Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Simplifying Object Detection: Classifier Grids for Learning Robust Adaptive Object Detectors

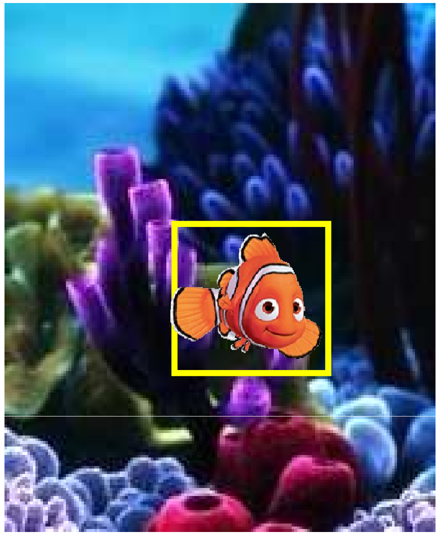
Helmut Grabner

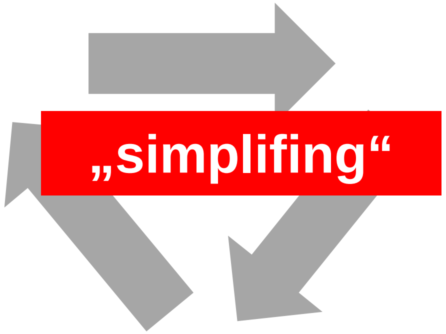
Simplifying Obj… Detection:
Classifier…
Robust A… s

…ut Grabner

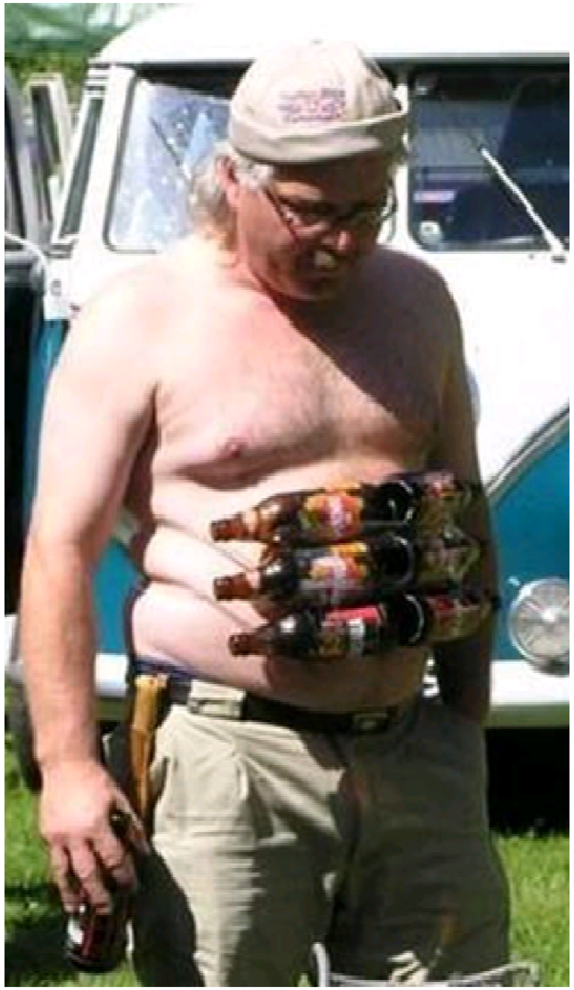
The talk of many cites and funny pictures…

Outline

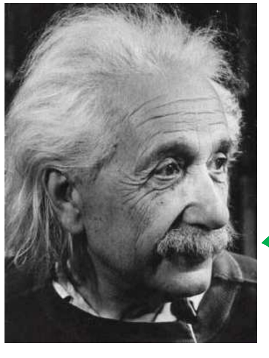


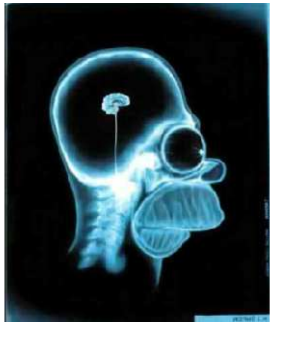

Approaches
„simplifing“
Visual Object Detection
Lessons Learned


D'OH!



Object Detection





Challenges 1: View Point Variation

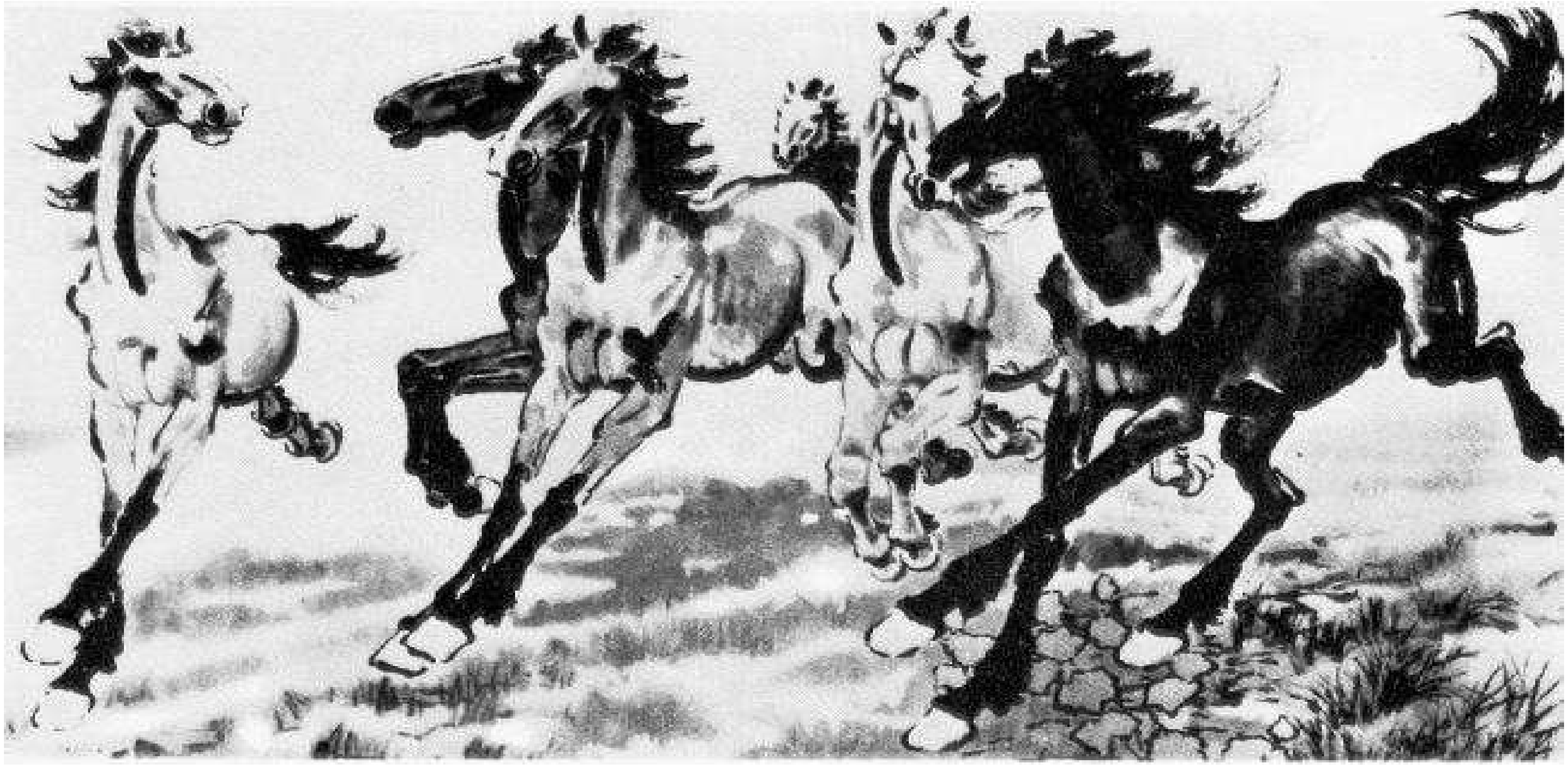
[Fei Fei, Torralba, Fergus 2005-2008]

Challenges 2: Illumination

Challenges 3: Occlusion

Challenges 4: Scale

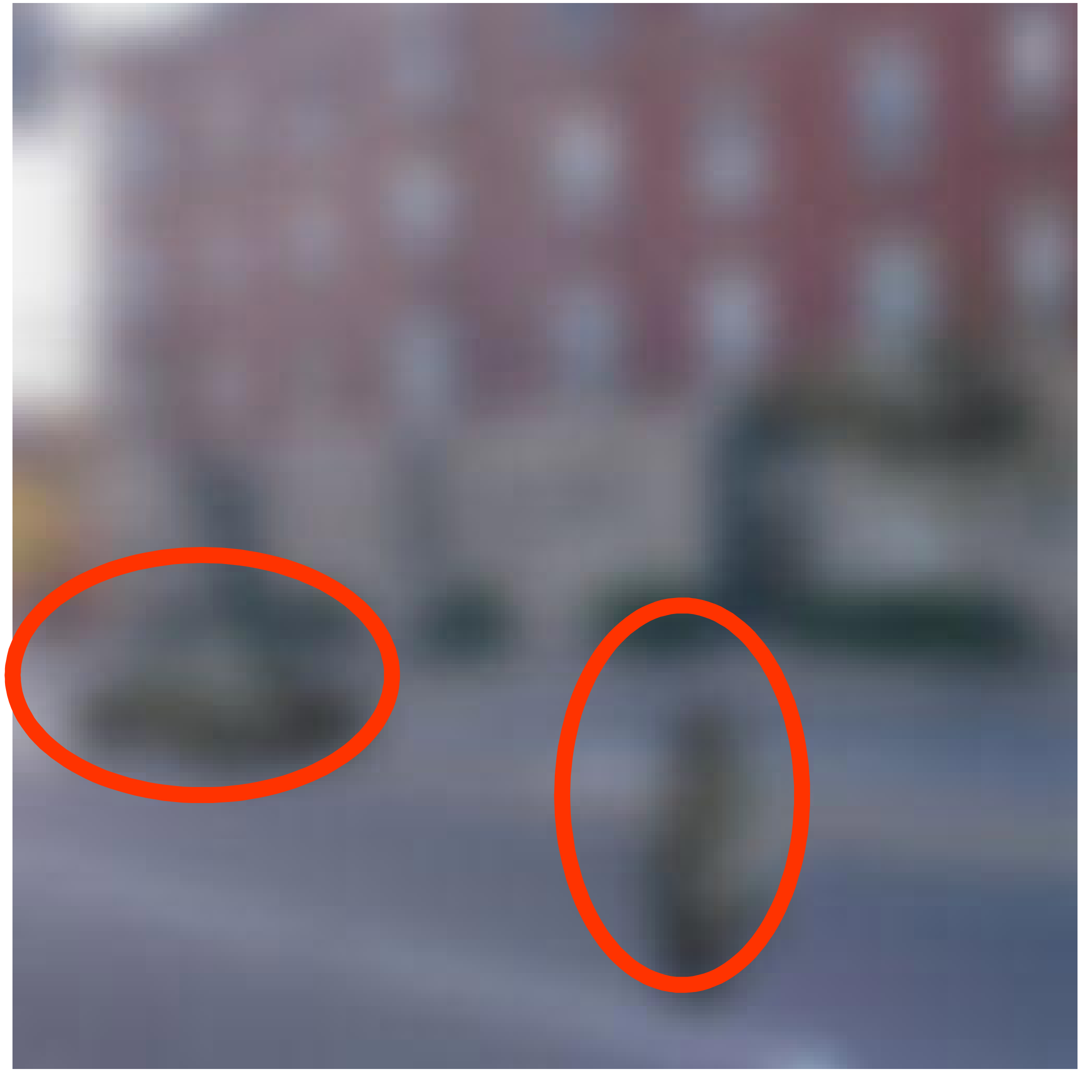
Challenges 5: Deformation



Challenges 6: Background Clutter

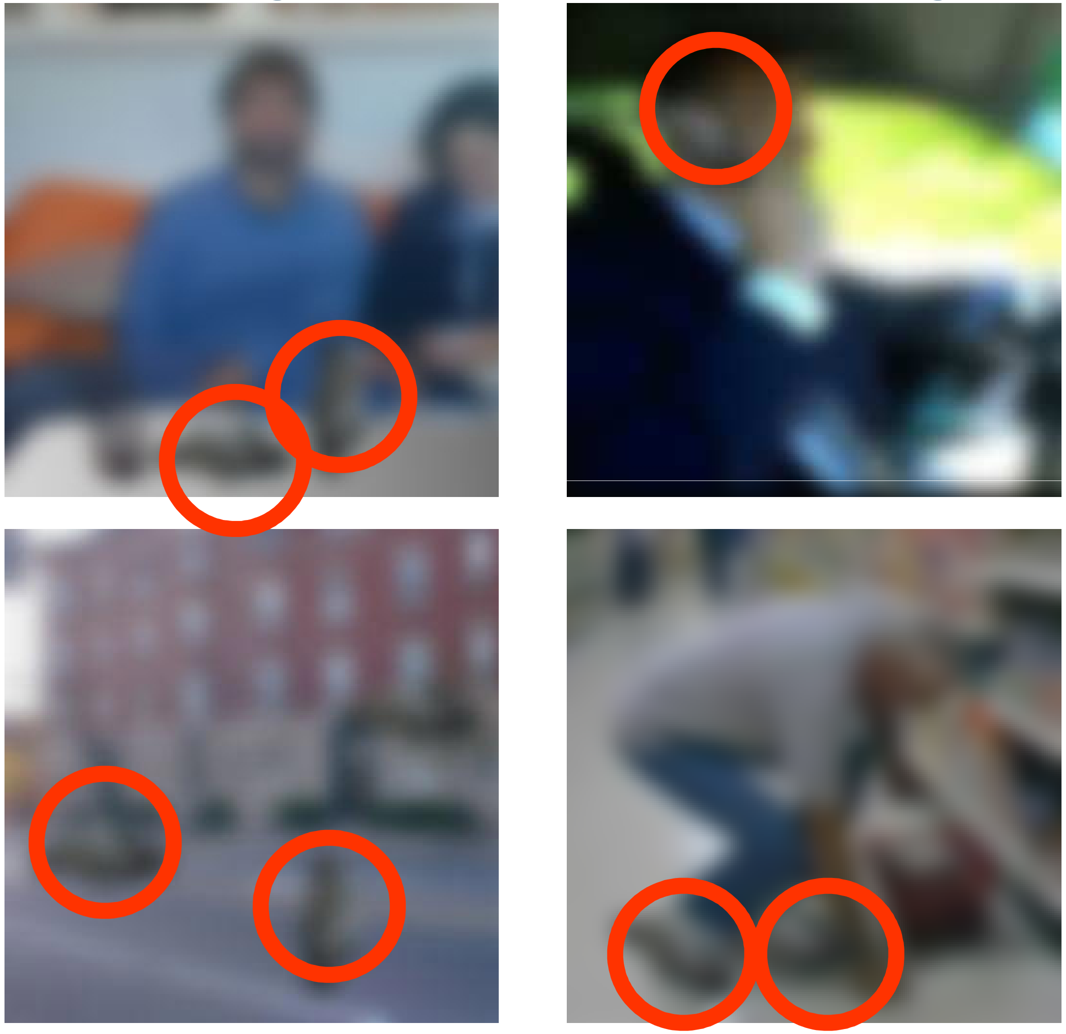
Challenges 7: Intra-class Variation

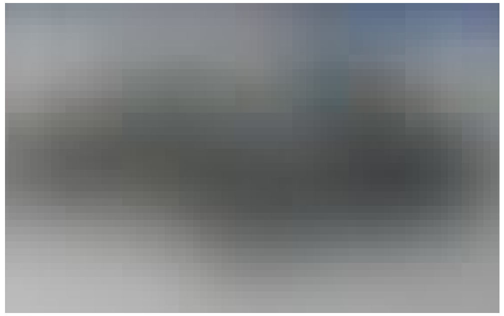
Challenges 8: Local Ambiguity



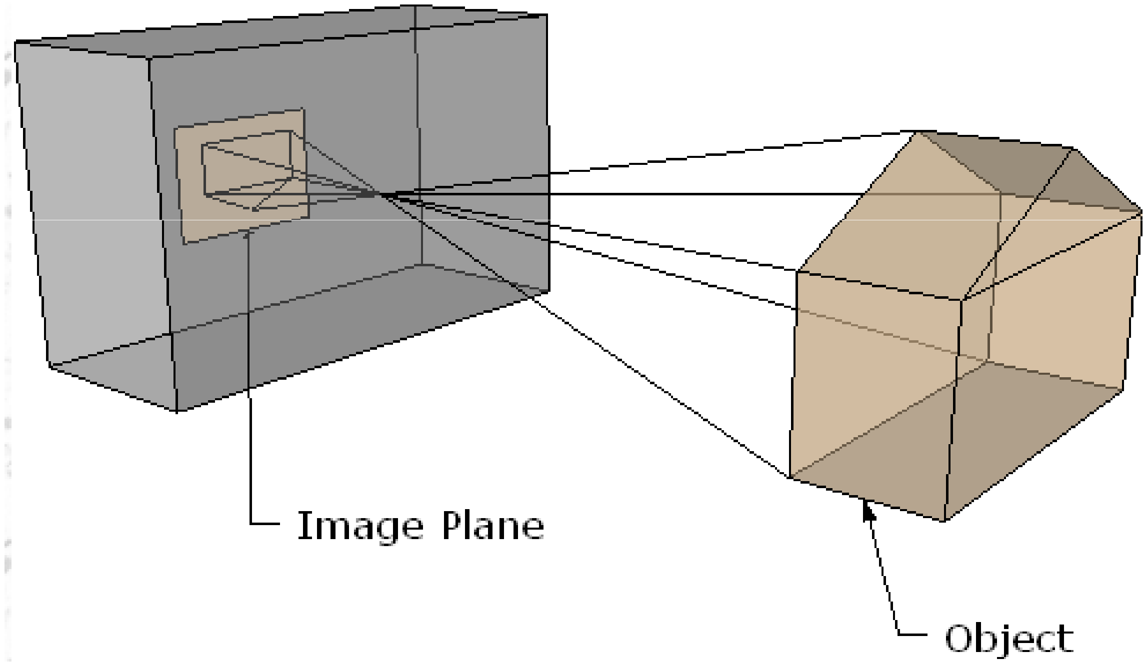
[Torralba]

Challenges 8: Local Ambiguity



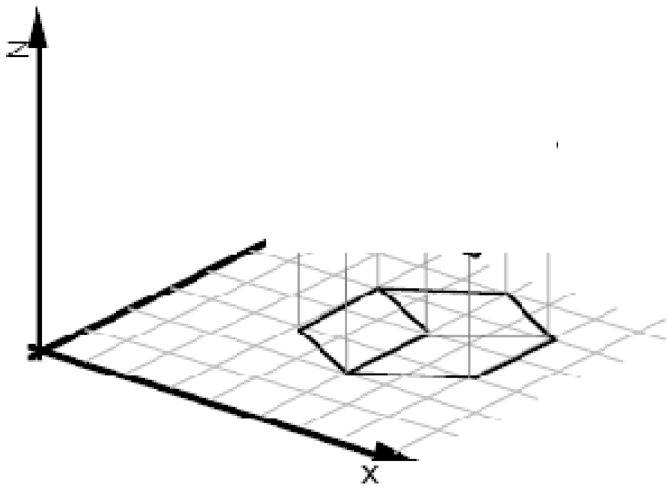


Challenges 9:
The world behind the image



Recovering 3D geometry from single 2D projection

Infinite number of possible solutions!



potential aggressive persons

Video Surveillance

people identification

left luggage detection

thieving

object monitoring

vandalism

people tracking

unusual event detection

loitering

house breaking

potential aggressive persons

Video Surveillance

people identification

left luggage detection

thieving

FOCUS: Object Detection

object monitoring

vandalism

people tracking

unusual event detection

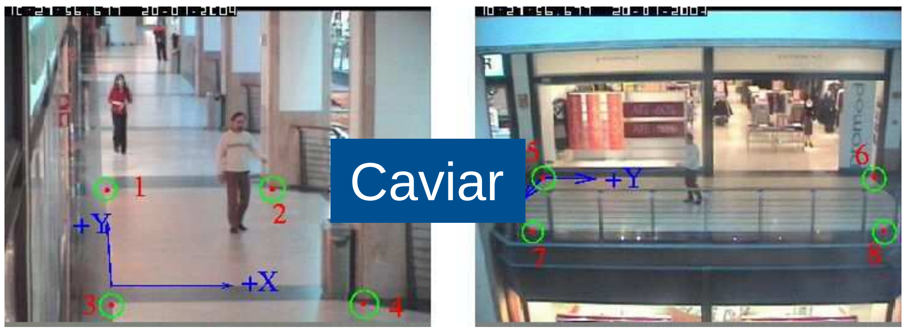
loitering

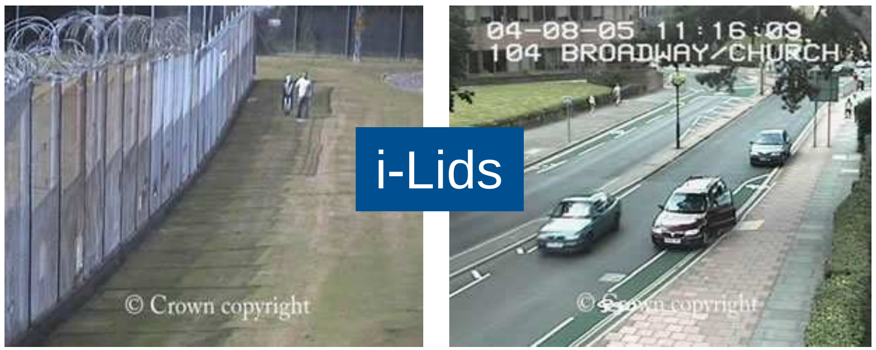
house breaking

Is this (nowadays) really our problem?

Datasets


PETS


Caviar
+Y
+X
1
2
3
4


i-Lids
© Crown copyright
04-08-05 11:16:09
104 BROADWAY/CHURCH

Crowds

Occlusions/Self Occlusions

Occlusions


Camera 4 1 4.33 fps , 6022 kBit/s 1 4:07:51 27
Camera 4 1 4.32 fps , 6020 kBit/s 1 4:07:50 27

Other moving objects

Similar Objects

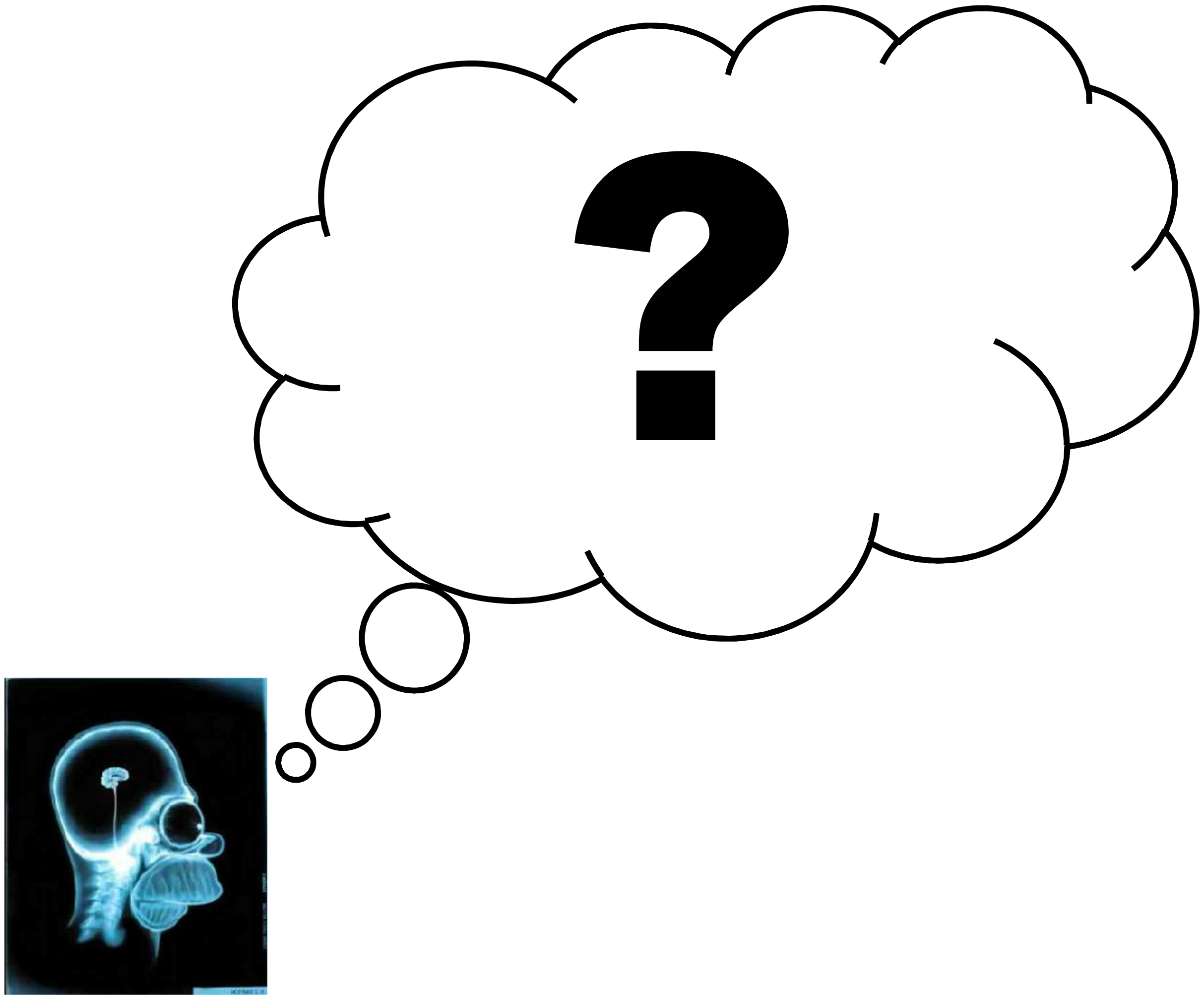
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Real-life Working conditions

Cameras not stabilized!

Real-Time

How can this be done?

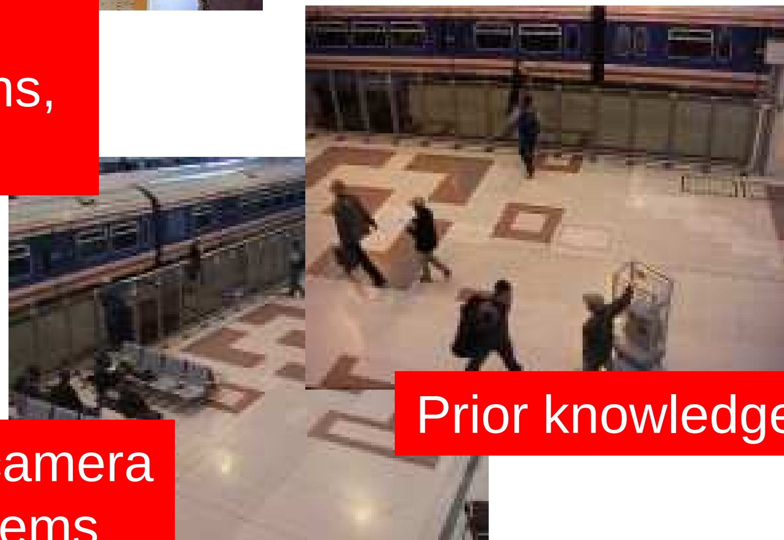


Benefits

Benefits

Static camera

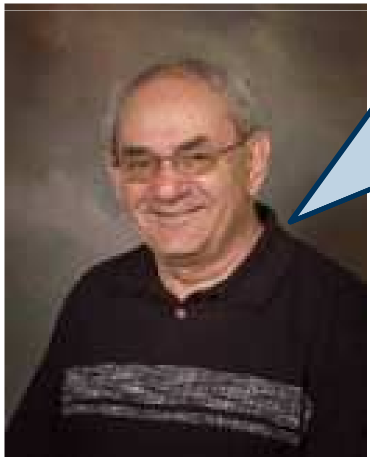
Few object classes (pedestrians, cars,…)



Multi camera systems

Prior knowledge

When trying to solve some problem, one should not solve a more difficult problem as an intermediate step.



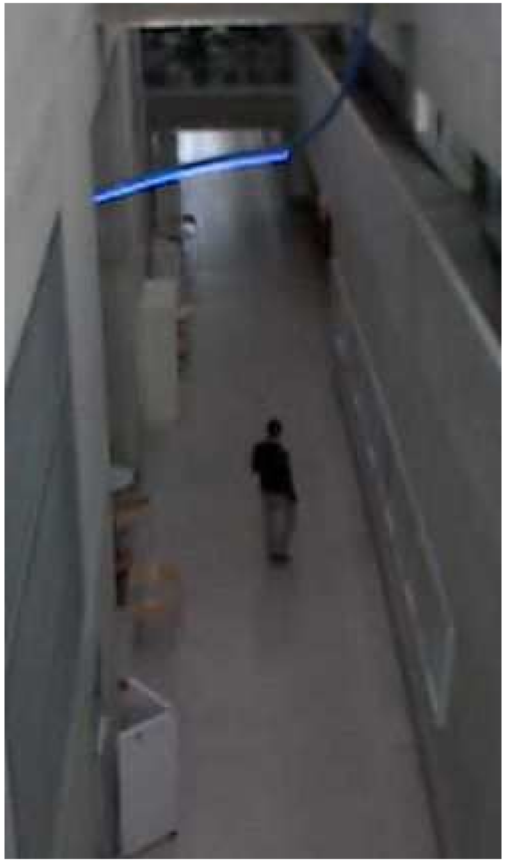
Vladimir Vapnik

PART I

Object Detectors

Object Detection

Input Sequence

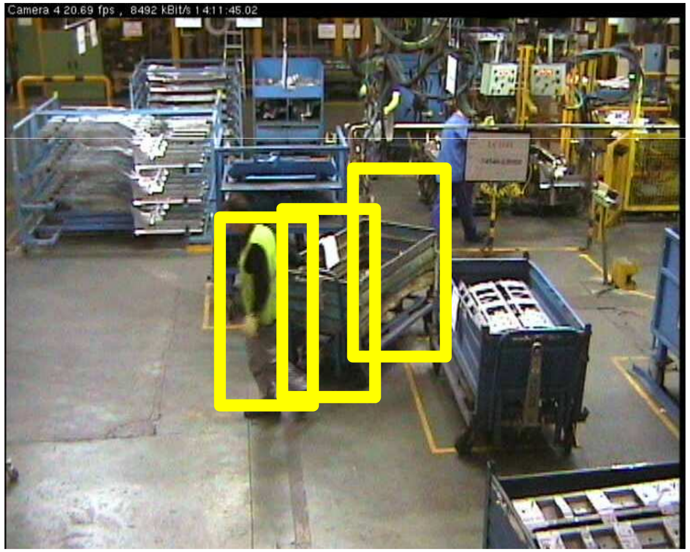


Fixed Camera Setup

Detection

Person Detection: (i) Background Modelling

Input image



Compare

Update

Current background model

Background Modeling

Update

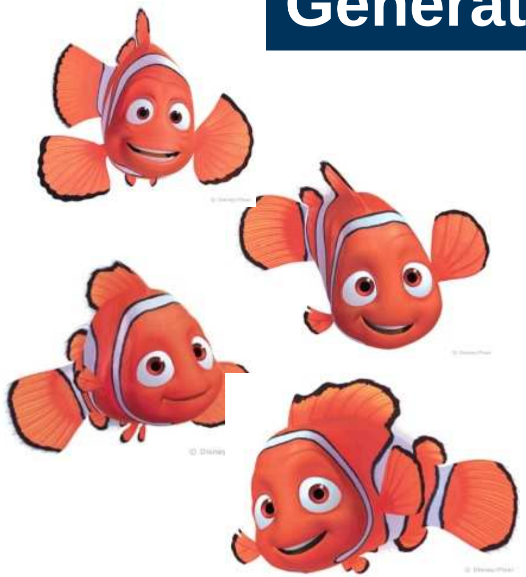
Compare

Easy to implement, good results but…

… strong assumptions (moving objects,…)

Appearance Model

Generative Model



Object

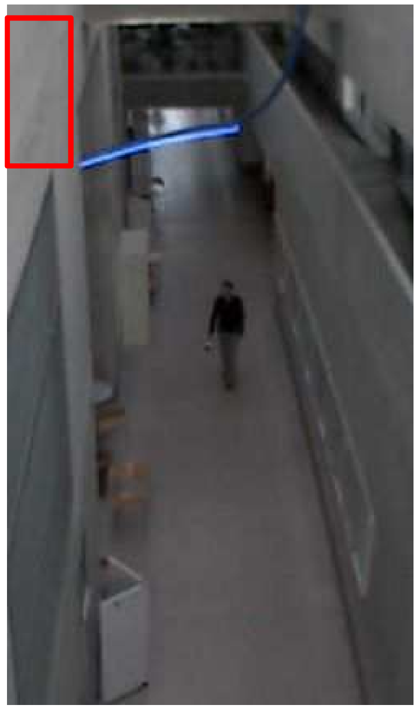
Background

Discriminative Model

Sliding Window

- Detection: Evaluation on many sub-images

Apply classifier exhaustively

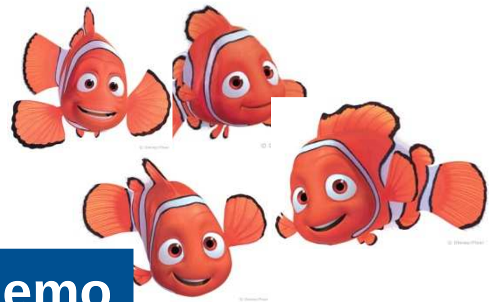


Post-processing →

[Viola/Jones, CVPR 2001]

Machine Learning

Chairs



Nemo

Supervised Learning

Three main issues

- Representation
  - How to represent an object (category)
- Learning
  - How to form the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Learning-based Methods in Vision

Learning as last resort.

We are trying to solve problems that do not have a solution!

Alyosha Efros

- This will be done using Data:
  - E.g. what happened before is likely to happen again
  - Google Intelligence (GI): The AI for the post-modern world!

Learning-based Methods in Vision

The aim of computer vision is to overfit to our visual world.

remark by Antonio Torralba
(after his third beer)

Person Detection

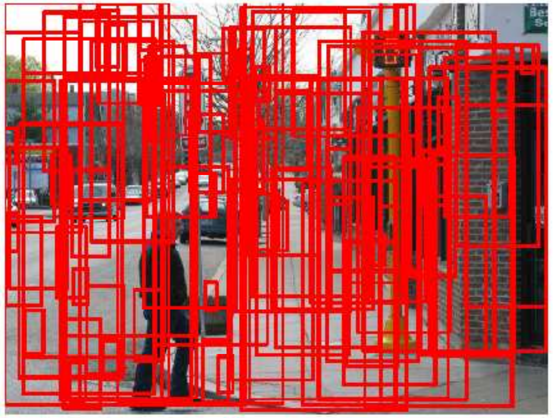
[Felzenszwalb et al. CVPR 2008]

Stable, always the same result

Has to cope with <u>ALL</u> possible situations!

Context

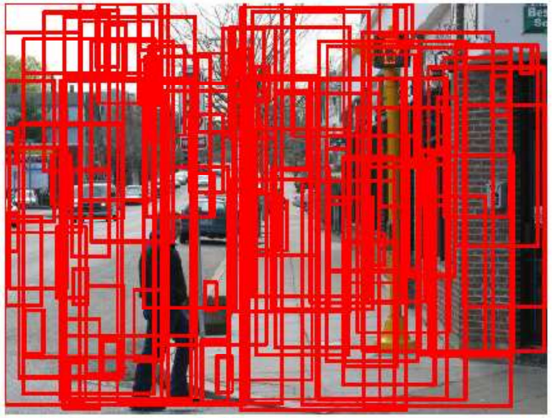
Context ⇨ Prior Knowledge



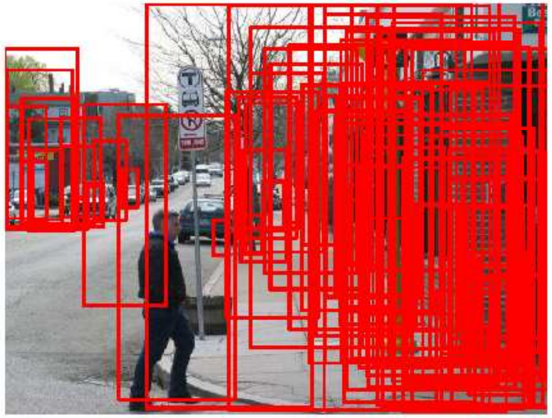
(b) P(person) = uniform

[Hoiem, Efros and Hebert, CVPR 2006]

Context ⇨ Prior Knowledge



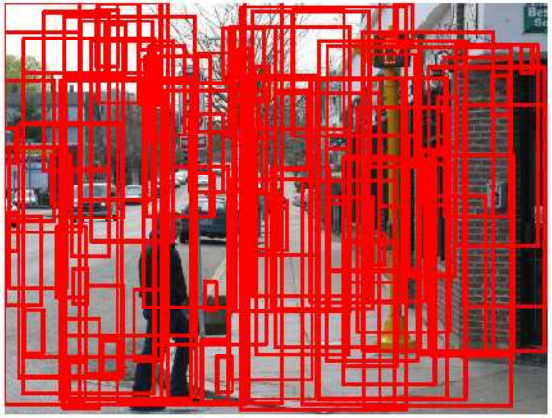
(b) P(person) = uniform



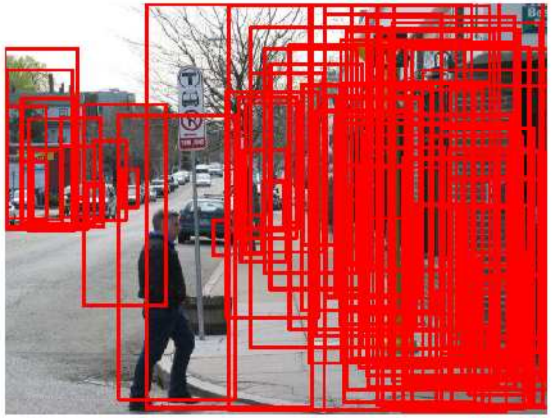
(d) P(person | geometry)

[Hoiem, Efros and Hebert, CVPR 2006]

Context ⇨ Prior Knowledge



(b) P(person) = uniform

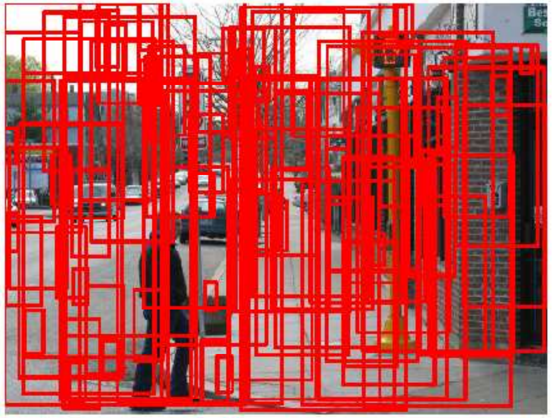


(d) P(person | geometry)

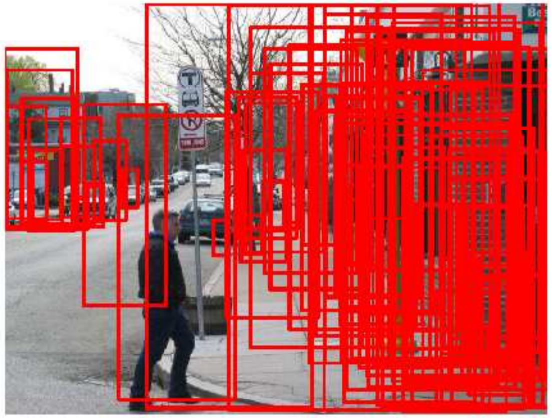
(f) P(person | viewpoint)

[Hoiem, Efros and Hebert, CVPR 2006]

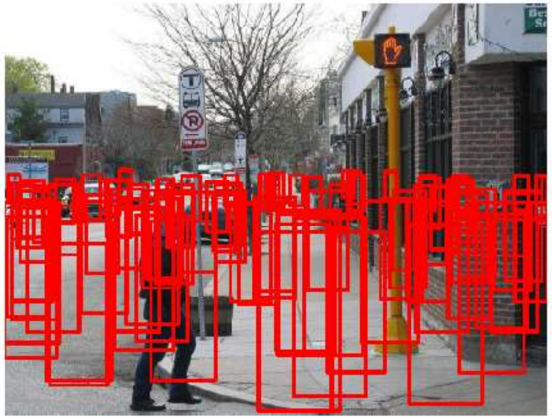
Context ⇨ Prior Knowledge



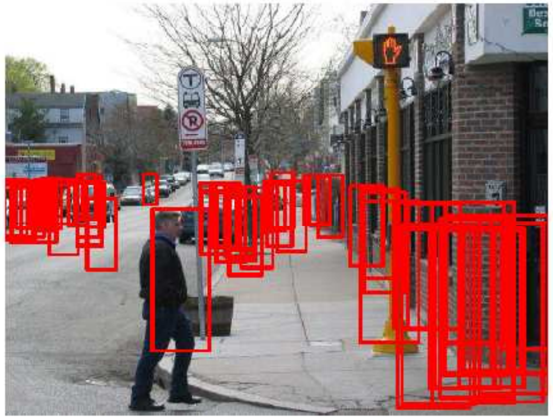
(b) P(person) = uniform



(d) P(person | geometry)



(f) P(person | viewpoint)



(g) P(person|viewpoint,geometry)

[Hoiem, Efros and Hebert, CVPR 2006]

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

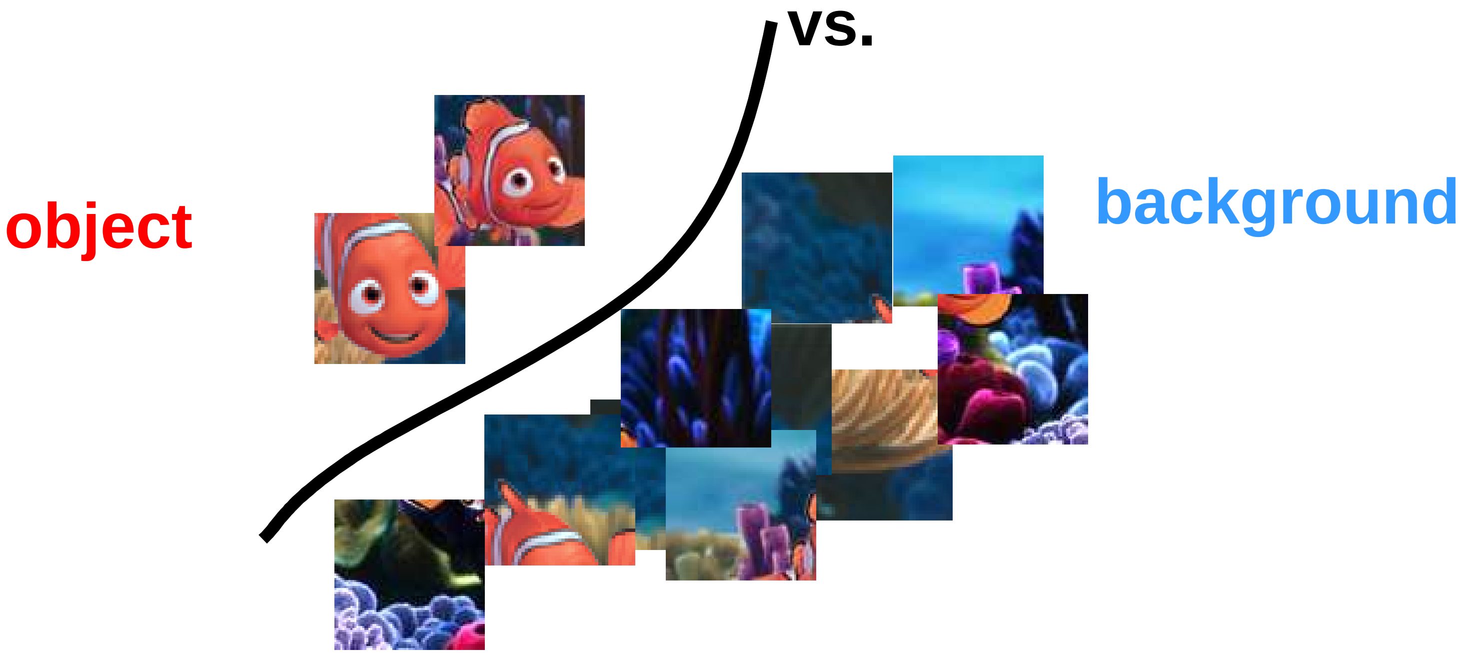
Simpler problem as before!

Some kind of post-processing (reduce false positive rate), classifier stays the same!

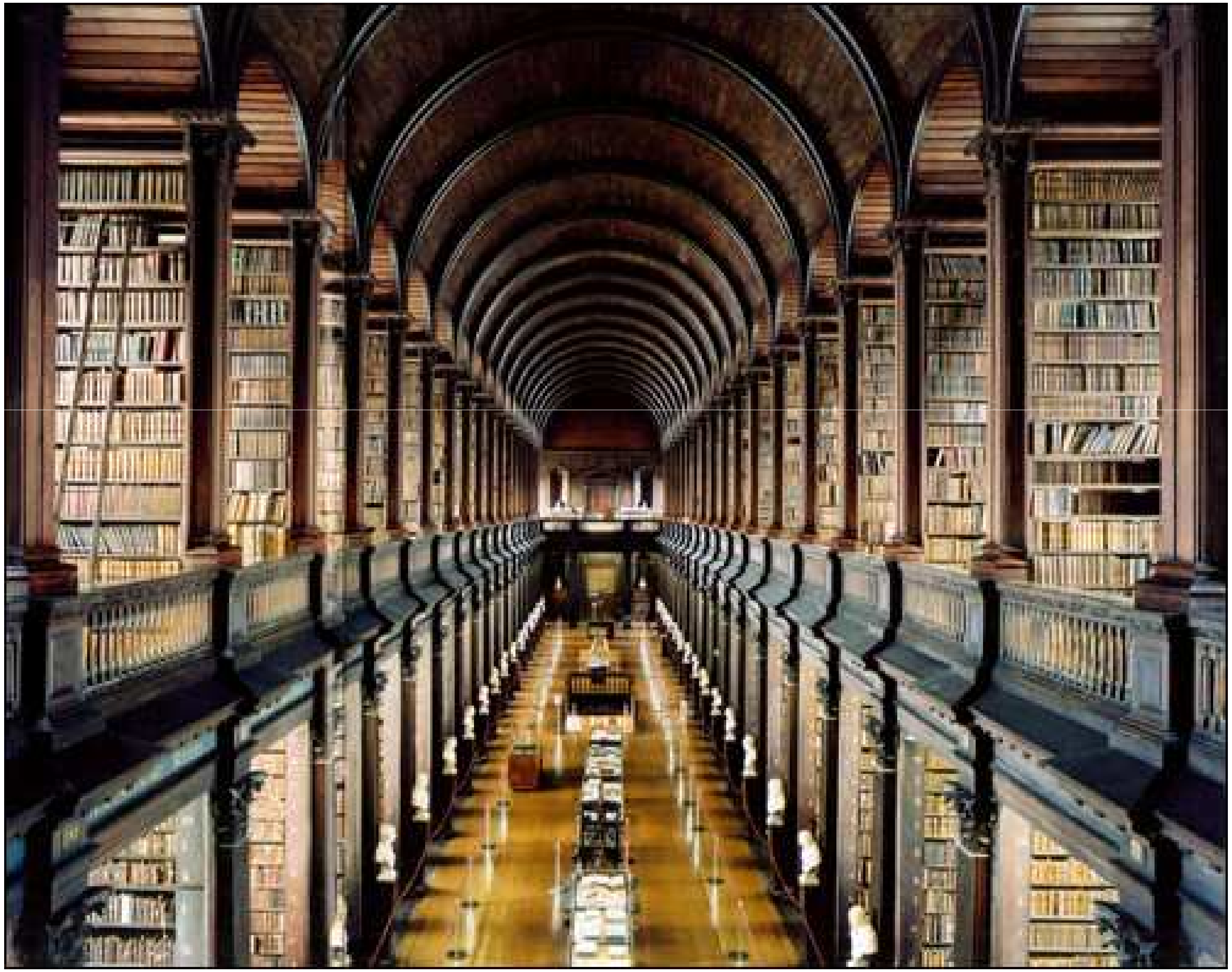
PART II

Adaptive Object Detectors

Object Detection as Binary Classification



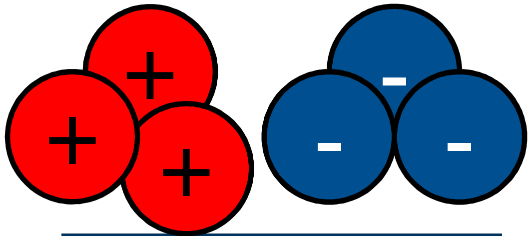
Off-line learning



On-line Learning

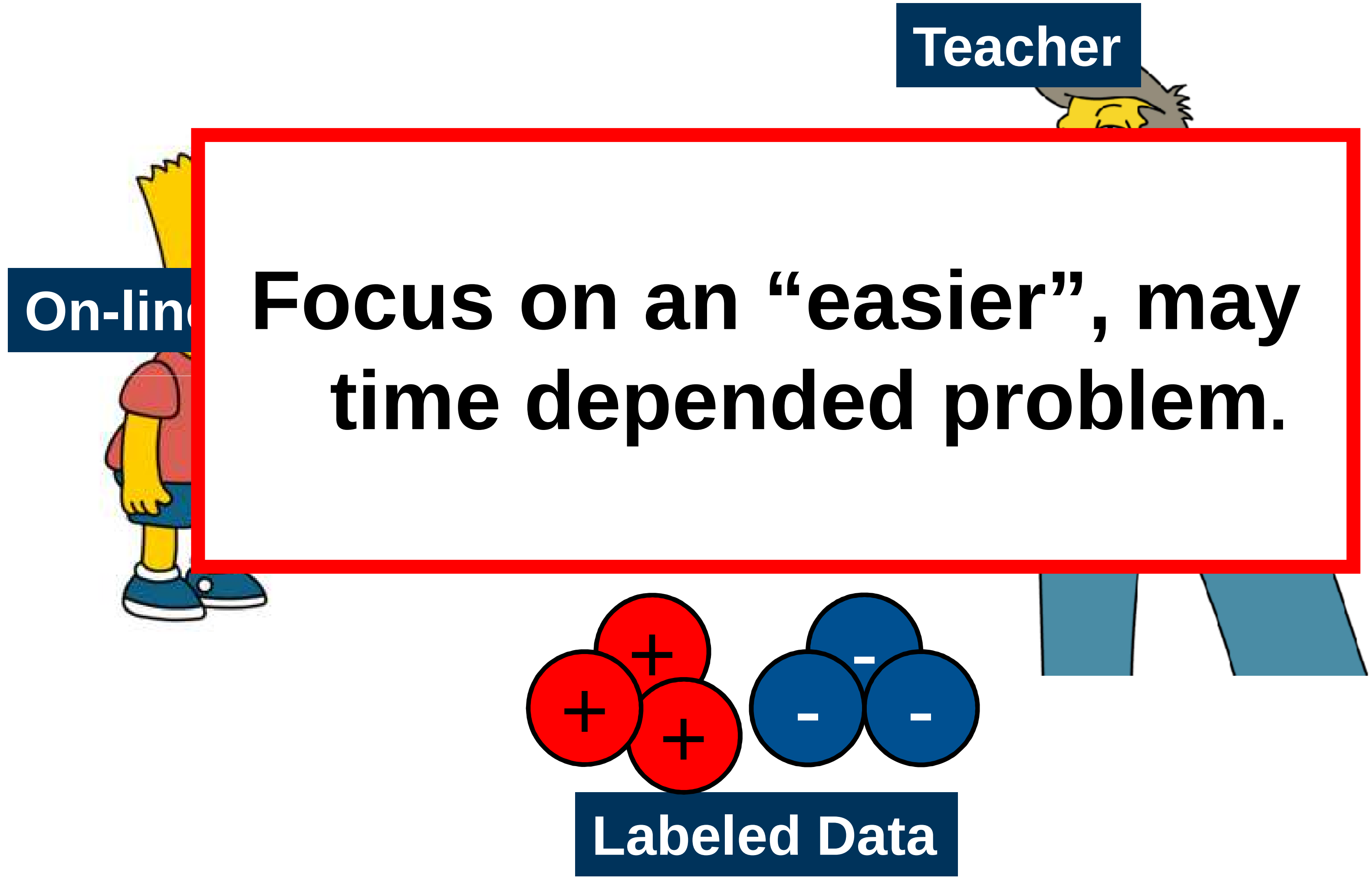
Teacher

On-line



Labeled Data

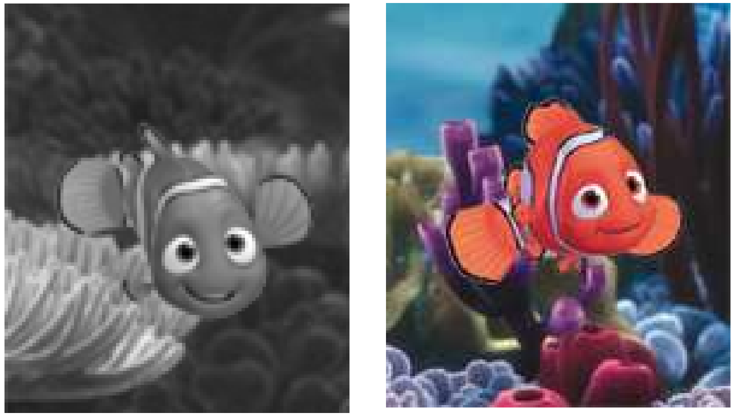
On-line Learning

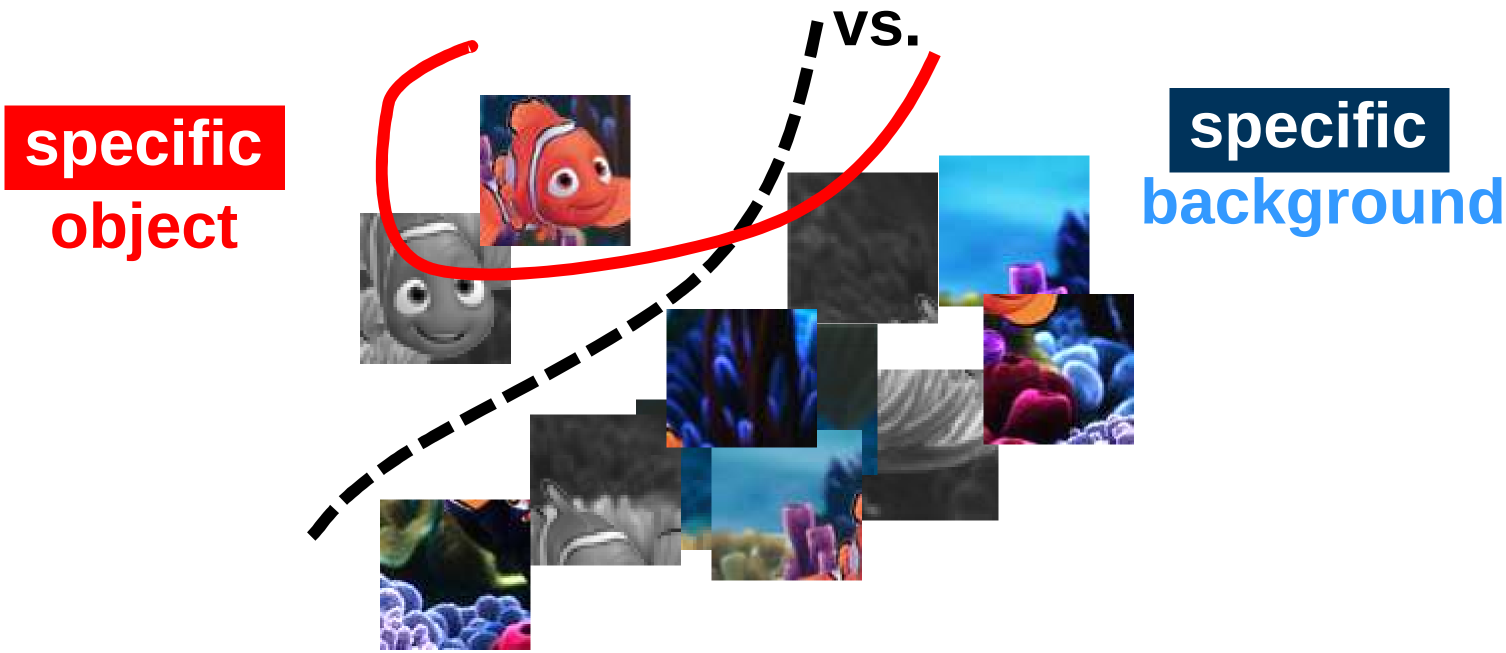


Focus on an “easier”, may time depended problem.

Object Detection

Adaptive Object Detection



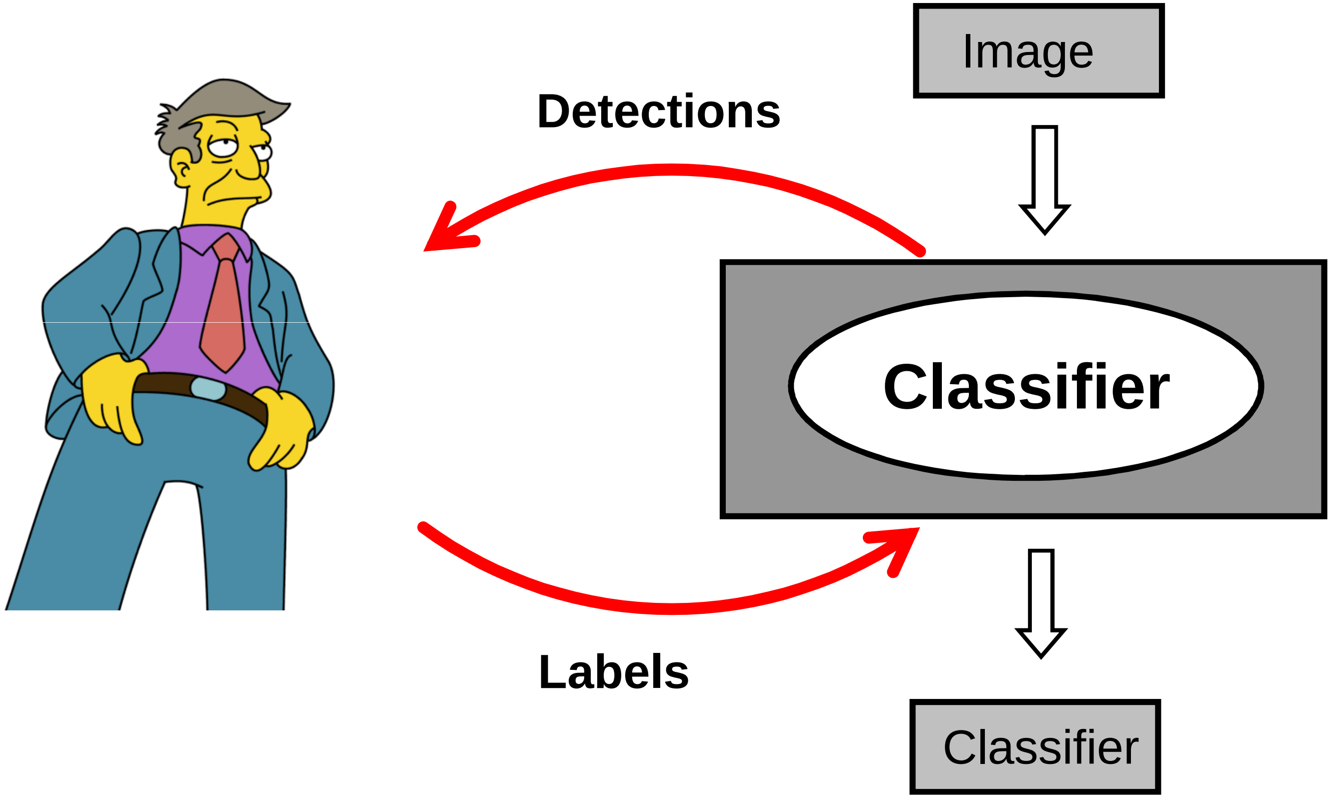


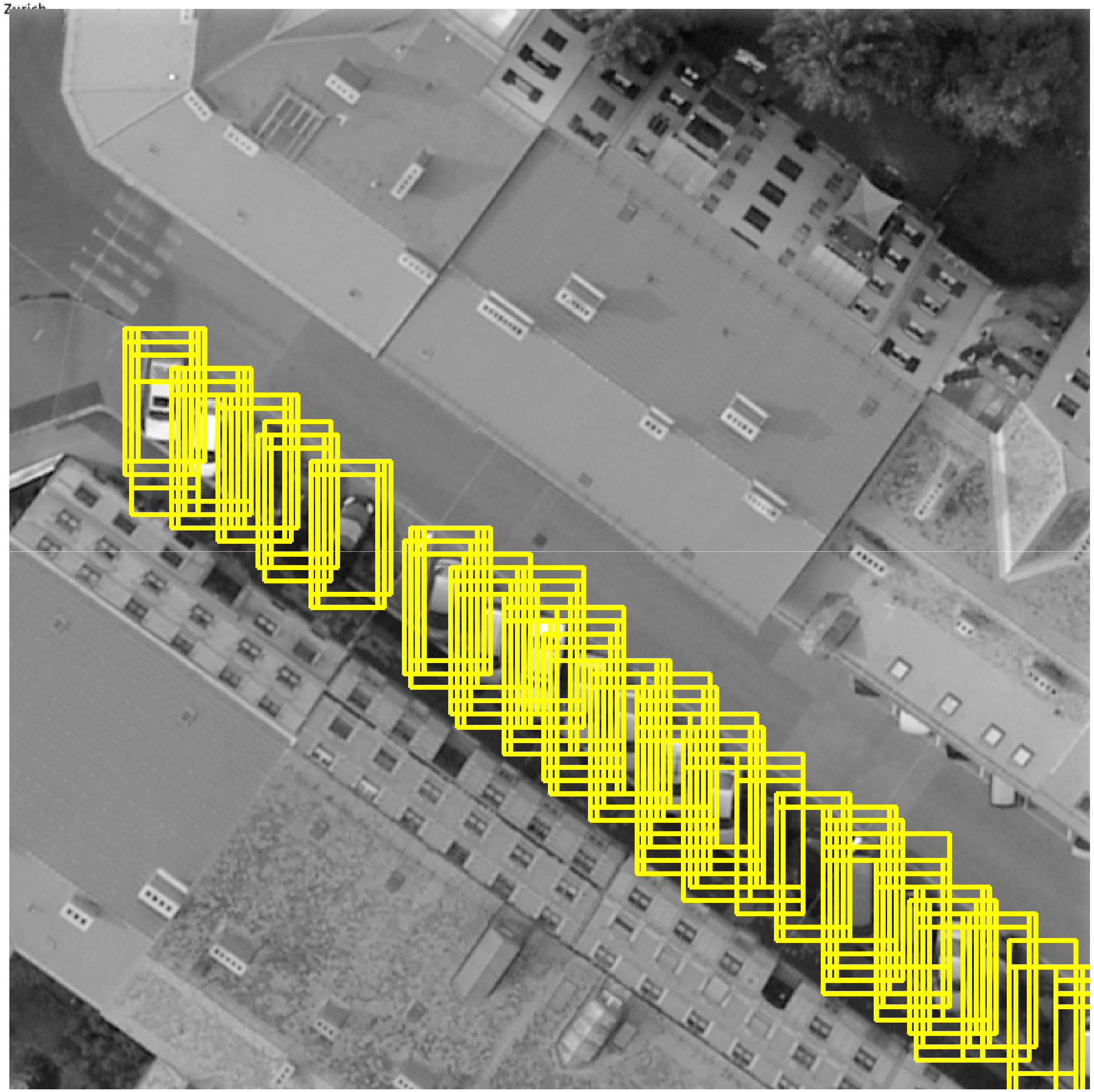
Simpler problem as before!

How to update the classifier?

"Fewer Clicks - Less Frustration"

- Active Learning

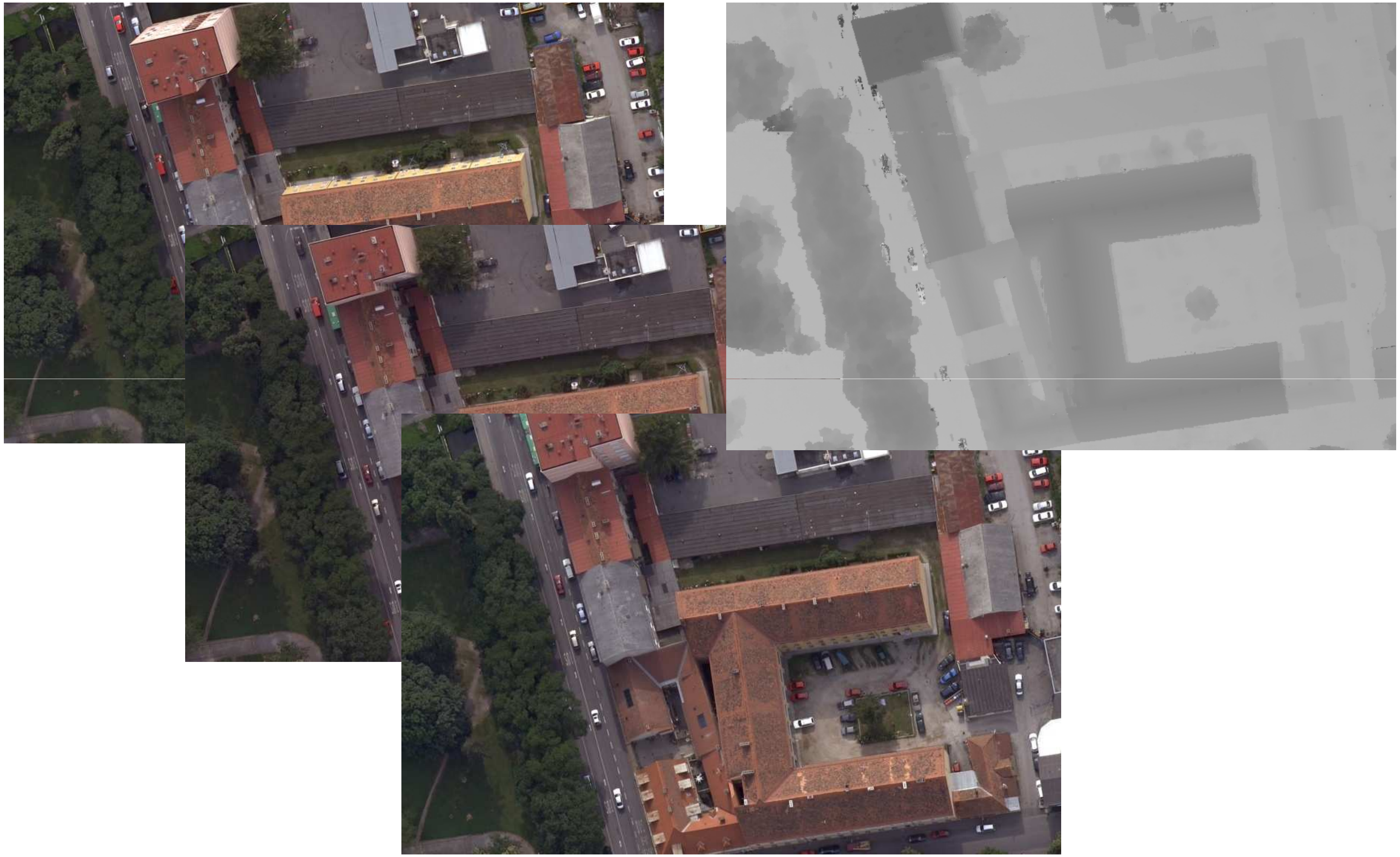




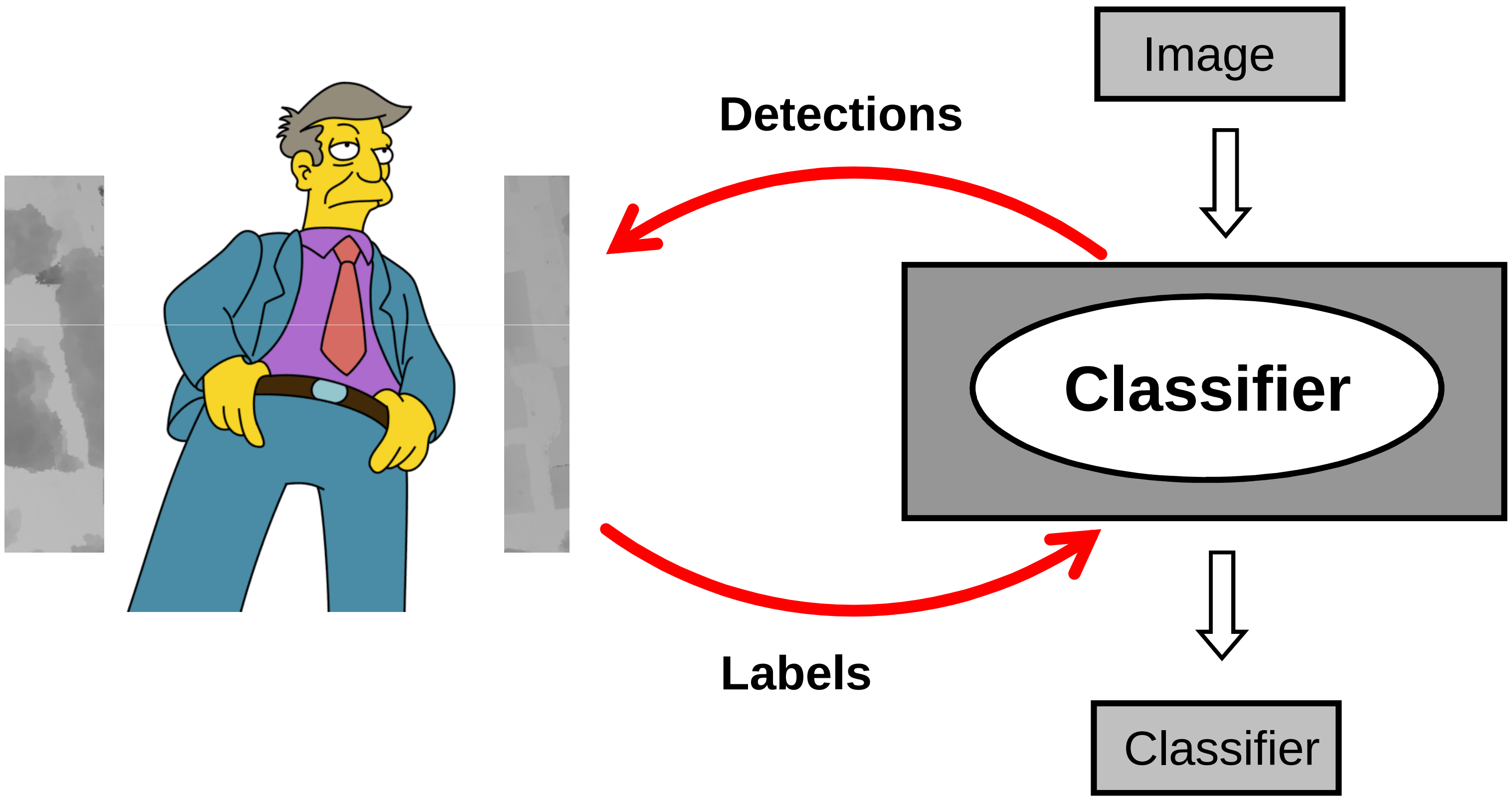
Minimize hand labeling

The Supervisor!

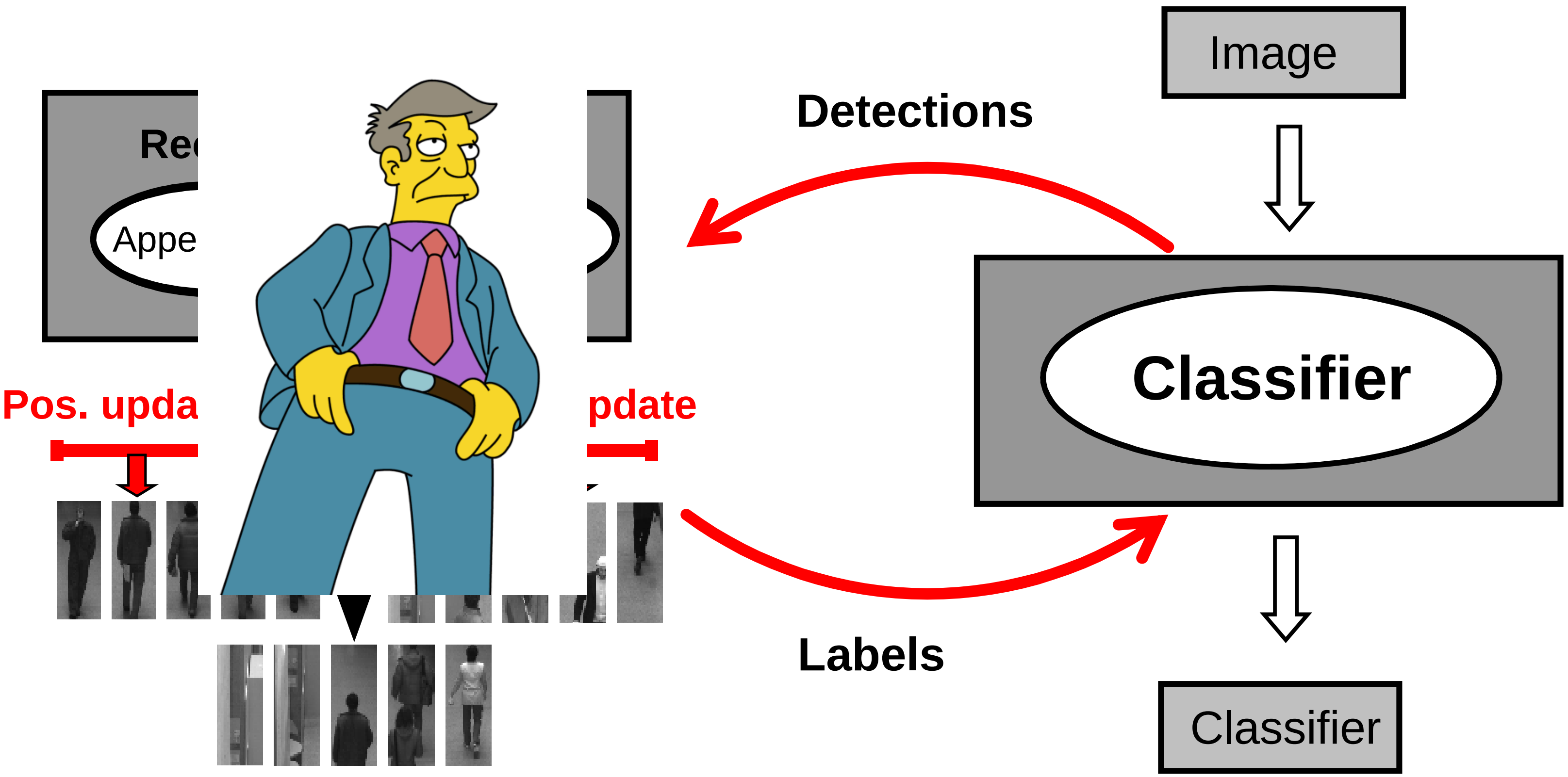
A 3D Teacher



A 3D Teacher

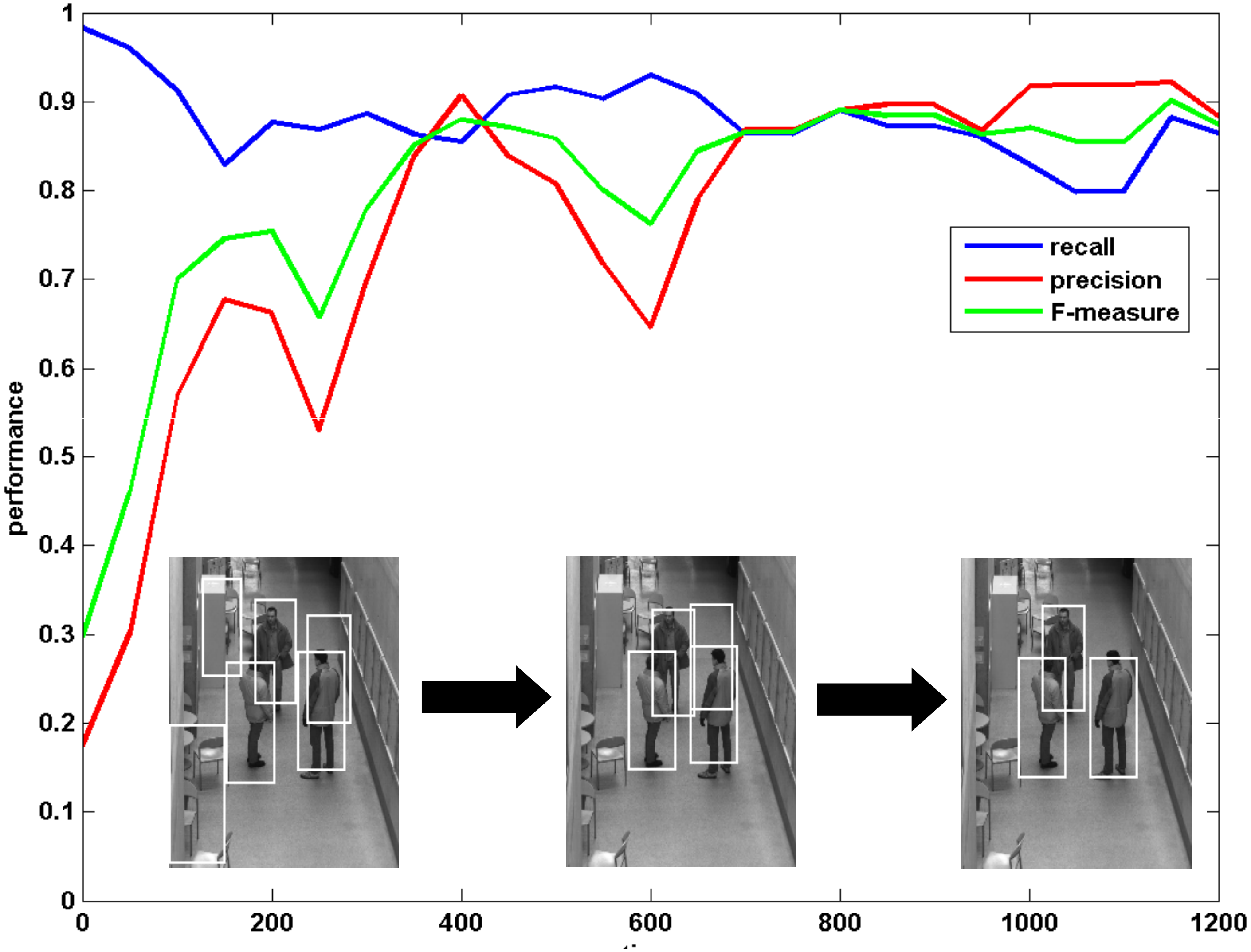


Conservative Learning



[Roth et al. VS-PETS WS 2005]

Improving Performance

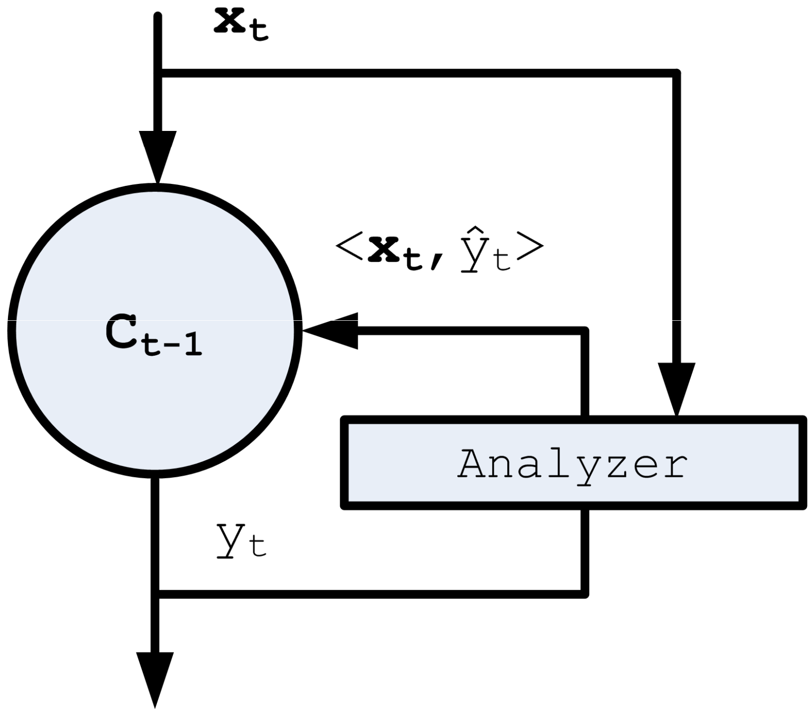


More or Less no Hand labeling, Improving Performance

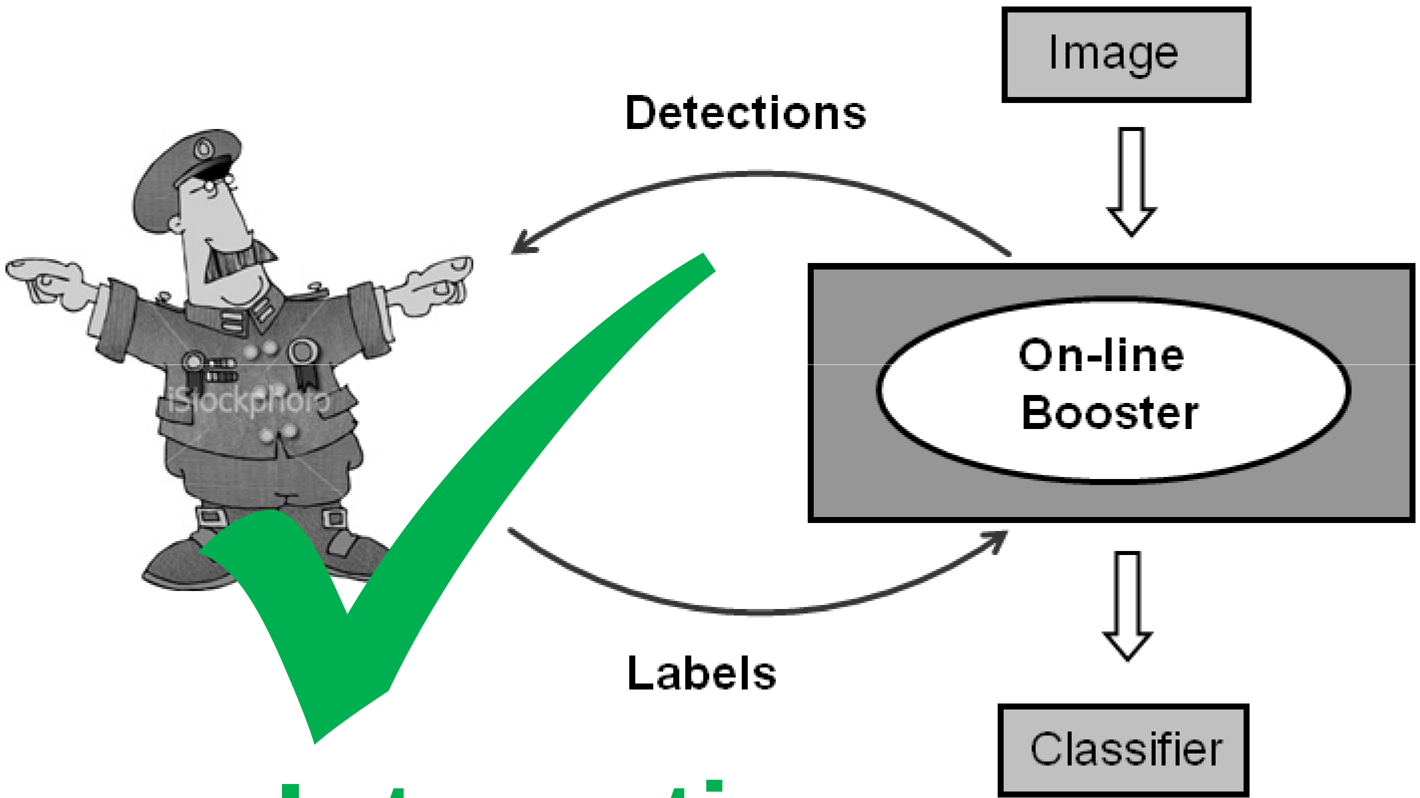
Does it fail?
If yes, when?

Two Issues

- Mainly Verification
  - Reduce false positive rate
- Update Strategies
  - Oracle
  - Verification
  - Co-training
  - Self-training



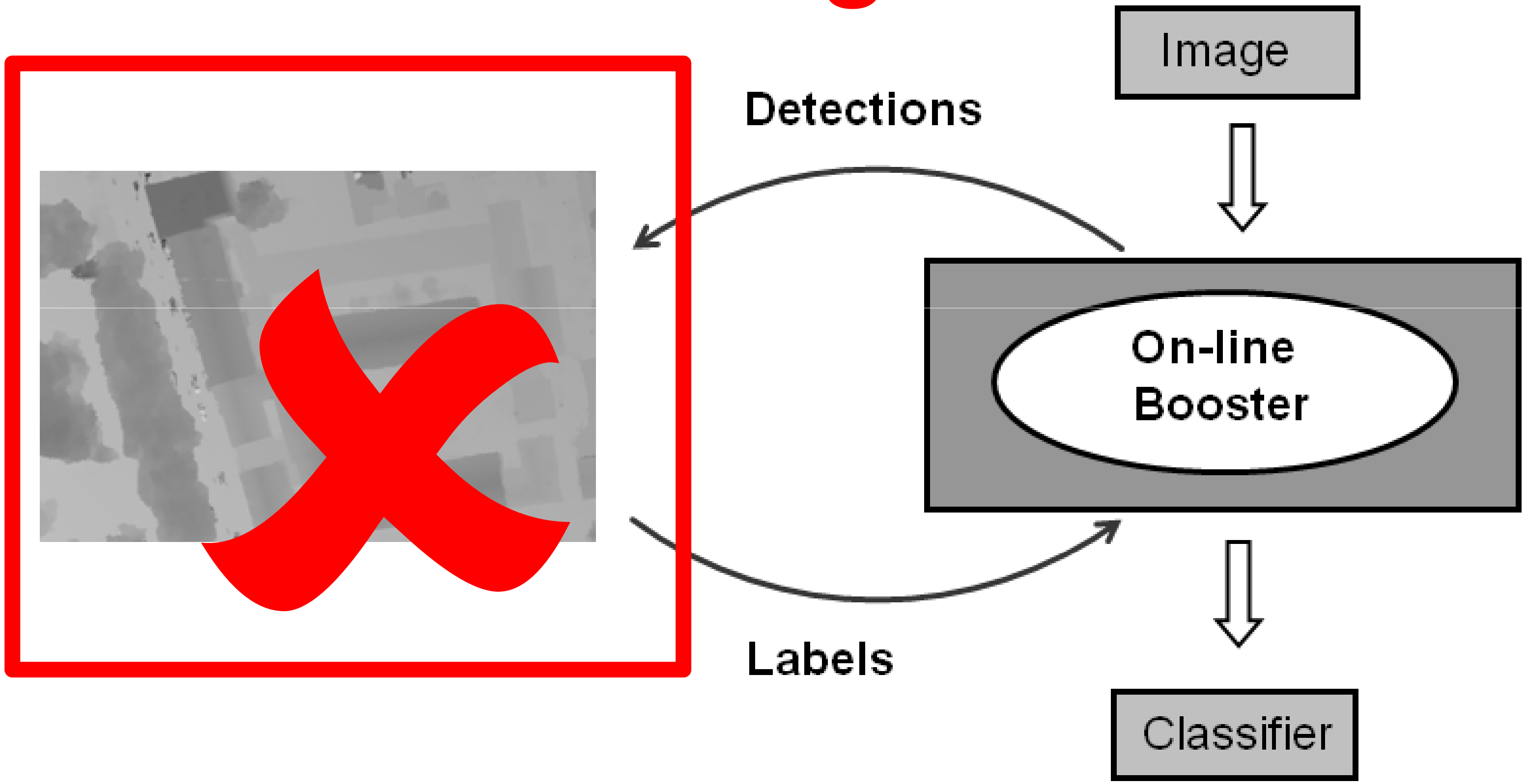
“Fewer Clicks - Less Frustration”



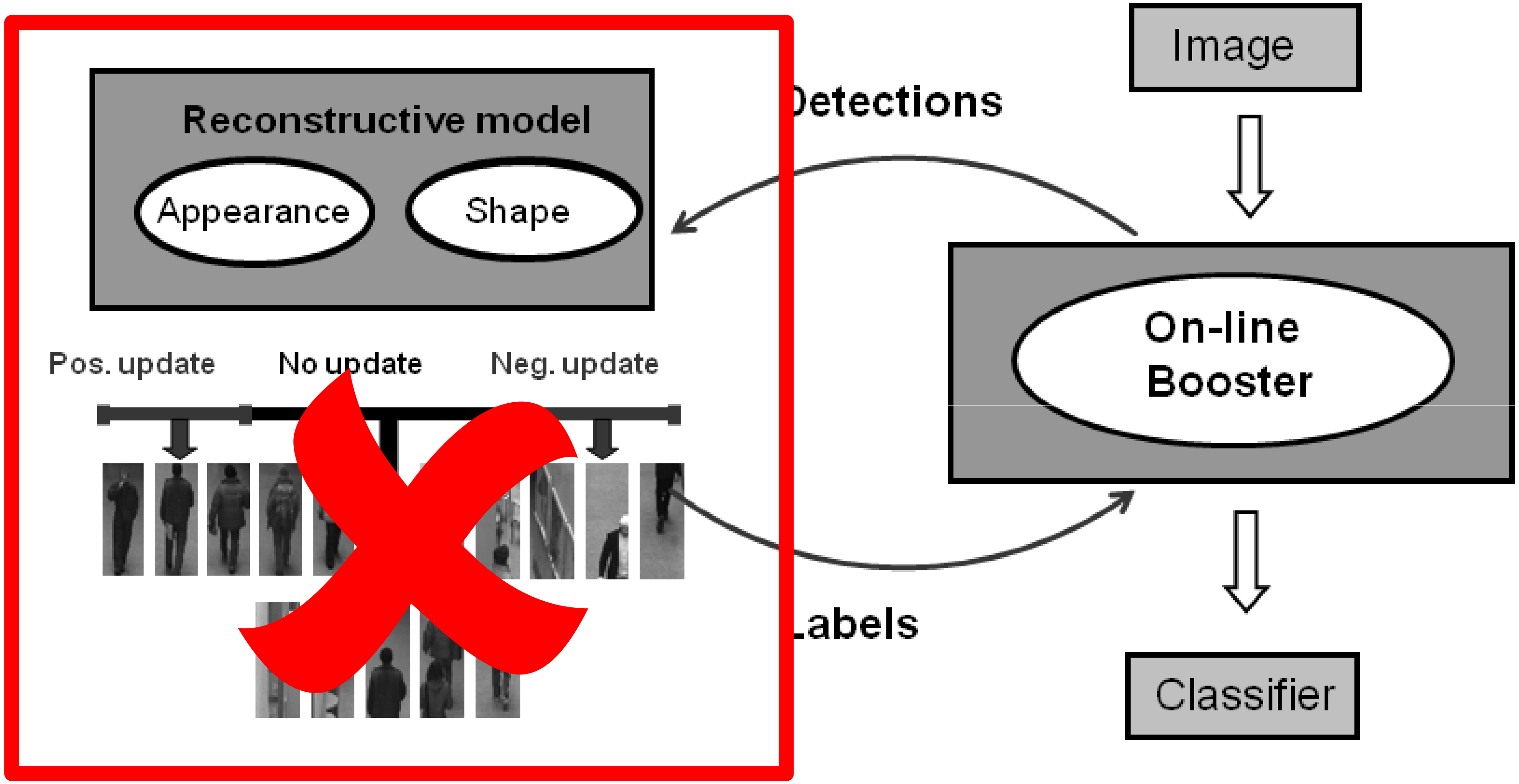
Human Interactions

3D Teacher

Verification using Redundancy

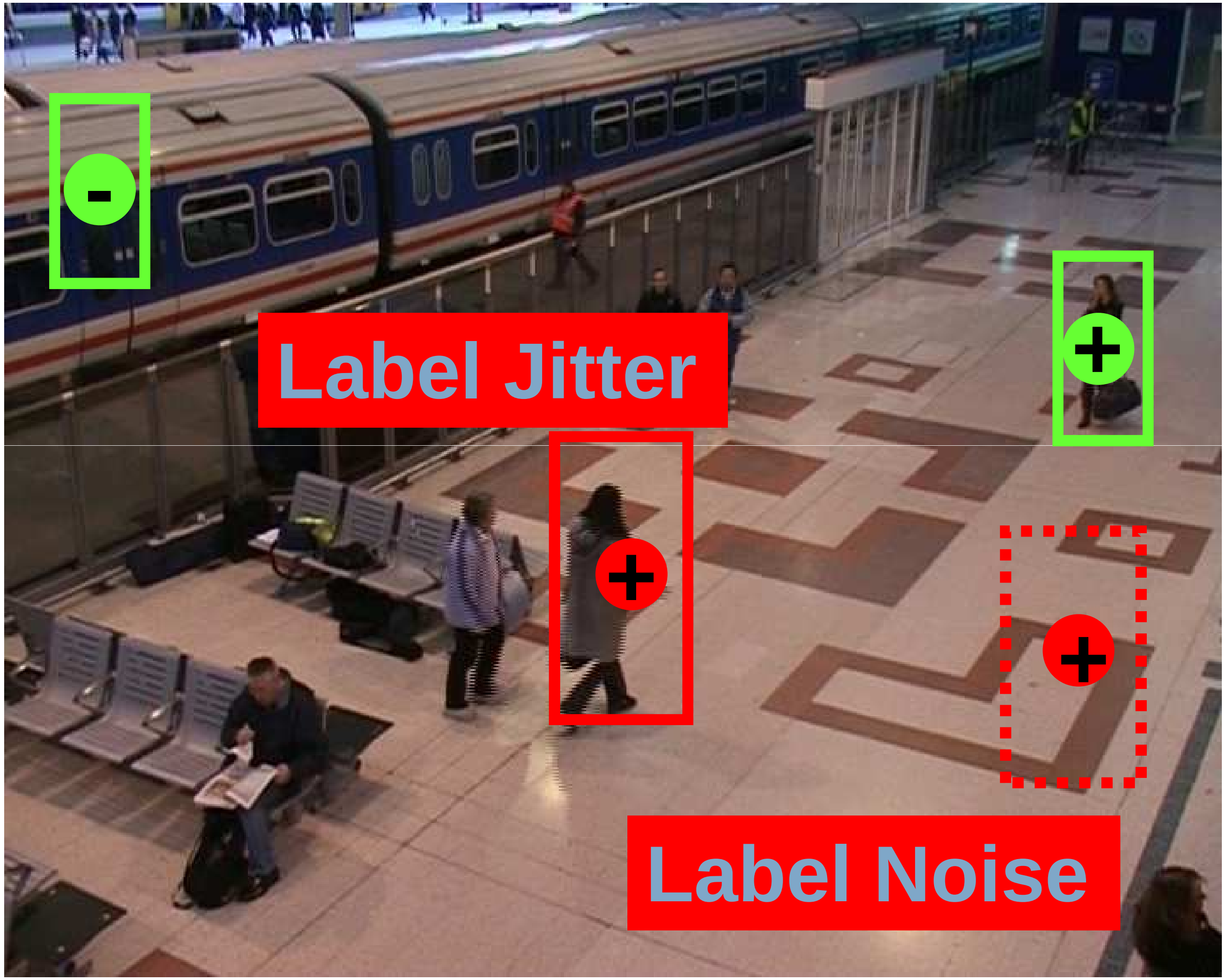


On-line Conservative Learning



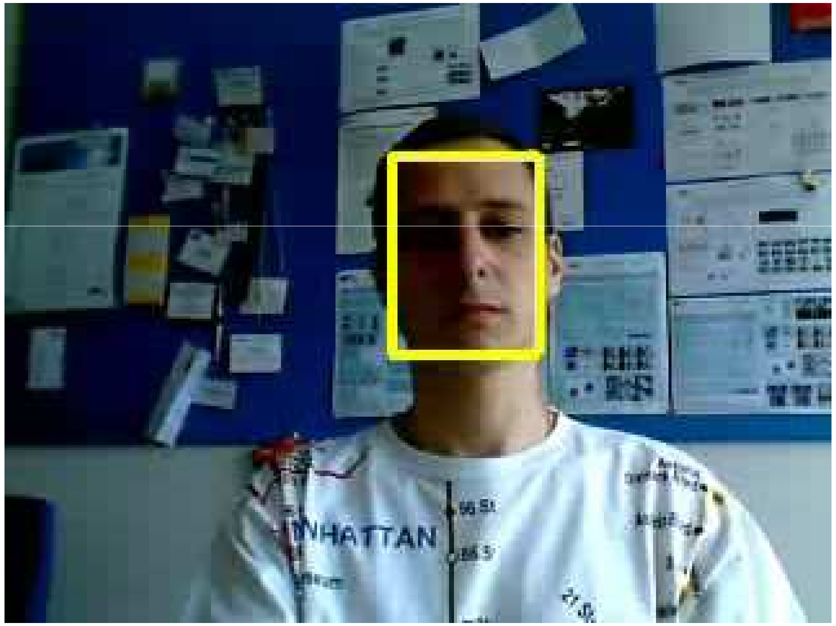
Verification using Classifiers

Problems of…



Btw: Relation to Object Tracking

… However, on-line adaption in model-free tracking faces one key problem: Each update of the tracker may introduce an error which, finally, can lead to tracking failure (drifting). …



PART II

Classifier Grid

We want to build a system which runs

24 hours a day, 7 days a week!

Further

simplifying the problem

in order that we can use a

fixed update strategy,

which do not suffer from the drifting problem.

Background Model as Binary Classification

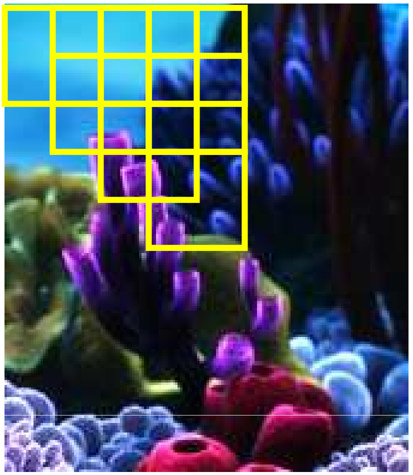
Current Frame

vs.

Image Statistics

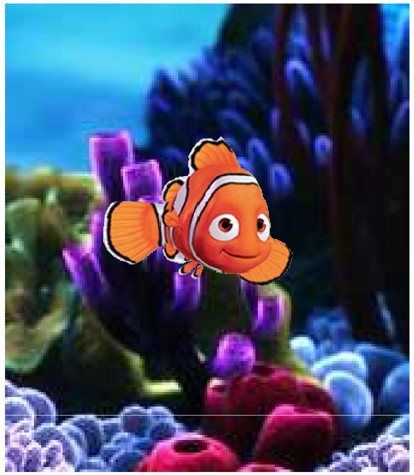
Background Model

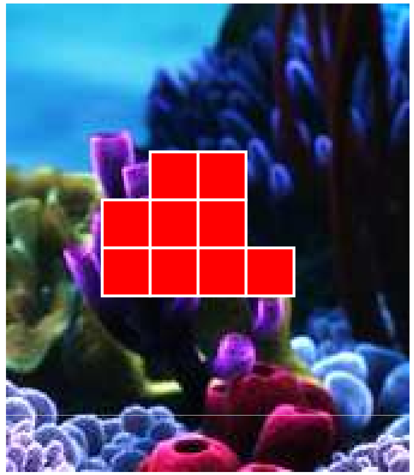
initial time

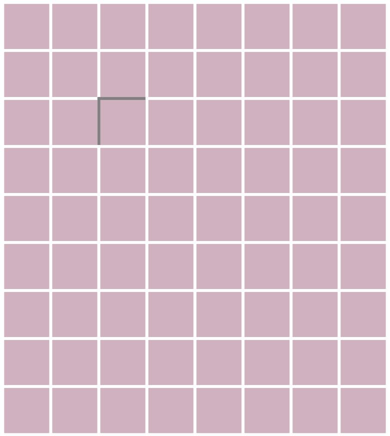


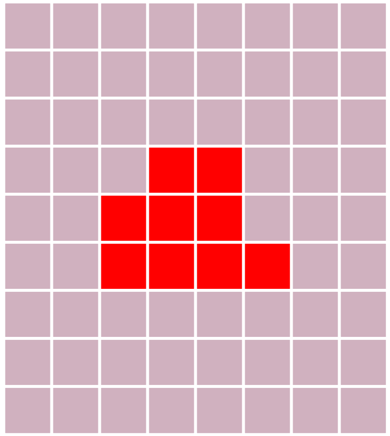
statistic predictable background

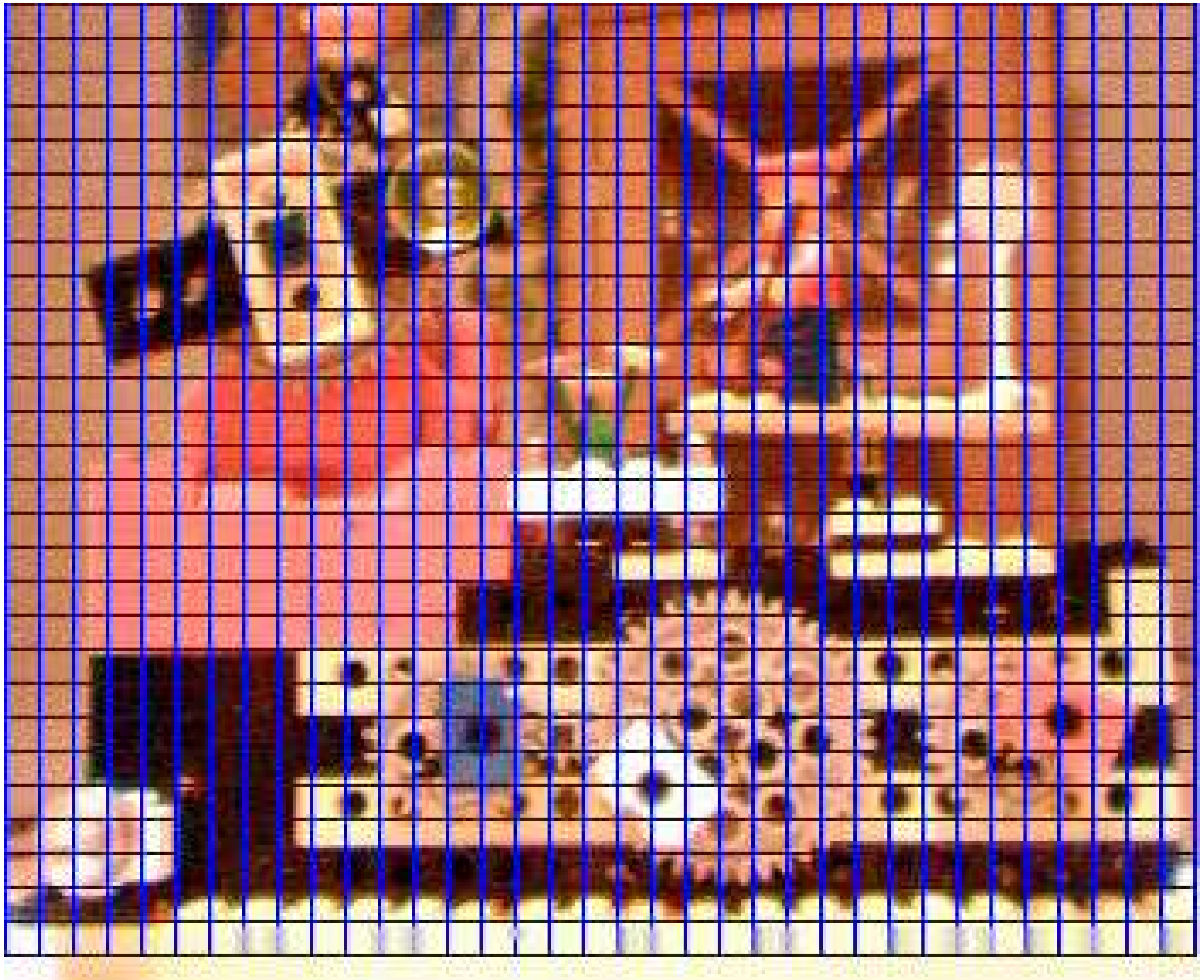
statistic non predictable background





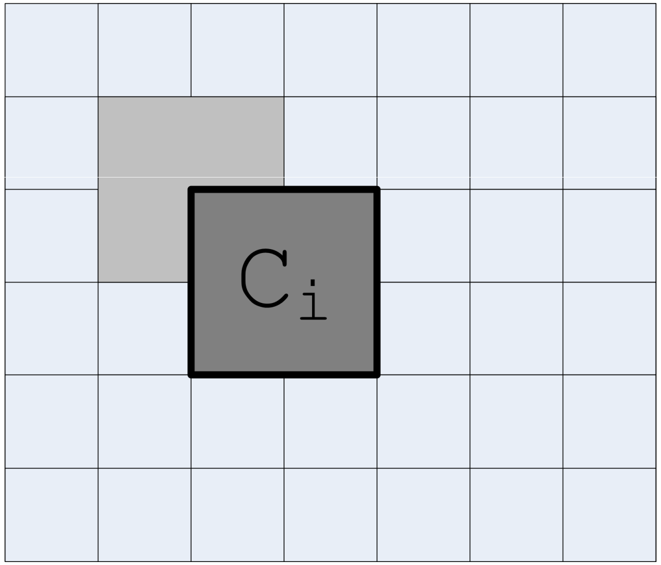




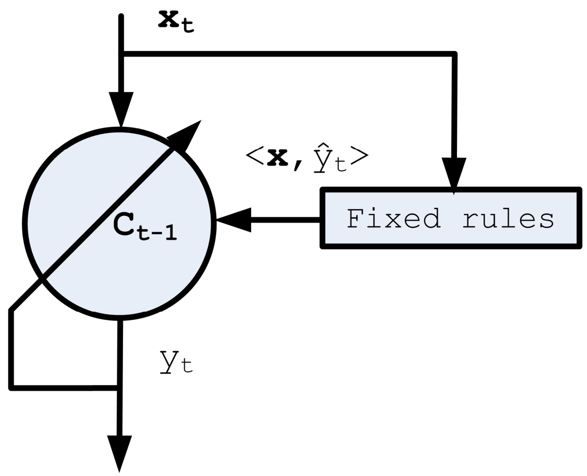


Basic Components

- Grid of classifier

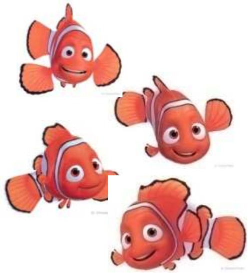


- Use a fix update rule
  - Pos Update: patch
  - Negative Update: “Any Image”



[Grabner et al.,VISAPP, 2007, PETS 2007]

Grid-based Object Detector

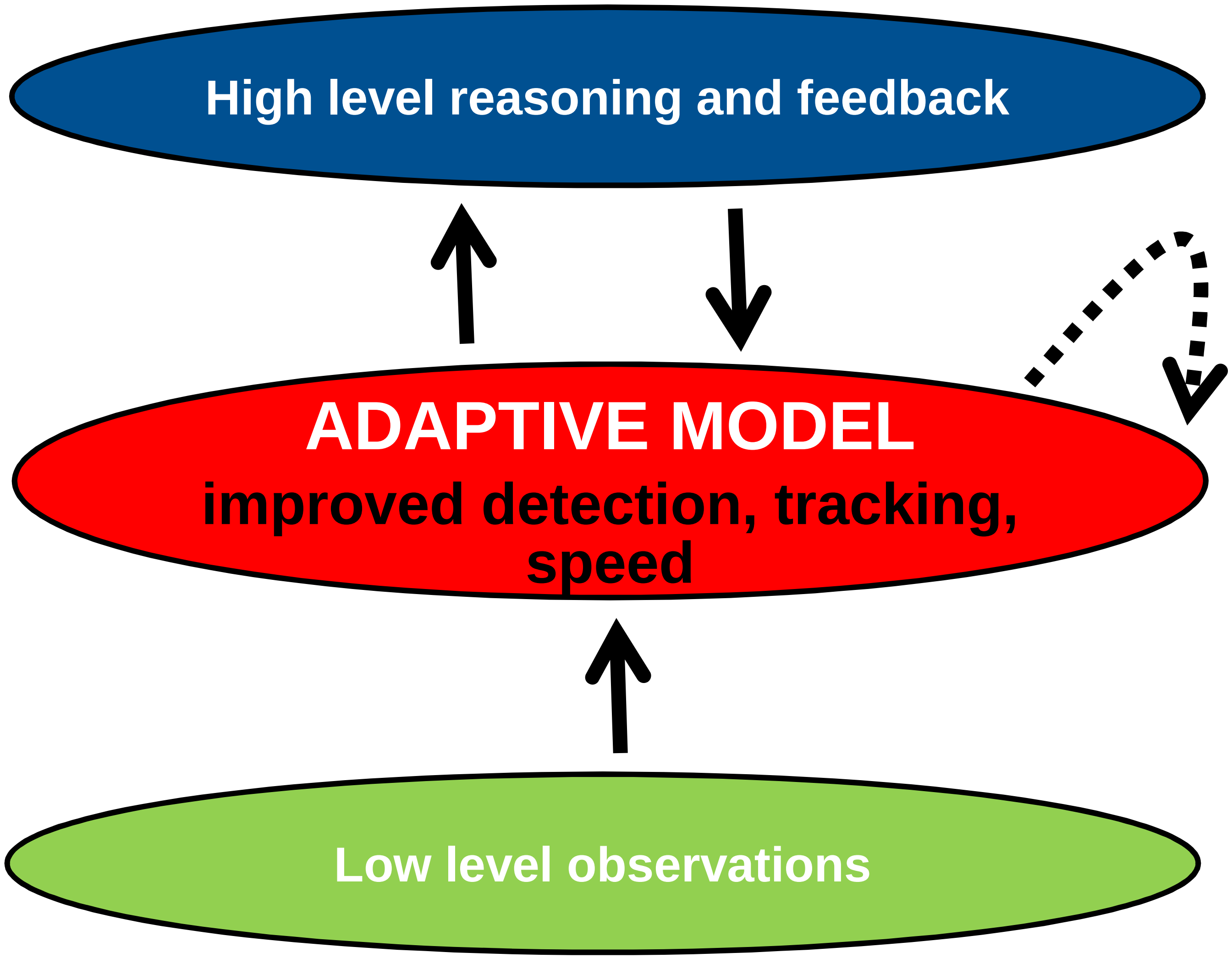


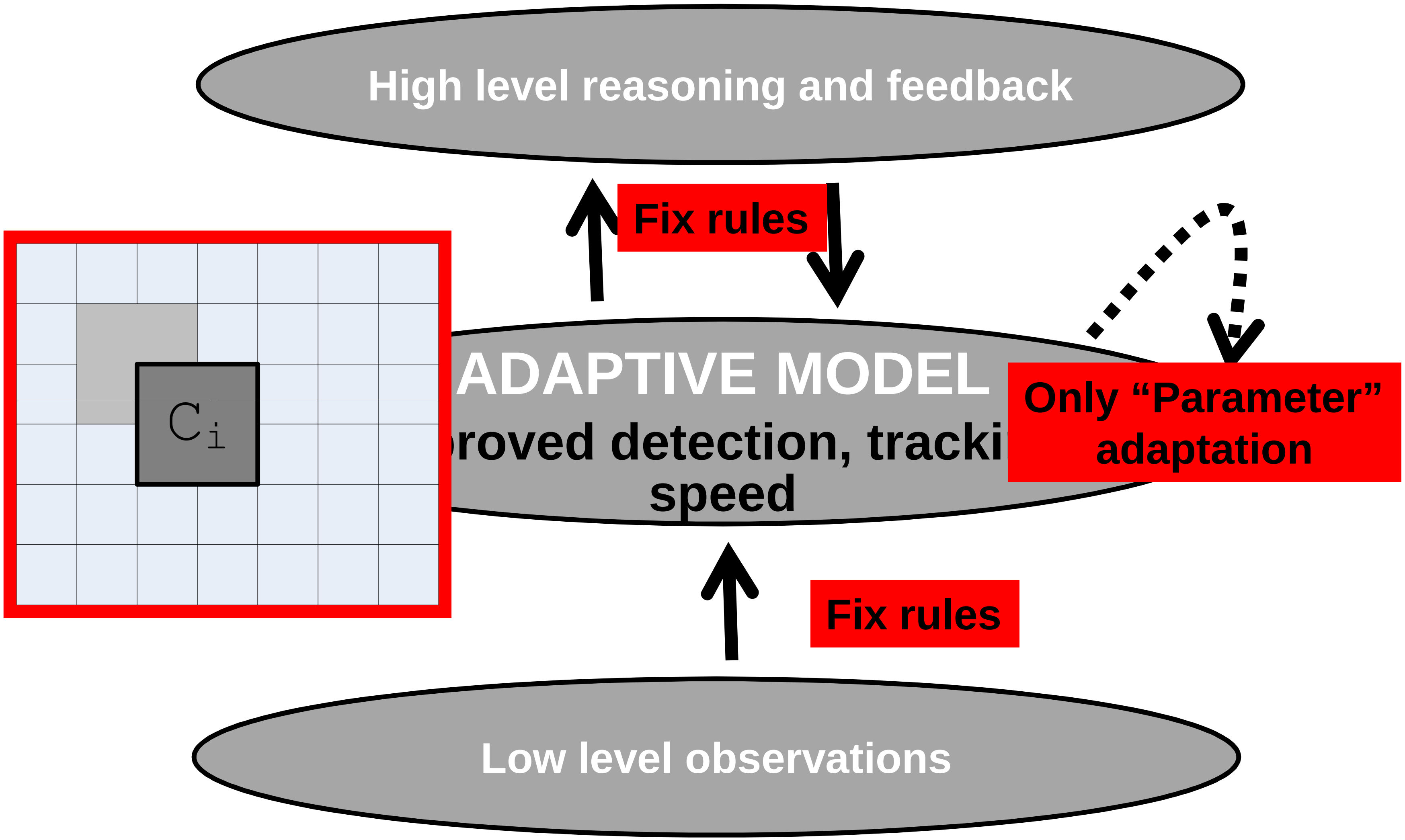
Object
vs.
Background

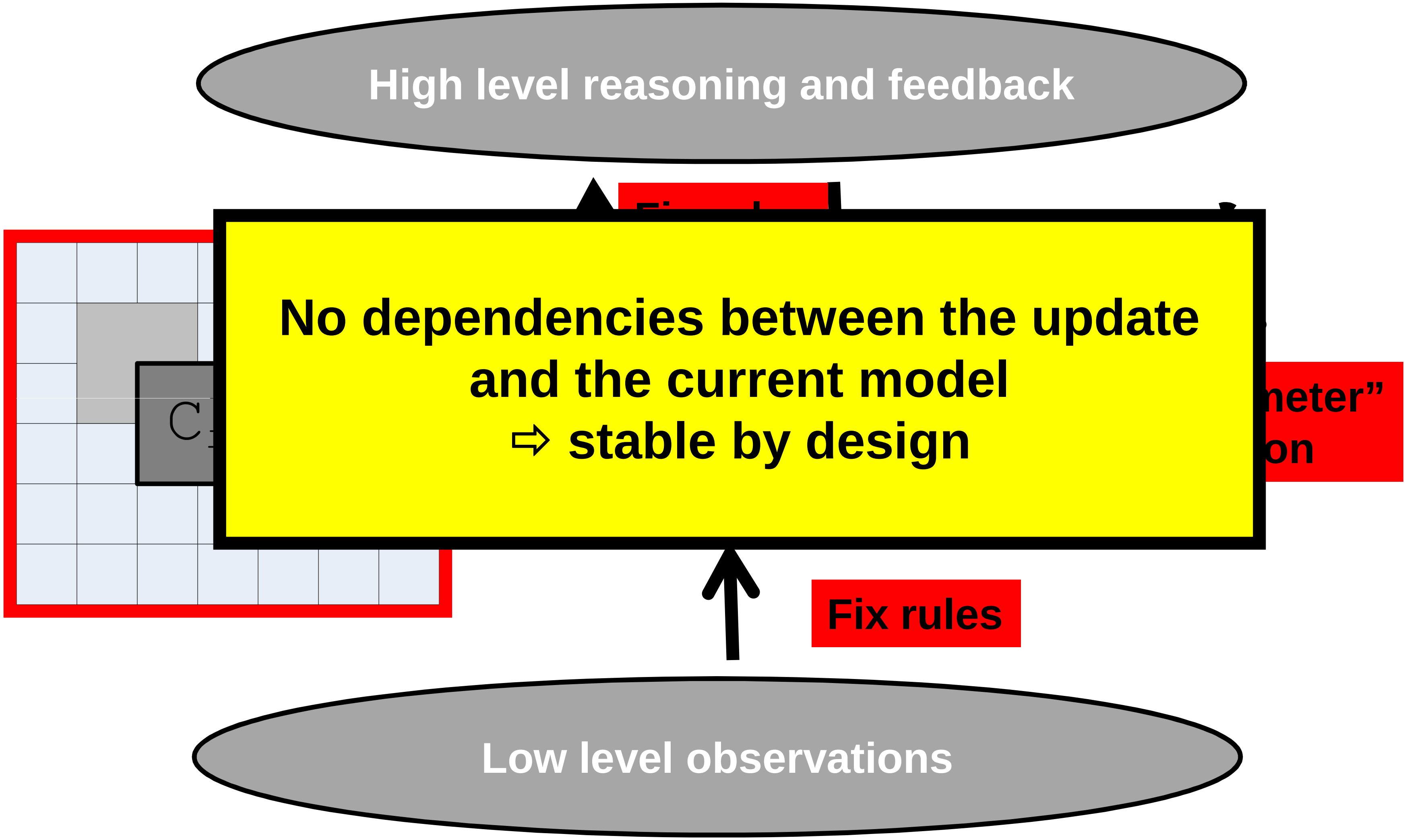
Object Detection

Background Modeling

Current Frame
vs.
Image Statistics

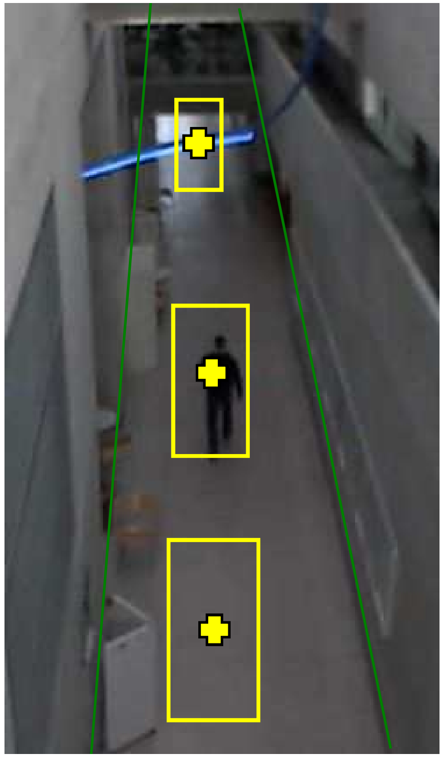

High level reasoning and feedback
ADAPTIVE MODEL
improved detection, tracking, speed
Low level observations

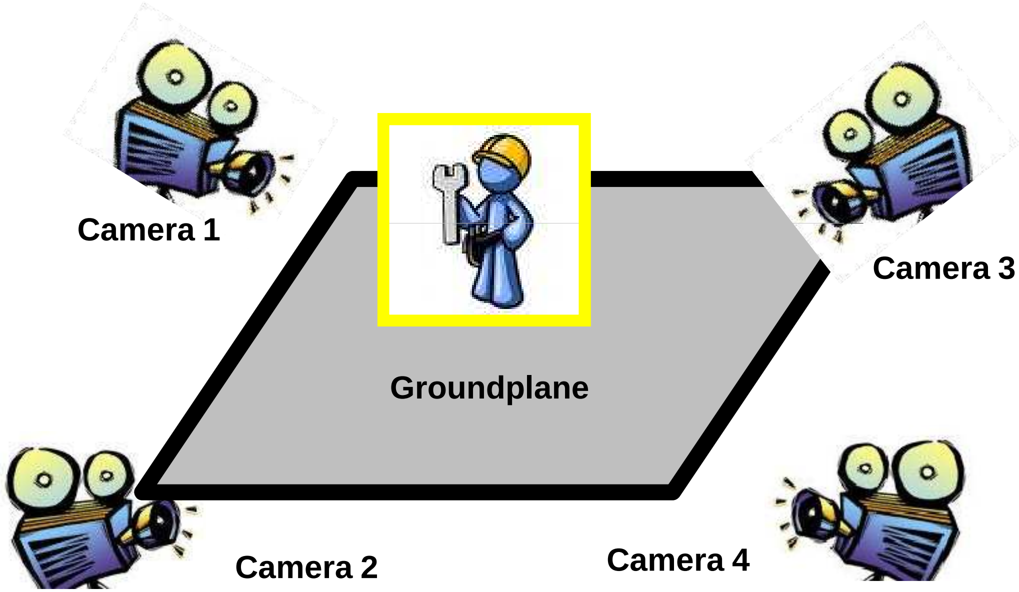

High level reasoning and feedback
Fix rules
ADAPTIVE MODEL
roved detection, tracki
speed
Only "Parameter"
adaptation
Fix rules
Low level observations
C_i


High level reasoning and feedback
No dependencies between the update and the current model
⇨ stable by design
Fix rules
Low level observations

High Level Knowlede

- Including prior knowledge, e.g., Scene Calibration…





PART III

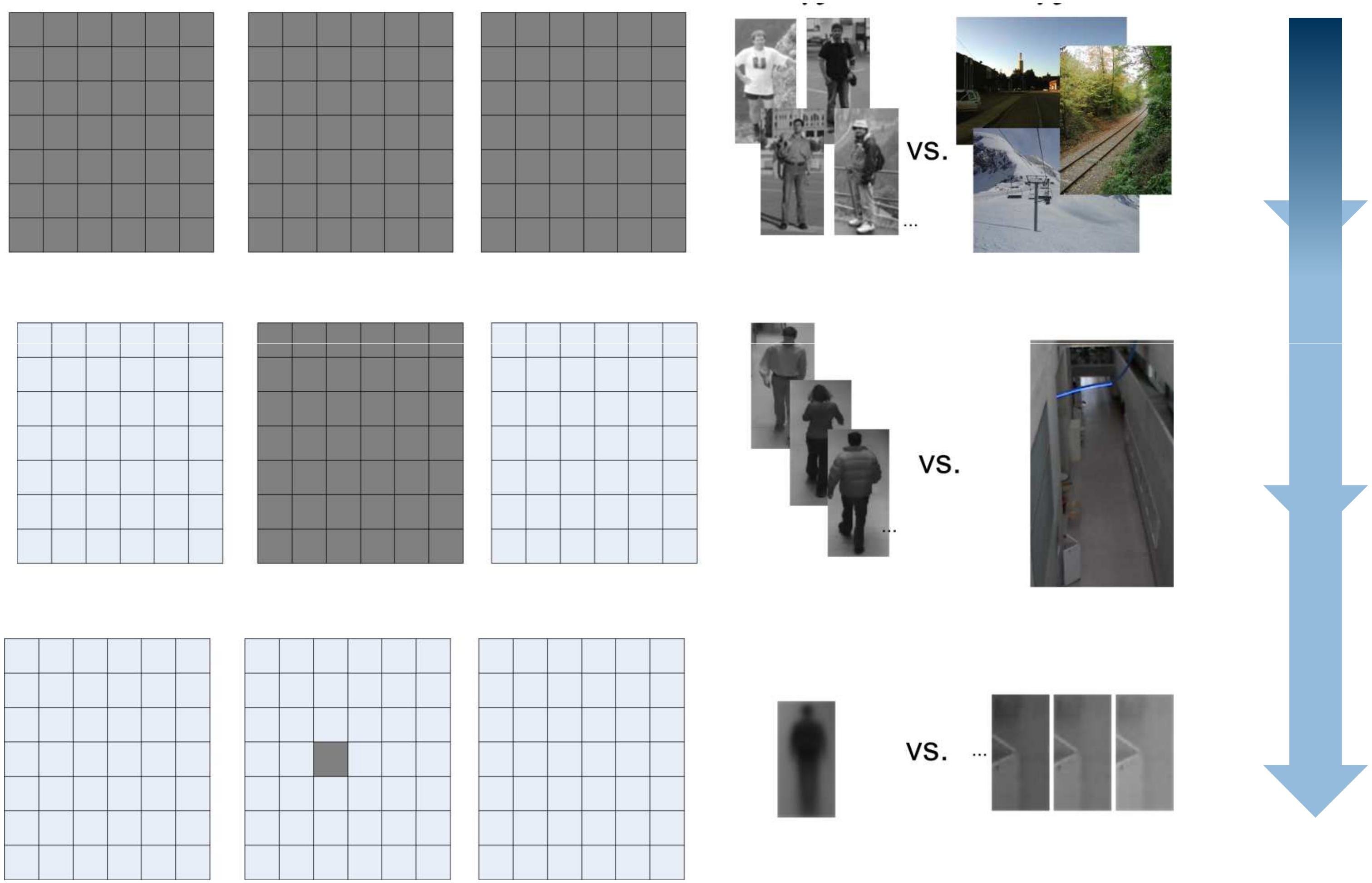
Using the Classifier Grid

Classifier should be applicable | Training set | Complexity

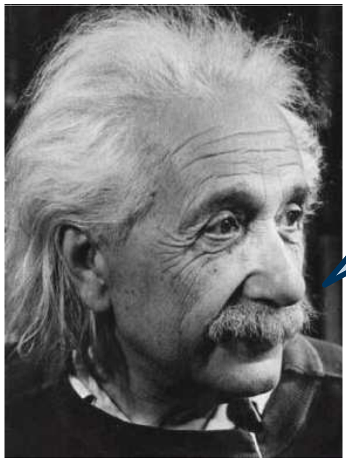
Fixed Detector

Scene specific

Grid based



Everything should be made as simple as possible, but not simpler.



Albert Einstein

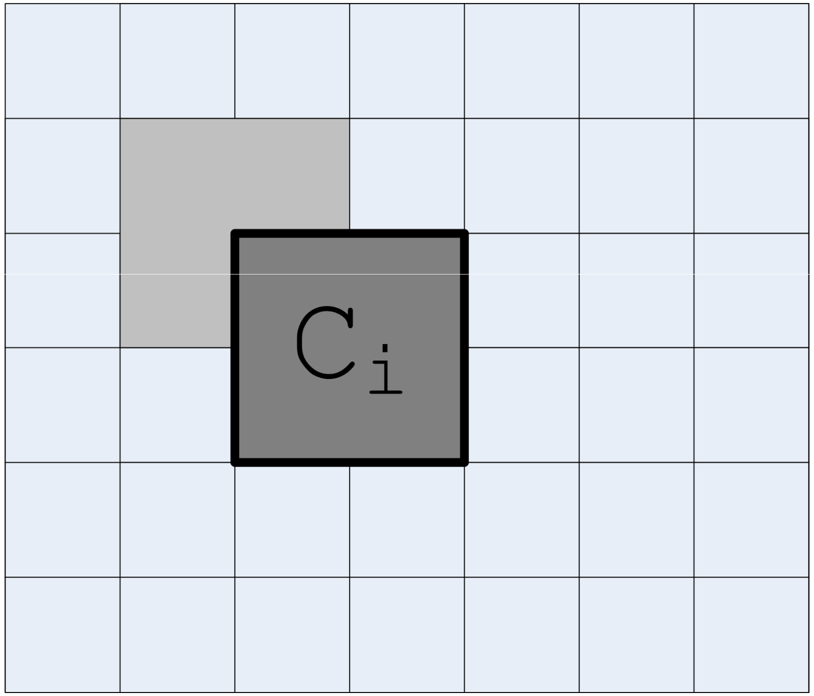
System 2007

“Is Pedestrain Detecion Really a Hard Task?”

Grabner, Roth, Bischof, PETS 2007

Classifier Grid

- One On-line classifier for each grid element



Simple Problem in time and space.

Fixed Update Rules for C_i

- 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

- **Any** On-line learning algorithm
  - Must cope with some (low) label noise
  - Fading memory (forgetting)
  - Good generalization (e.g. maximum margin classifier)

- Updates
  - Only **ONE (mean person)** pos. path was used for the experiments

Result and Comparison

[Dalal and Triggs, CVPR 2005.]

[Hoiem, Efros and Hebert, CVPR 2006]

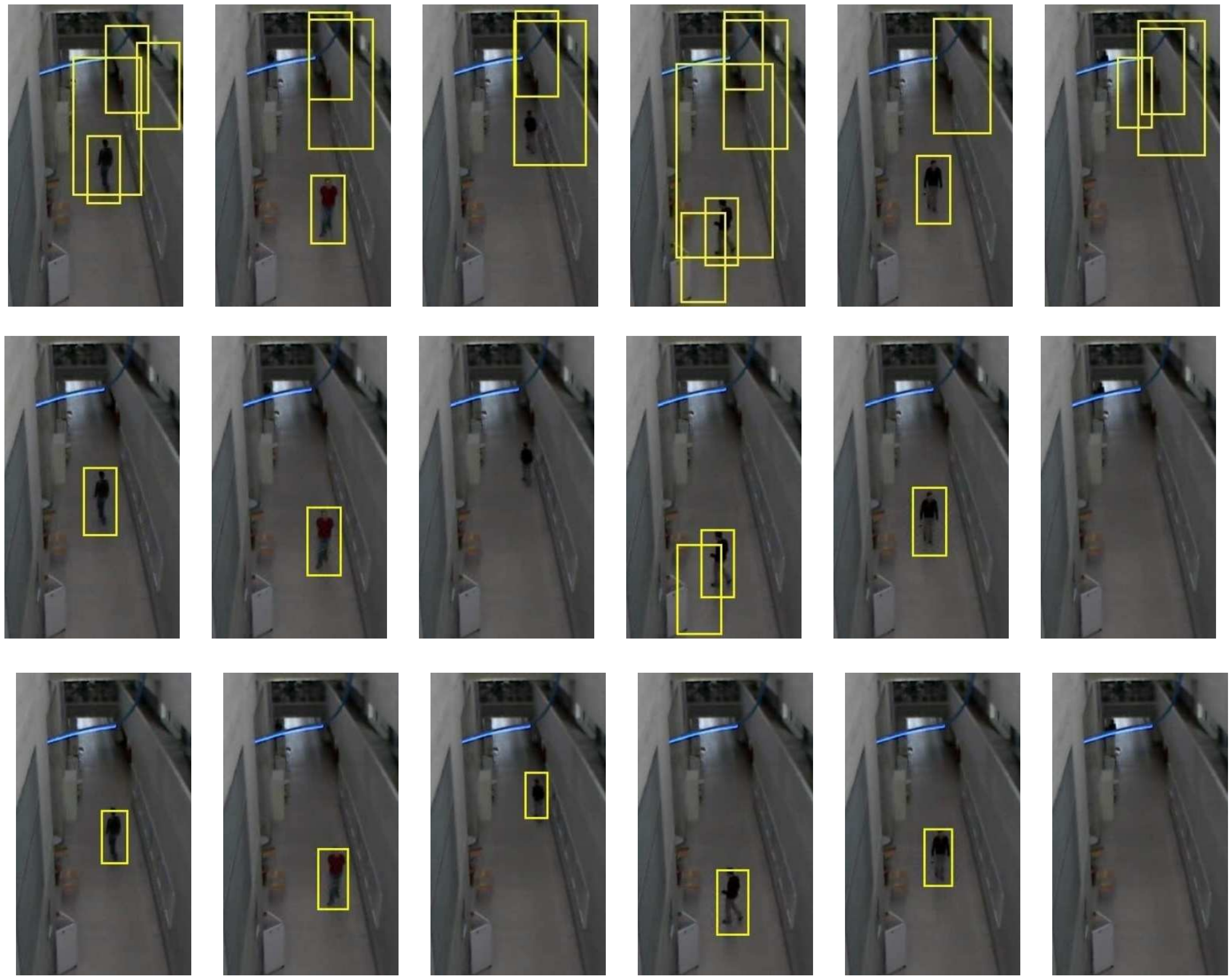
This approach

Result: Qualitative Comparisons

Static object detector

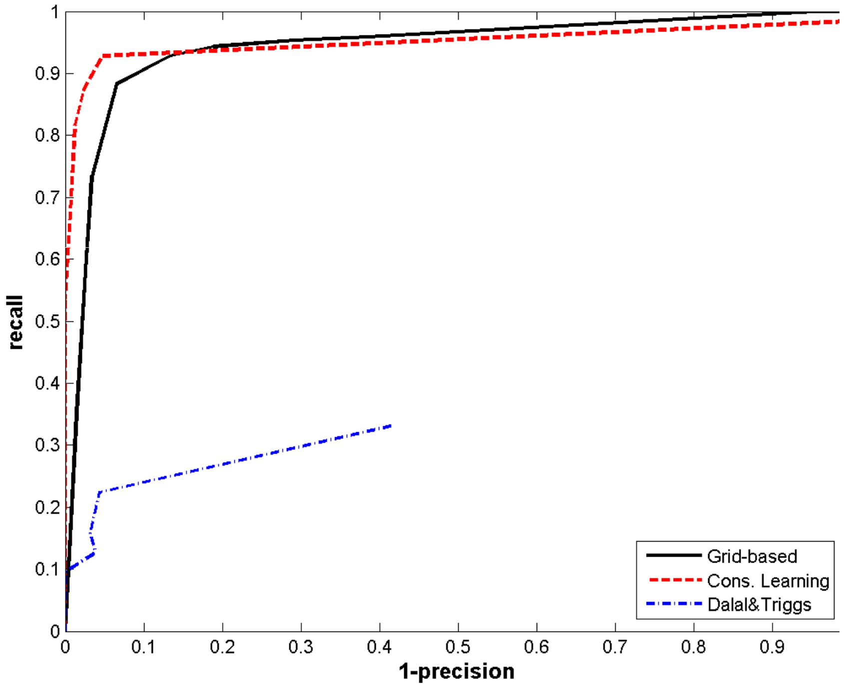
Scene Specific

Adaptive object detector

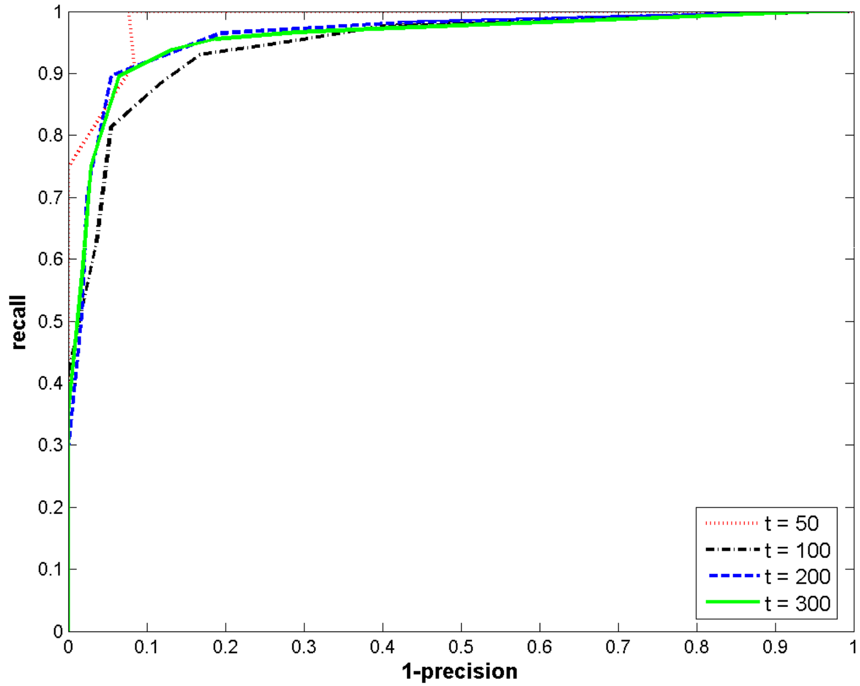


Toy Example Results

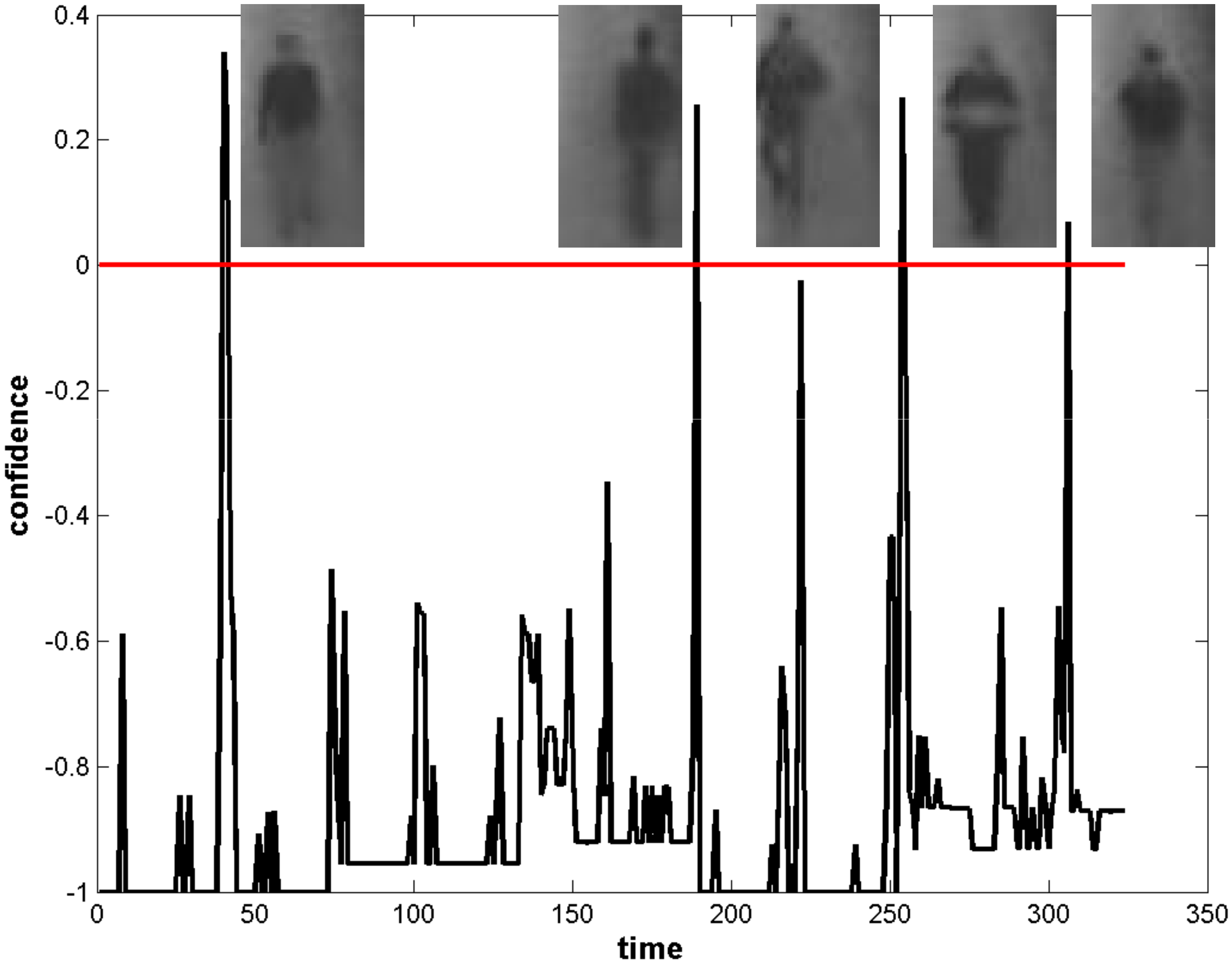
ROC



Convergence Speed



Confidence of a Patch over Time



I‘m NOT a simple Background Model

Extension: System 2009

“Classifier Grids for Robust Adaptive Object Detection”,

Roth, Sternig, Grabner, Bischof, CVPR 2009

Fixed Update Rules

- Positive updates
  - From a fix set

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

  - Correct by definition

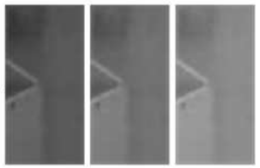
- Negative updates
  - Current patch

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

  - Correct most of the time, wrong with

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

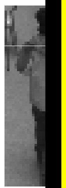
vs. ...



Fixed Update Rules

- Positive updates
  - From a fix set
- Negative updates
  - Current patch

$\langle \mathbf{x},$



ne,

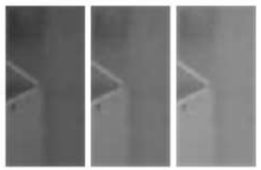
- Corre

$= \frac{\#p_i}{\Delta t}$

Deeper look into the underlying principles…

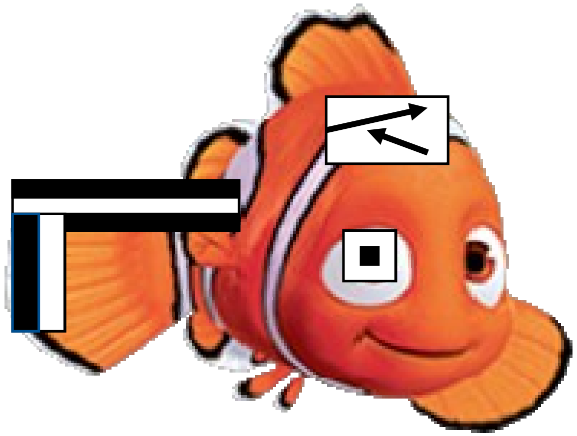
⇨ More Stable and robust.

vs. …



On-line Classifier

Combination of simple image features
using on-line Boosting as Feature Selection



$$\mathrm{sign}(\alpha_1 \cdot \quad + \alpha_2 \cdot \quad + \alpha_3 \cdot \quad + \ldots)$$

[Viola, Jones CVPR 2001]
[Grabner, Bischof, CVPR 2006]

Off-line boosting

Strong classifier

Weak classifier

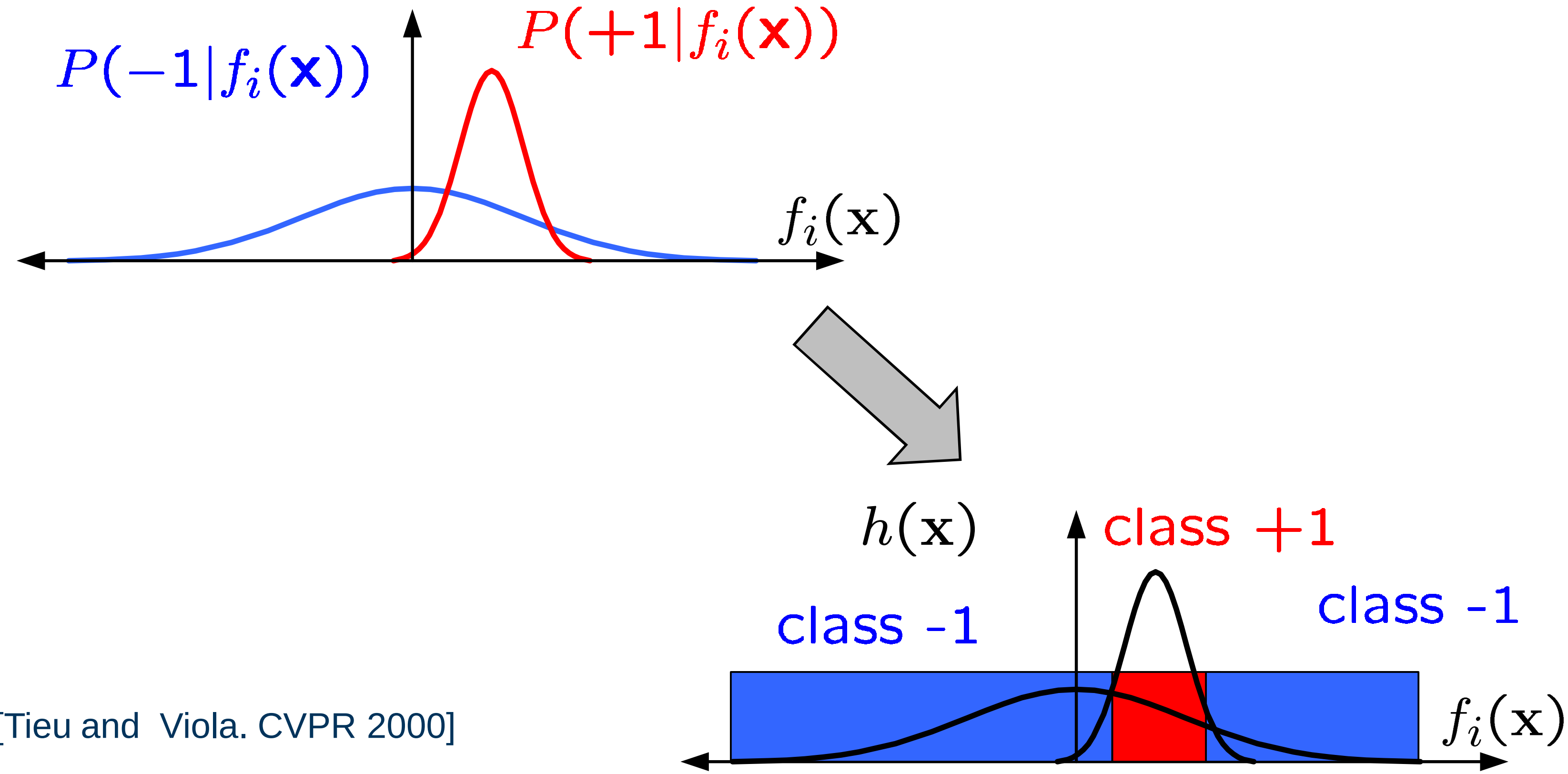
$$H(\mathbf{x}) = \operatorname{sign}\left(\sum_{n=1}^{N} \alpha_n \cdot h_n(\mathbf{x})\right)$$

Reweighting the training examples

[Freund and Schapire 1997, Oza and Russell, 2001]

Boosting for Feature Selection

Each feature corresponds to a weak classifier



[Tieu and Viola. CVPR 2000]

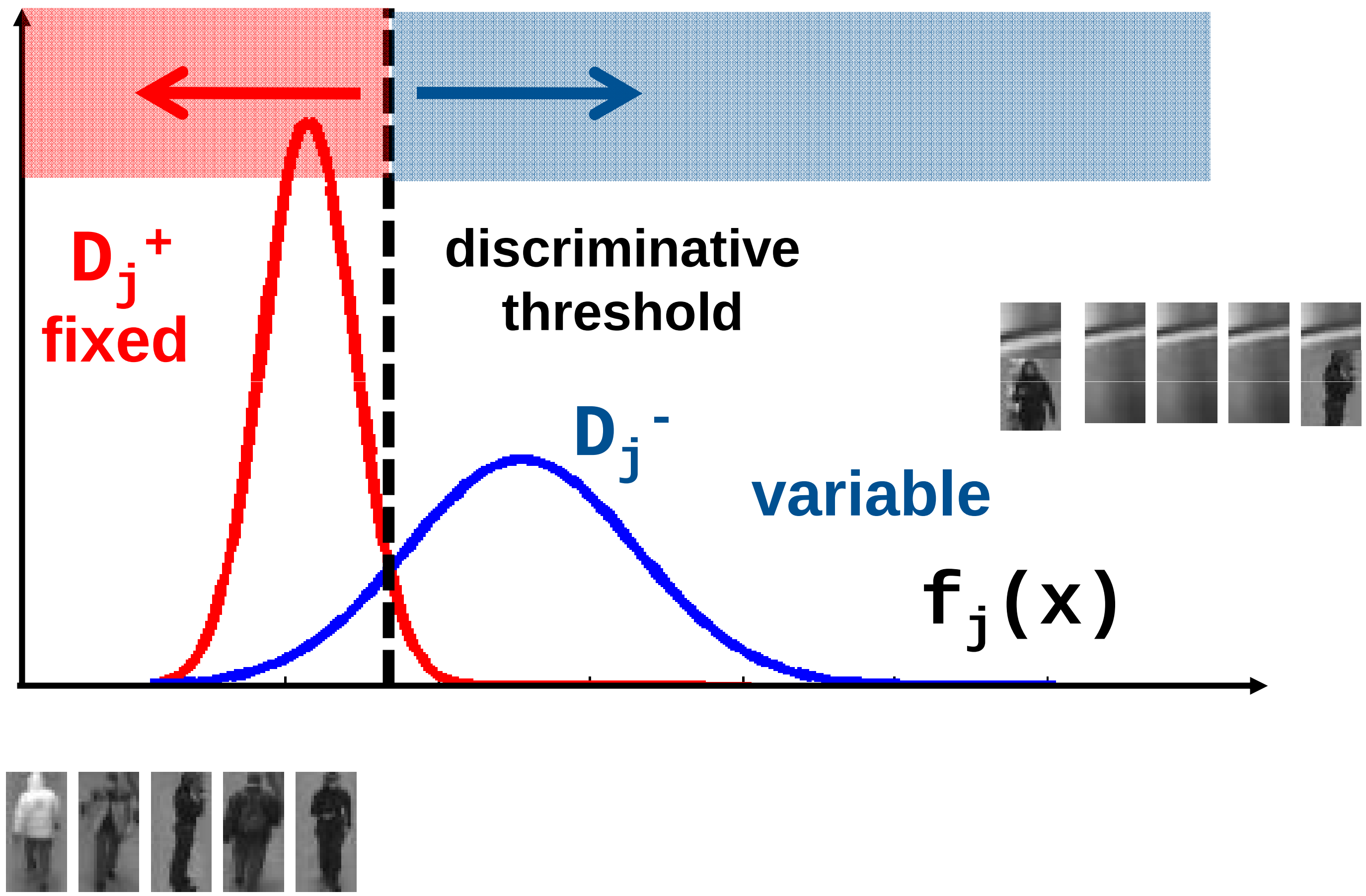
Principle: Modification of the Feature Selection Process

Generative model for Object of interest

Generative model for Background

Discriminate Classifier On Feature Level

Feature Selection



Summary

Method	Positive Updates	Negative Updates
General Object Detector	No	
Adaptive Detector	Some sort of supervision	
Background Model	Natural Image Statistics	Current Patch
Classifier Grid 2007	Predefined Positive Set	Current Patch
Now	**No (Precalculated Statistics)**	**Current Patch**

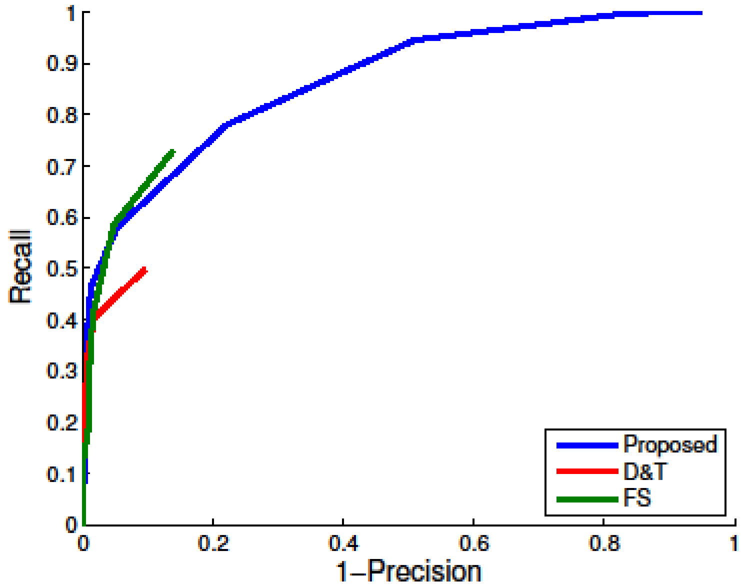
Positive Class specifies the object of Interest and is fix!
Adaptation is done only be negatives (Scene/location/time specific)

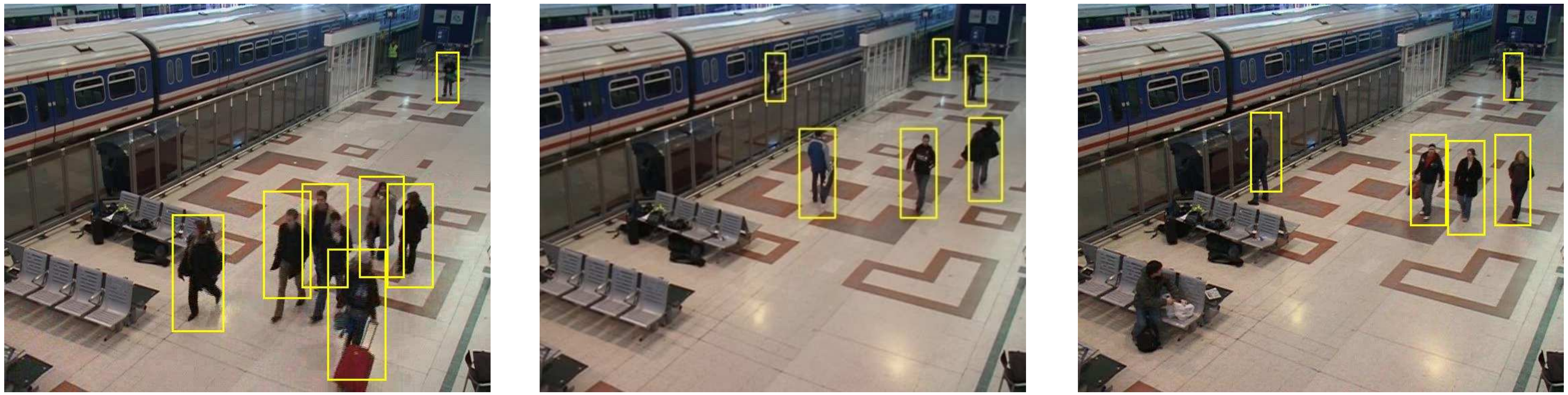
O.k., Imaging all the clever thinks you can do – we did them all.

Andrew Zisserman

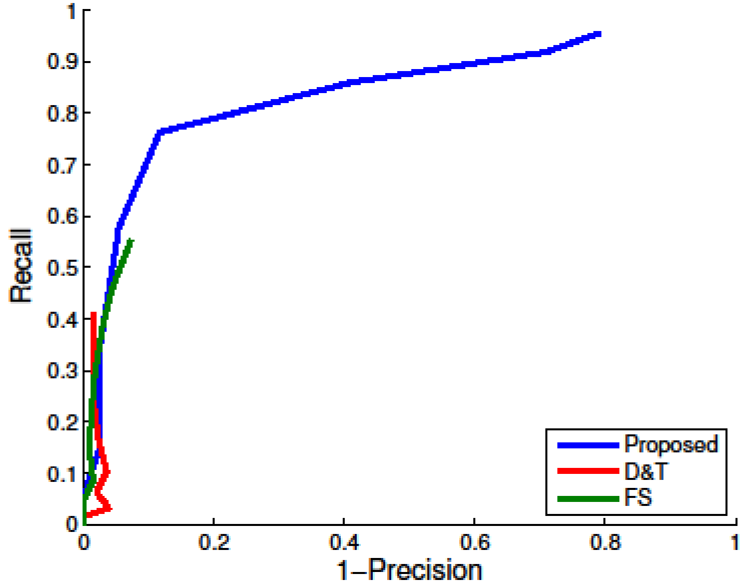
☺; AZ’s response to “a” question ask in our seminar.

Results: PETS 2006

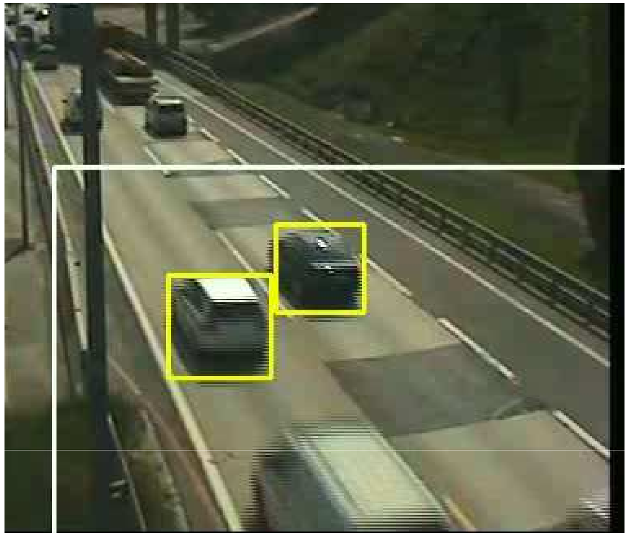


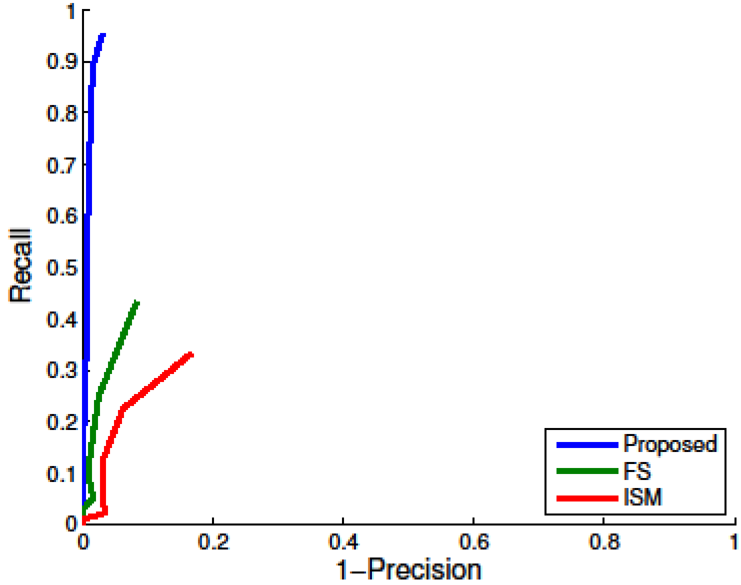


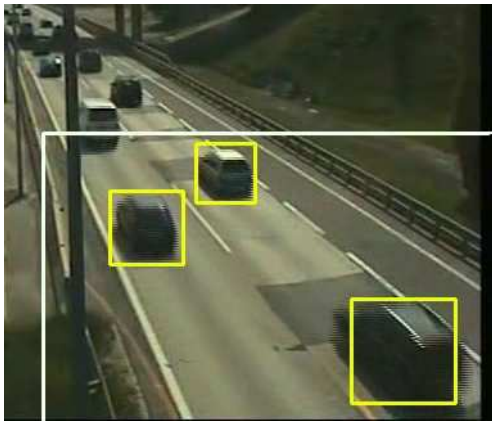
Results: Caviar

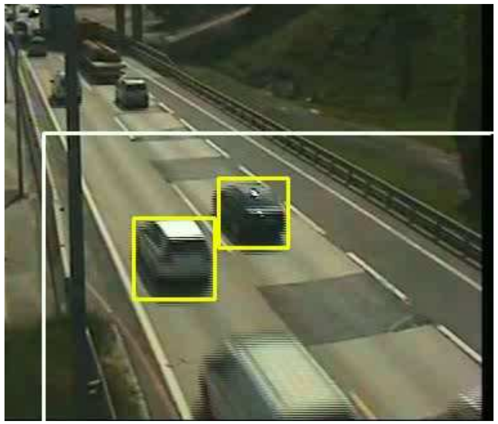


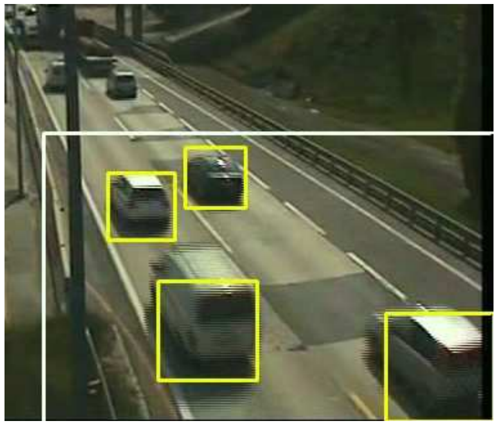
Results: Cars

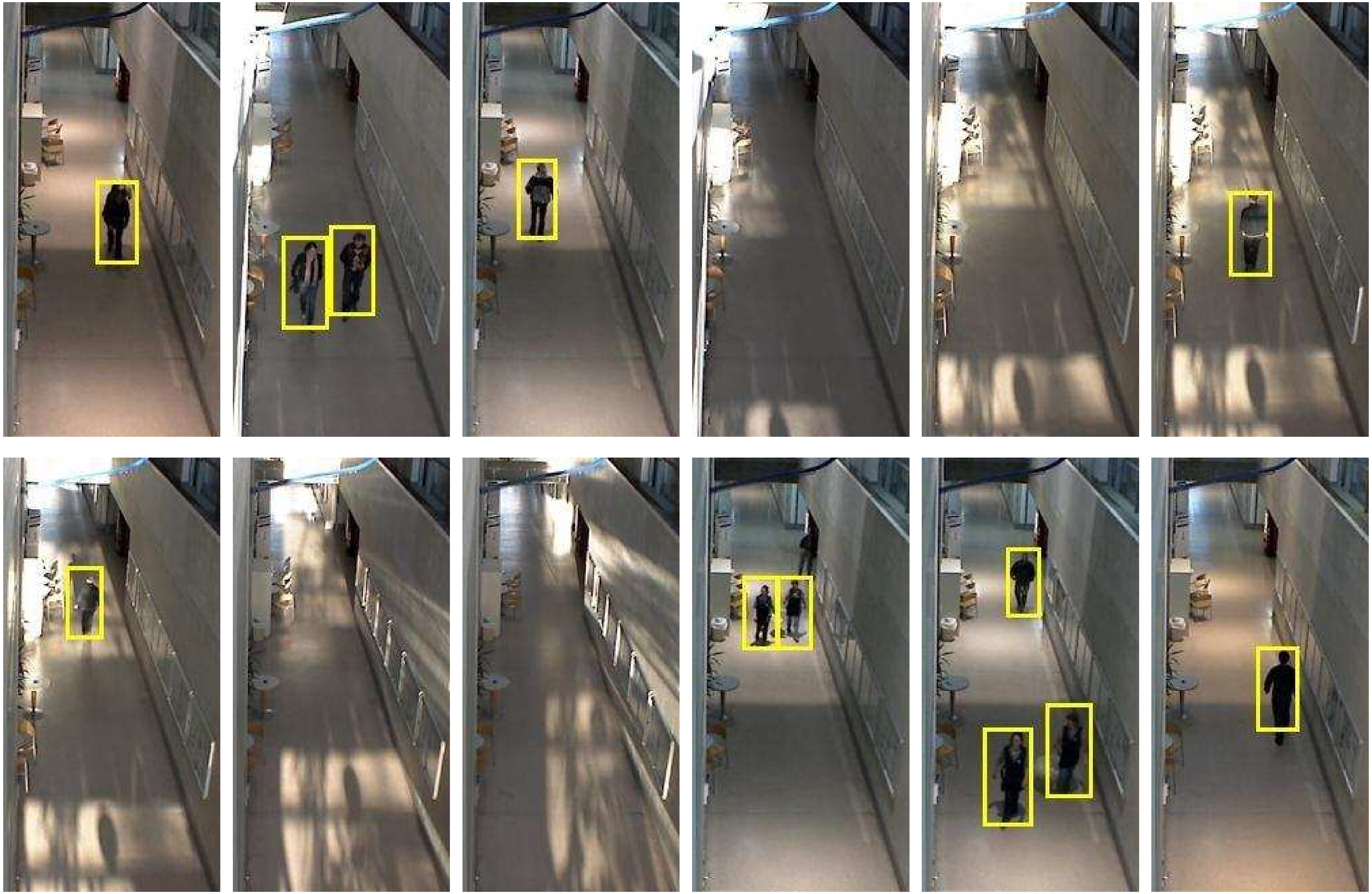


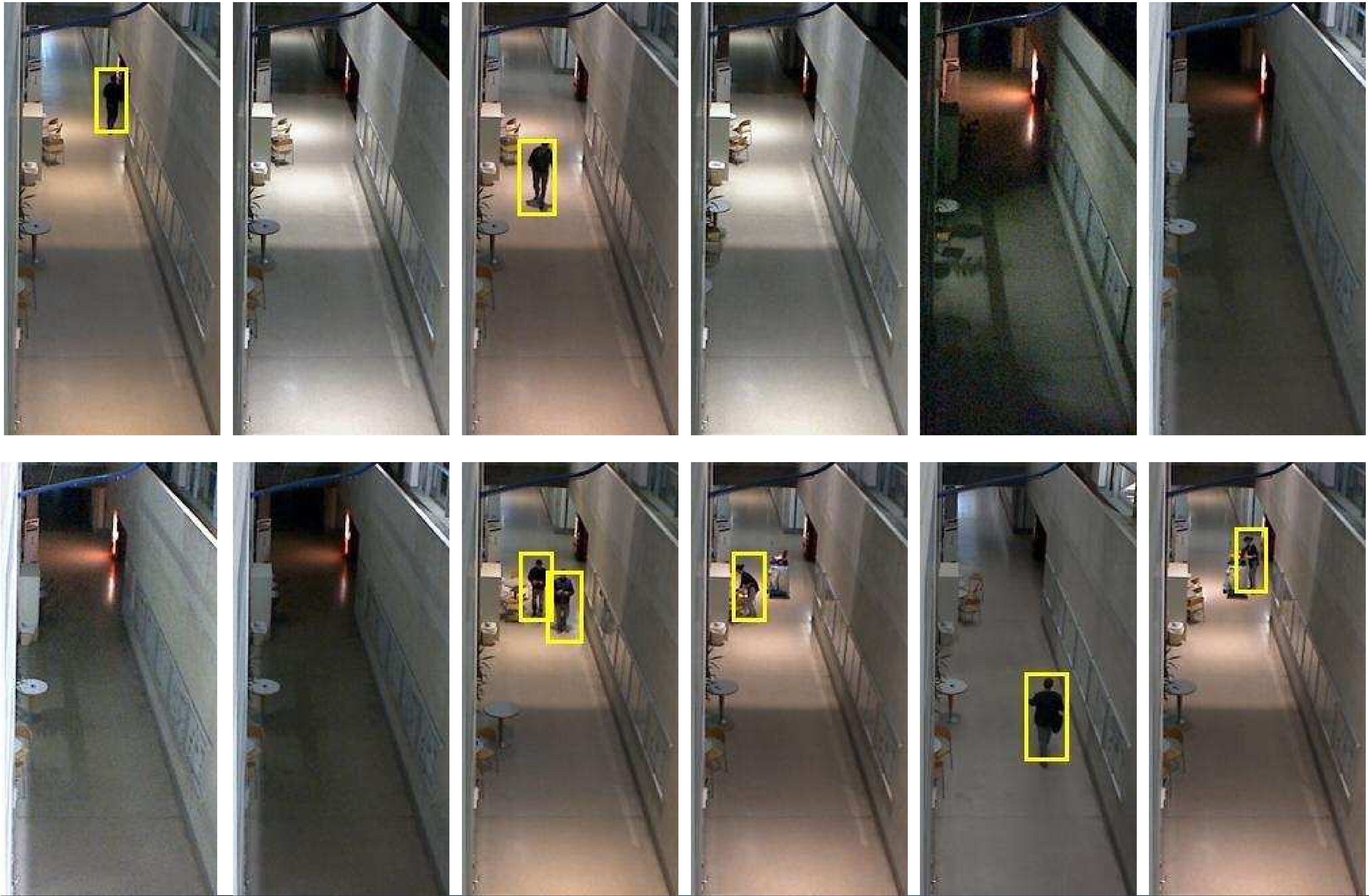












Long Term Experiment (several days)

NO DRIFT!

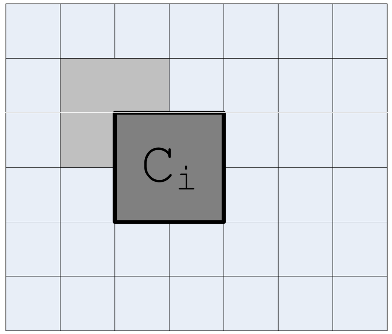
Object Detection Solved.

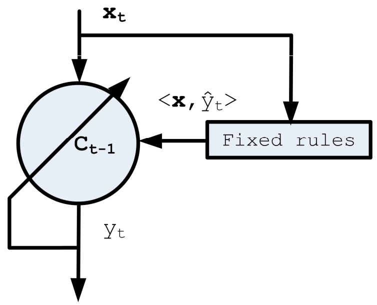
No.

Conclusion

- Focus on the problem

- Making things simple
  - Classifier Grid
  - Fix update rules





- ABC
  - Approaches, Benefits and Cares

Your choice…

Thank you for your attention

Comments and current work

Object Detection Solved.

No.

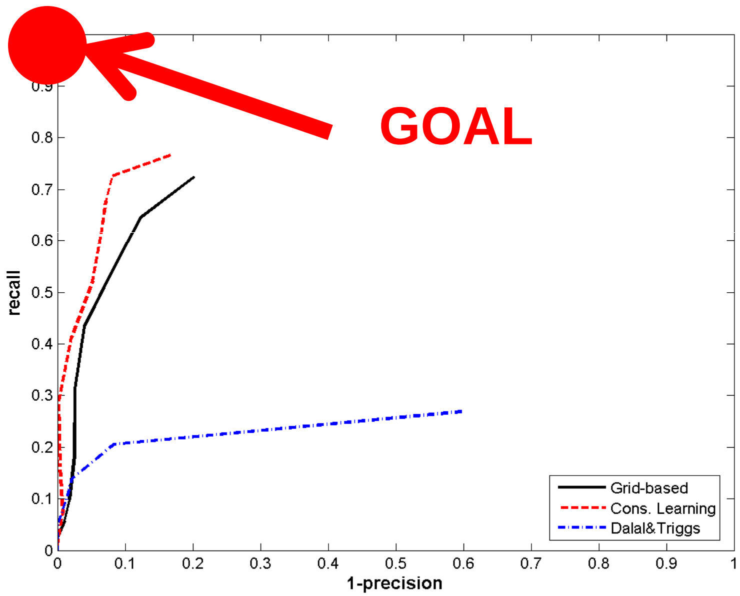
Object Detection Solved?

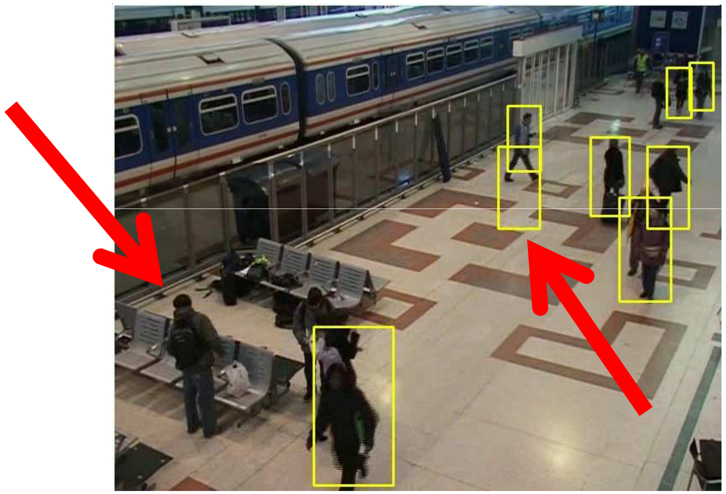
Does it fail?
If yes, when?

Are there limitations?
If yes, what are the main limitations?

Problems

- Hehe, there are still many misses and false detections!

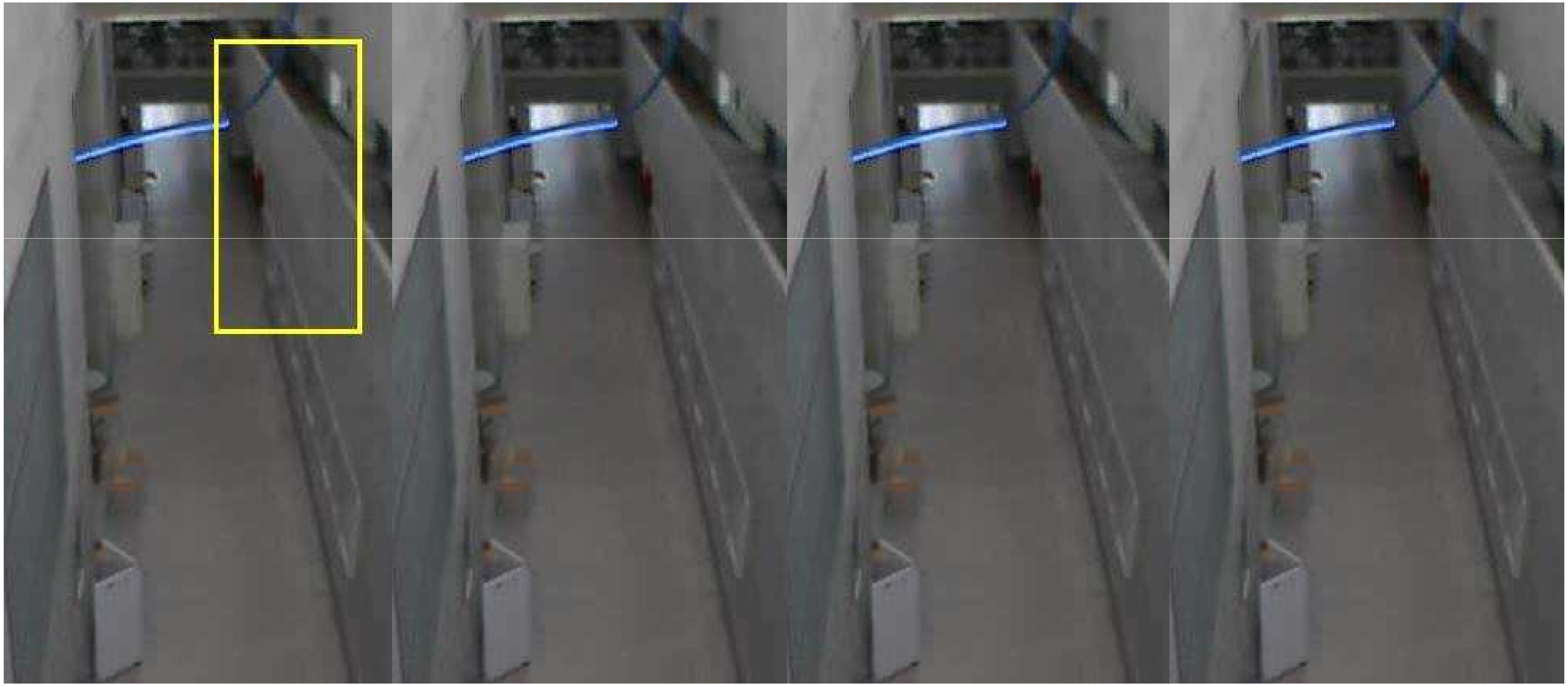




- Reasons: Occlusion, Context,…

Result and Comparison

[Dalal and Triggs, CVPR 2005.]	[Hoiem, Efros and Hebert, CVPR 2006]	**This approach**	Conservative learning (scene adaption)



$$P(y=1|\mathbf{x}) = \frac{p(\mathbf{X}|y=1)p(y=1)}{p(\mathbf{X})}$$

Unlabeled data

Thomas Bayes

Transfer learning, Semi-Supervised Learning

Summary

Method	Positive Updates	Negative Updates
General Object Detector	No	
Adaptive Detector	Some sort of supervision	
Background Model	Natural Image Statistics	Current Patch
Classifier Grid 2007	Predefined Positive Set	Current Patch
Classifier Grid 2009	No (Precalculated Statistics)	Current Patch
Now	**Using an Object Tracker**	**Background Image**

Really Focusing on the Specific Scene!

Coping with occlusions

Tracking, Detection and Recognition should be seen as ONE problem!

Any problems?

Acknowledgments

Luc van Gool

Severin Stalder

SCOVIS

EU-project SCOVIS under grant agreement no 216465.

Institute for Computer Graphics and Vision
Graz, University of Technology, Austria

Horst Bischof

Peter M. Roth

Sabine Sternig

In conjunction with ICCV 2009

3rd On-line Learning for Computer Vision Workshop 2009

Kyoto, Japan, October 3, 2009

CALL FOR PAPER

SUBMISSION DEADLINE: June 19, 2009

Organizer:
Fatih Porikli, MERL
Horst Bischof, TU-Graz
Helmut Grabner, ETHZ

Invited Speacker:
Pietro Perona, CALTECH

Program Committee:
Matt Brand, Tat-Jen C[illegible] Cetin, Rama [illegible] Davis, Ahmed El[illegible], Guoliang Fan, Riad Hamm[illegible]ud, Omar Javed Qiang Ji, Jiri Matas, Peter Meer, Nikunj Oza, Peter Roth, Venkatesh Saligrama, Stan Sclaroff, David Suter, Oncel Tuzel, Lior Wolf

We invite you to participate in the 3rd *On-line Learning for Computer Vision Workshop* (OLCV'09) which will be held in junction with ICCV [illegible] Kyoto, Japan. The workshop will bring together computer vision rese[illegible]ested in providing solid foundations to this promising and challe[illegible]

http://vision.ee.ethz.ch/olcv2009

Topics:
The topics of inte[illegible]ot limited to:

- On[illegible]ject detection and tracking,
- [illegible]g for object identification and recognition,
- [illegible]mental fusion of multi-modal data,
- Applications using on-line classification methods, and
- Theoretical characterizations.
- Work towards a solid framework for benchmarking on-line learning algorithms

Important Dates:

Submission of full papers	June 19, 2009
Notification of acceptance	July 20, 2009
Submission of camera ready papers	August 31, 2009
Workshop	October 3, 2009