-x) \sum_{i<a} w(i)\ i\right]}{\sum_a \left[K''(a-x) \sum_{i<a} w(i)\right]} - x$$



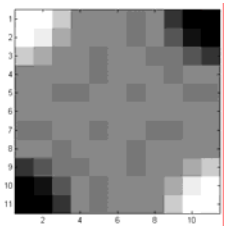




Speedup factors (Matlab):

6 window size: 30-by-50 pixels

30 window size: 90-by-90 pixels



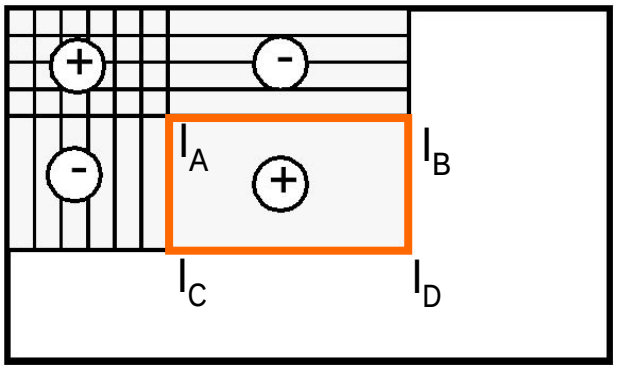
Fast Mean Shift Computation using Integral Images

3 integral images (SAT) are computed:

- Fast computation of integral images

$$I_{int}(x,y) = \sum_{x' \le x, y' \le y} I(x',y')$$

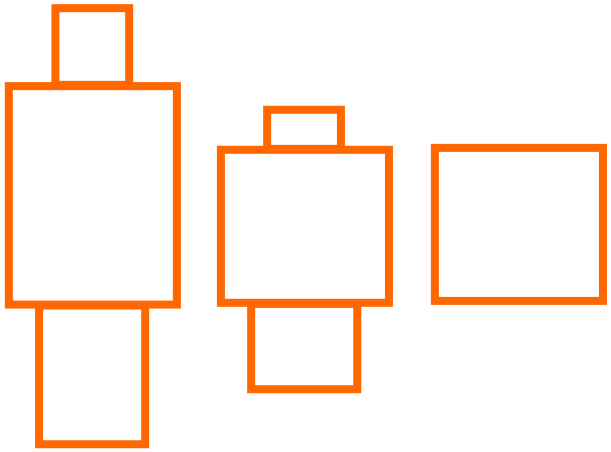
$$I_{\text{int}}(x,y) = I_{\text{int}}(x,y-1) + I_{\text{int}}(x-1,y) + I(x,y) - I_{\text{int}}(x-1,y-1)$$



Integral image

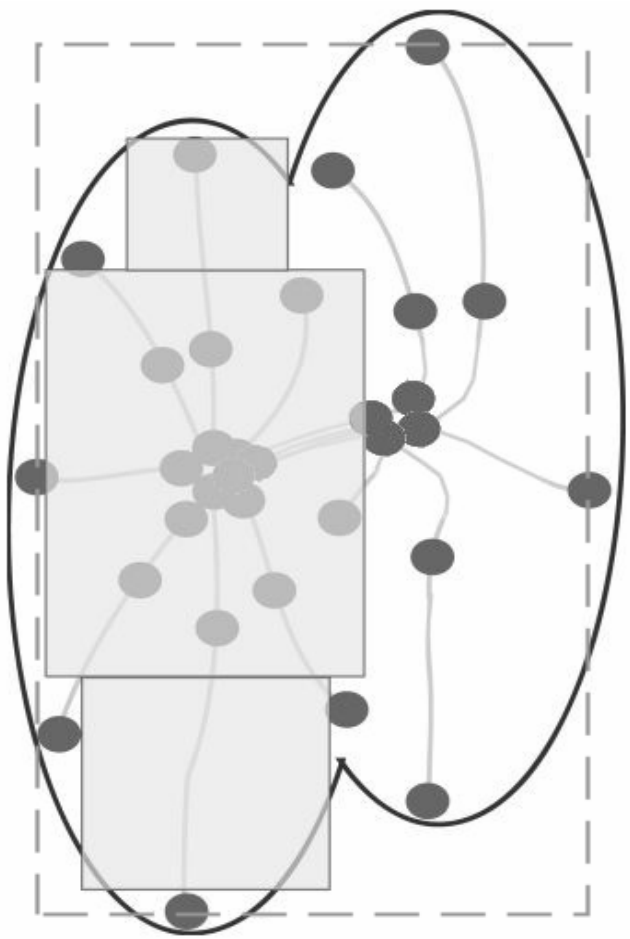
Sum of intensities within a region:

$S_{area} = I_A + I_D - I_B - I_C$



- Simple human model consisting of rectangular regions
- Fast *Hypothesize-and-Test* steps are possible

Model-based validation



- Penalized likelihood criterion

1. Model insertion at detected mode
2. Cost computation

$$C(\theta_z) = (1 - P(I|\theta_z))e^{\beta Z}$$

using

$$P(I|\theta) = \exp\Big(- a \Big[1 - \frac{1}{A_{R_\theta}} \sum_{x,y \in R_\theta} I(x,y) \Big] - b \Big[\frac{1}{A_{R_U}} \sum_{x,y \in R_U} I(x,y) \Big] \Big)$$

3. Model insertion at most probable location.
4. Insertion stopped upon cost increase

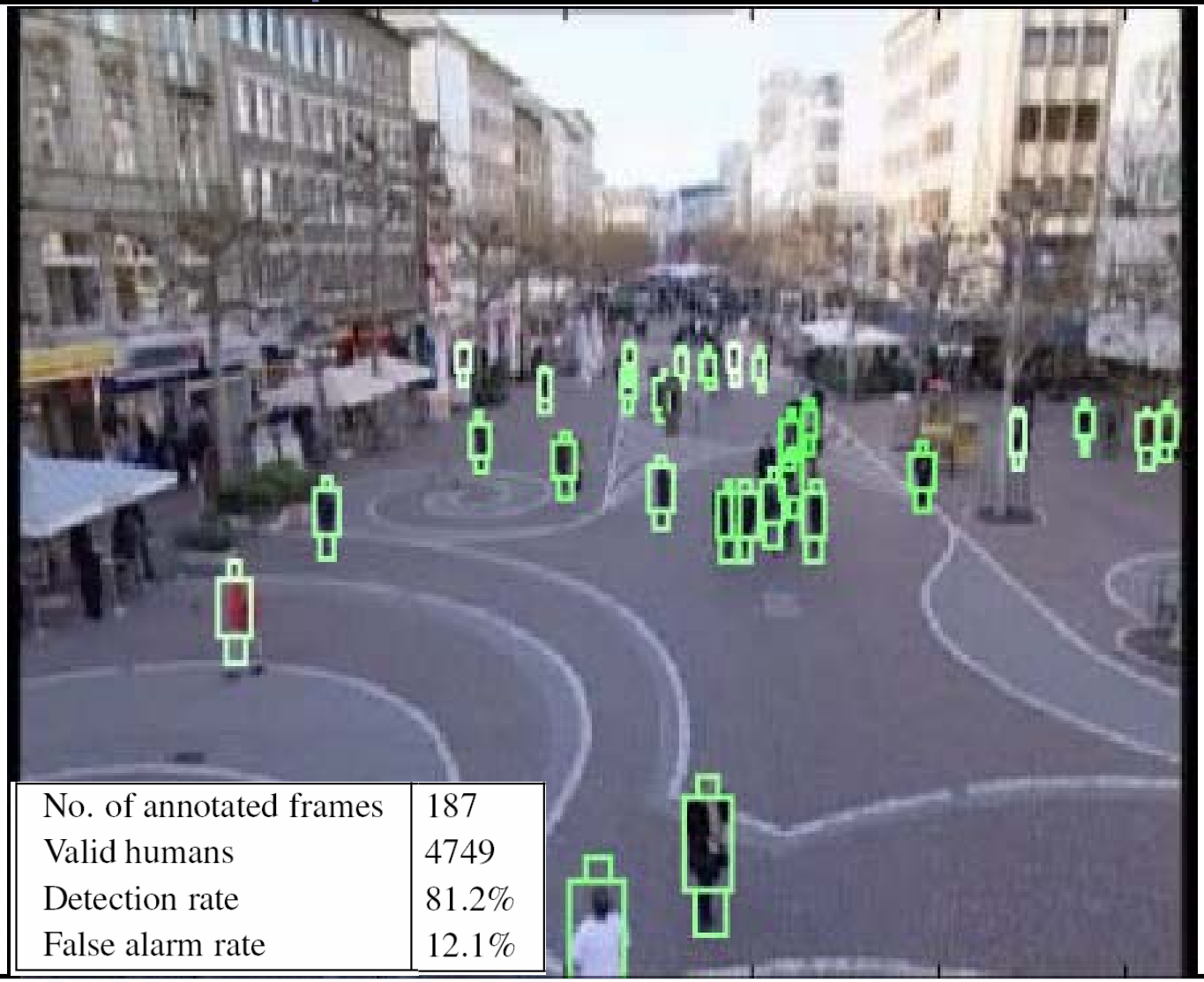
Evaluating detection results

Ground truth
(manual annotation)

Moving humans with
more than 50% visibility

One-to-one mapping
between ground truth
and detection results

Detection experiments (independent processing of each frame)



No. of annotated frames	187
Valid humans	4749
Detection rate	81.2%
False alarm rate	12.1%

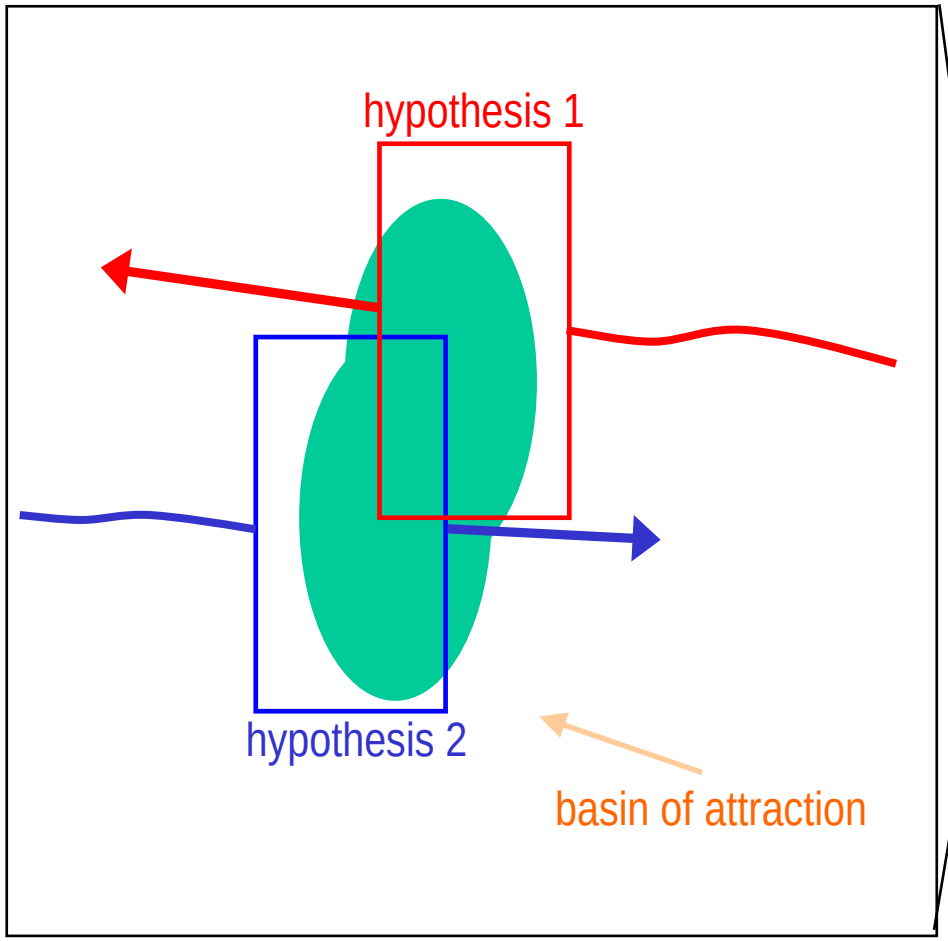
Detection experiments (independent processing of each frame)

Evaluation of detection performance

Two evaluation sequences:

	Sequence	A	B
	No. of annotated frames	470	879
	Valid humans	6147	5380
blob-based method	Correct detections	3096	2762
	Hit rate	50.3%	51.3%
	False alarm rate	23%	4%
proposed approach	Correct detections	5400	4533
	Hit rate	87.9%	84.3%
	False alarm rate	29%	19%

Combining mean shift-based detection with tracking



Detected humans generate hypotheses

Hypothesis-object association on frame-to-frame basis.

Occlusions:

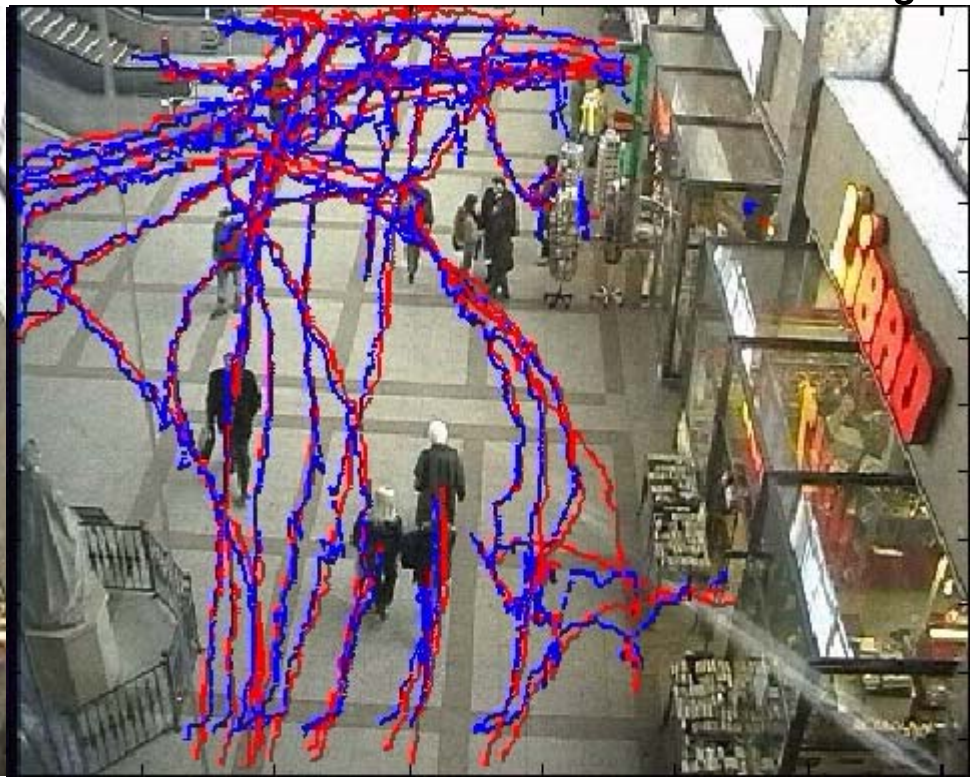
Priors on the number of objects constituting the group are available.

Model-based validation using priors

Tracking experiment: crowded indoor scene

Blob tracking

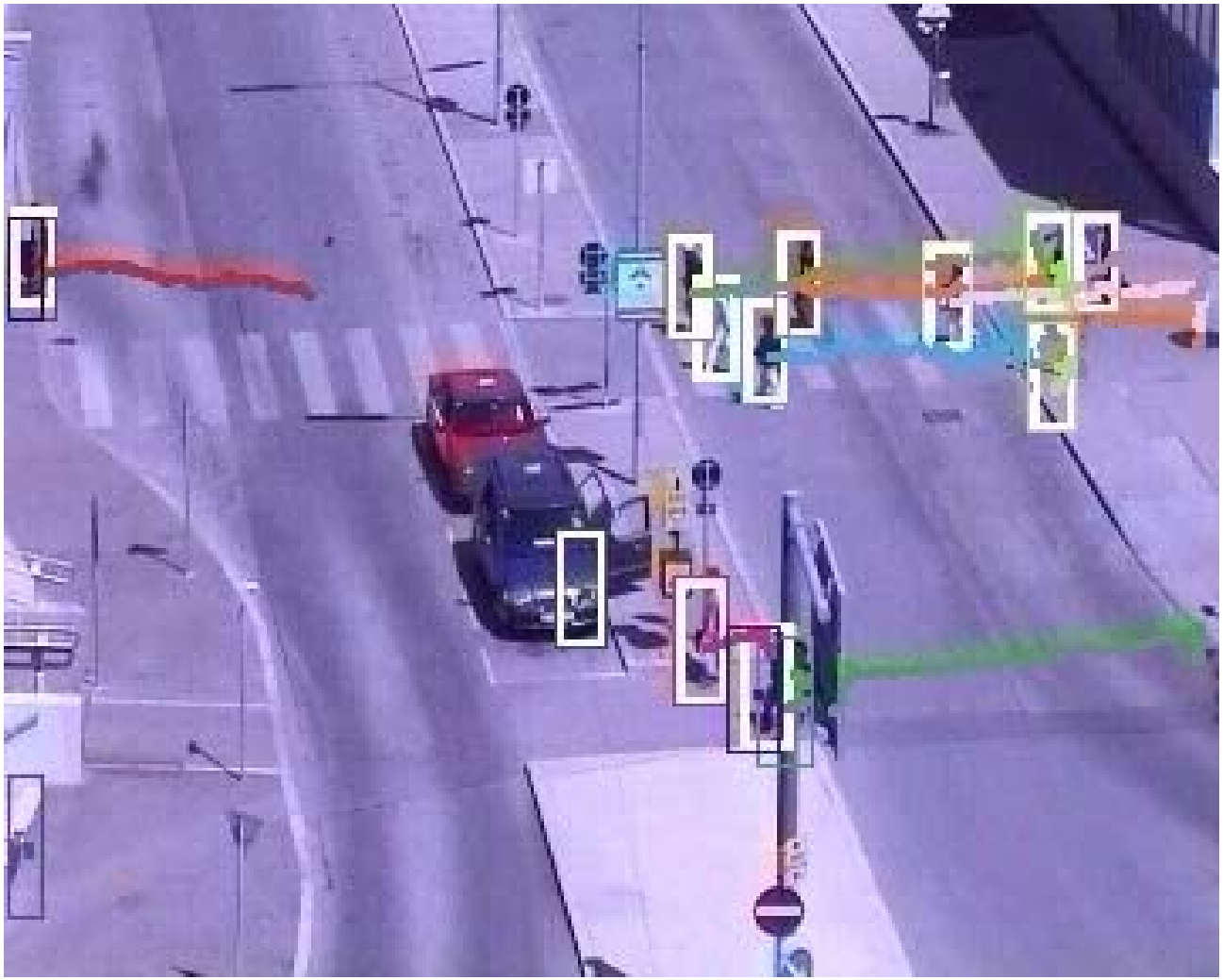
Mean shift-based detection and tracking



Evaluation of tracking performance:

Number of annotated frames: 1013
Number of unmatched tracks: 292 (83.2%)
Track integrity : 1.4

Tracking experiment



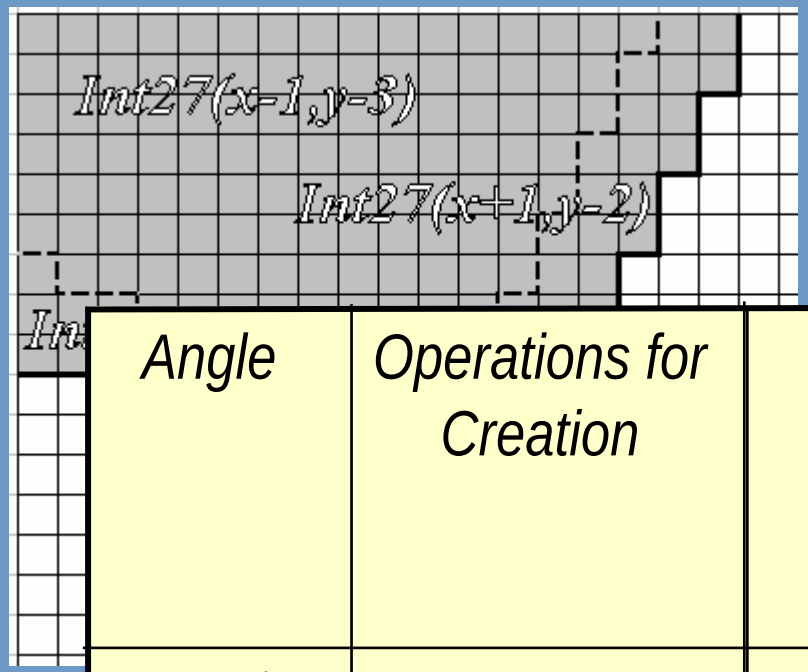
Tracking experiment

Mean Shift Mode Tracking

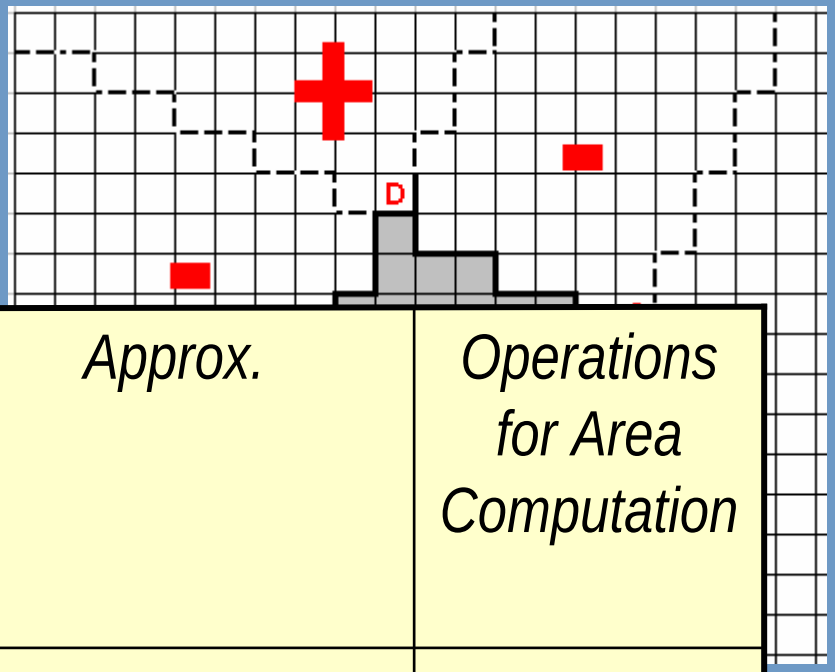
Mean shift procedure using oriented kernels

Mean shift procedure using oriented kernels

Integral image computation in a single pass

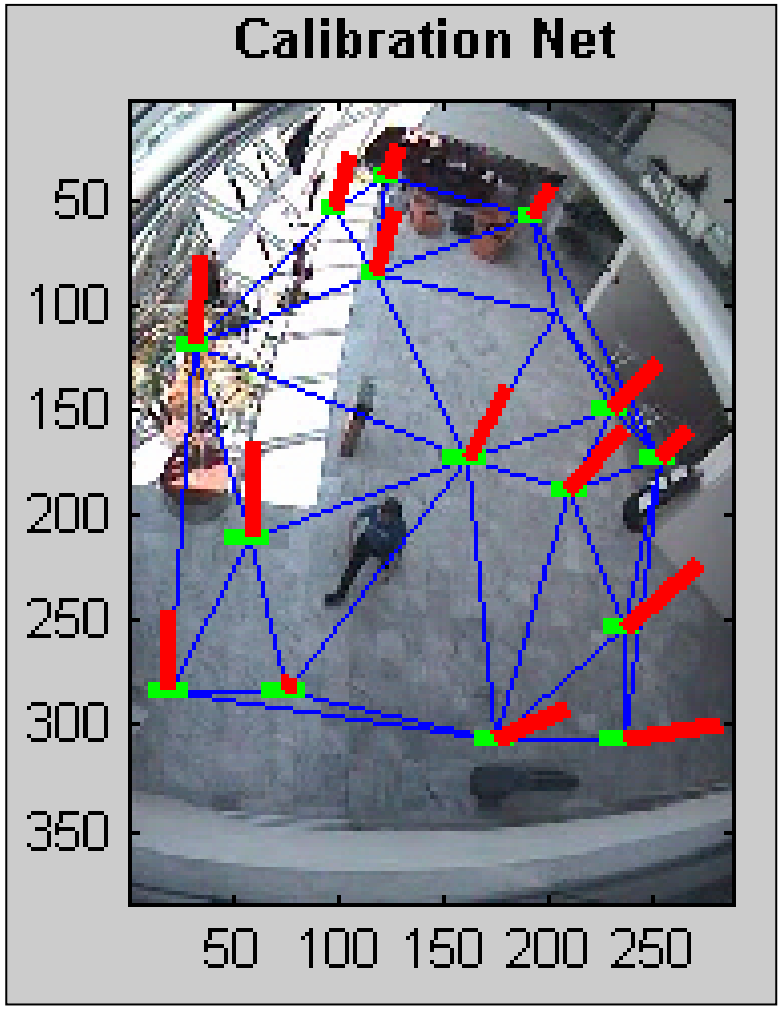


Area sum computation



Angle	*Operations for Creation*	*Approx.*	*Operations for Area Computation*
45° *	4/pixel	Exact	3
27°, 63°	7/pixel	Even side lengths	

Mean shift procedure using oriented kernels



Fast mean shift-based clustering

Unconstrained clustering: kernel size – unknown parameter

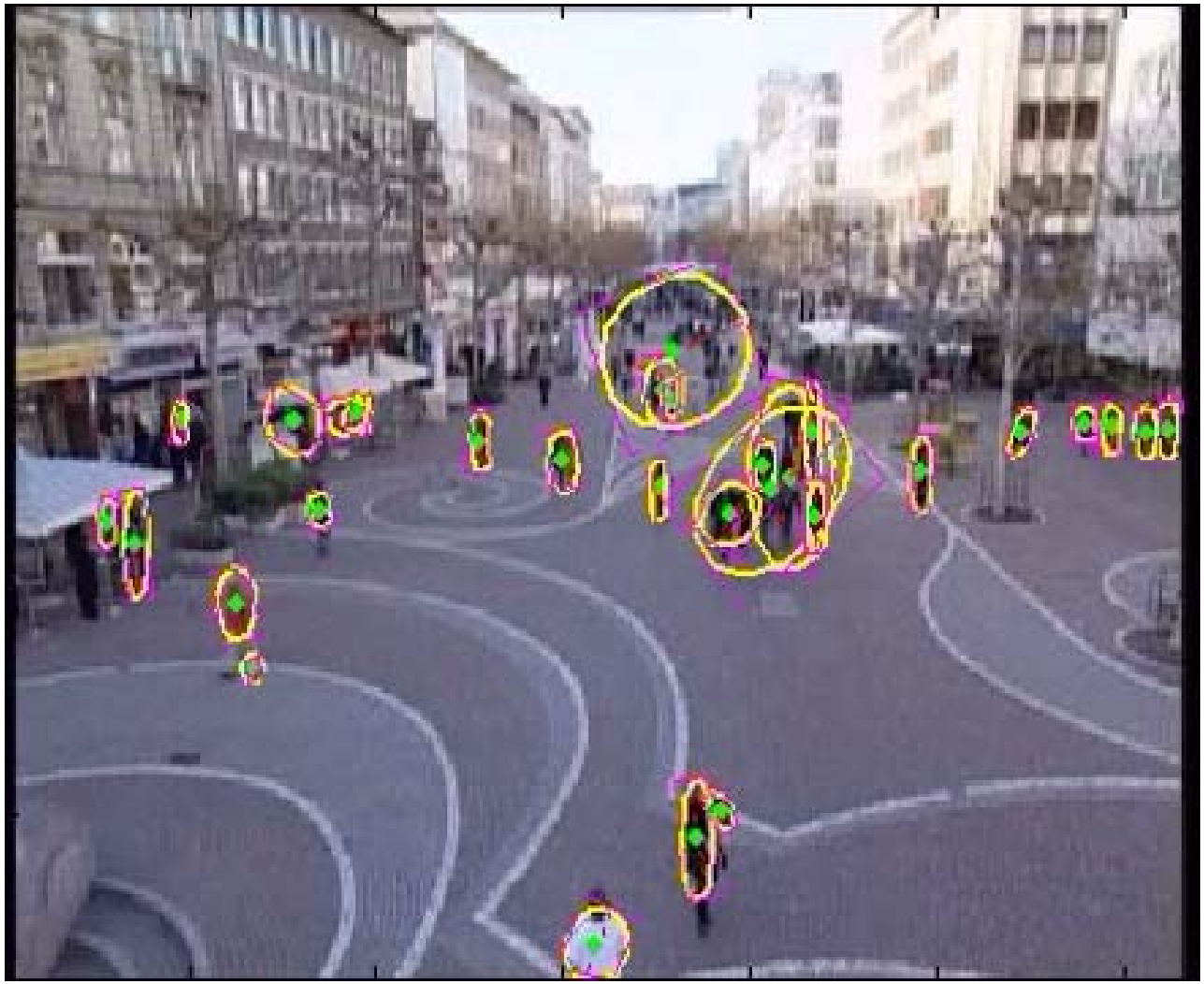
Distribution: multi-modal, multi-scale patterns

Known: range of scales at which structures appear

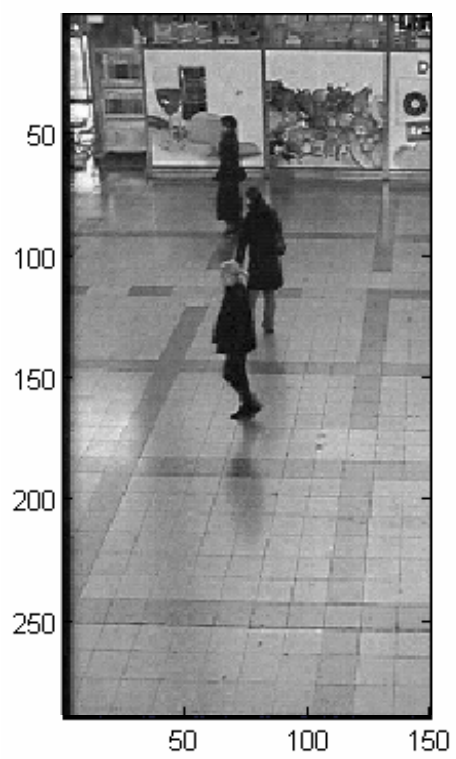
Fast mean shift-based clustering

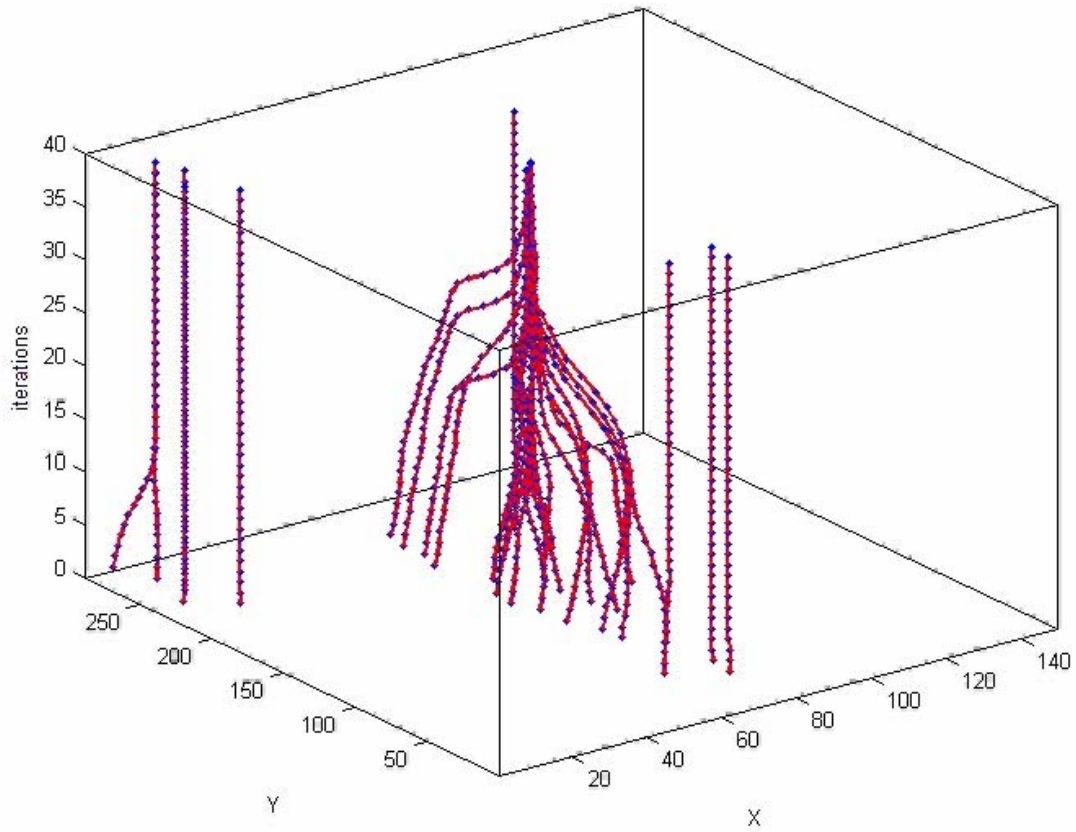
1. Generating sample set with (x_0, σ_0)
2. Mean shift mode seeking
3. Estimating local covariance
4. Orienting kernel (discrete angles)
5. Stop upon convergence

Fast mean shift-based clustering



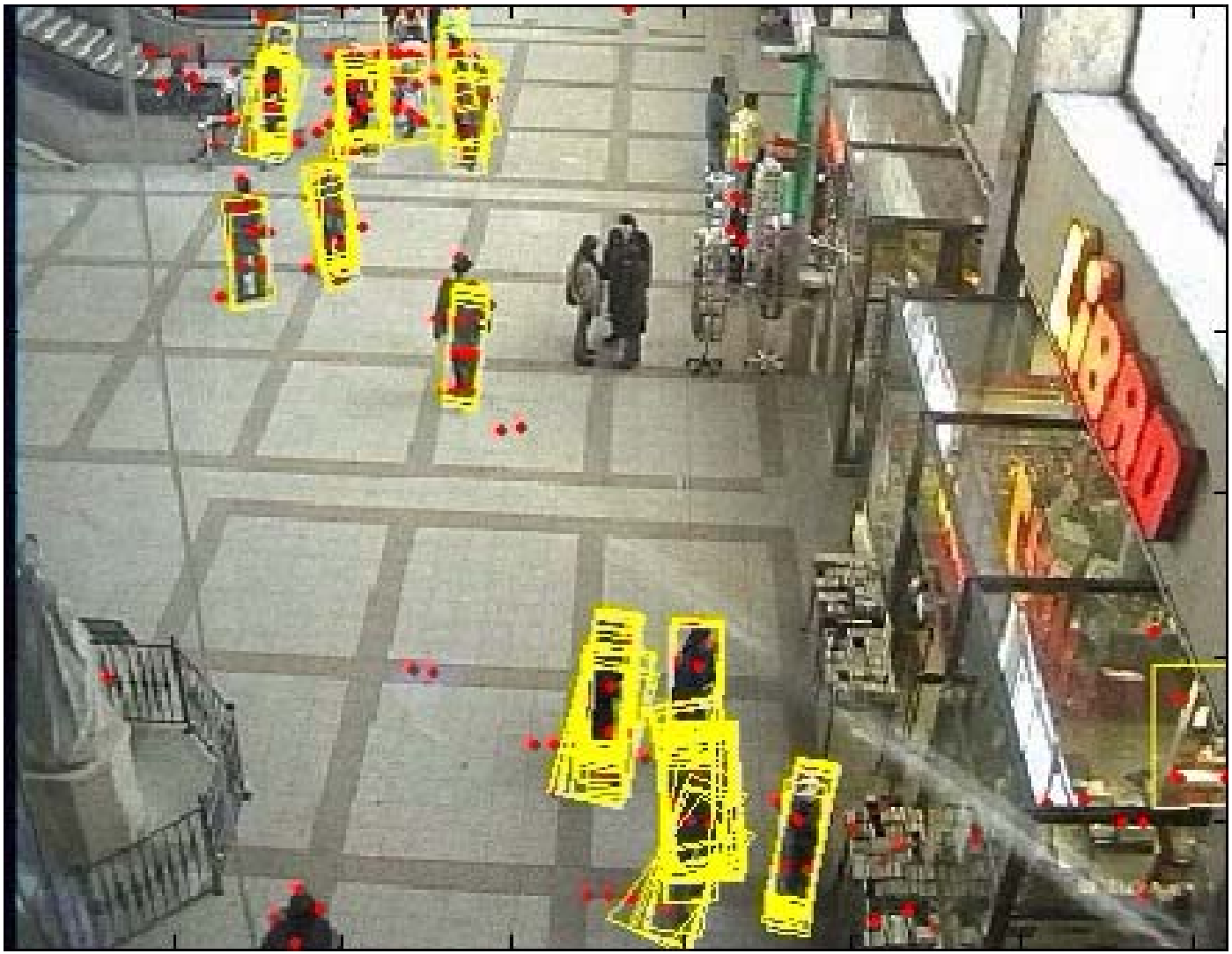
Fast mean shift-based clustering





Scale selection problem: *most-stable-over-scales* criterion

Fast mean shift-based clustering - Experiment



Fast mean shift-based clustering

Applied to:

- Clustering regions within cornerness measure map
- Texture similarity measure
- Clustering the output of boosted cascade classifier
- Finding significant modes in 2D color histograms
- Intensity template correlation

Summary

- Fast mean shift-based clustering:
  Relying on unfiltered data, number of clusters is unknown
- Efficient combination of low-level information and independent high-level knowledge
- Promising (real-time) performance on challenging data
- Data-driven model selection - detecting arbitrary objects

Future Work

- Mean shift-based feature point tracking
- Data association using spatio-temporal reasoning
- Inferring model from clustering