Robust Person Detection for Surveillance using Online Learning

Horst Bischof

Inst. for Computer Graphics and Vision

Graz University of Technology

Take Home Message

Person Detection in surveillance scenarios

is feasible

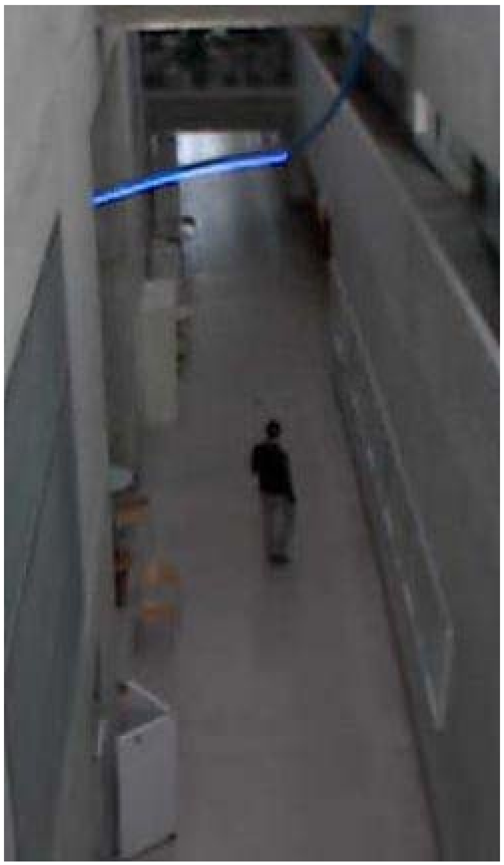
using online learning methods

Outline

- Person Detection in Surveillance Scenario
  - Simple methods (background modeling)
  - Person detection approaches
  - The need for adaptivity & learning
  - **Conservative Learning applied to person detection**
  - **Grid of classifiers**
- [Tracking by detection]

Person Detection

Input Sequence



Video Surveillance:
Fixed Camera Setup

Algorithm

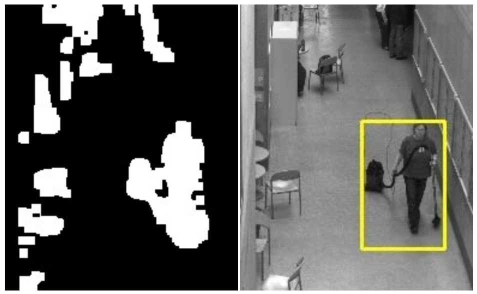
Ground Truth

Person Detection Approaches

1. Background Modeling + Classification
2. Generic Person Detector
3. Generic Detector + Scene Information
4. Adaptive Model

Background Model

- Model „empty scene“
  - Median, Approx. Median, Gaussian Mixture Model, AR models etc.
- Difference BG scene
- Postprocessing + Blob classification



- Classification is hard
- Shadows, illumination, dynamic backgrounds etc.

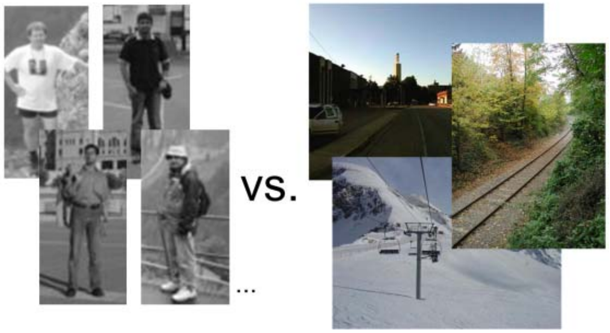
Person Detection

Object detector

- Not using a background model
- Training a classifier
- Detection: Evaluation on many sub-images

Generic Object Detector

(Large) Trainingsset



Fixed classifier trained via an off-line Learning-algorithm

Has to cope with all possible situations!

N. Dalal and B. Triggs. **Histograms of Oriented Gradients for Human Detection**. CVPR 2005.

Improvement using Scene Information

Use of a Generic Classifier

Include Scene Specific Information (i.e. Scale of the Object)

Simpler problem as before!

Some kind of post-processing, classifier stays the same!

D. Hoiem, A. Efros and M. Hebert. **Putting Objects in Perspective**. CVPR 2006.

Adaptive Models

Trainigsset

vs.

Classifier updated with an on-line learning algorithm

Even a simpler problem!

How to update the classifier?

P. Roth, H. Grabner, D. Skocaj, H. Bischof Ales Leonardis. **On-line Conservative Learning for Person Detection**. VS-PETS WS 2005

Motivation for Conservative Learning

Mount a camera, take the huge amount of data that arrives from the video stream

Ultimate Goal

Learn in an unsupervised manner an object representation

Learning with minimal samples

Can we avoid hand labeling (in a robust way)?

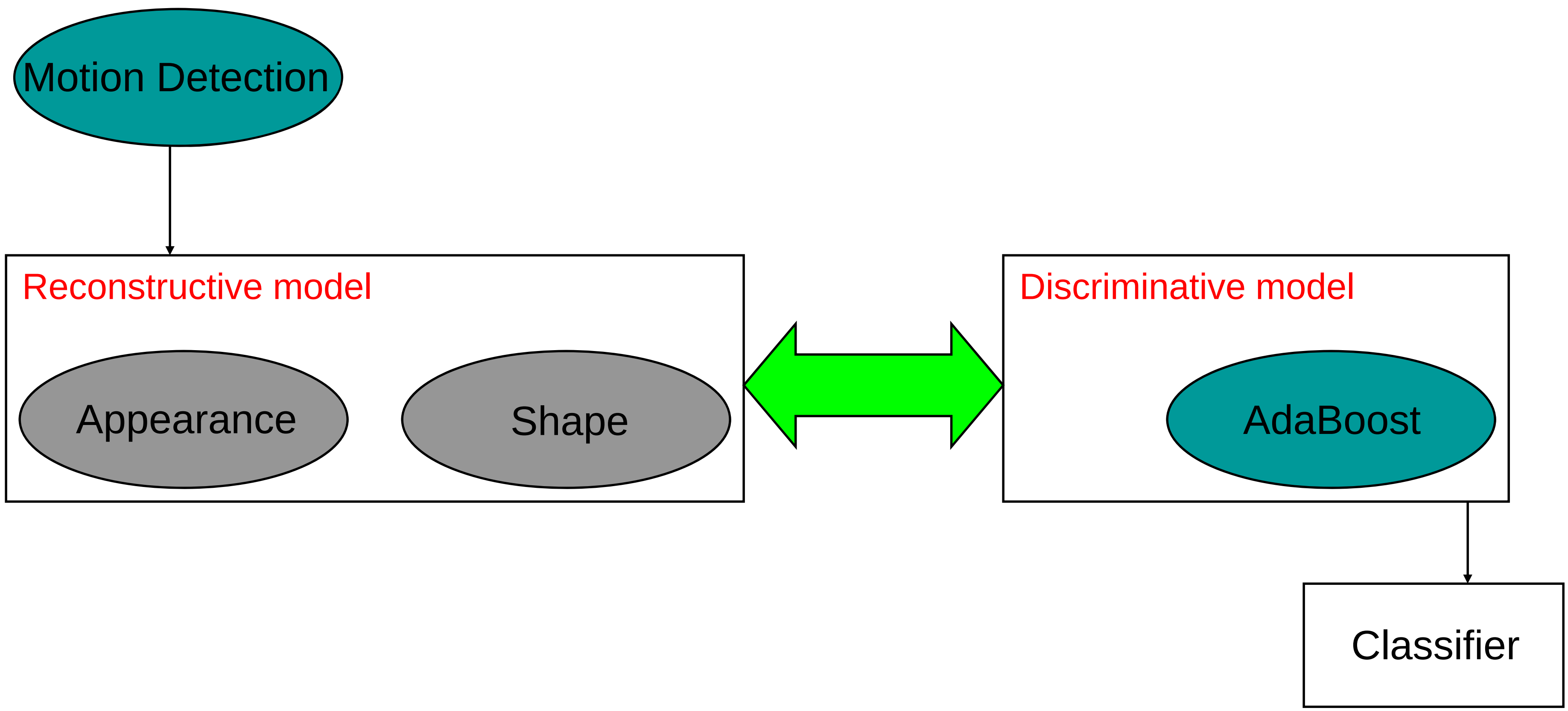
YES, …

Key features of our approach

- Start with simple detector
- Huge amount of video data => learn in a conservative manner
- Combine several detectors: reconstructive and discriminative methods
- Robust training
- Use on-line algorithms

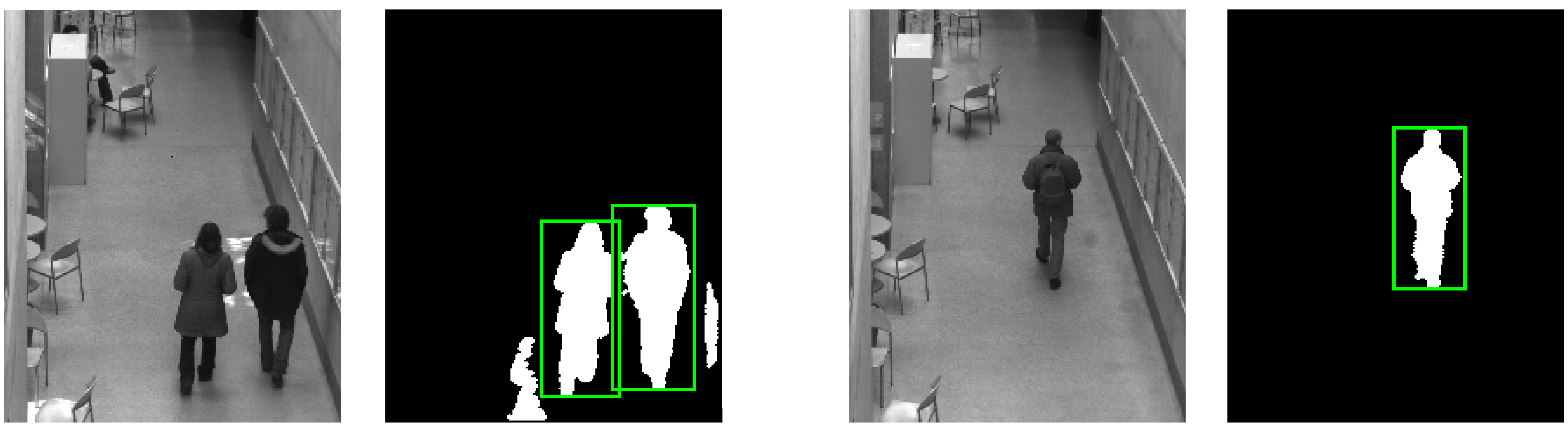
Conservative Learning Framework

Person Detection without hand labeling



Motion Detection

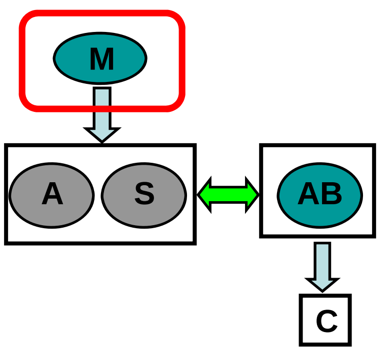
- **Approximated Median (McFarlane & Schofield, 1995)**
- **Aspect ratio and size within limits**
- **Conservative threshold**



Motion based Training

Pos. examples: patches obtained by motion detection

Neg. examples: patches where no motion was detected



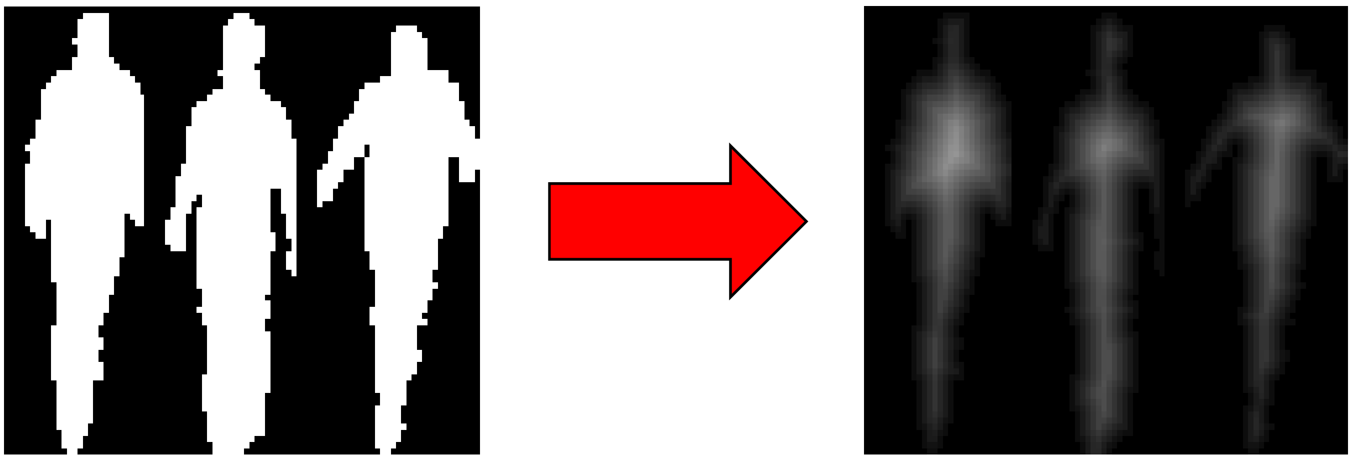
Reconstructive Classifier

- Robust PCA on appearance
- Robust PCA on shape
- Compute reconstruction error
- Conservative threshold

Reconstructive Classifier

PCA on shape

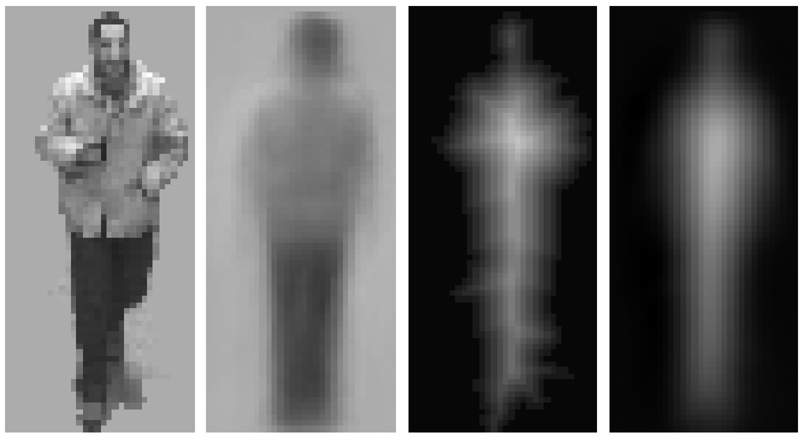
– **distance transform on motion blobs**



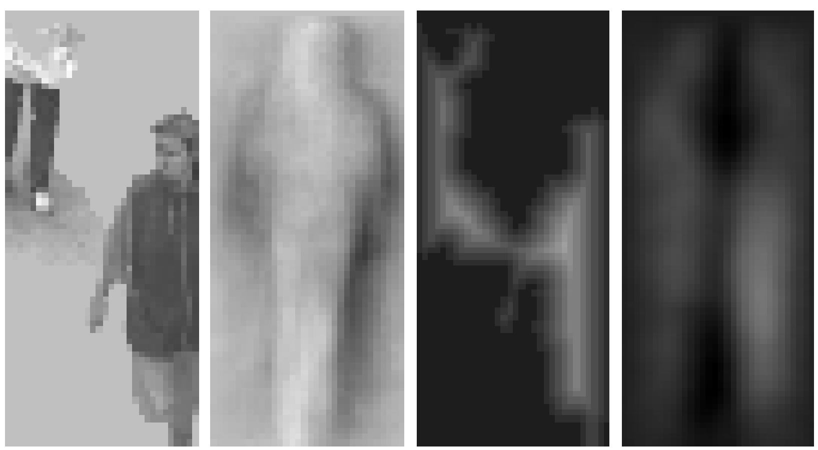
– **Compute shape model**

Reconstructive Classifier

Reconstructions

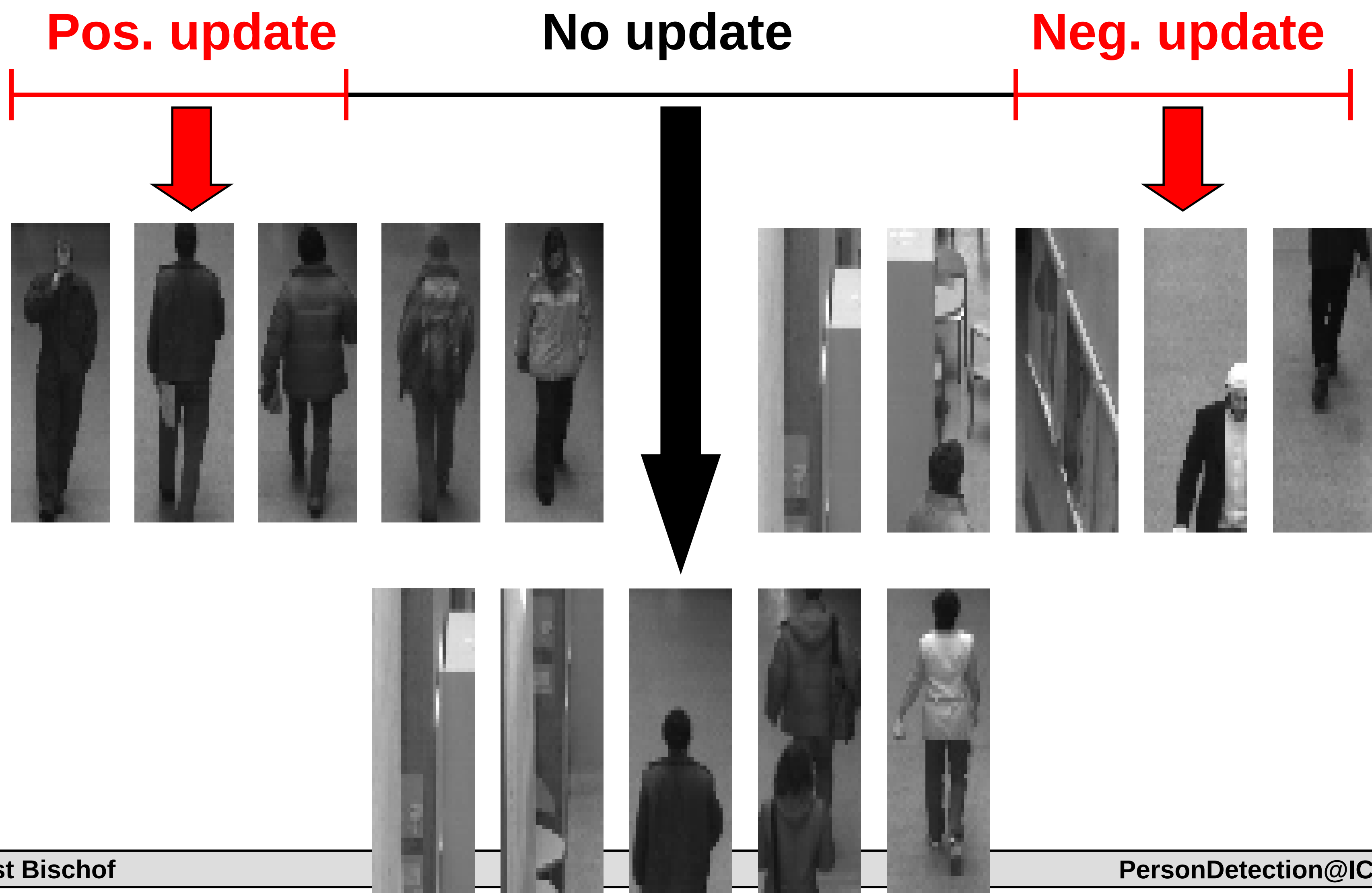


positive



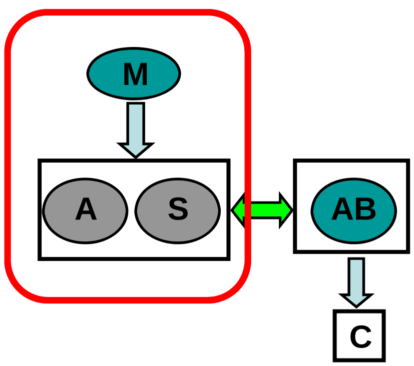
negative

Reconstructive Classifier



Reconstructive Training

- **Pos. examples: patches obtained by motion detection (verified by reconstructive model)**
- **Neg. examples: patches where no motion was detected**



Discriminative Classifier

Discriminative classifier

- **AdaBoost (60 weak classifiers)**
- **Haar Wavelets (Viola & Jones)**
- **Local Orientation Histograms (e.g., Levi & Weiss, Dalal & Triggs, etc)**

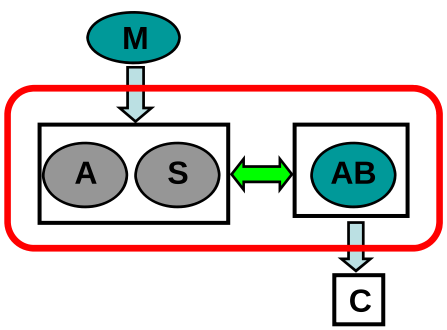
→ Use integral images to calculate features fast

Final Classifier

New positive and new negative examples are collected during evaluation

Classifier is retrained with new data

We sample the decision border !



Experiments

Train sequence: 4500 frames (2 „days“)

Test sequence: 300 frames (1 „day“)

Trained 3 classifiers based on different training data on the train sequence

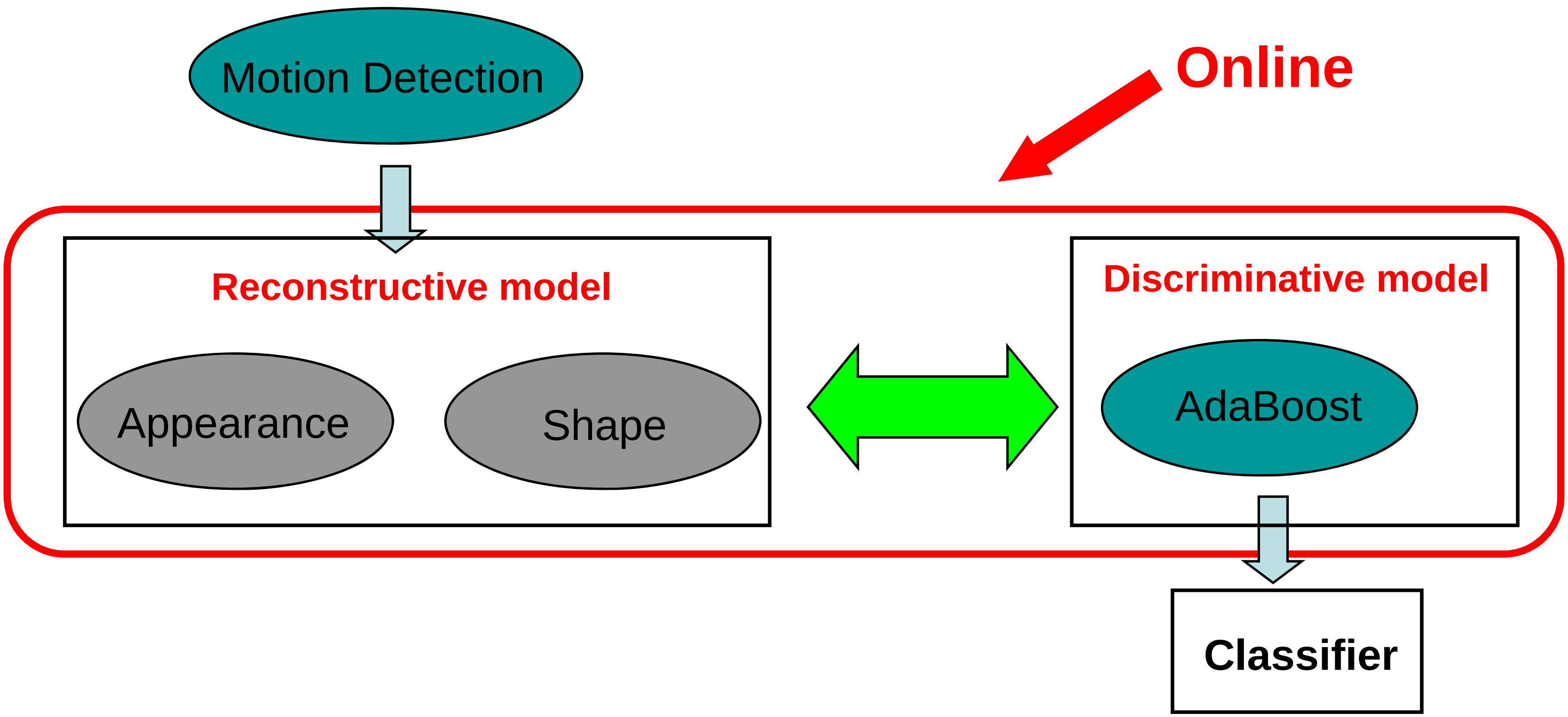
Evaluate classifier on the test sequence

Expected result: number of false positives can be clearly decreased

Results

Method	det.	false-pos.	prec.
AdaBoost1	229	605	27.5%
AdaBoost2	216	160	70.7%
AdaBoost3	220	12	94.2%

Online Training



On-line Training

- Motion Detection

  → no Problem

- PCA On-line

  → no Problem (e.g. Skocaj et.al.)

- Boosting Online
  - Online Boosting see

    H. Grabner and H. Bischof. **On-line Boosting and Vision**. CVPR 2006.

Generalized Model

Summary of Conservative Learning

Extensions

- Framework is general

  Exploit more modules to further increase the robustness

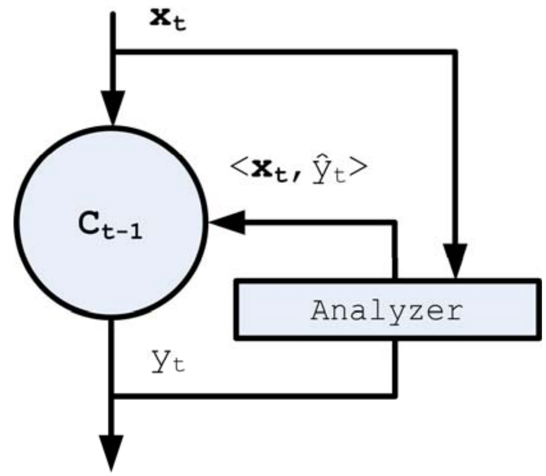
- Use different cues

  Co-training

- Use a tracker to get more positive examples

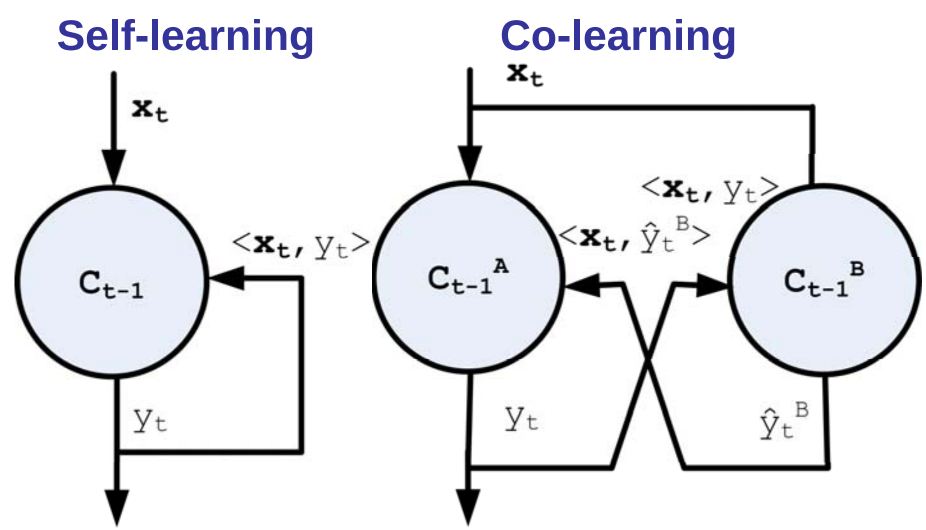
Update Strategies for On-line learning

- Unlabeled sample $\mathbf{x}_t$, who to choose $\hat{y}_t$?
  - Self-training
  - Co-training
  - Conservative Learning



– What is if $\hat{y}_t$ is wrong?

! DRIFTING !
when running for a long time



Grid-based Person Detector

We want to build a system which runs 24 hours a day, 7 days a week!

Simplify the problem
such that we can use a
fixed update strategy,
no drifting problem.

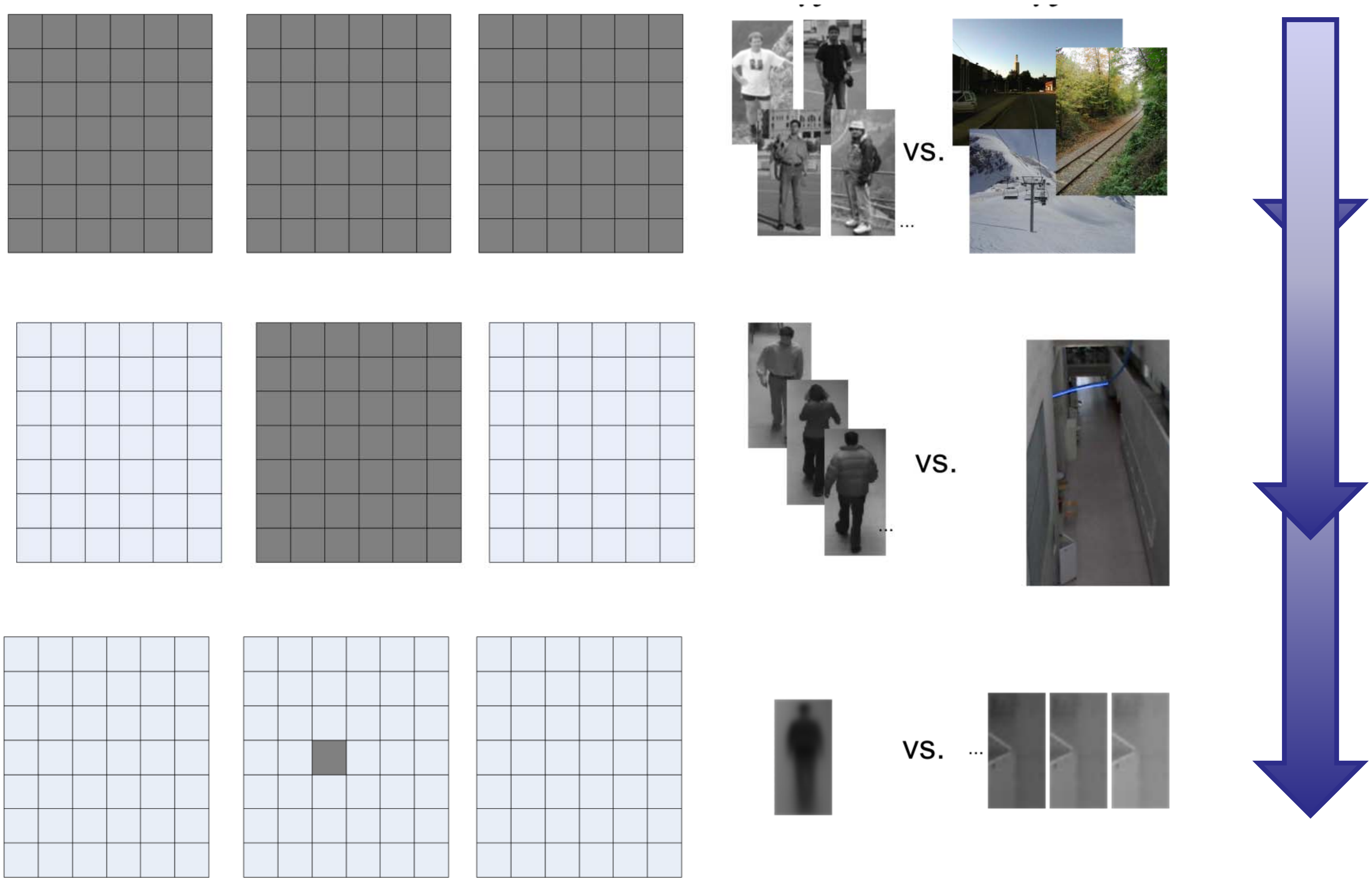
Simplifying the Problem

Classifier should be applicable | **Training set** | **Complexity**

Fixed Detector

Scene specific

Proposed Grid based



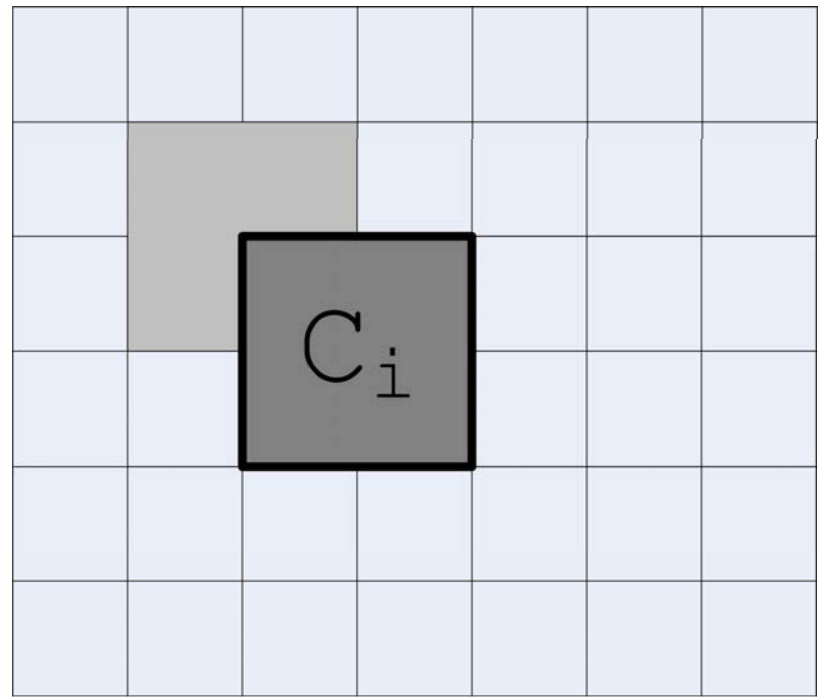
Classifier Grid

Trainingset

One On-line classifier for each grid element

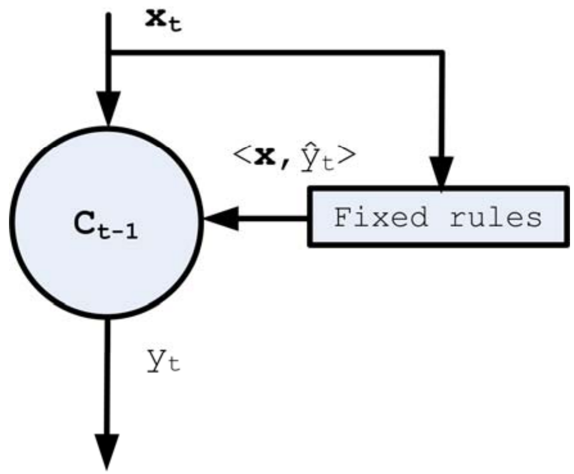
Relations to background modeling

Simple Problem in time and space.



H. Grabner, P. Roth, M. Grabner and H. Bischof. **Autonomous Learning a Robust Background Model for Change Detection**. PETS WS 2006.

Fixed Update Rules



No dependencies between the update and the current model ⇨ stable by design

Positive updates

– From a fix set

$$\langle \mathbf{x}, +1 \rangle, \qquad \mathbf{x} \in \mathcal{X}^+$$

– Correct by definition

Negative updates

– Current patch

$$\langle \mathbf{x}_{i,t}, -1 \rangle$$

– Correct most of the time, wrong with

$$P(\mathbf{x}_i = \text{person}) = \frac{\#p_i}{\Delta t}$$

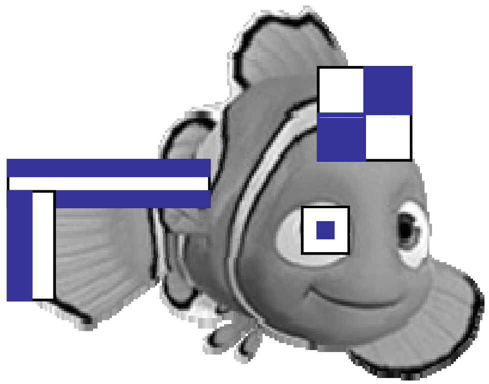
On-line Classifier

On-line learning algorithm

- Must cope with some (low) label noise
- Fading memory (forgetting)
- Good generalization (e.g. maximum margin classifier)

On-line boosting for feature selection

- Simple Haar wavelets
- Simple extension to outlier detection



H. Grabner and H. Bischof. **On-line Boosting and Vision**. CVPR 2006.

Experiments and Results

Toy Example

Public Sequences

- PETS 2006
- and Caviar Sequences

Parameter (constant)

- Ground plane estimation (scale estimation)
- Positive dataset contains only **1** sample (mean image)
- 10 selectors each with 20 weak classifier features

Result and Comparison

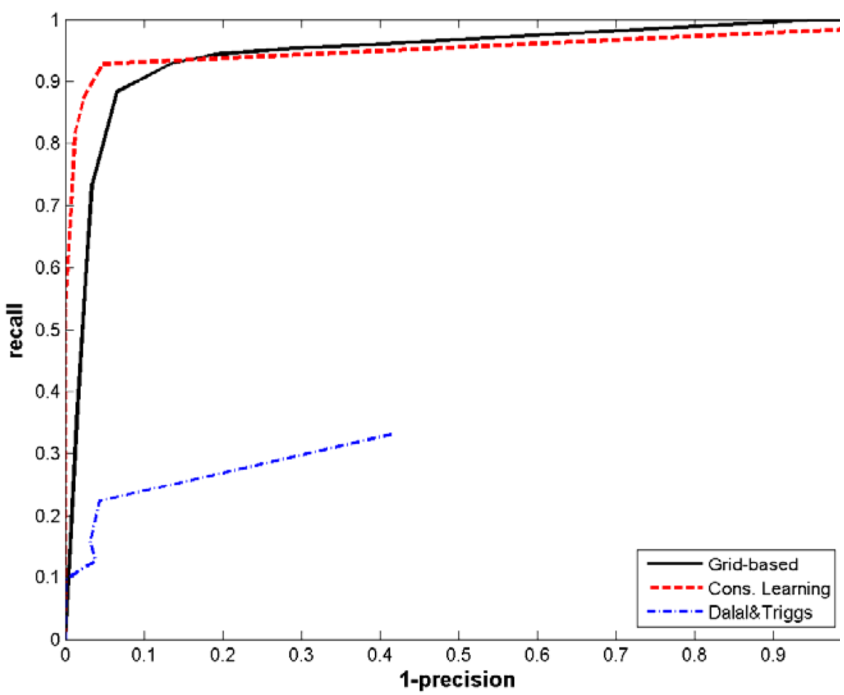
Dalal/Triggs (generic detector)

This approach

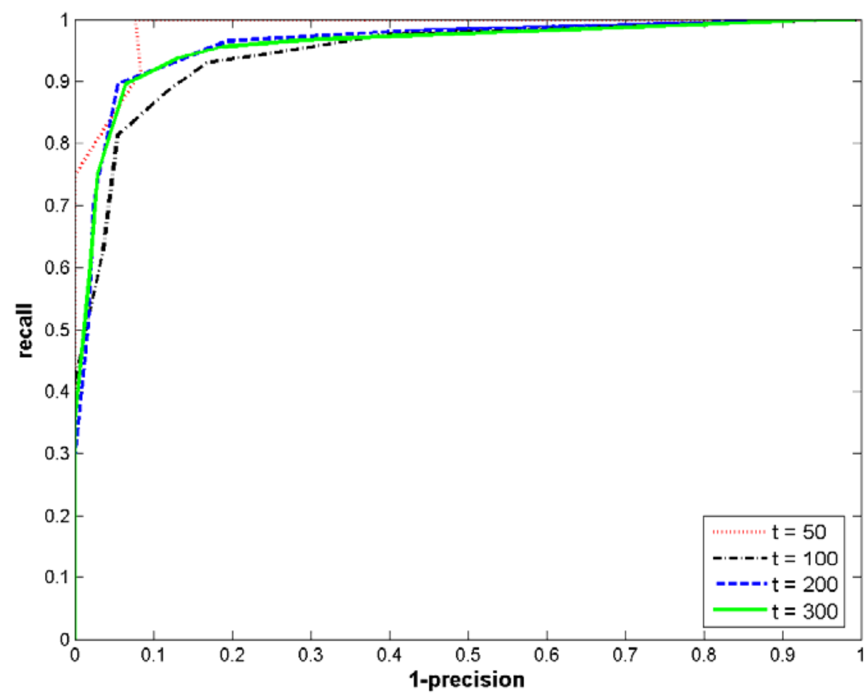
Conservative learning (scene adaption)

CoffeeCam Results

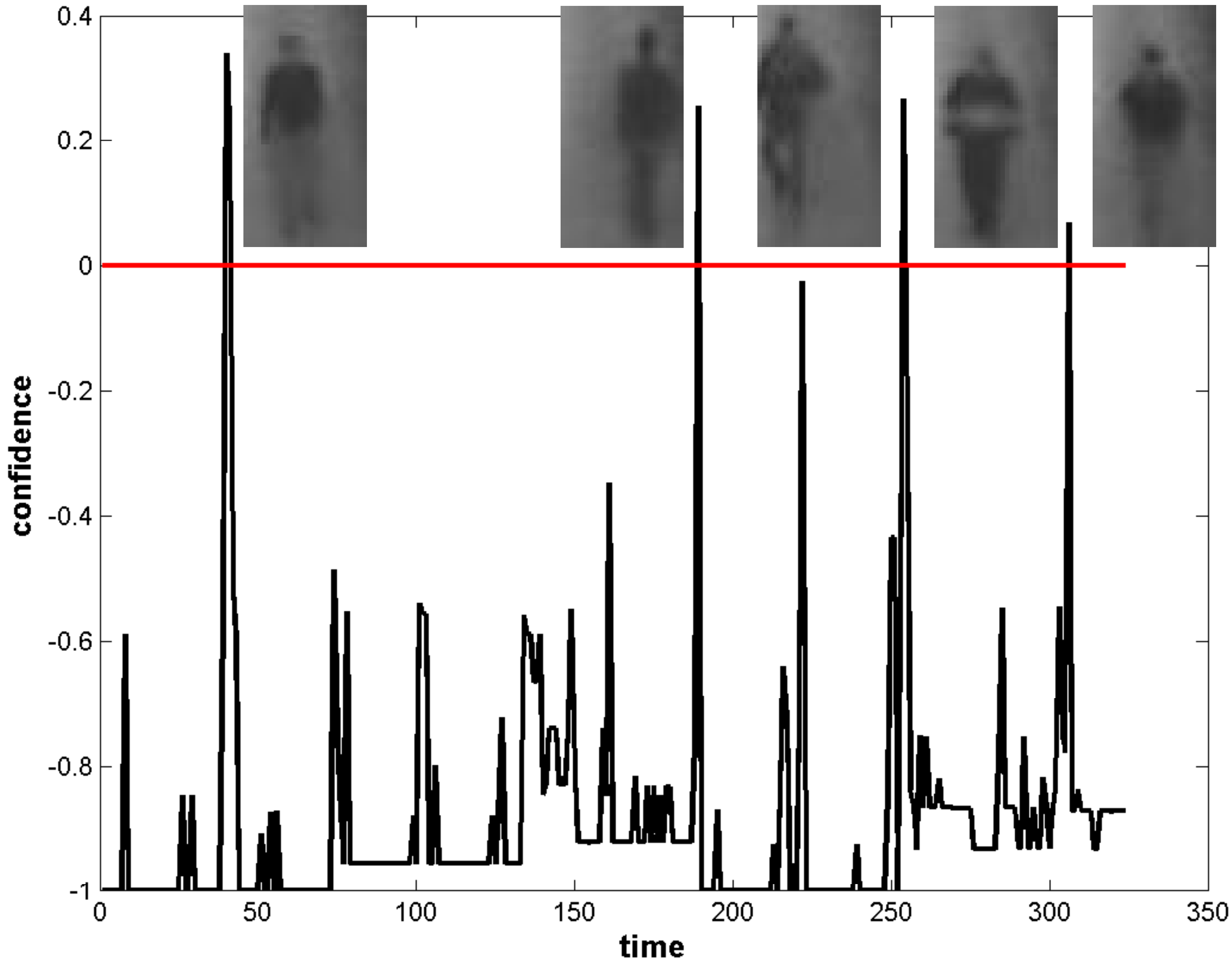
ROC



Convergence Speed

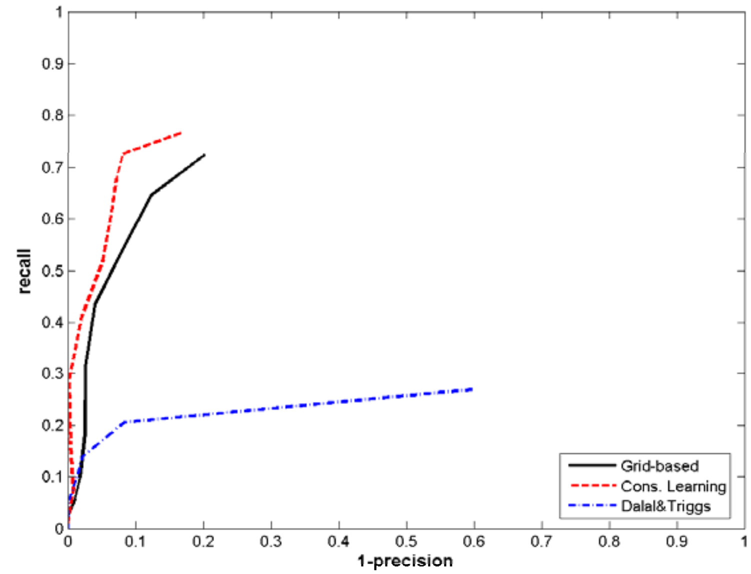


Confidence of a Patch over Time

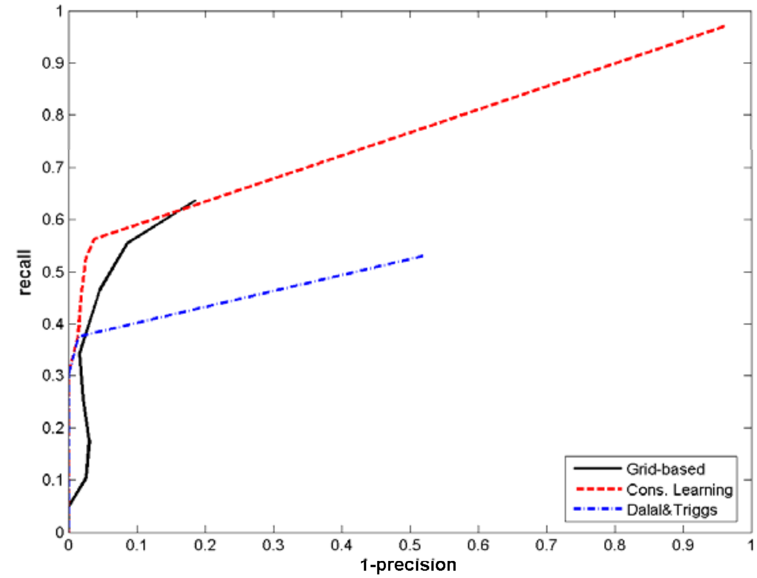


I'm NOT a simple Background Model

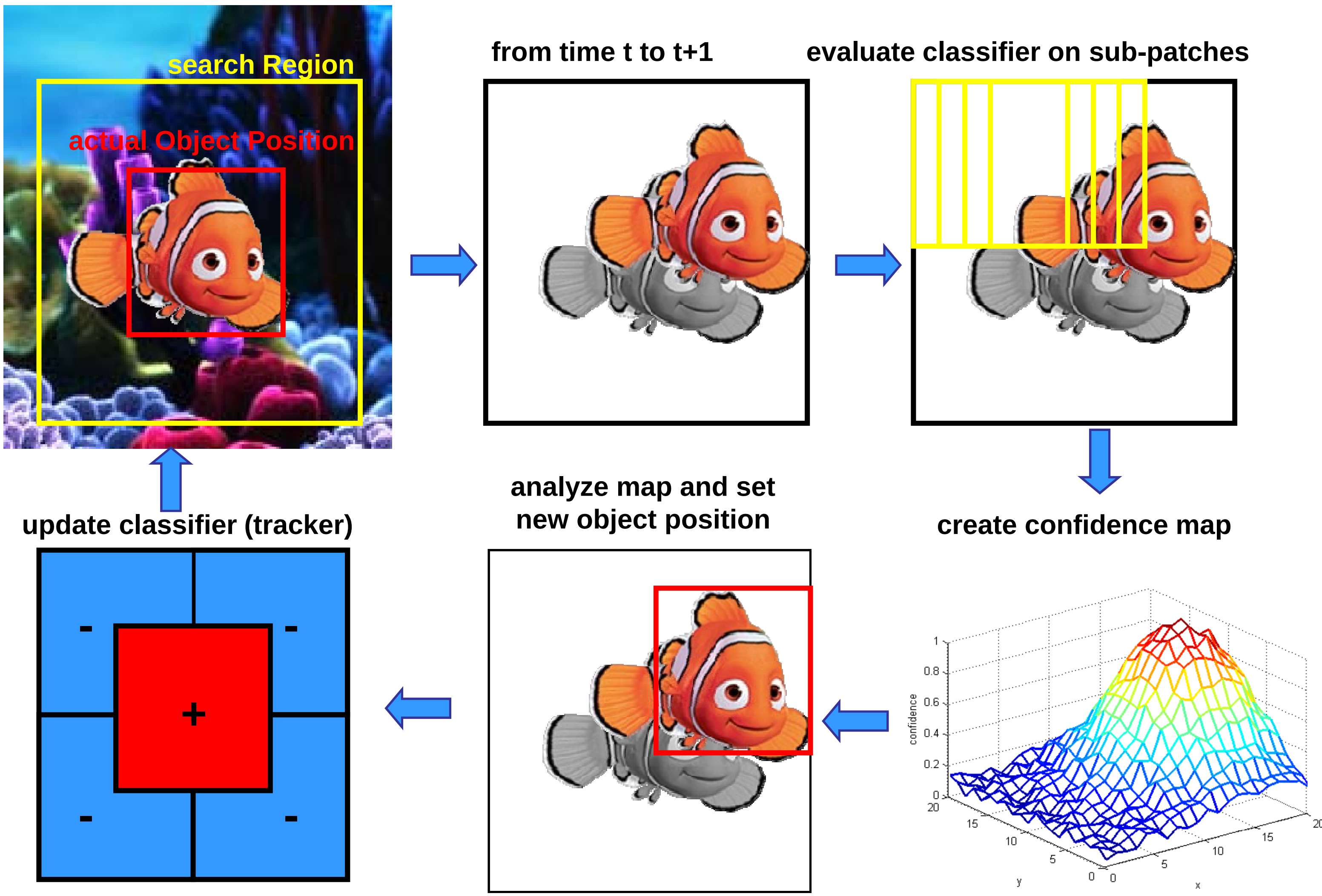
Results: PETS 2006



Results: Caviar



Object Tracking



Features for Tracking

- Haar Wavelets
  - Learned Threshold or
  - Bayesian decision criterion based learned mean/Variance

- Integral Orientation Histograms
  - Cluster of positive and negative samples for each bin, modeled as Gaussian

- Simplified Version of Local Binary Patterns
  - Cluster of positive and negative samples for each bin, modeled as Gaussian

Tracking Examples

Acknowledgments

Material presented is based on:

Joint work with my students: H. Grabner, M. Grabner, P. Roth

Funding provided by:

Austrian Joint Research Project Cognitive Vision under sub-projects S9103-N03 and S9104-N04

FIT-IT Program funded by BMVIT under Project AUTOVISTA, EVis

Doctorial College: Confluence of Vision & Graphics