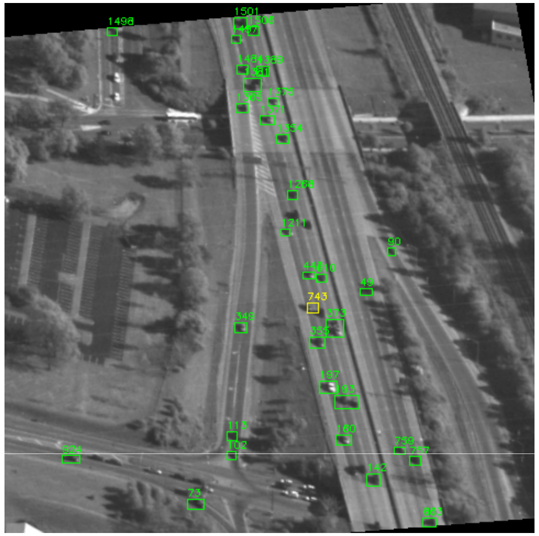
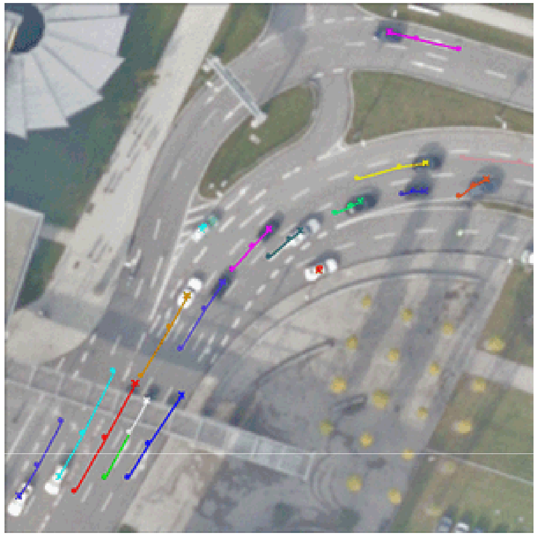
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

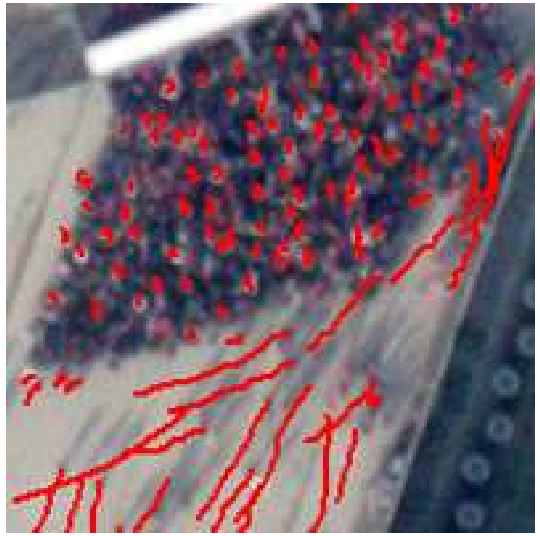
# Robust Tracking using Visual Constrains and Context

**Helmut Grabner**

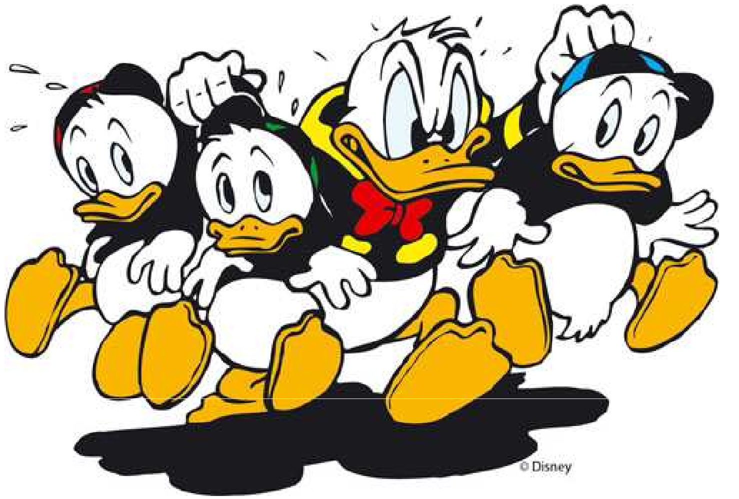
Ftan, 2011/07/19

# Put cards on the table: I will NOT talk about …

# BUT, I will talk about …

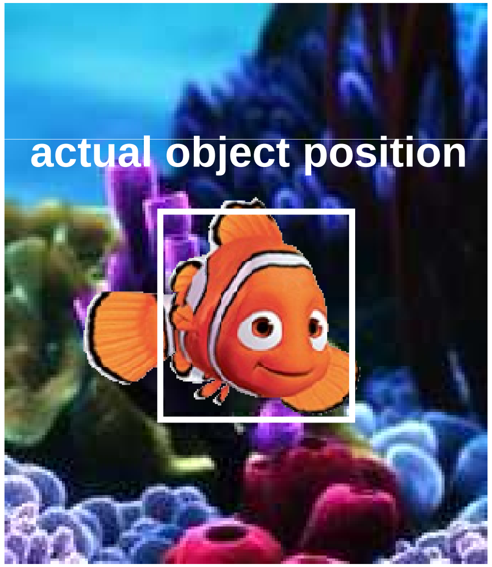


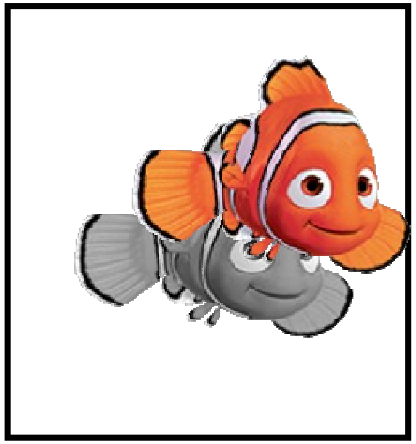

# We all know what tracking is, right?

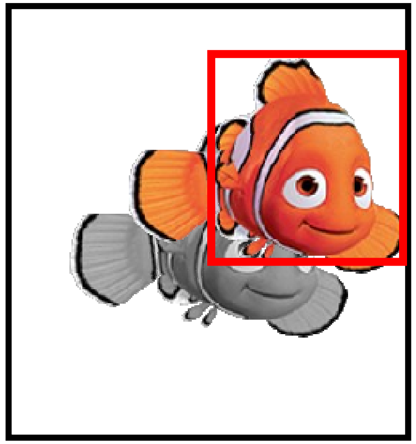
[Grabner et al., VideoProc CVPR'06]

# We all know what tracking is, right?

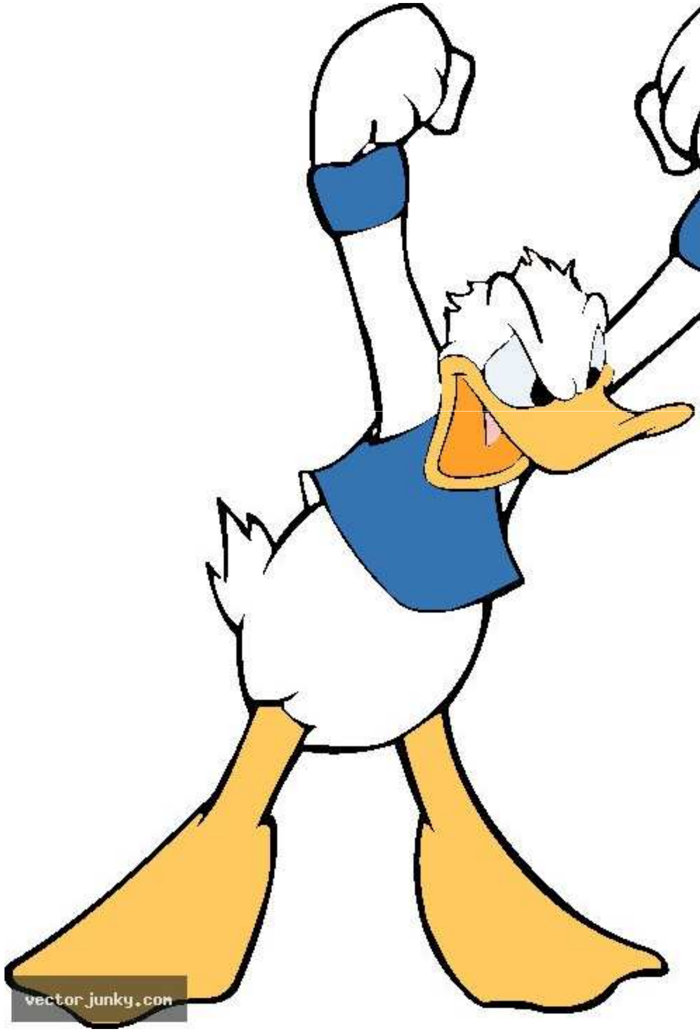
**Time t** | **from time t to t+1** | **„find“ again**



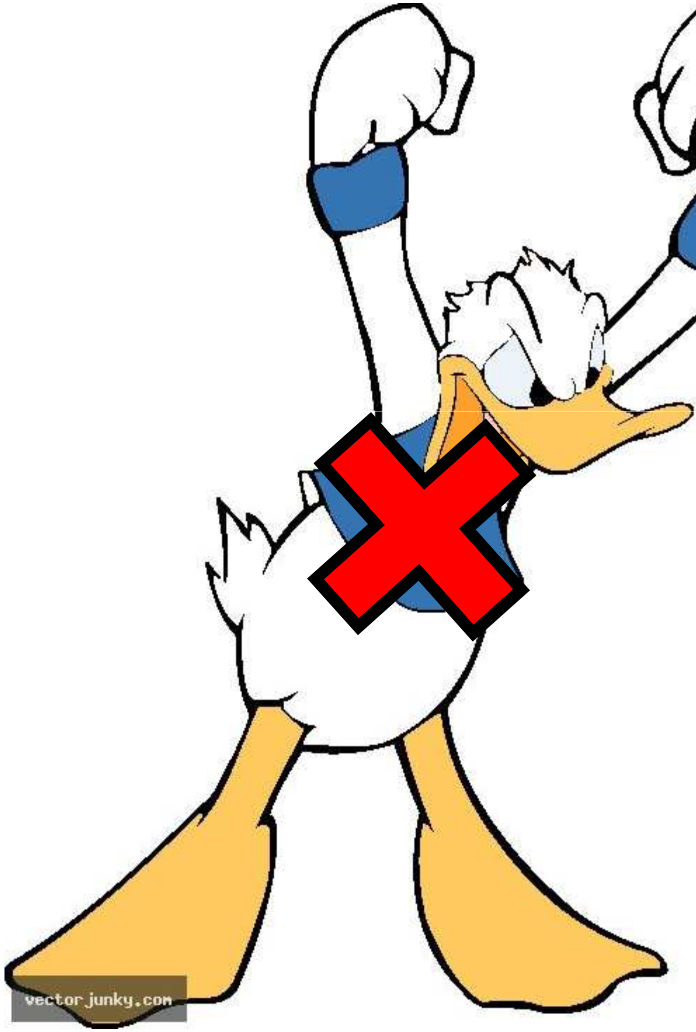





# What to track?

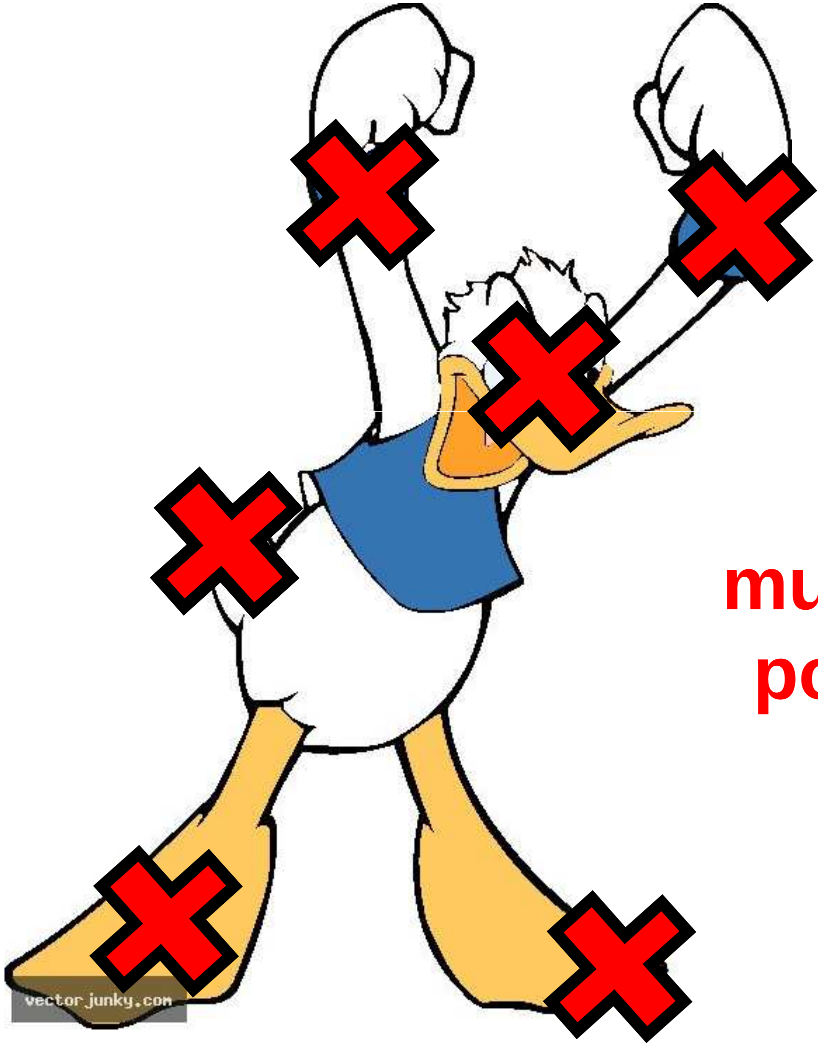


# What to track?




center point

# What to track?

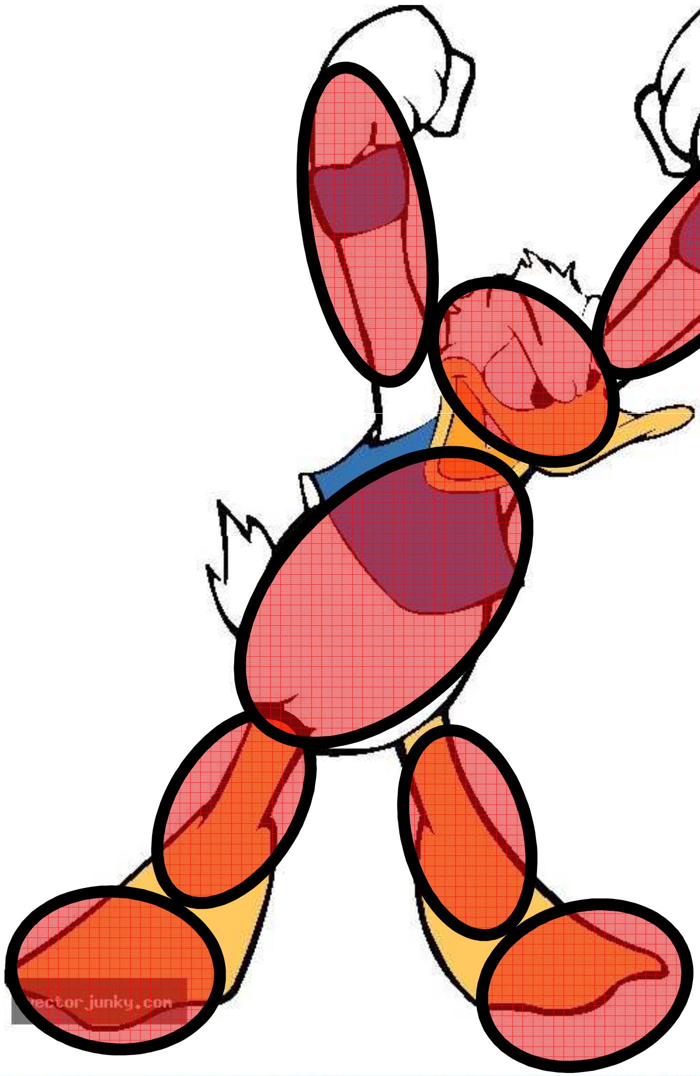



**multiple points**

# Example application: Structure From Motion

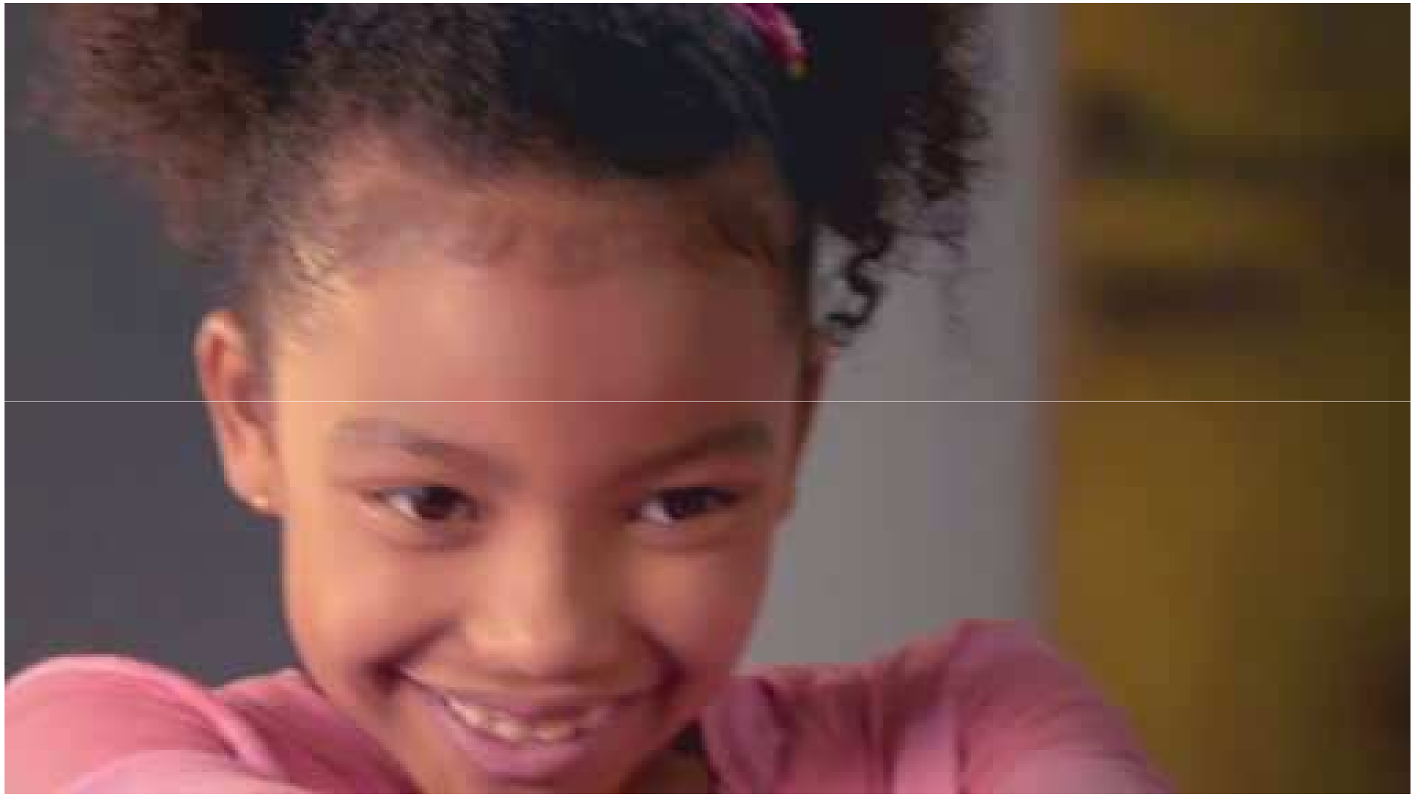
[Pollefeyes et al., IJCV 2004]

# What to track?




**(body) parts**

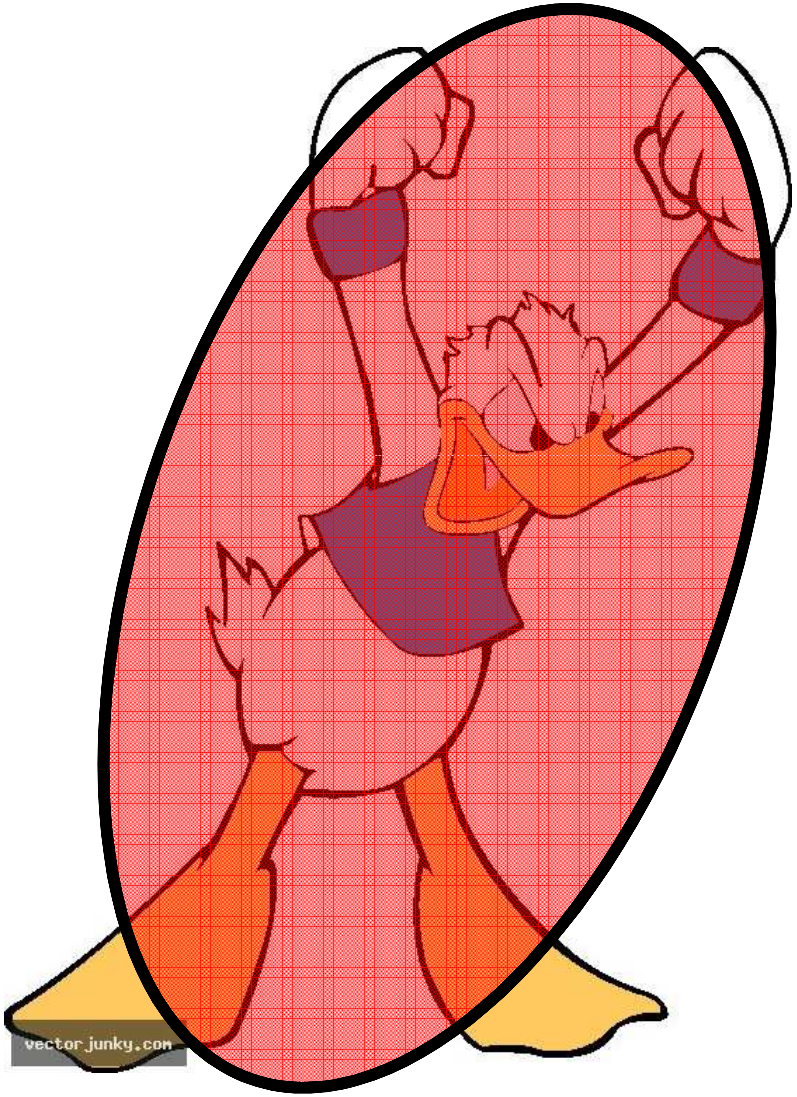
# Example application: Game interface



KINECT for XBOX 360

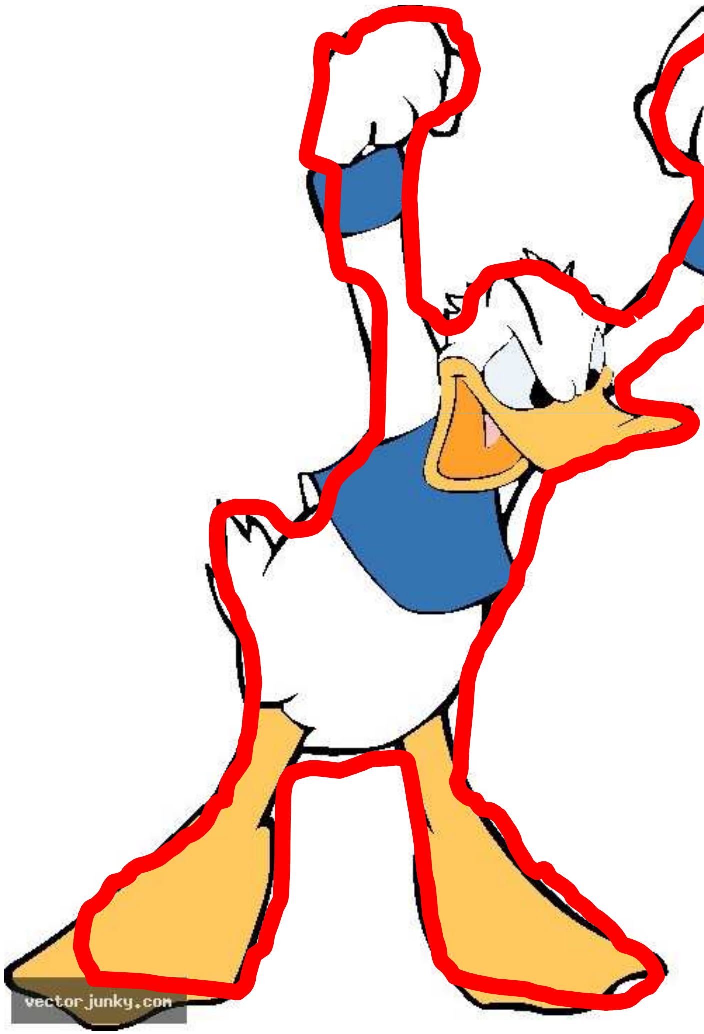
[Shotton et al., CVPR 2011]

# What to track?



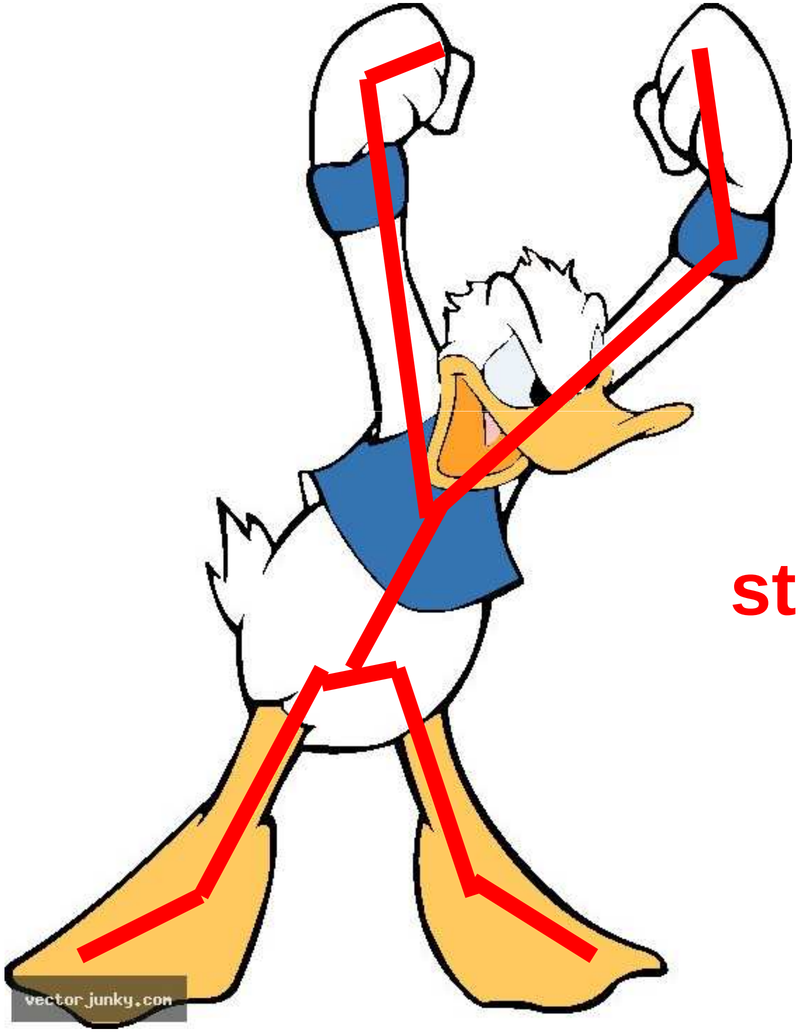

region

# What to track?




outline

# What to track?



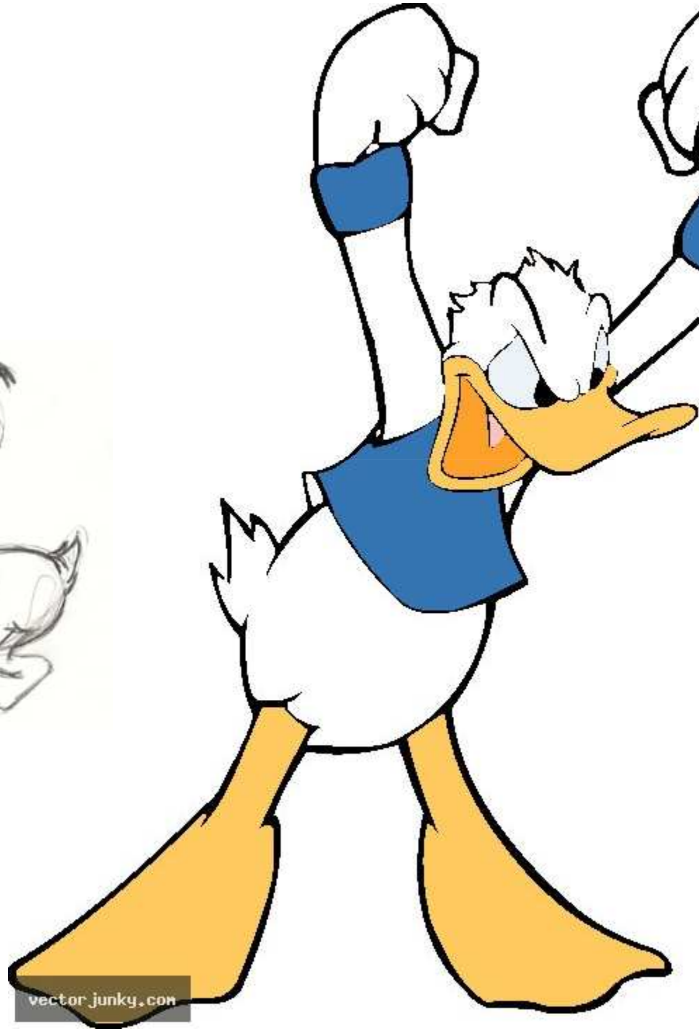

structure

# Example application: Augmented reality

[Wagner et al., ISMAR'09]

# In this talk…

# "OBJECT"



# Why is it hard?

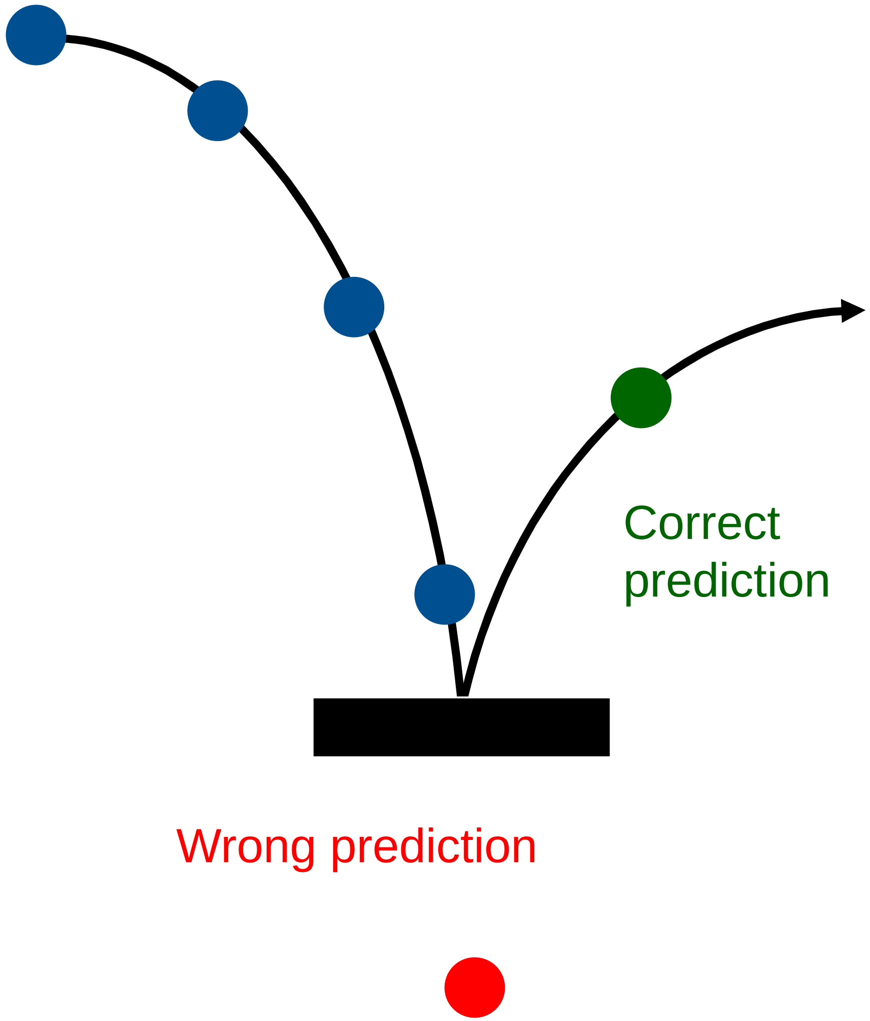
- **initialization**

# Why is it hard?

- initialization
- **fast movement**

# Why is it hard?

- initialization
- fast movement
- **nonlinear dynamics**



# Why is it hard?

- initialization
- fast movement
- nonlinear dynamics
- **multiple (similar) objects**

# Why is it hard?

- initialization
- fast movement
- nonlinear dynamics
- multiple (similar) objects
- **background clutter**

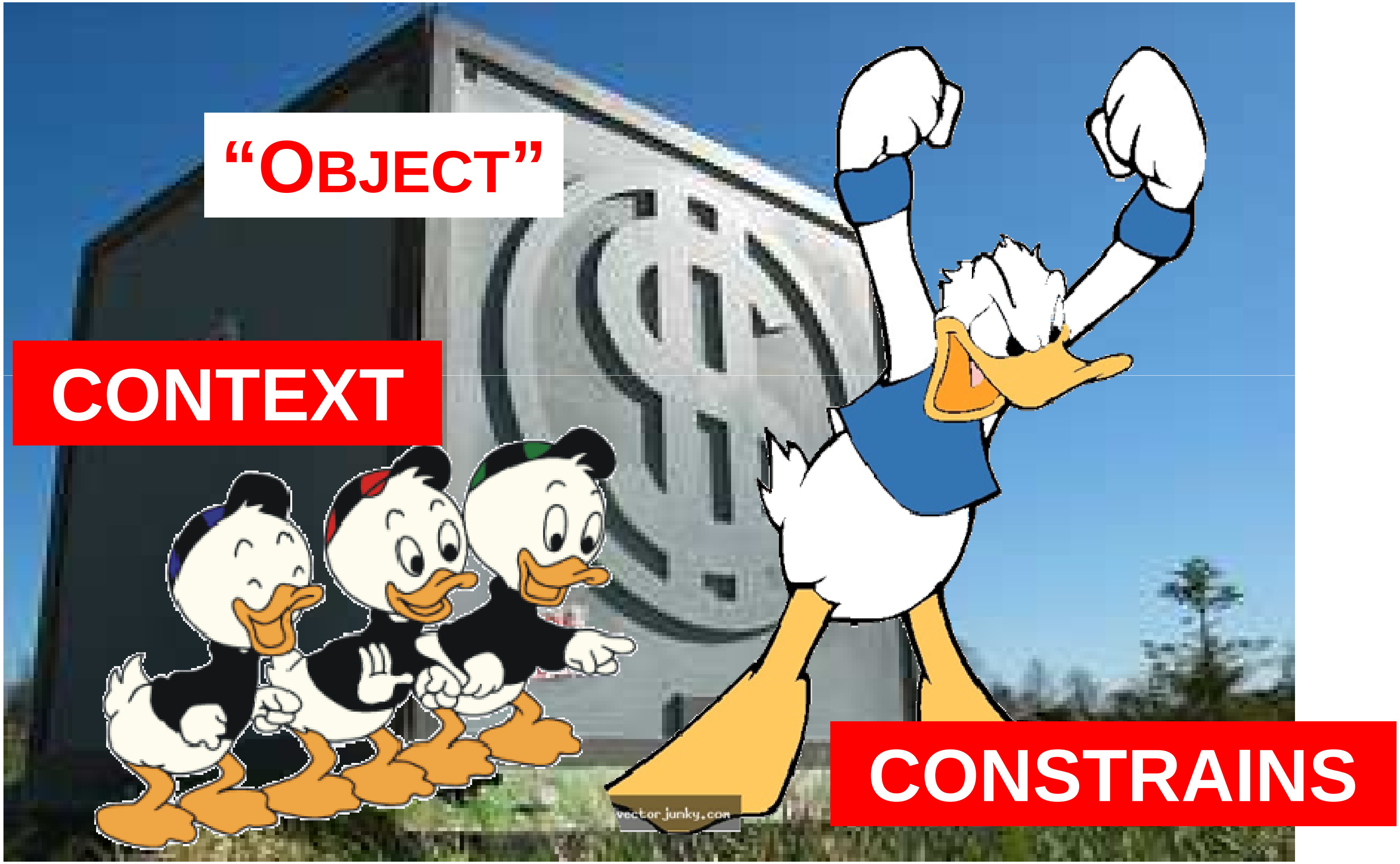
# Why is it hard?

- initialization
- fast movement
- nonlinear dynamics
- multiple (similar) objects
- background clutter
- **appearance changes (model drift)**

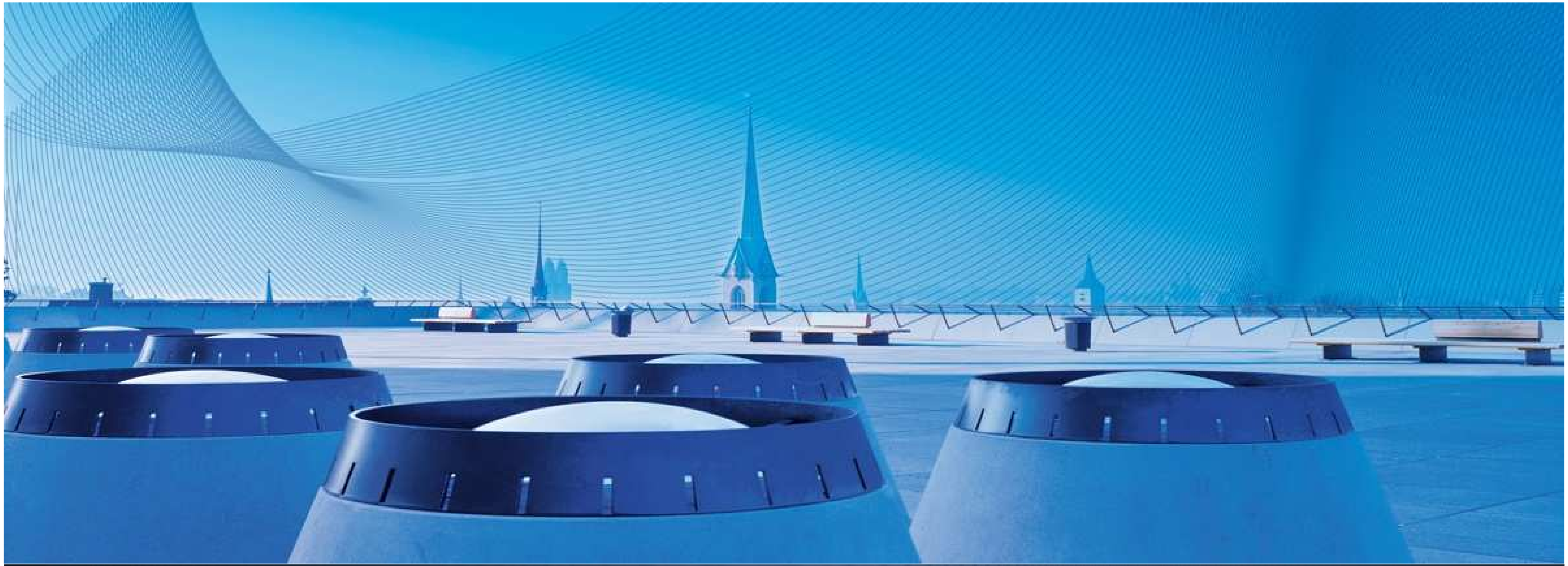
# Why is it hard?

- initialization
- fast movement
- nonlinear dynamics
- multiple (similar) objects
- background clutter
- appearance changes (model drift)
- **occlusions/ self occlusions**

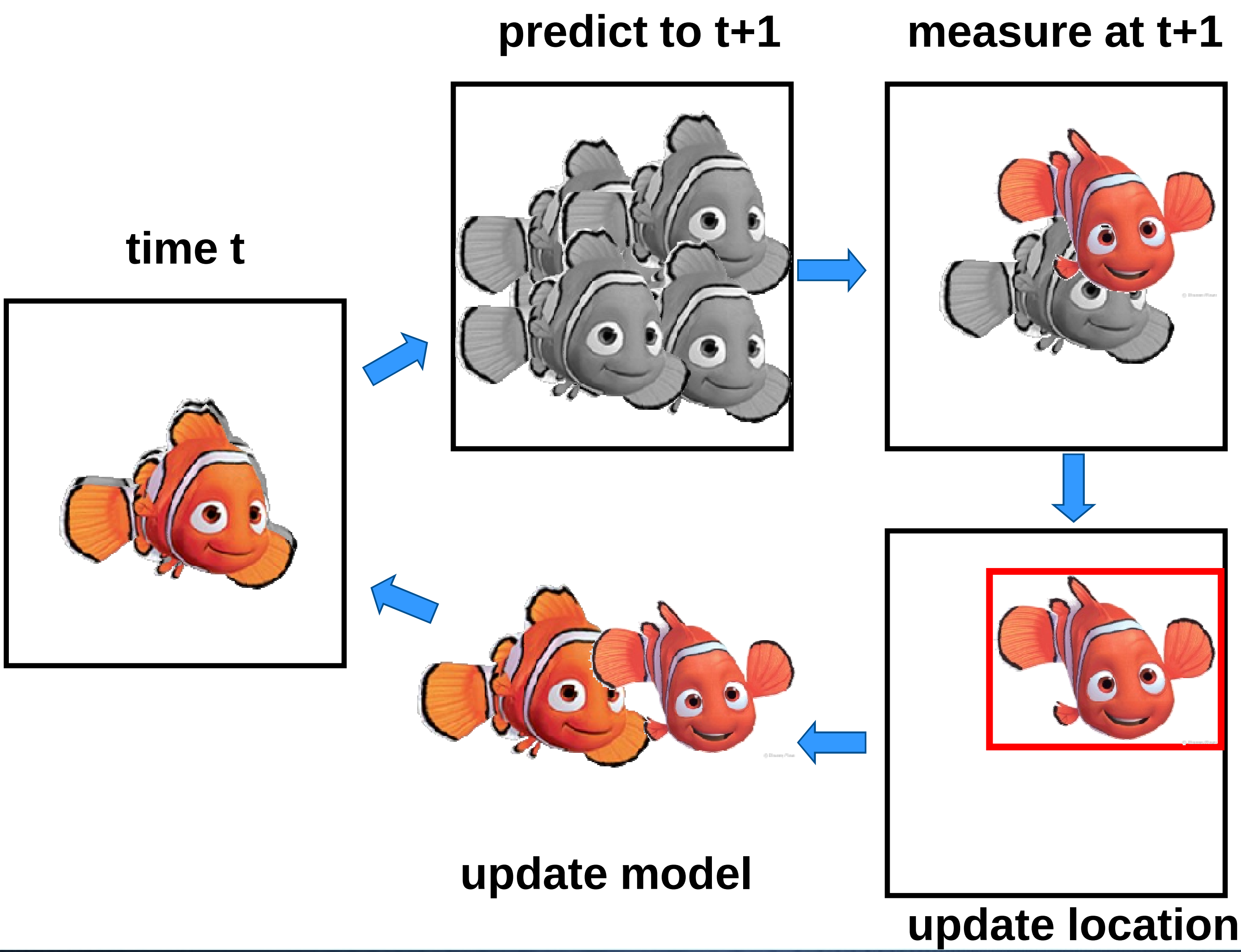
# In this talk…


"OBJECT"
CONTEXT
CONSTRAINS
vectorjunky.com

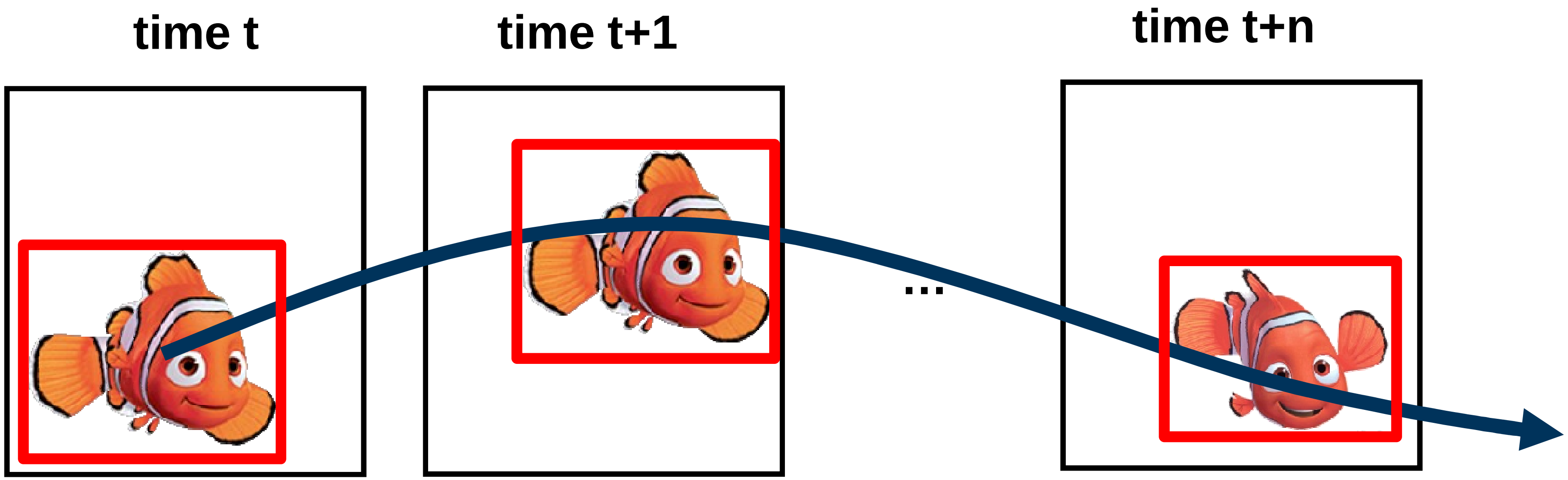
# Before we actually start:
# Two Tracking paradigm



# (i) Traditional tracking loop



# (ii) Tracking-by-detection




(i) detect object(s) independently in each frame

(ii) associate detections over time into *tracks*

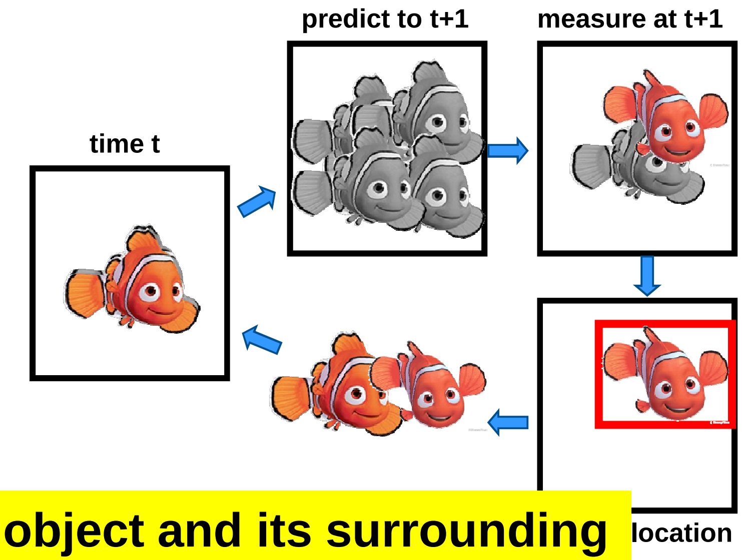
# Using constrains and context

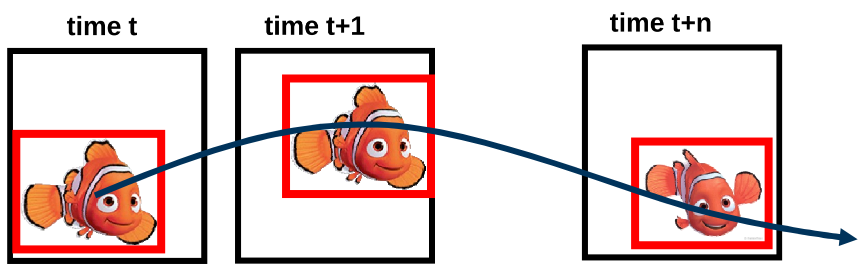
## Prediction/Correction
(model-free-tracking)

## Tracking-by-detection

information about the object dynamics

scene context (e.g., geometry)




object and its surrounding

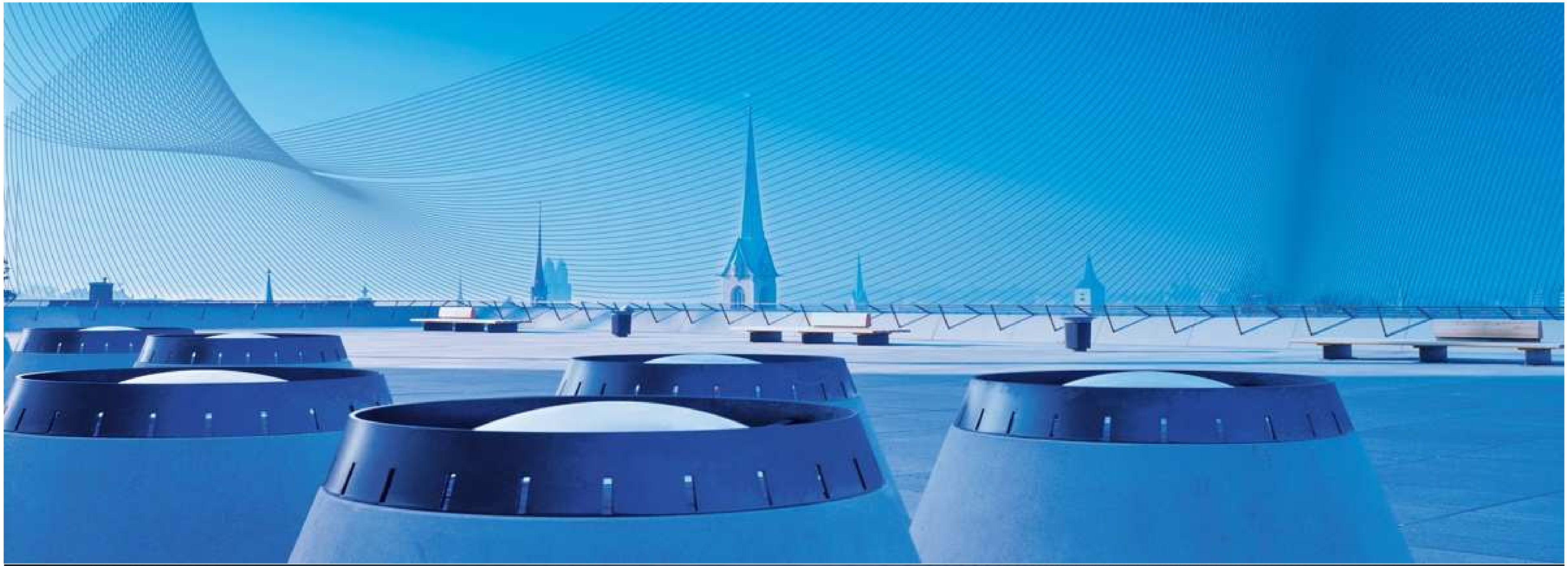
build and adapt the detector

learning (refining) a model during tracking

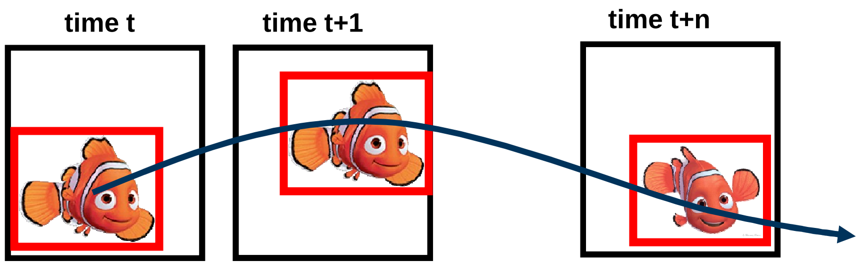
exploit relations to other objects

# Tracking-by-detection

model of the object (object category) is given



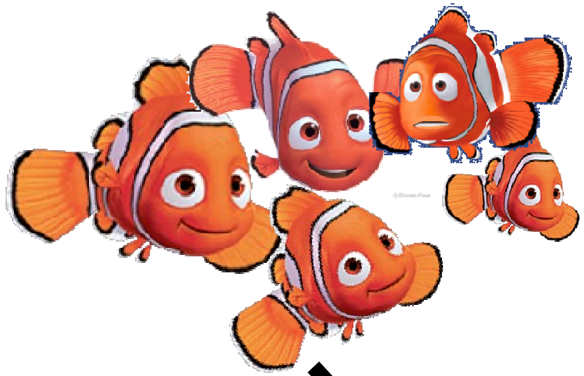
# SCENE CONTEXT




build and adapt the detector

[Grabner et al., Photogrammetry & Remote Sensing'07]
[Klucker et al., ICCV'07 WS on 3dRR]

# How to get the detections?

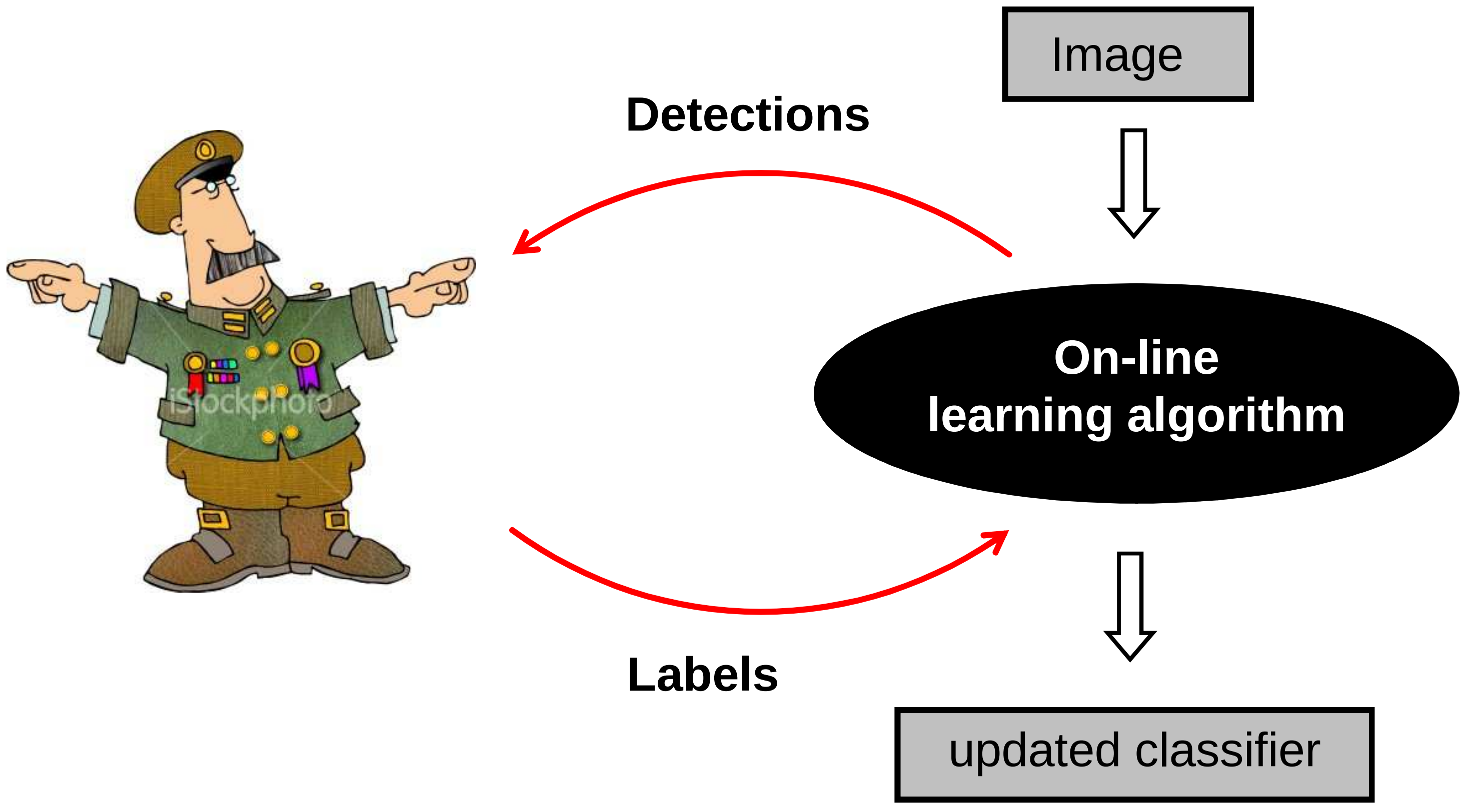


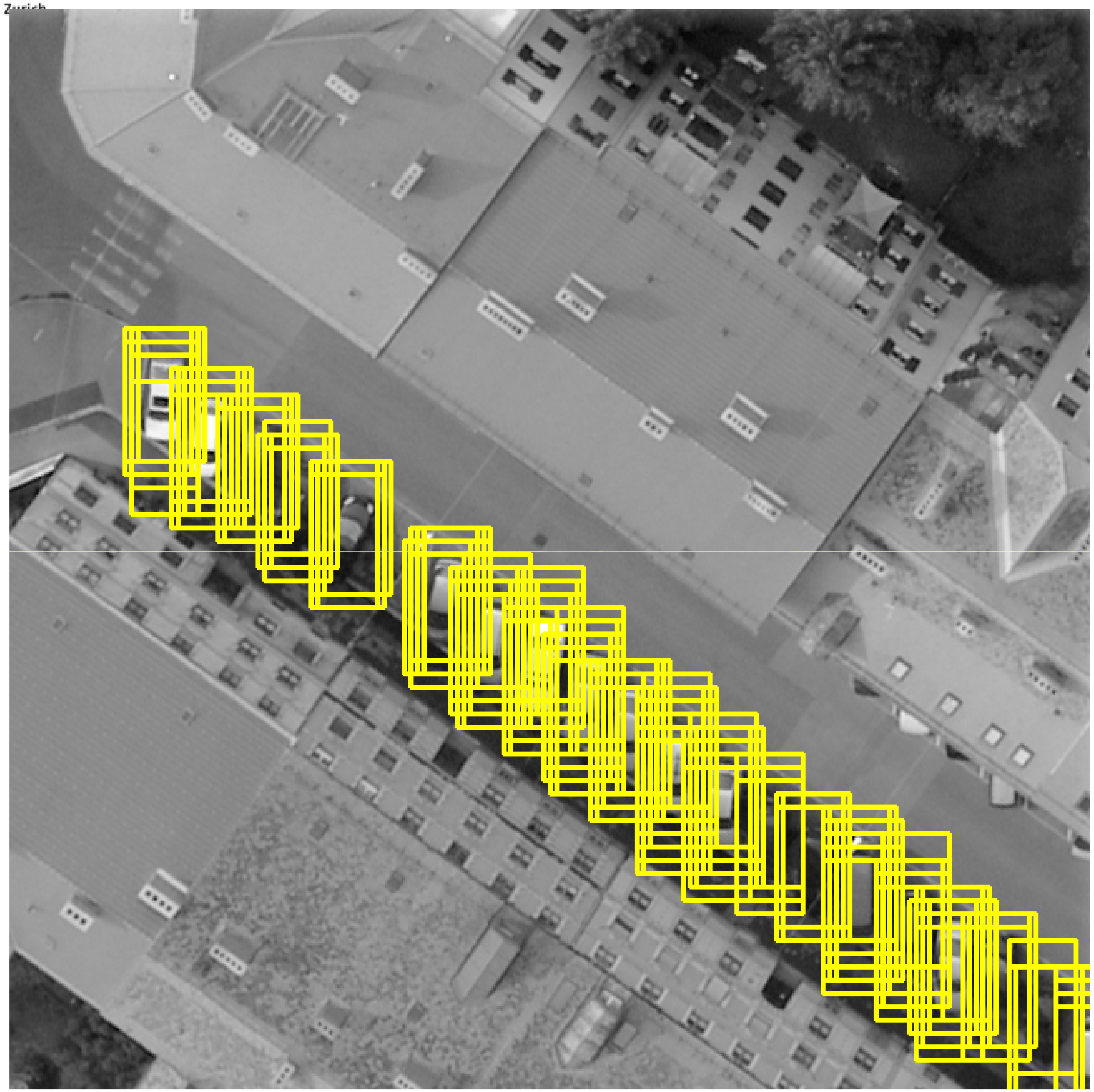

Nemo‘s

Background

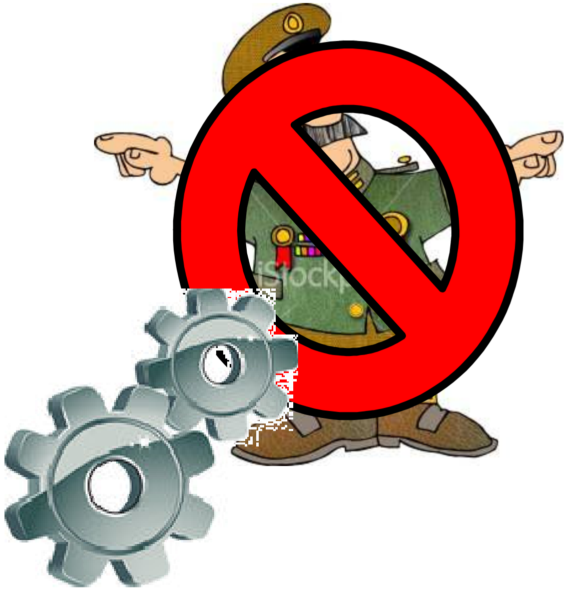
**Supervised Learning**

# “Fewer Clicks - Less Frustration”





# Minimize the human labeling effort



Image

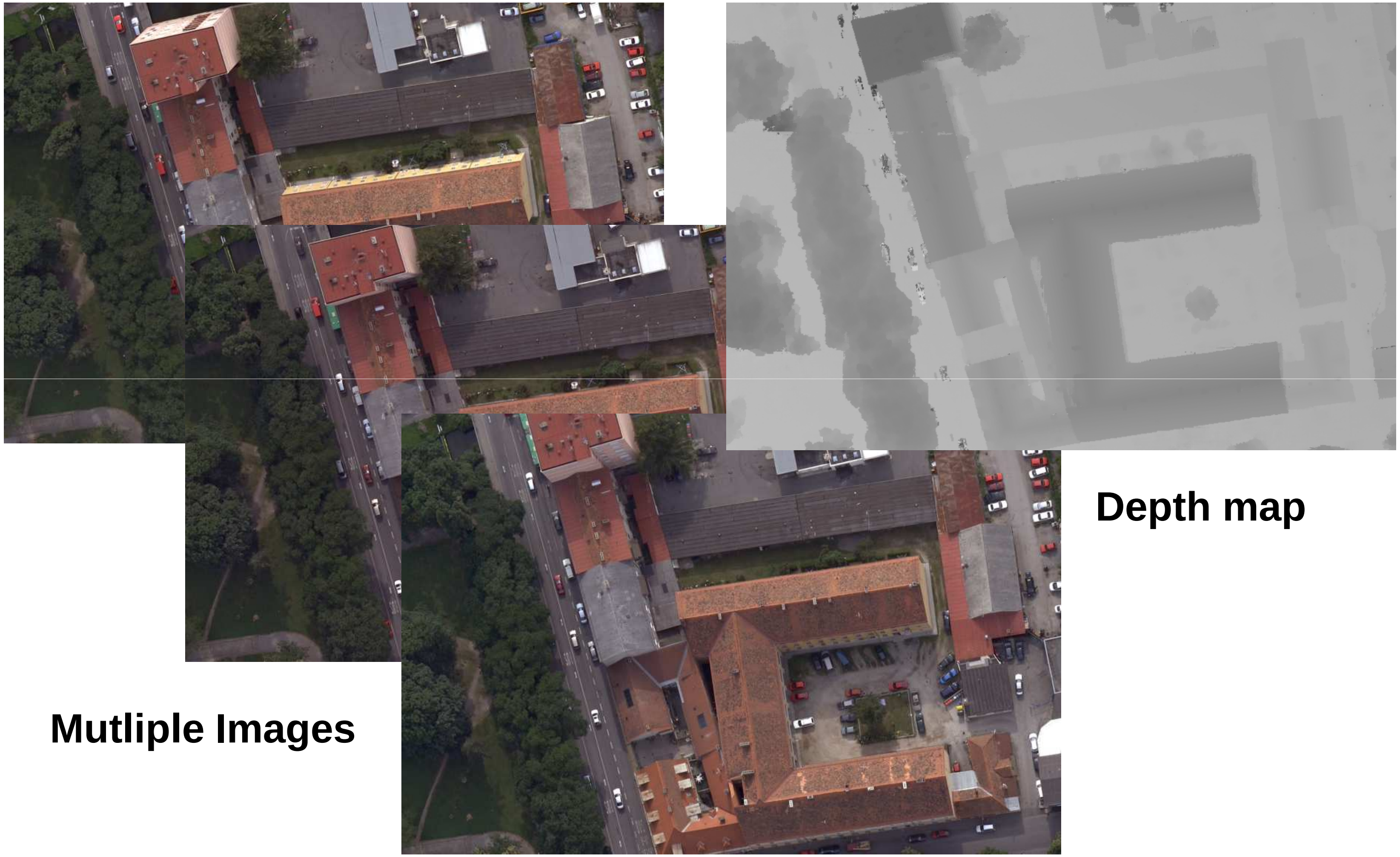
Detections

On-line
learning algorithm

Labels

updated classifier

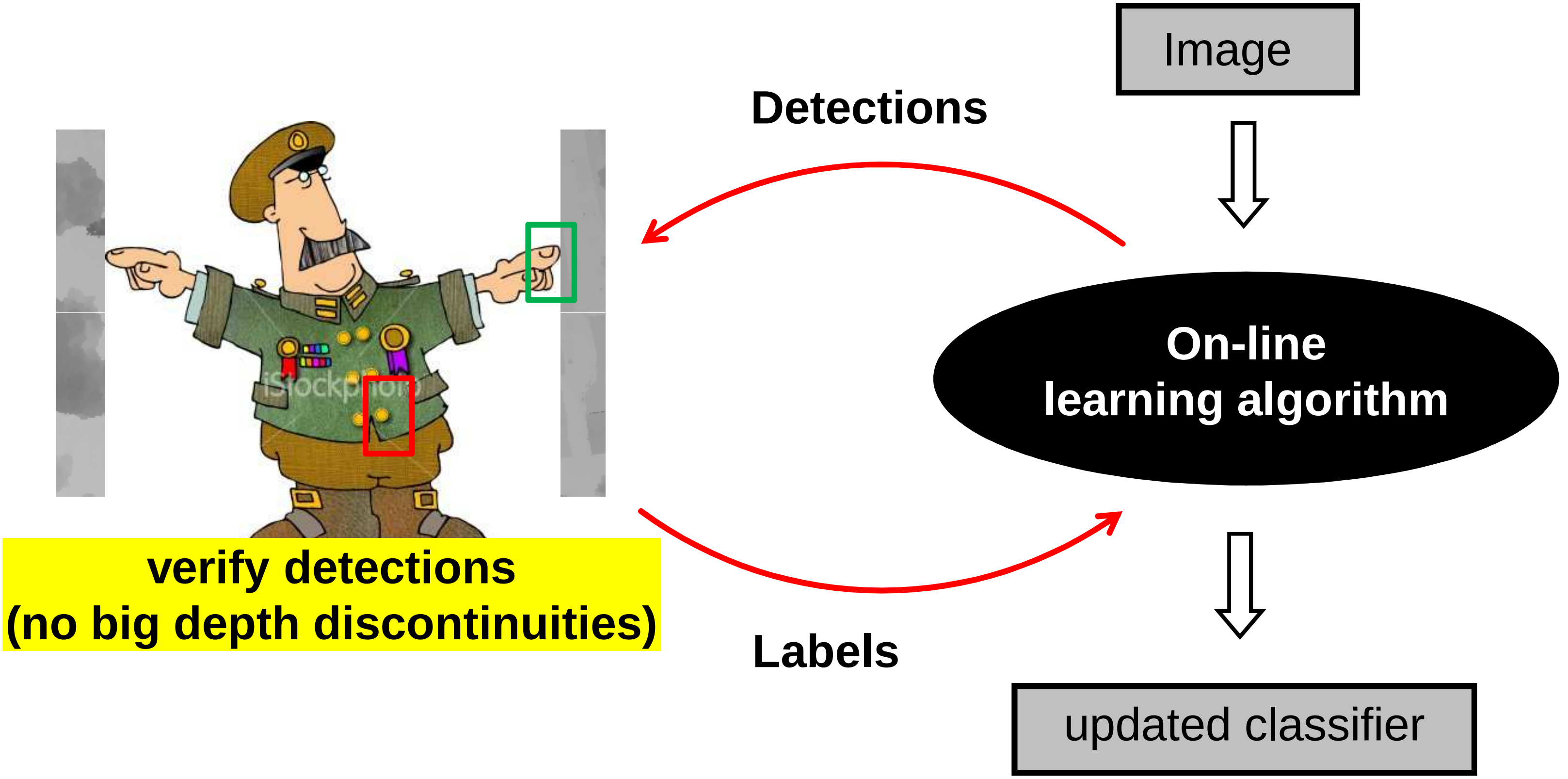
# Exploiting the 3d structure



Depth map

Mutliple Images

# A 3d teacher


Detections
Image
On-line
learning algorithm
verify detections
(no big depth discontinuities)
Labels
updated classifier

# Results

# Results

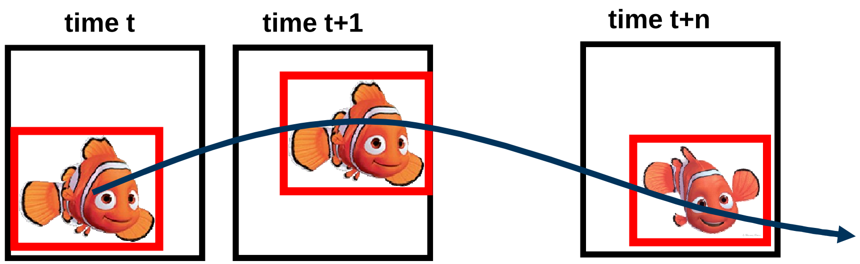
# Results

# LESSON LEARNED 1

**Visual constrains saves you time!**

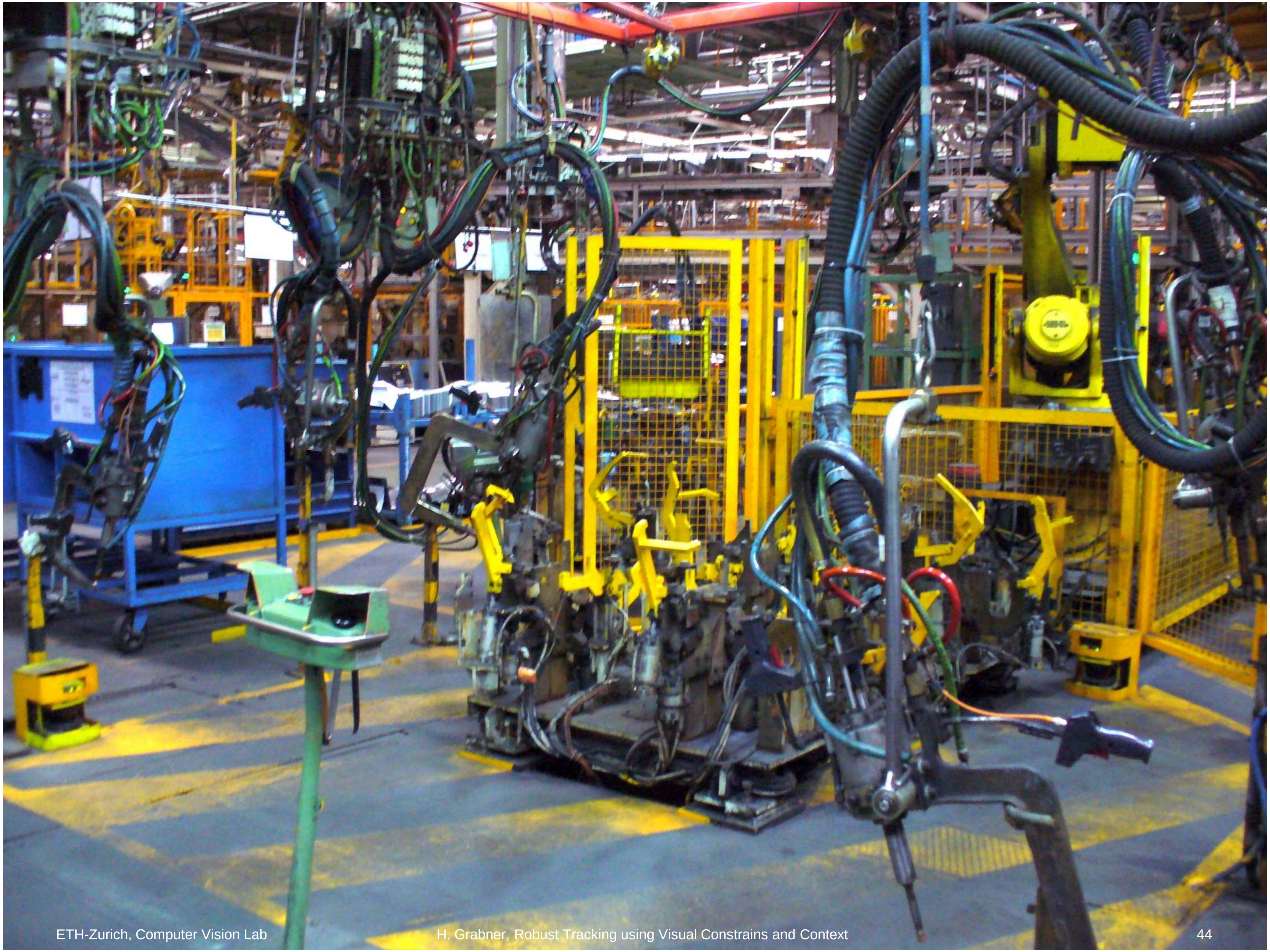
# SCENE CONTEXT

information about the object dynamics




scene context (e.g., geometry)

[Stalder et al., ECCV'10]




Input image

Detection

Tracking

Behaviours

Workflow

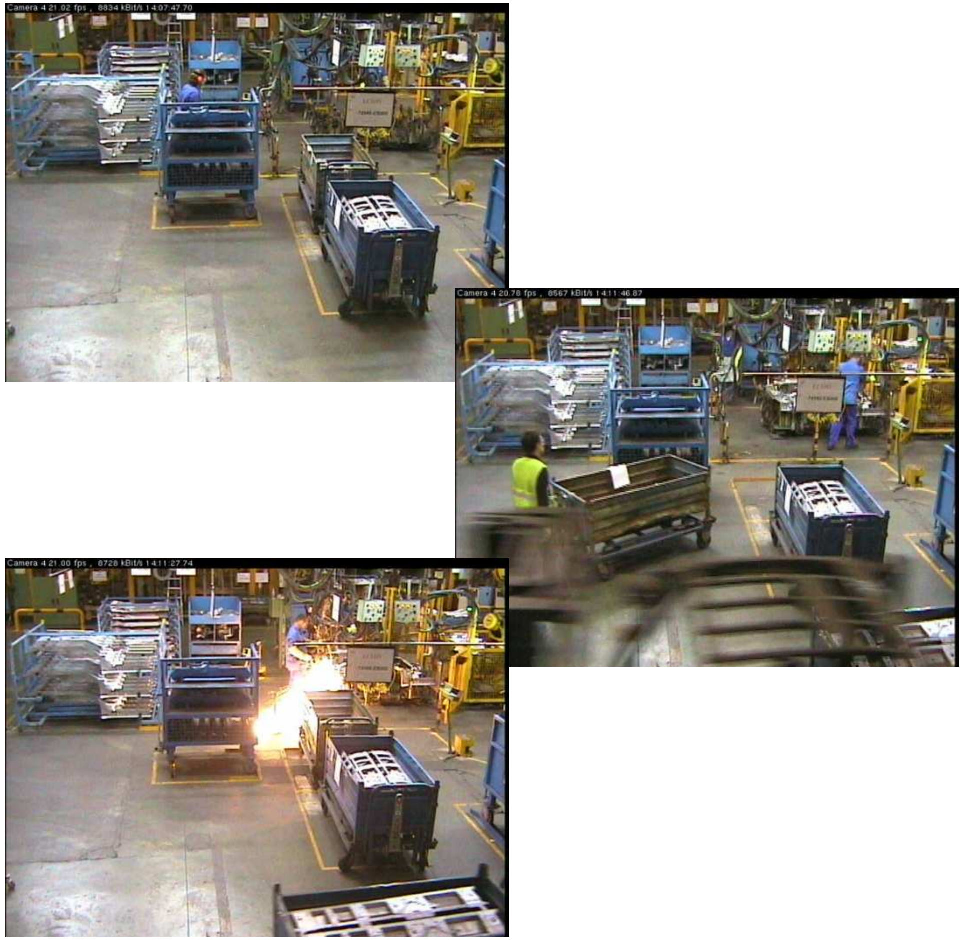
Reasoning

Statistics

# Challenges of an industrial environment

# Challenges of an industrial environment

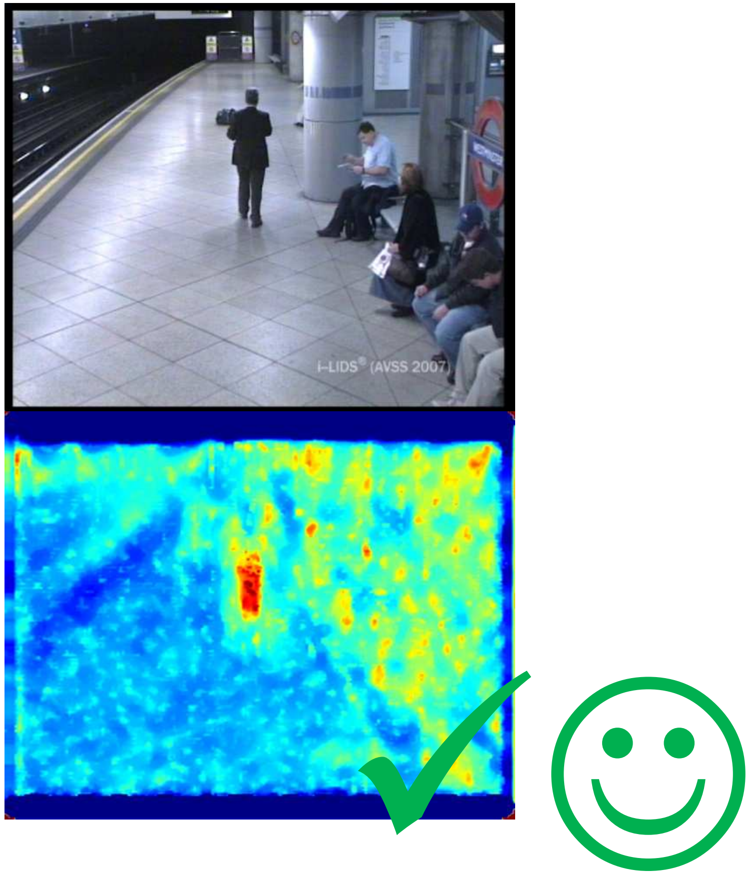
- Occlusions
- Self Occlusions
- Other moving Objects
- Similar Objects
- Industrial working conditions
- Real-time



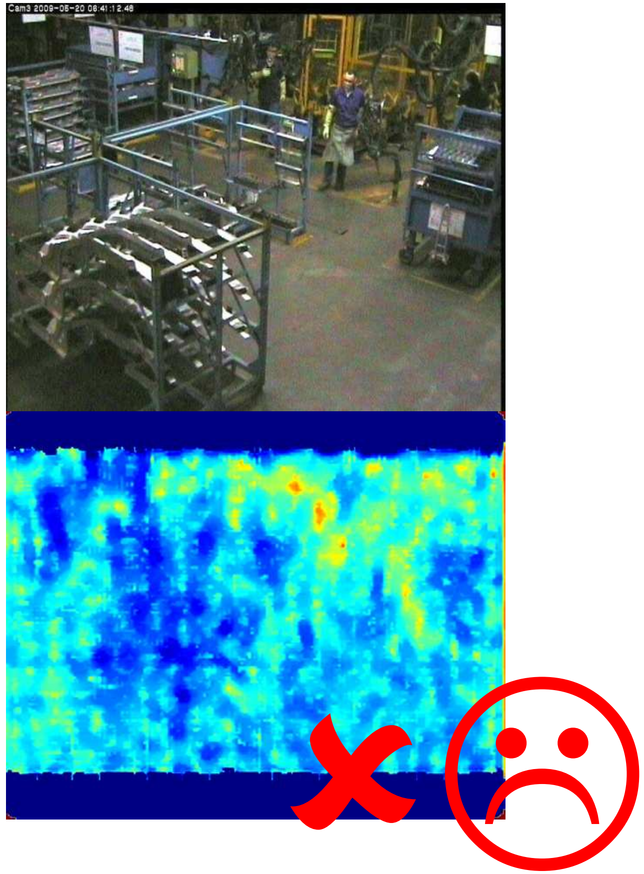
# State-of-the-art person detector

**i-Lids**

P(person)

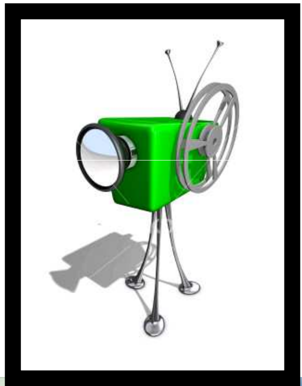



**Scovis-NISSAN**



“reasonably good detector”

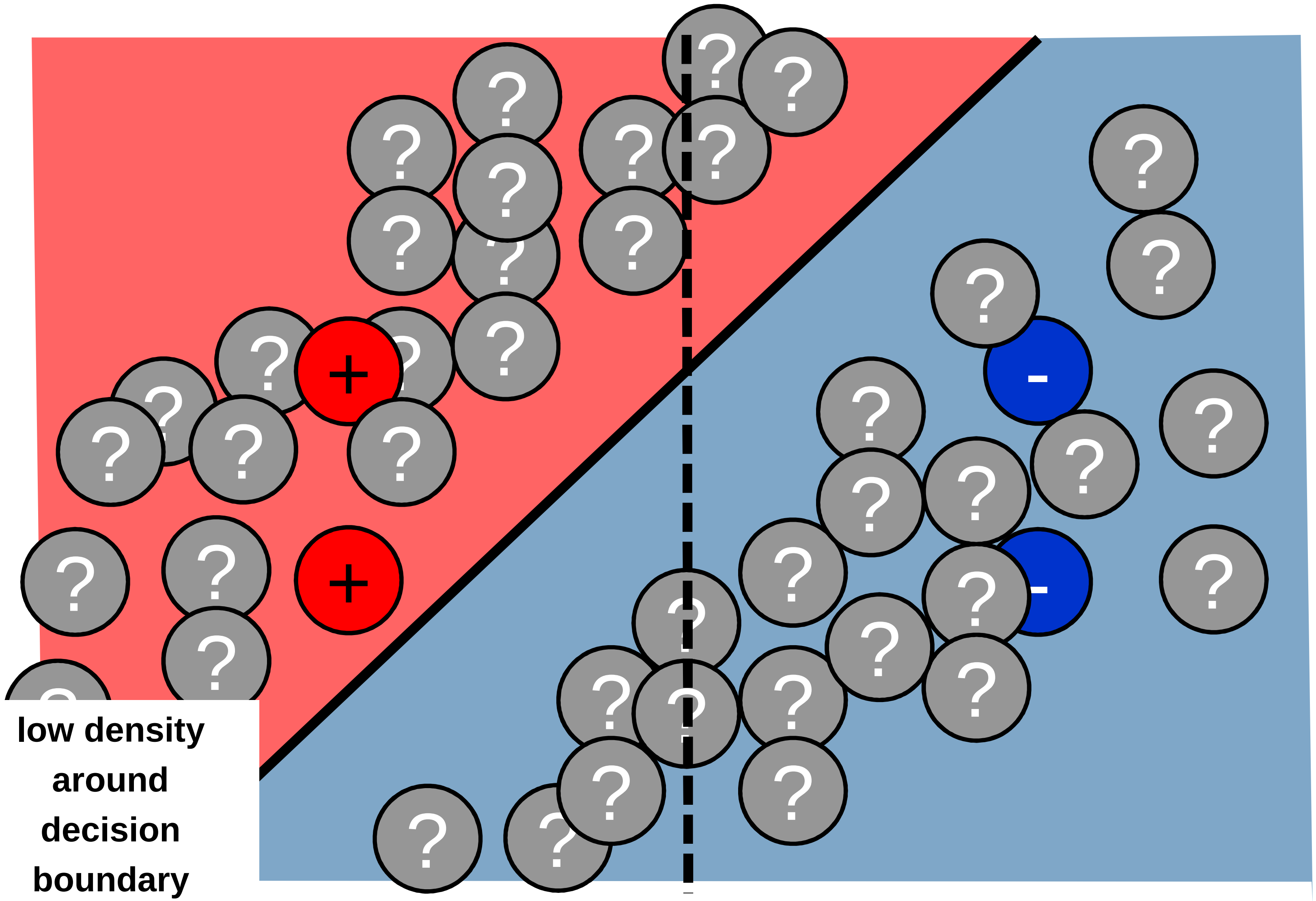
**Use unlabled data to specialize!**



# How to interpret unlabeled data?


- Machine Learning?
- Semi supervised learning?
- Transfer learning?
- „Learning to Learn“?

# Can unlabeled data help?



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

**Unlabeled data**

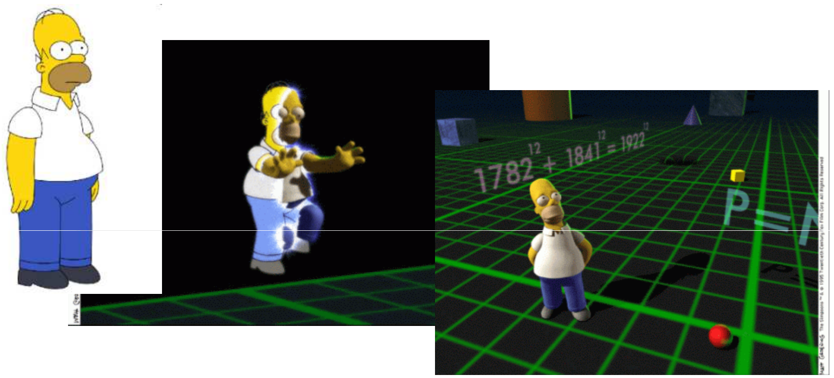
**Thomas Bayes**
**1702-1761**

**Further Assumptions!**

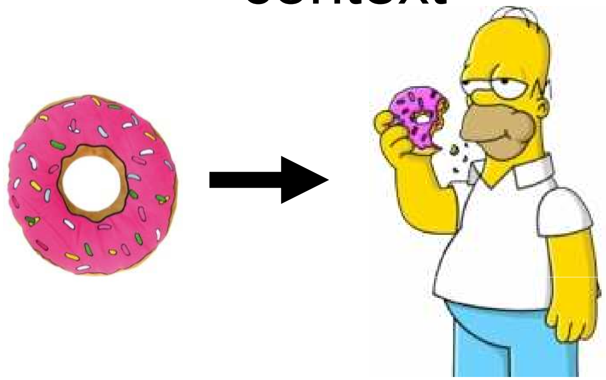
**(e.g., manifold assumption, cluster assumption,…)**

# Additional information/constraints

Geometry and structure

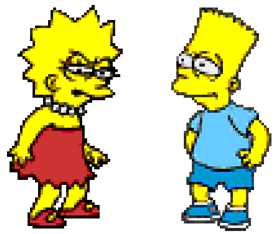


Objects are embedded into context



Temporal consistency

Typical behavior



# LESSON LEARNED 2

**Vision is more than pure machine learning!**

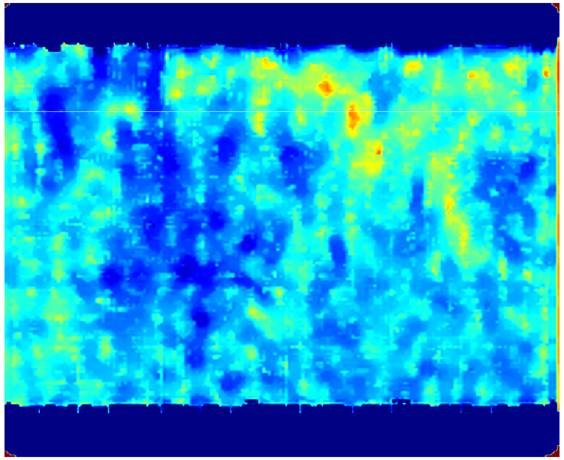
# Input to our System

**Input image**


**Object detector**

S(I,x,y,s)

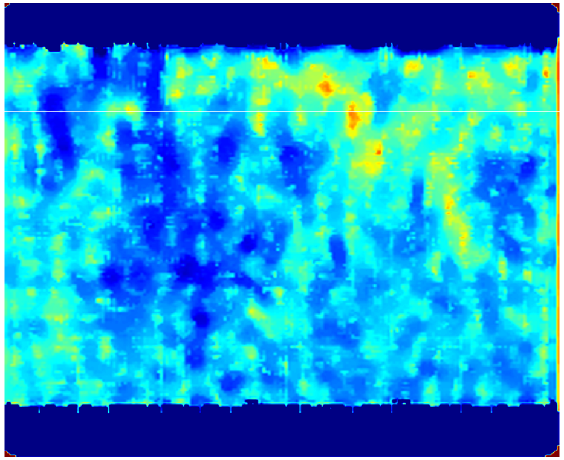
**Confidence map**



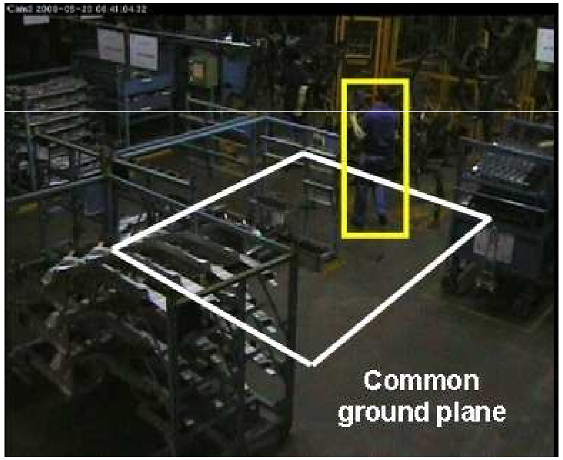
[Dalal and Triggs, CVPR'05]

# Geometric Filter

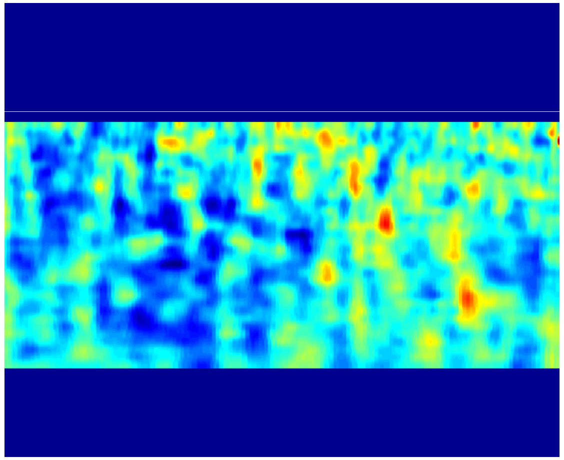
**Confidence map**



**Geometric constrains**

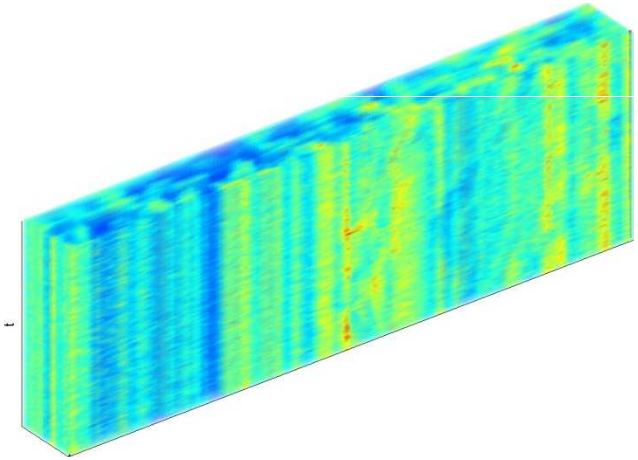



**Confidence map**

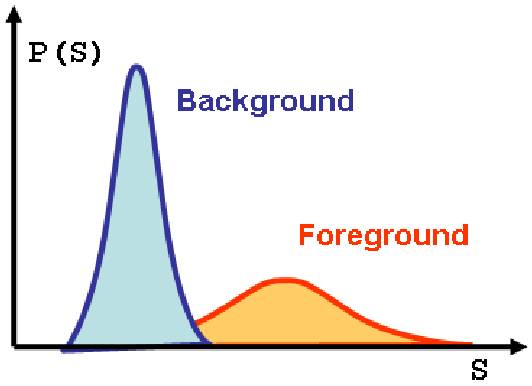


# Background Filter

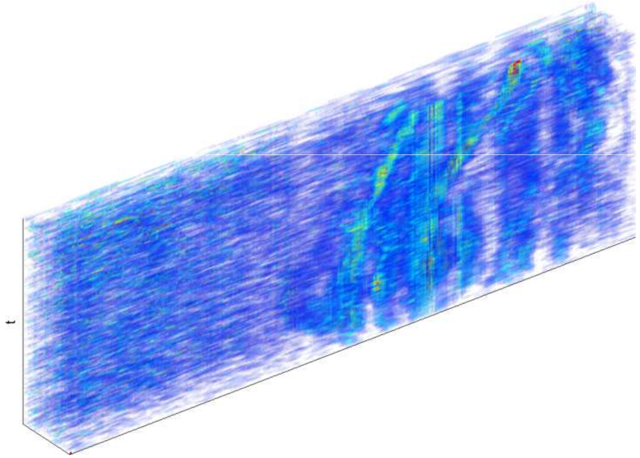
**Confidence volume**



**Background modeling**



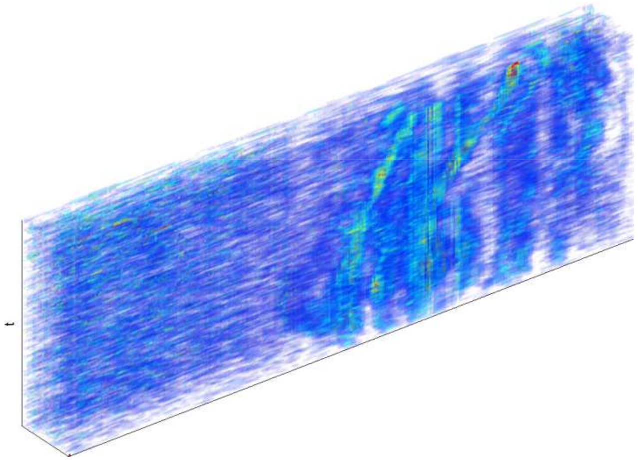

**Confidence volume**

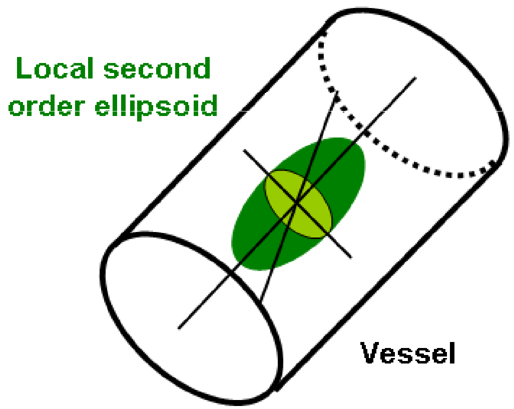


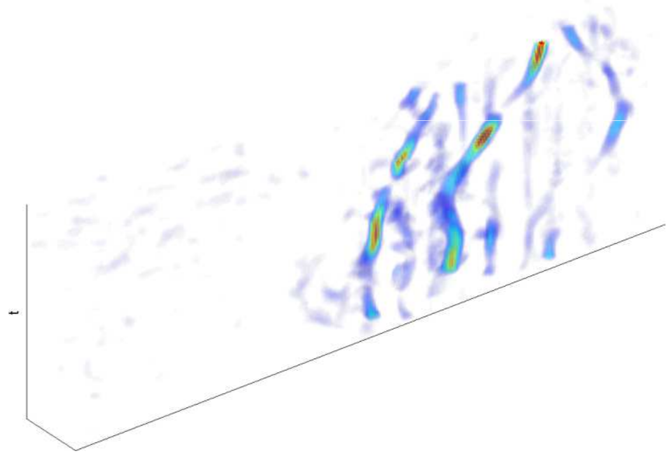
[Stauffer and Grimson, CVPR'99]

# Trajectory Filter

**Confidence volume** | **Vessel-Filtering** | **Confidence volume**



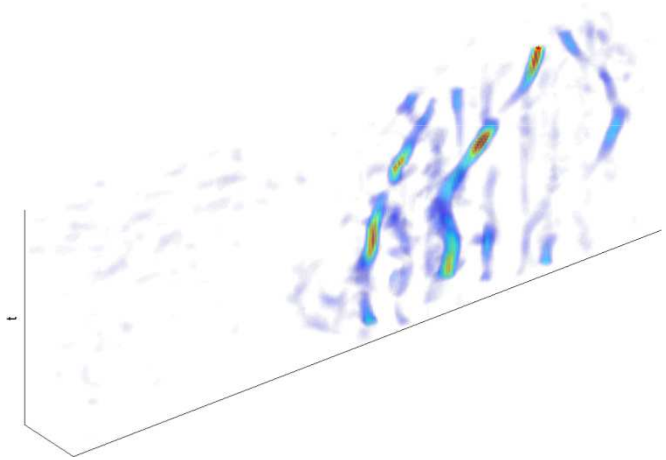


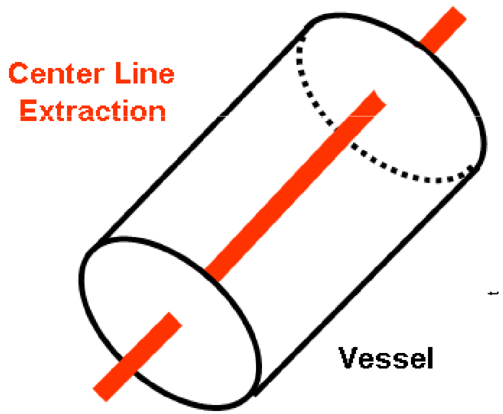
[Frangi et al., '98]

# Particle Filter (cont. Trajectories)

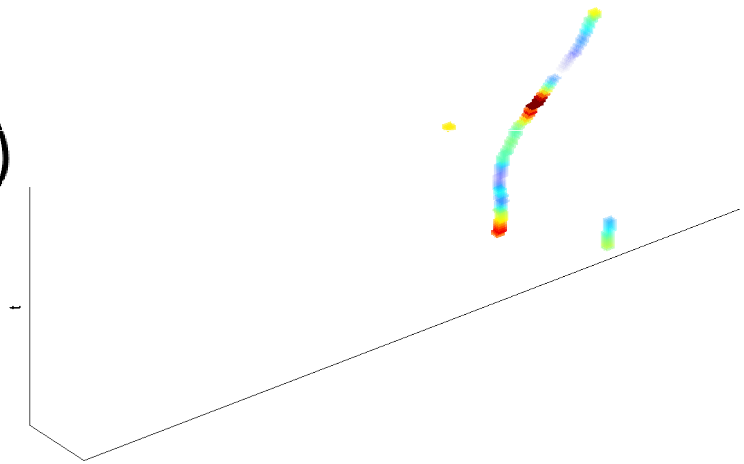
**Confidence volume**



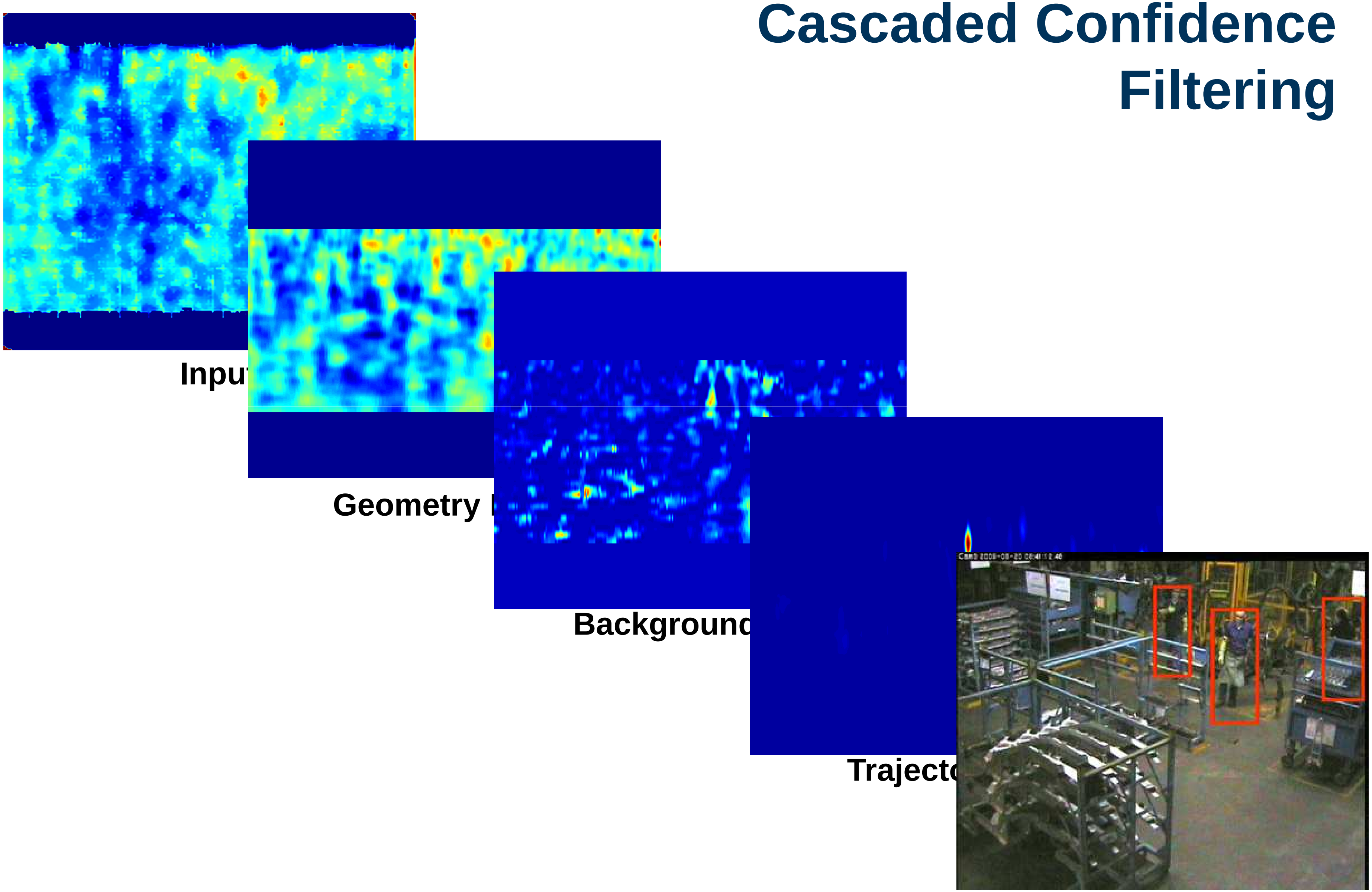
**Particle filter**



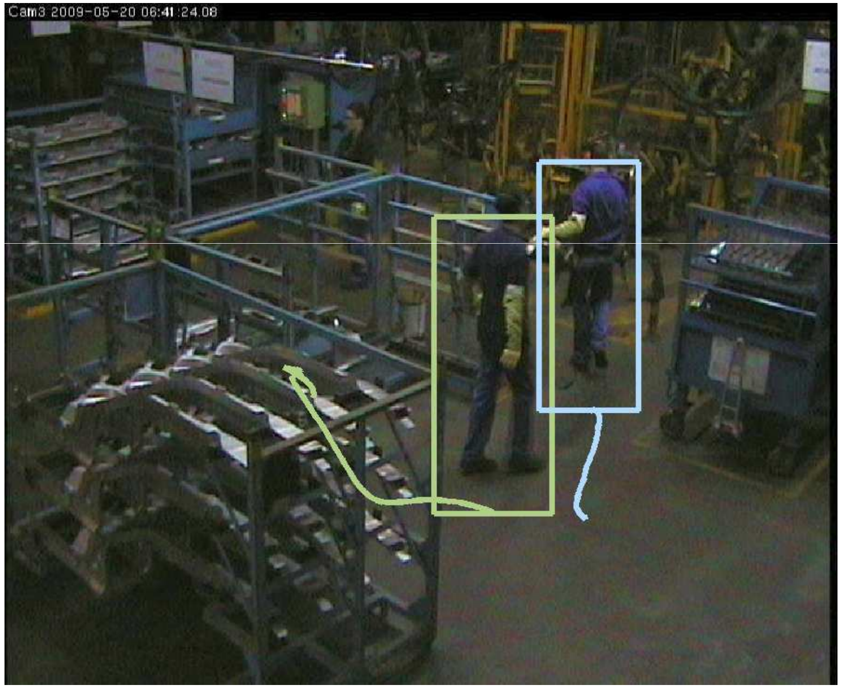

**Confidence volume**

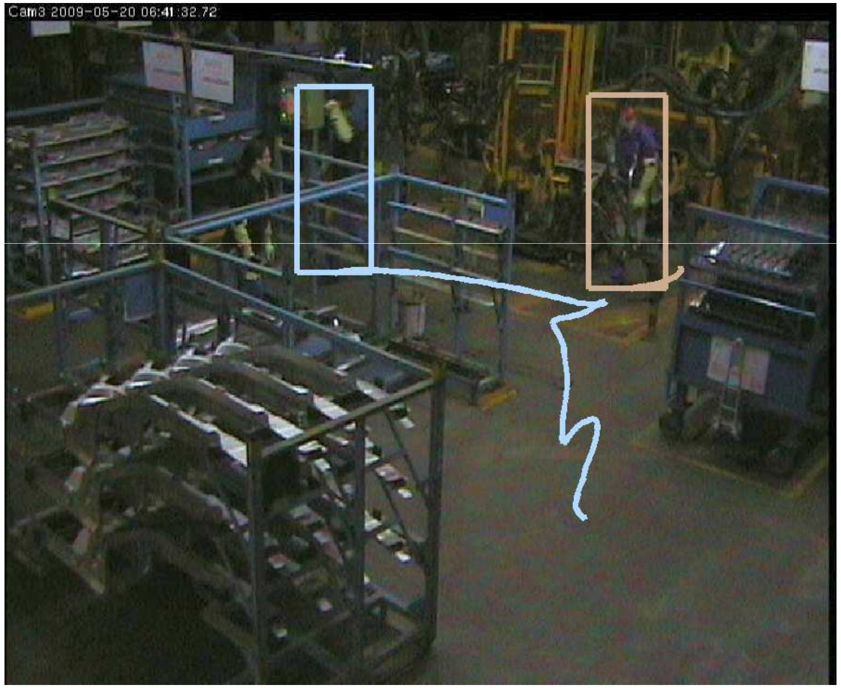


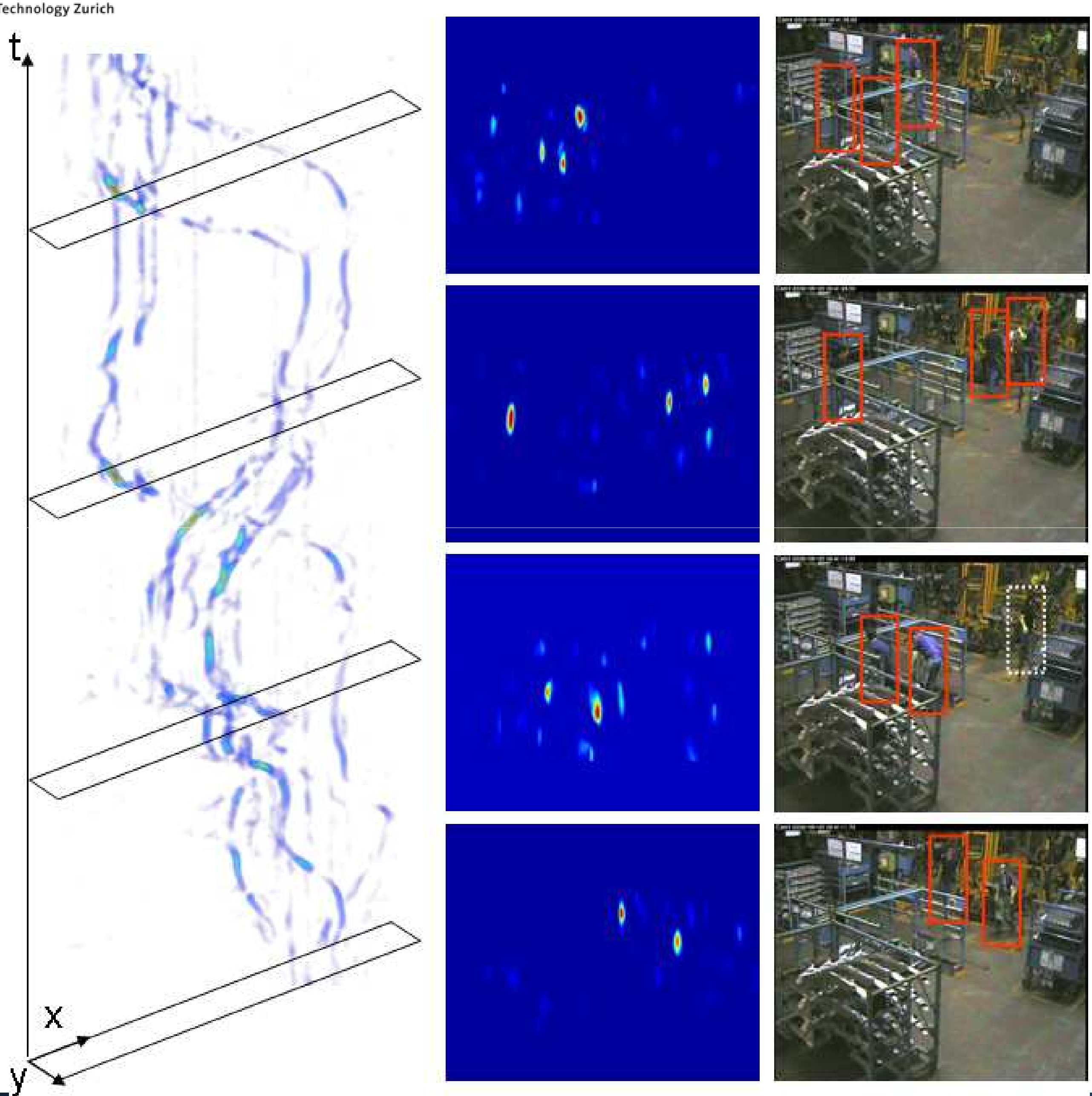
# Cascaded Confidence Filtering

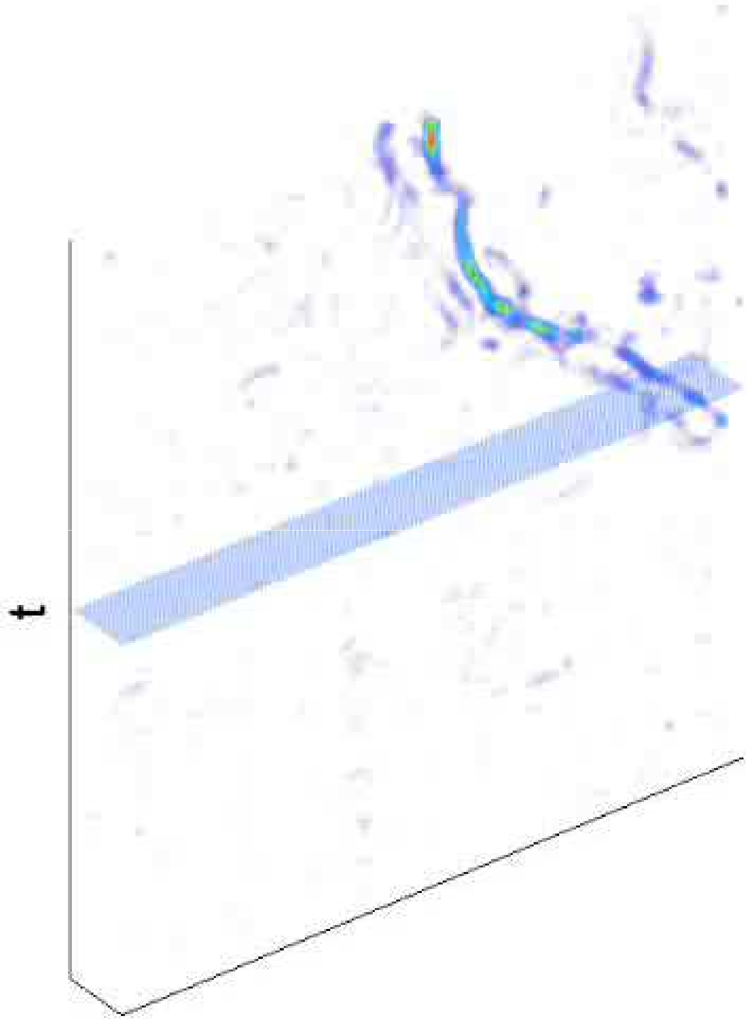


# Results




t
x
y



# Tracking-by-Detection

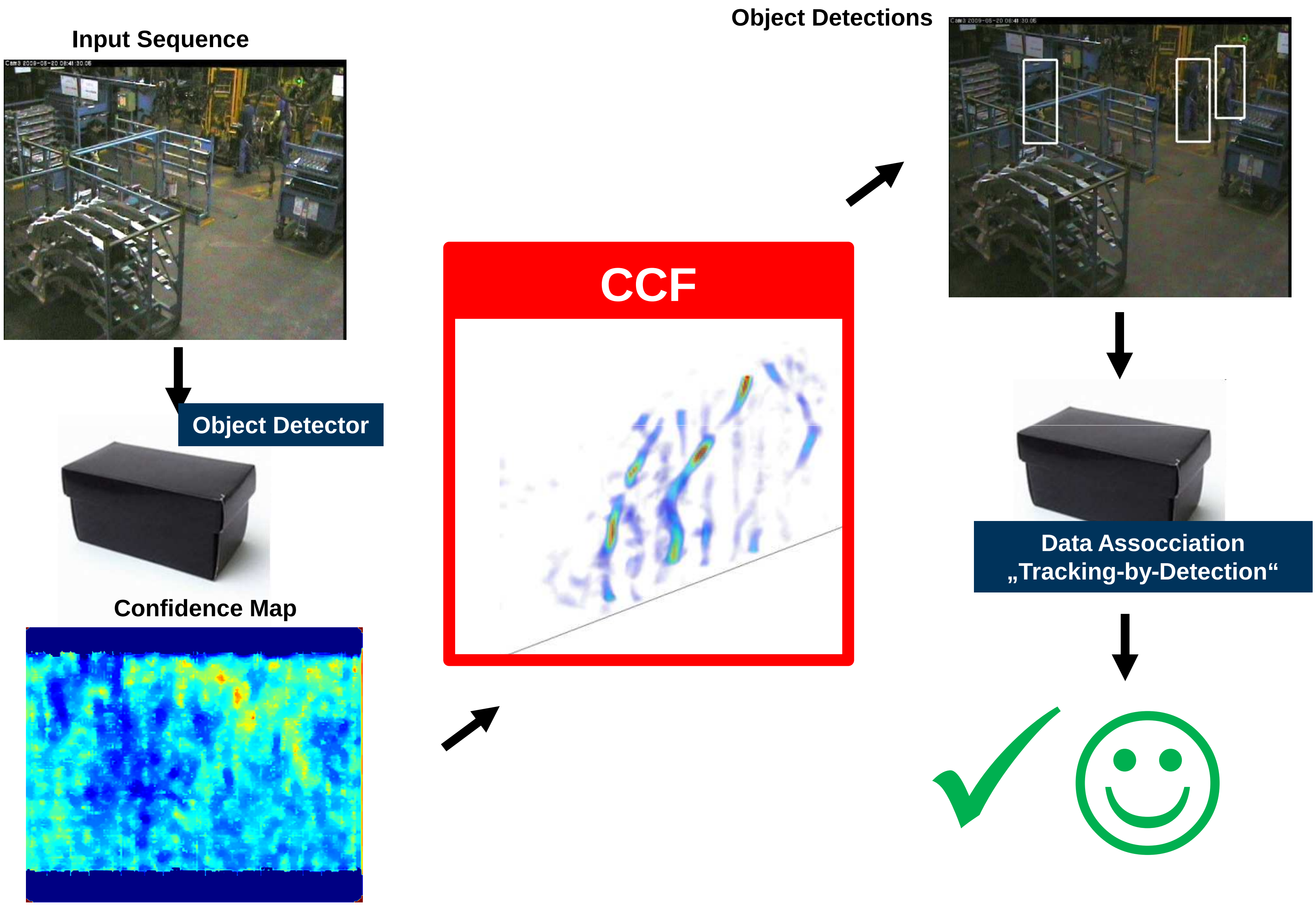
**Input Sequence**

**Object Detector**

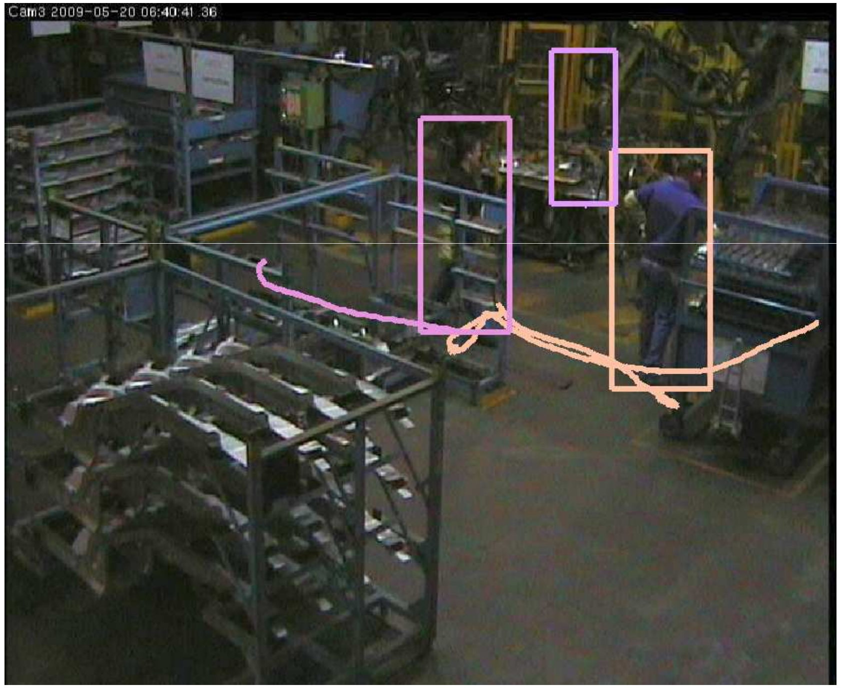
**Object Detections**

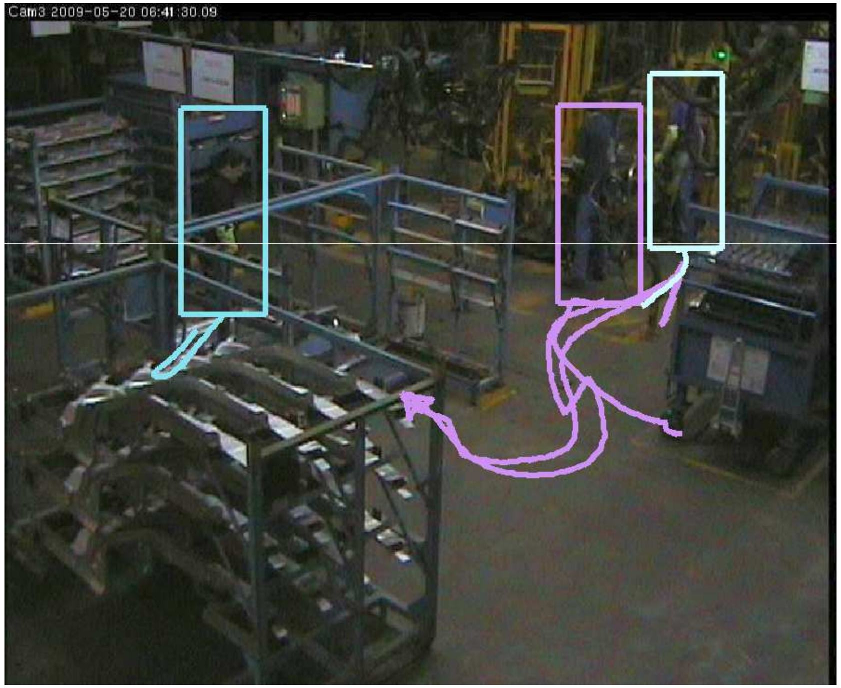
**Data Assocciation „Tracking-by-Detection“**

[Huang et al., ECCV’08]


Input Sequence
Object Detector
Confidence Map
CCF
Object Detections
Data Associaation
„Tracking-by-Detection“

# Improved tracking-by-detection



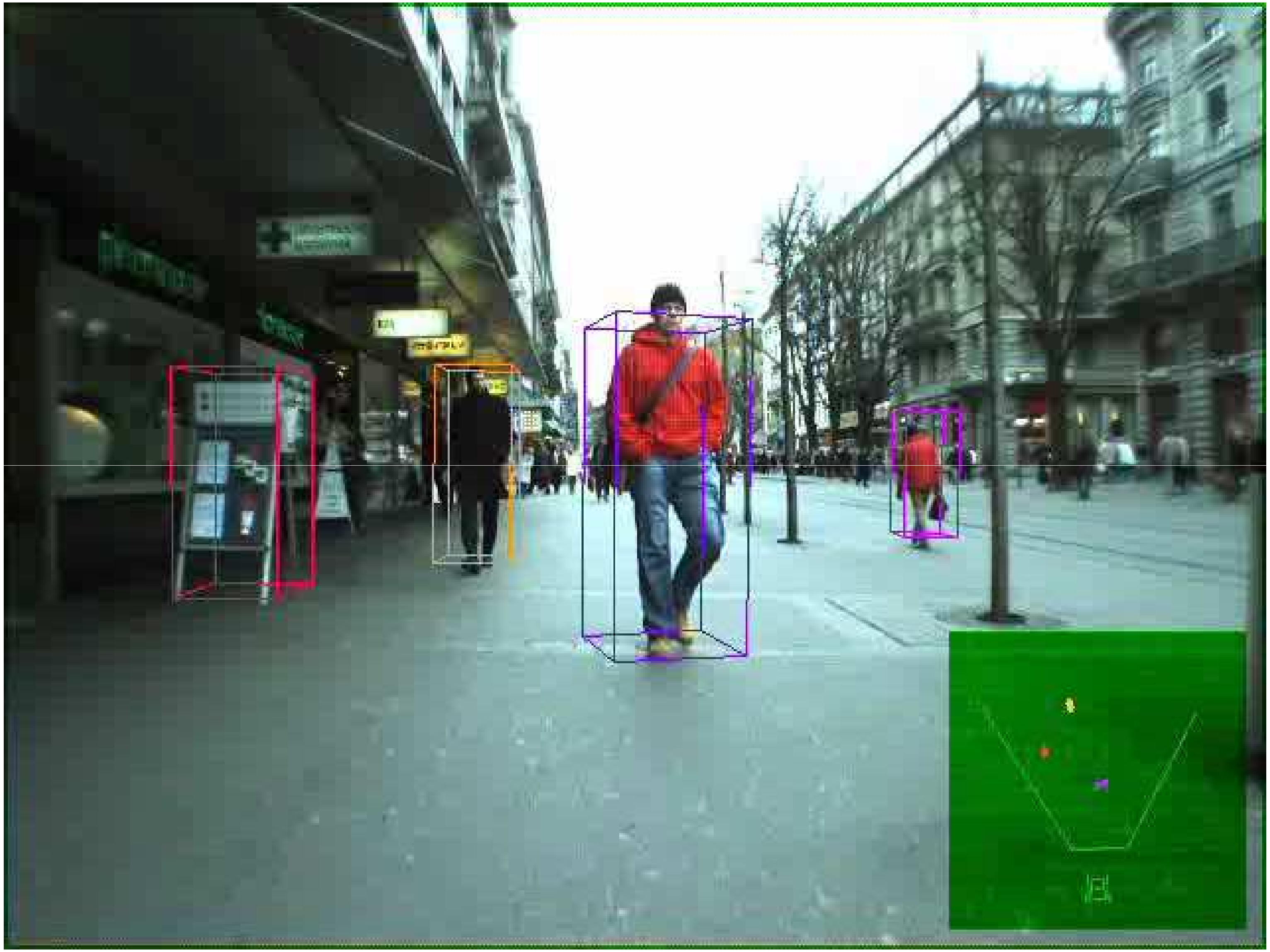




Cam3 2009-06-20 08:40:21.08

# LESSON LEARNED 3

**Use any available constrains!**

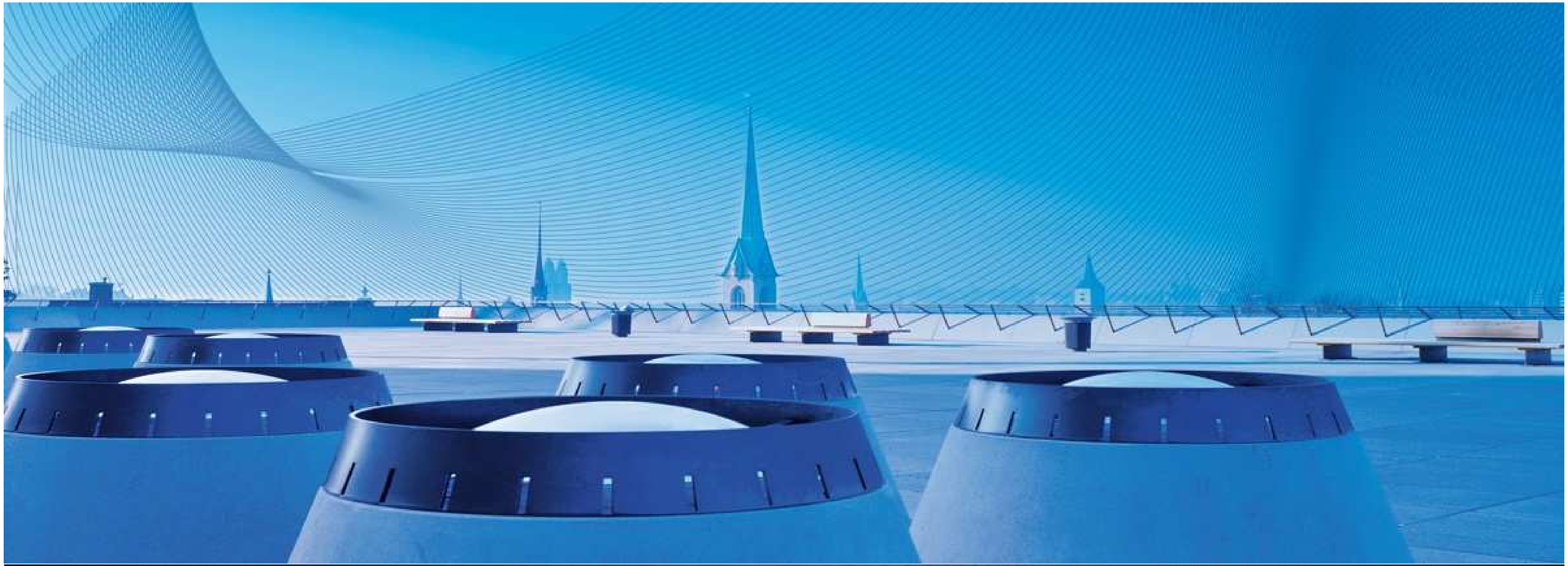
**common ground-plane, static camera, expected dynamics,…**



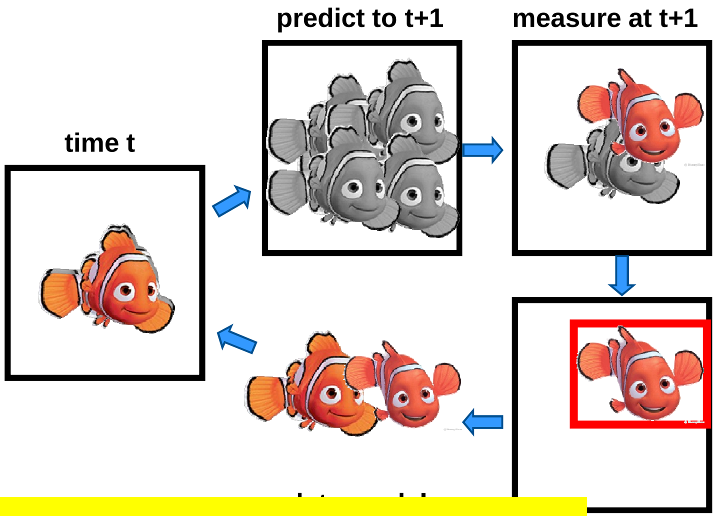
[Leibe et al., PAMI’08]

# Model-free-Object Tracking

No object model is given



# LOCAL CONTEXT



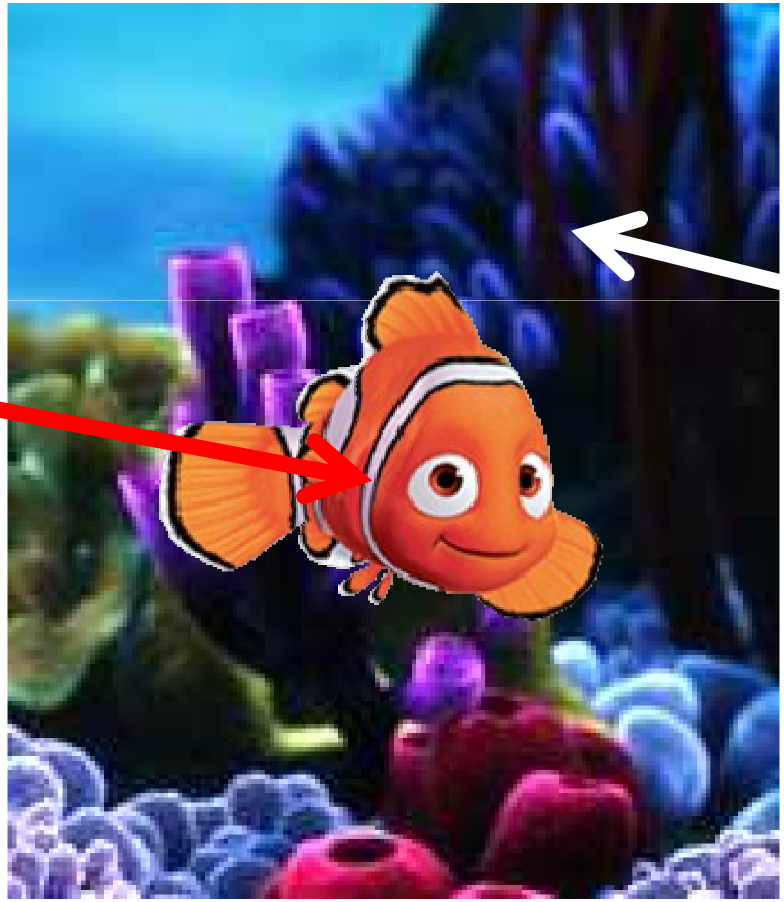

**object and its surrounding**

[Grabner et al., CVPR’06, BMVC’06]

# Objects are not isolated!

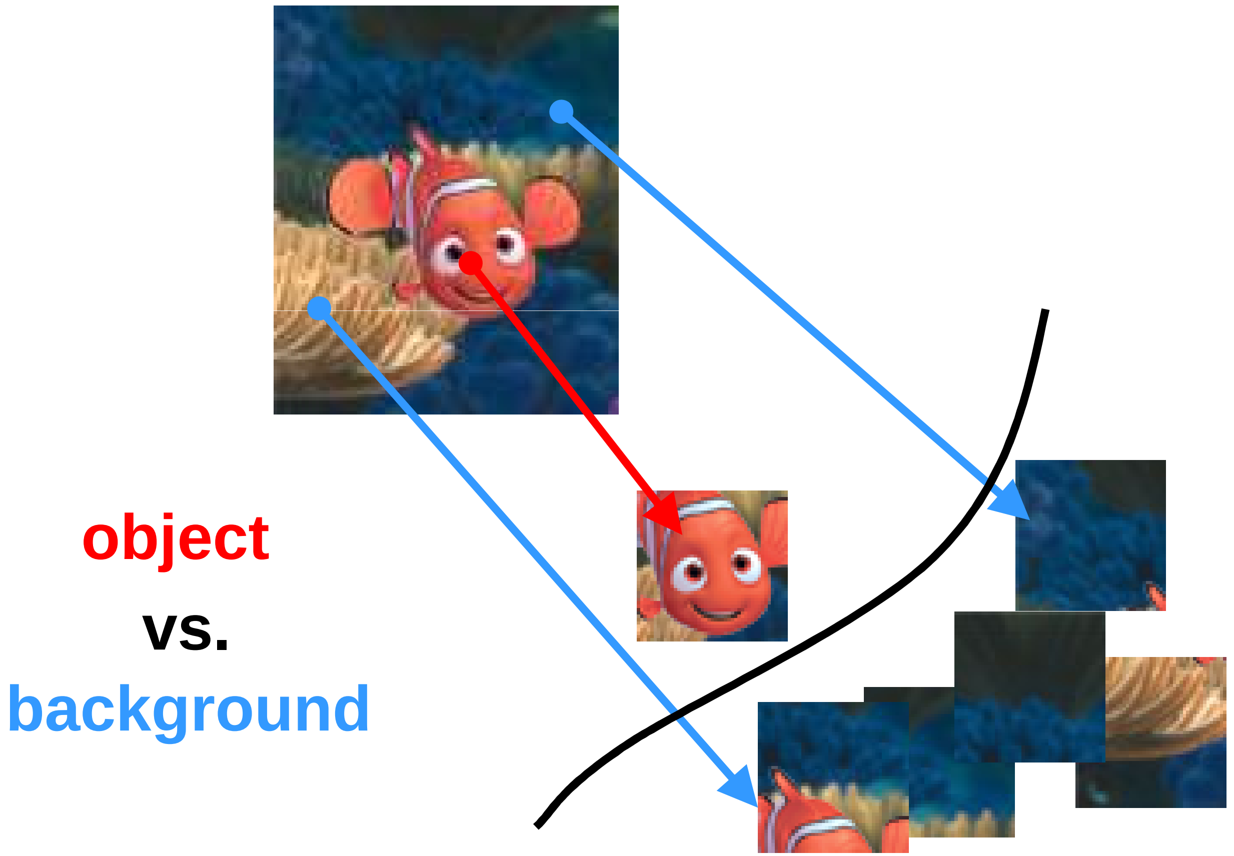
- Learning current object appearance vs. local background.

current
obj. appearance

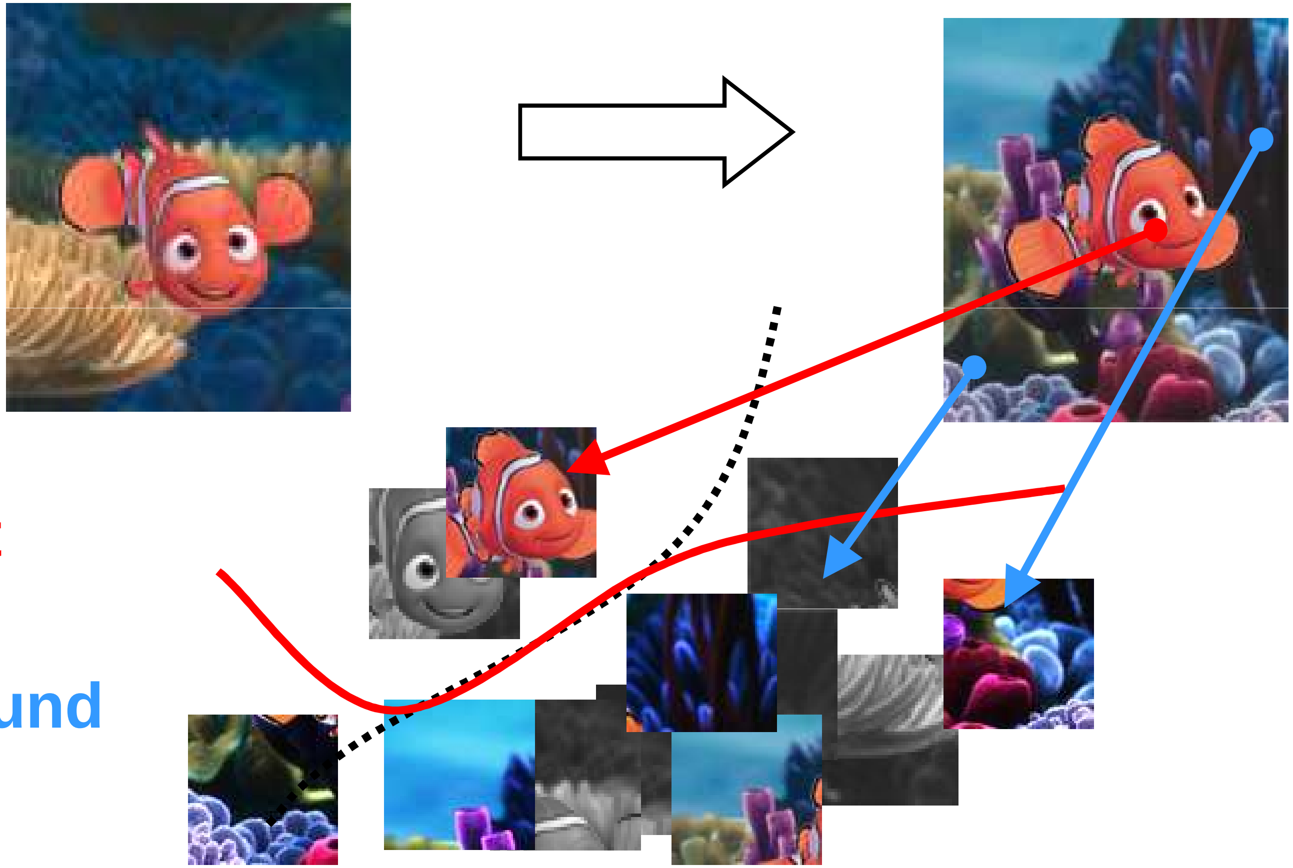


current
background

# Tracking as binary classification probelm



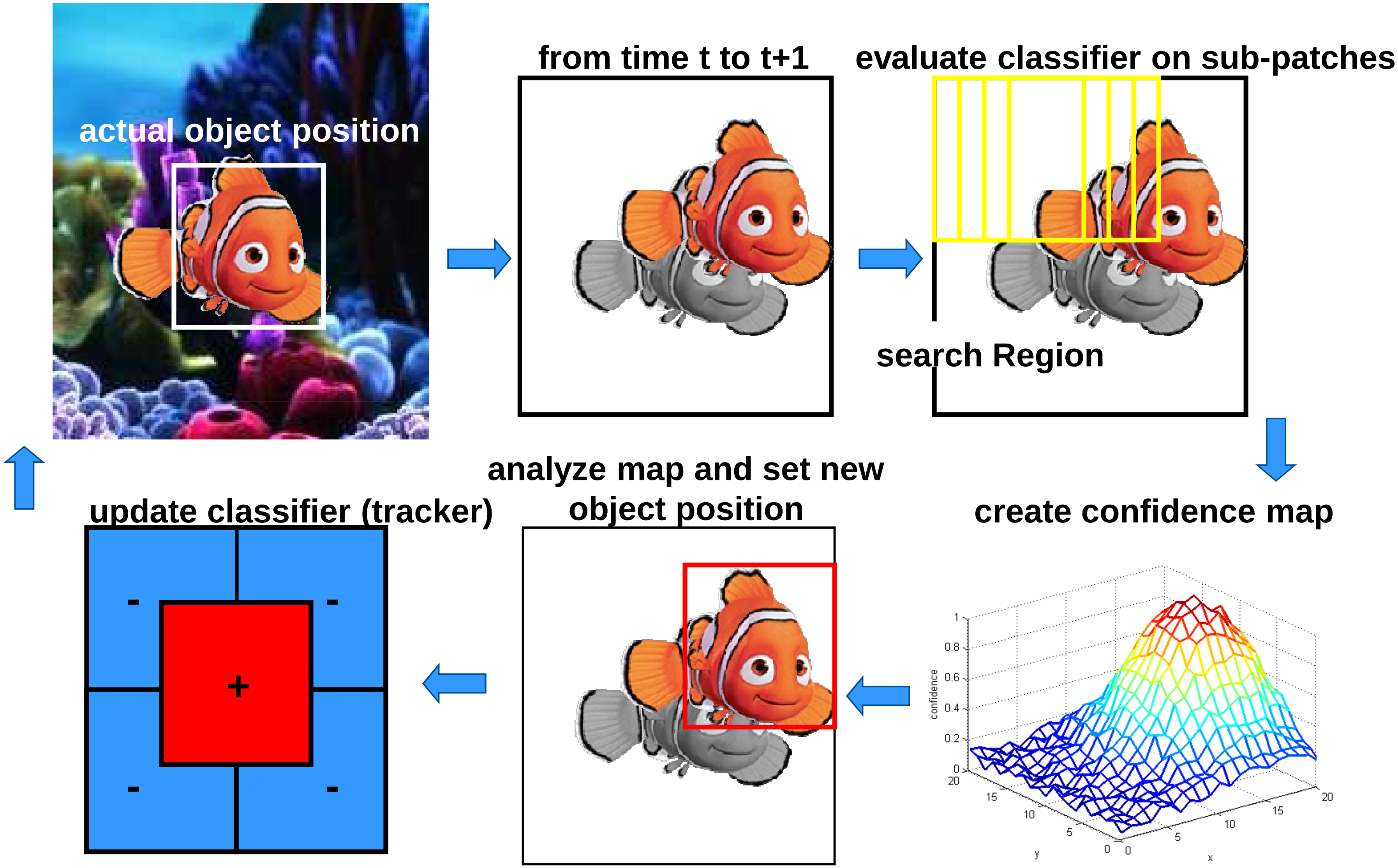
# Tracking as binary classification probelm

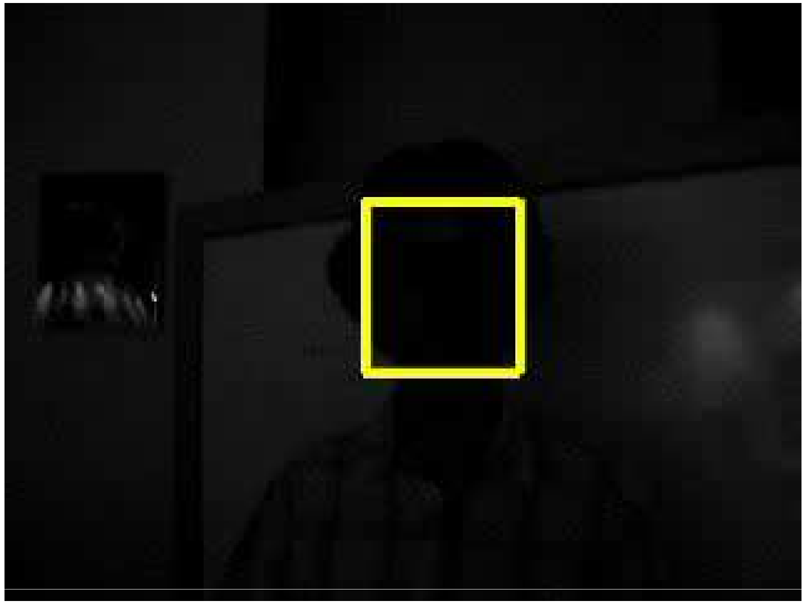


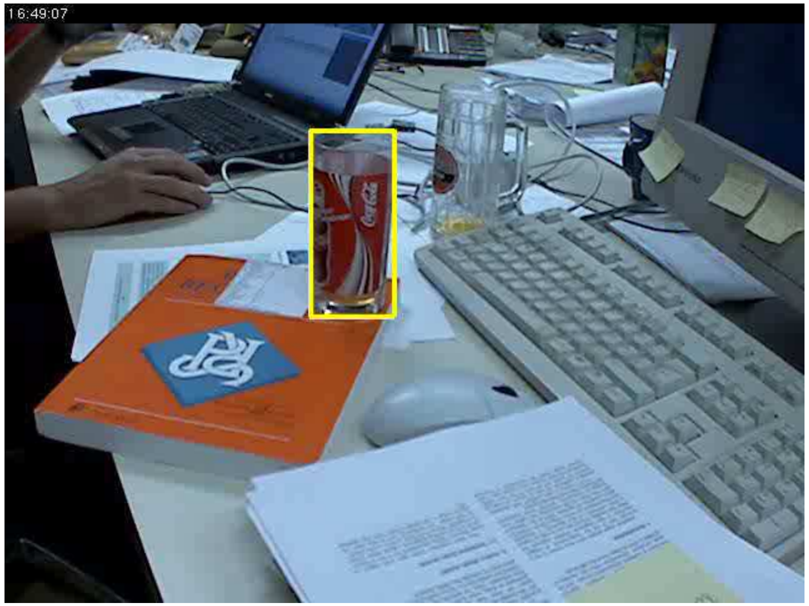
object

vs.

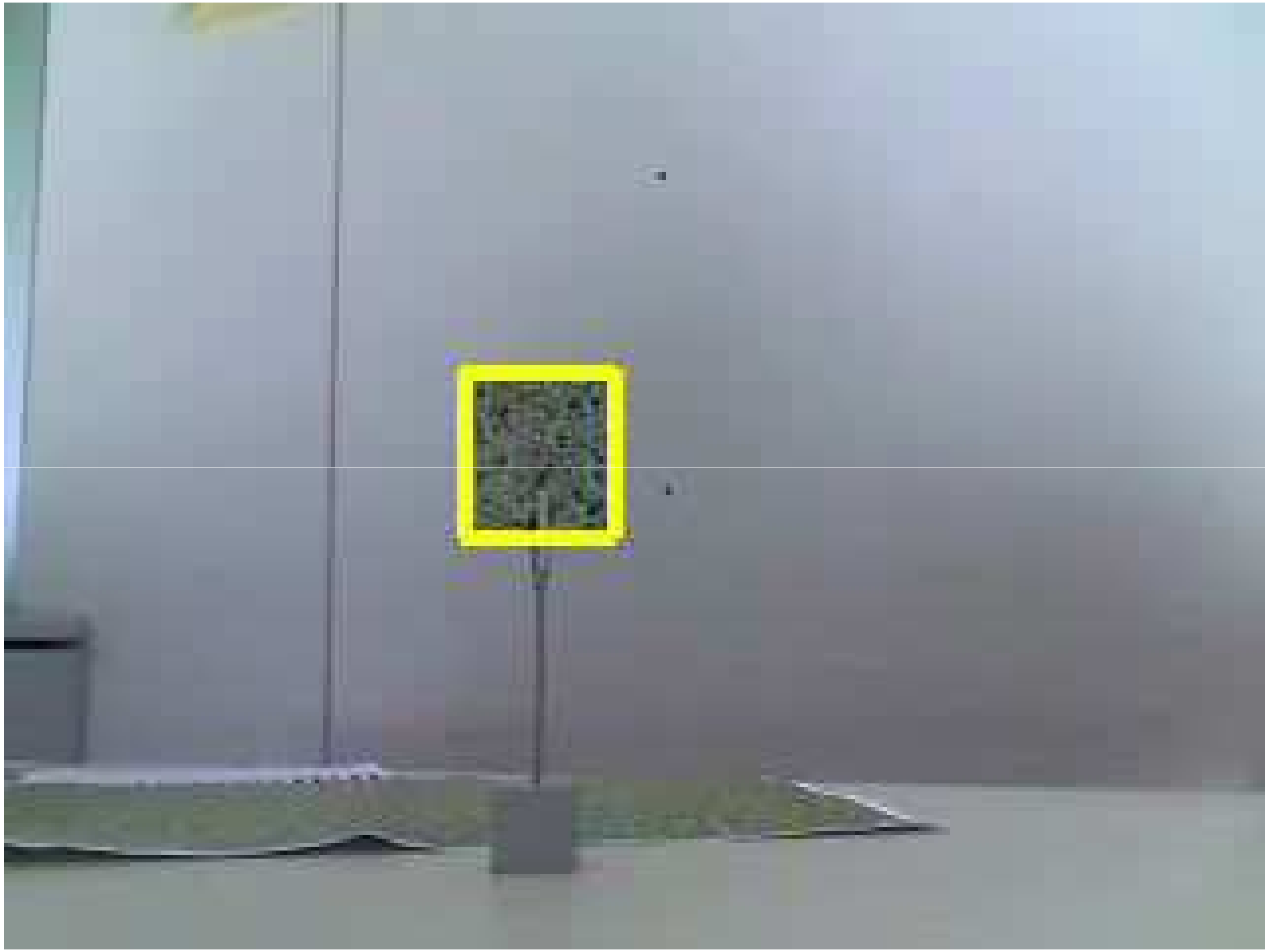
background


actual object position
from time t to t+1
evaluate classifier on sub-patches
search Region
create confidence map
confidence
analyze map and set new object position
update classifier (tracker)
+
-
-
-
-

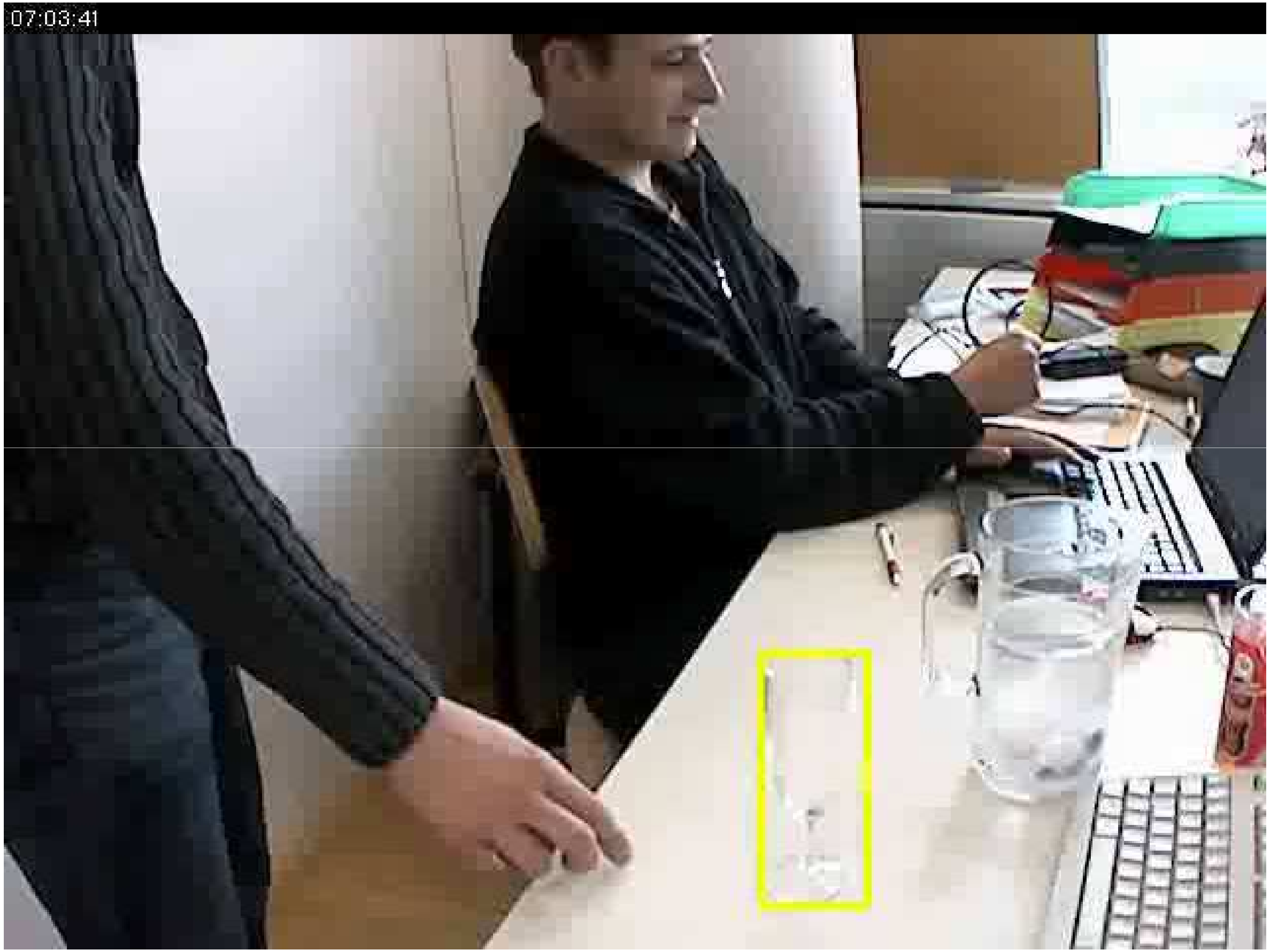



16:49:07

# “Simple tracking”



# “Tracking the invisibvle”



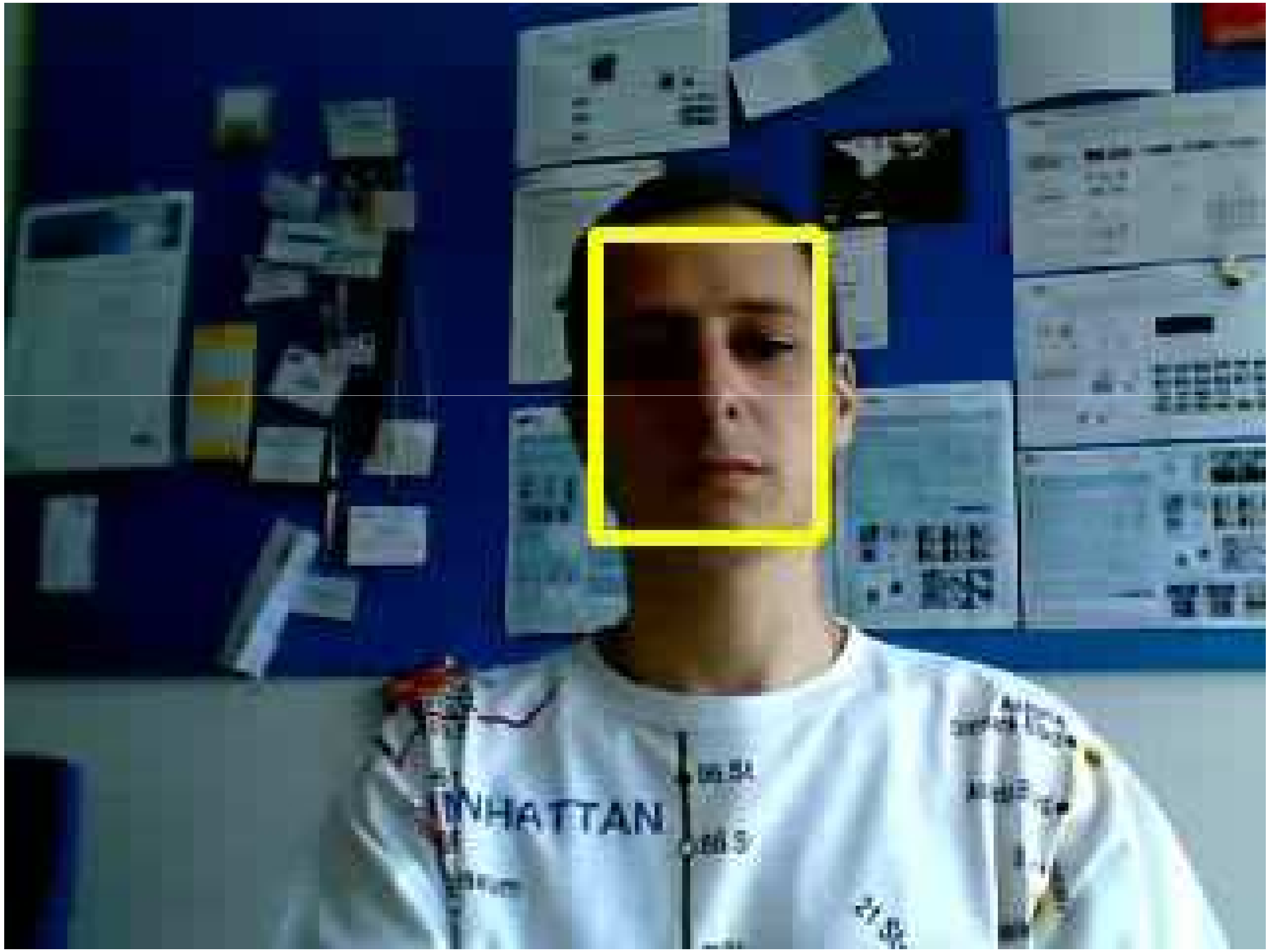
# LESSON LEARNED 4

**Foreground/ background discrimination reduces ambiguities.**

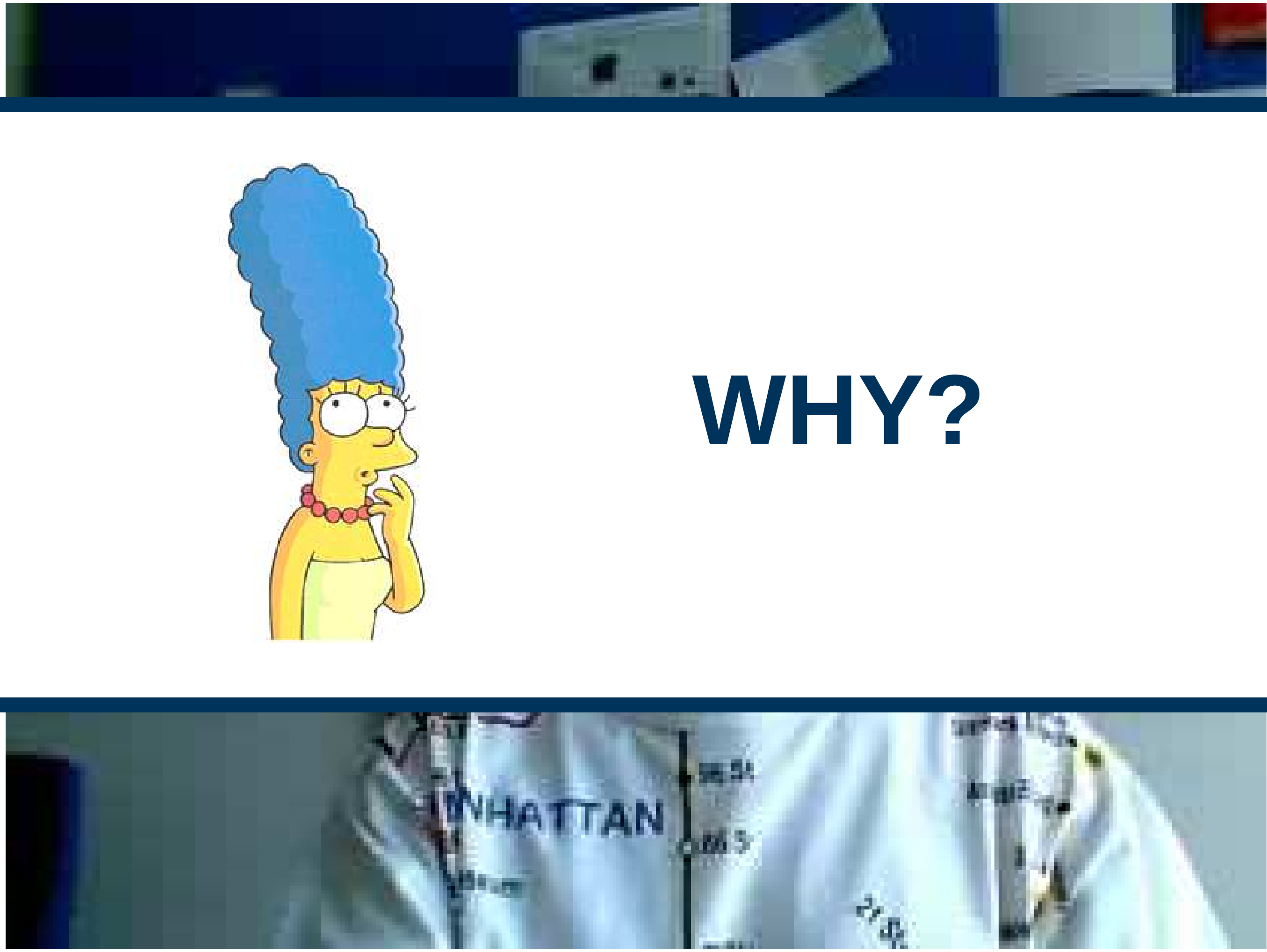
# Tracking Solved ☺


**Does it fail?**
**If yes, when?**

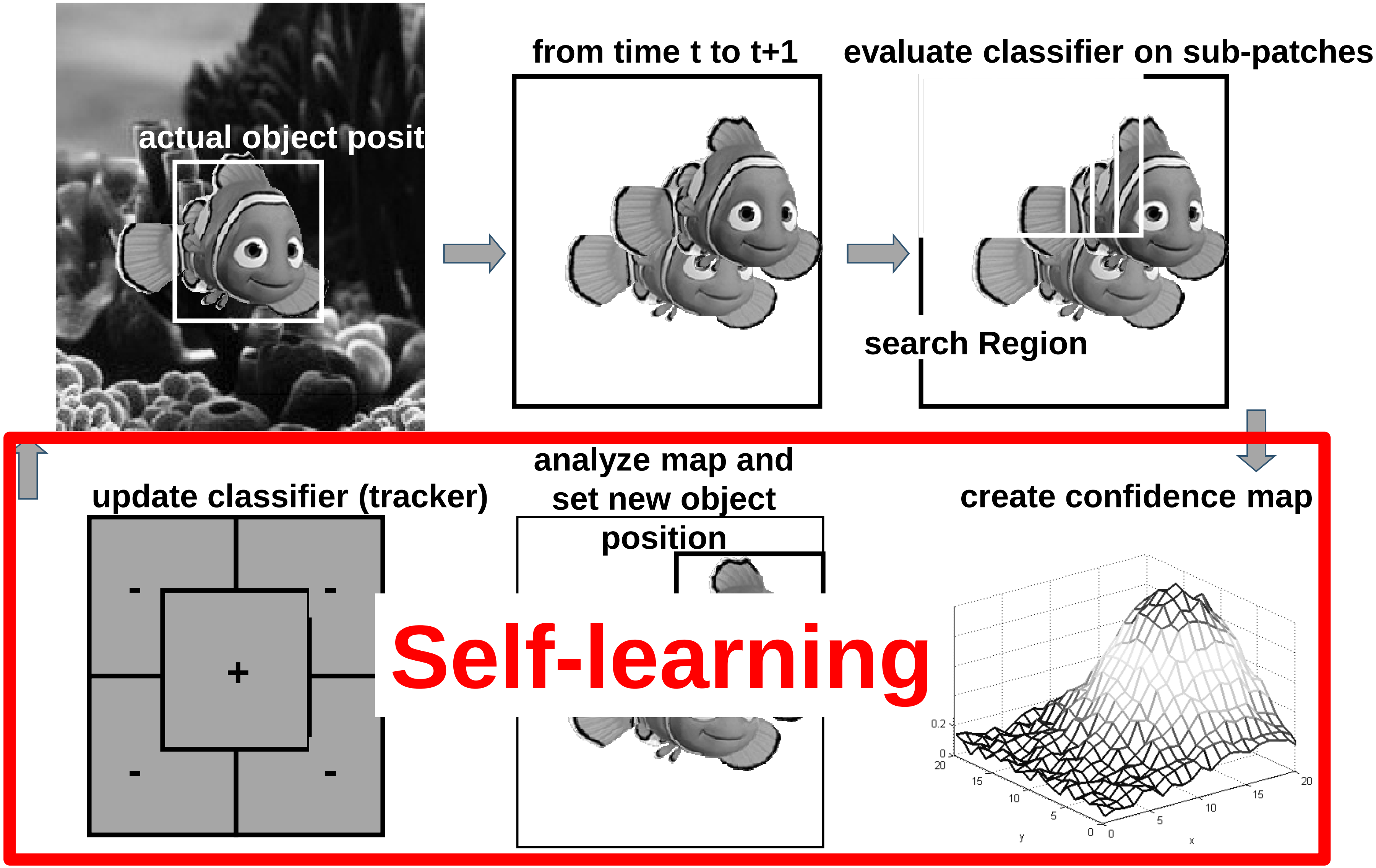
# When does it fail…



# When does it fail…

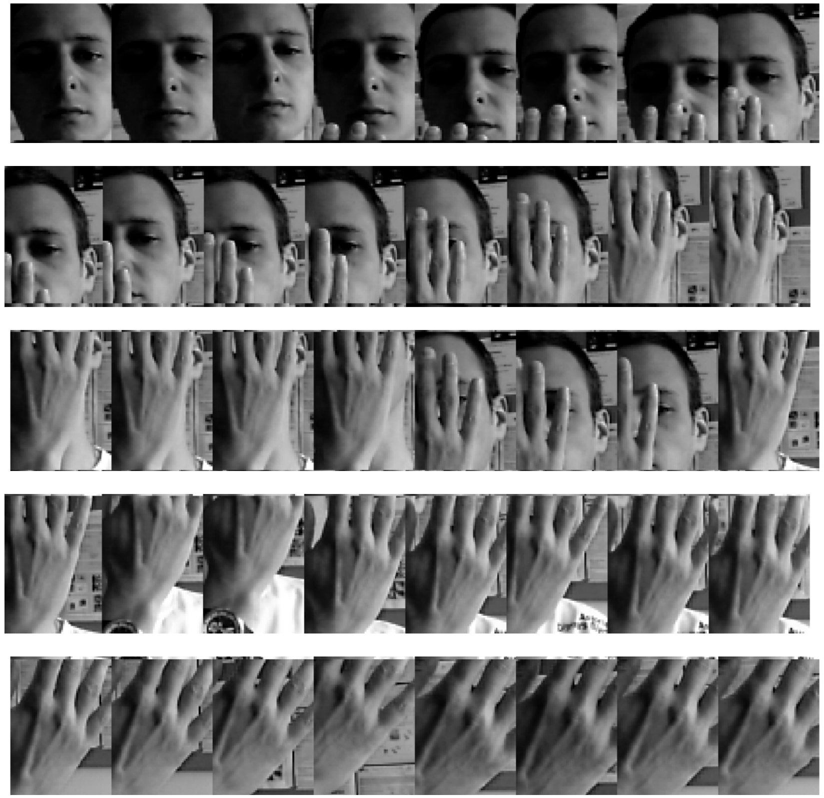



**WHY?**

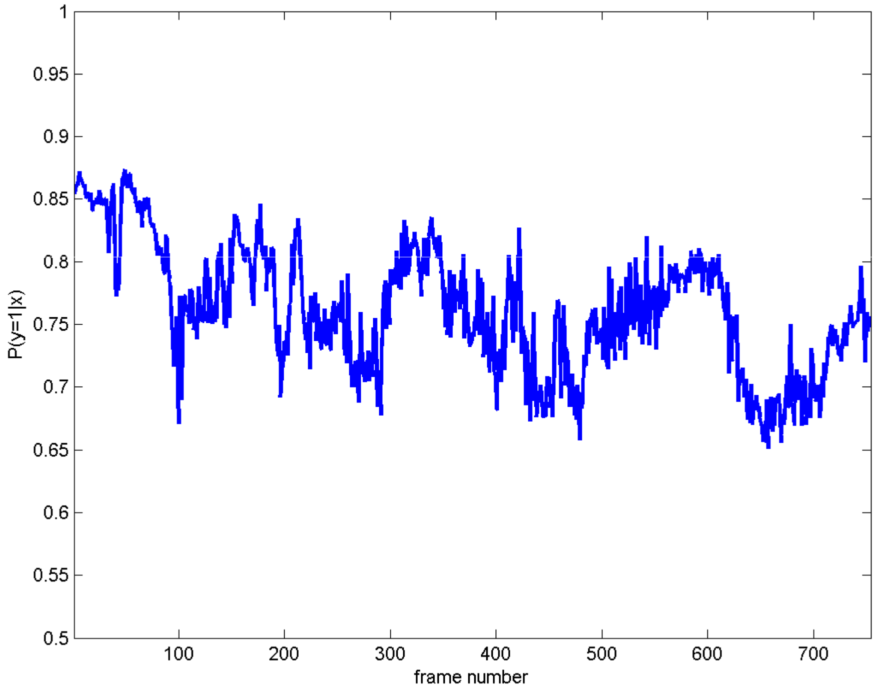

actual object posit
from time t to t+1
evaluate classifier on sub-patches
search Region
update classifier (tracker)
analyze map and
set new object
position
create confidence map
Self-learning
+
-
-
-
-
0.2
0
20
15
10
5
0
y
0
5
10
15
20
x

# Drifting

## Tracked Patches



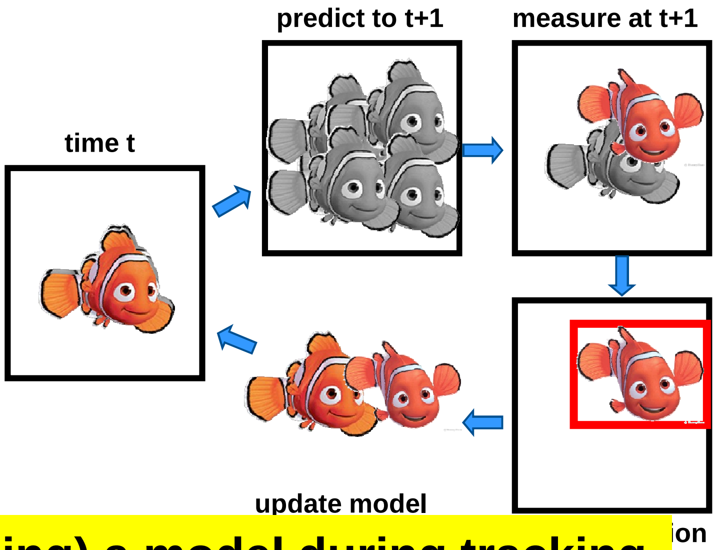
## Confidence



# LESSON LEARNED 5

**Self-learning $\rightarrow$ drifting!**

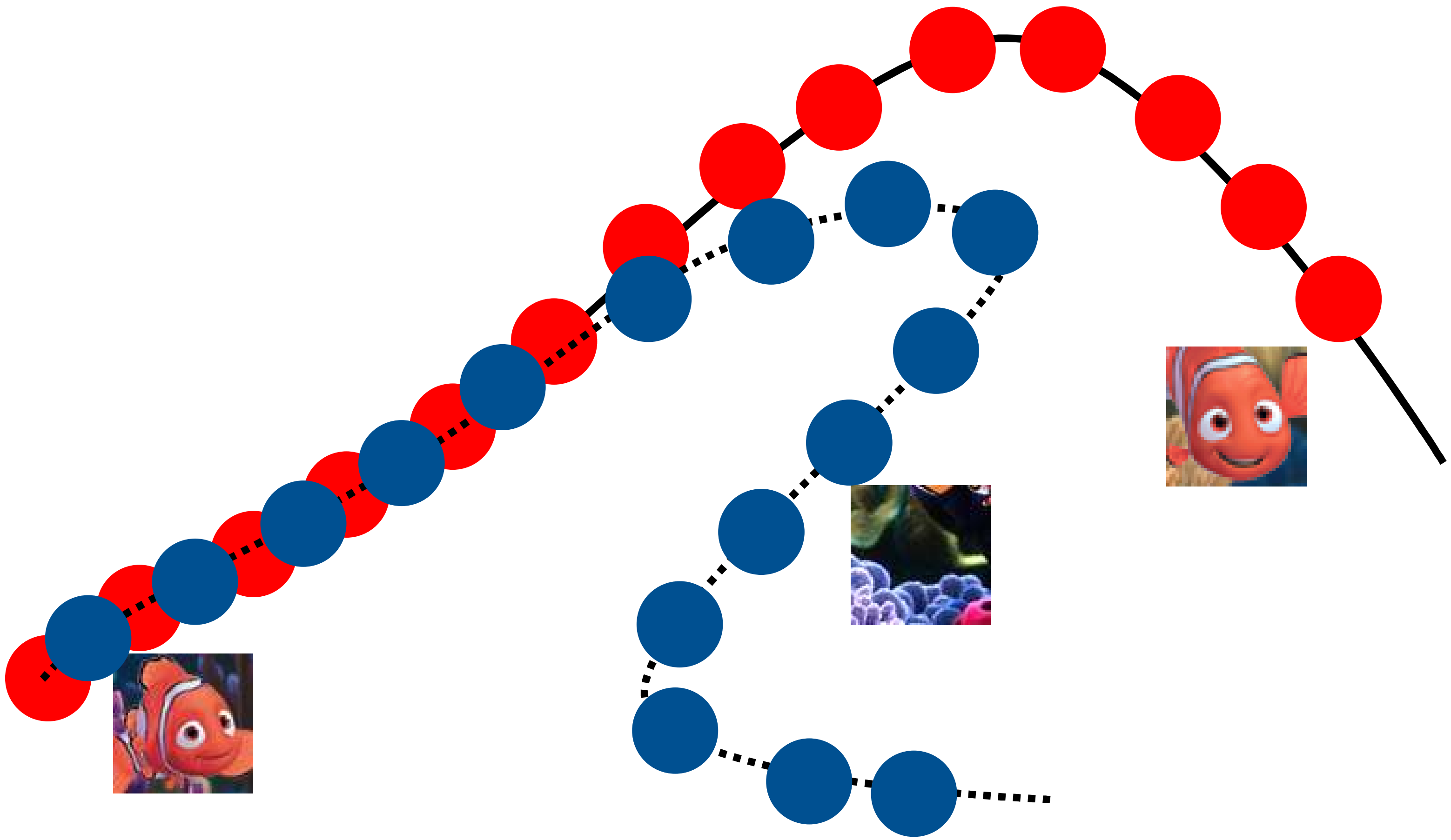
# TRACKING AND DETECTION



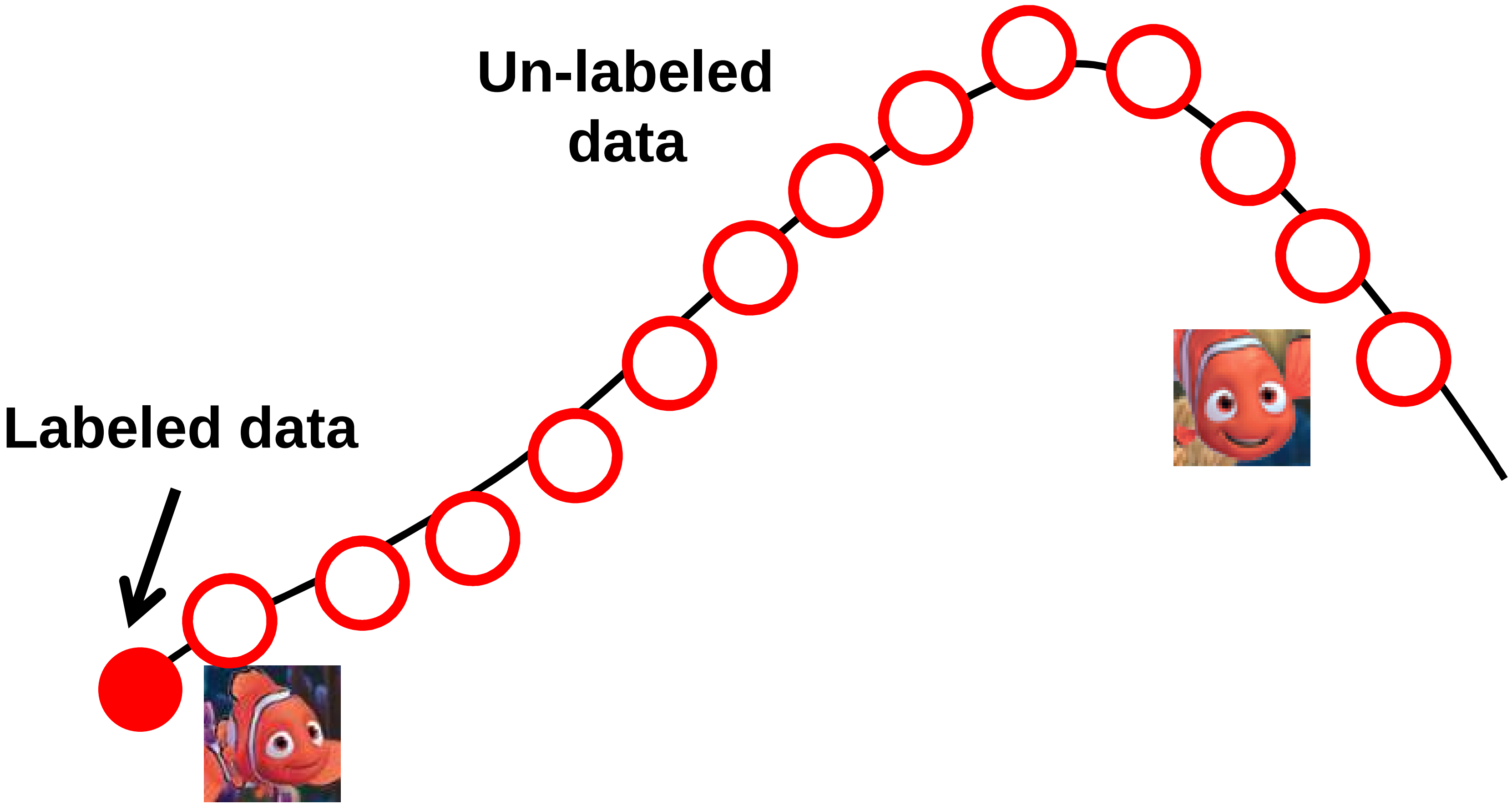

**learning (refining) a model during tracking**

[Grabner et al., ECCV'08]

# Review: Self-learning

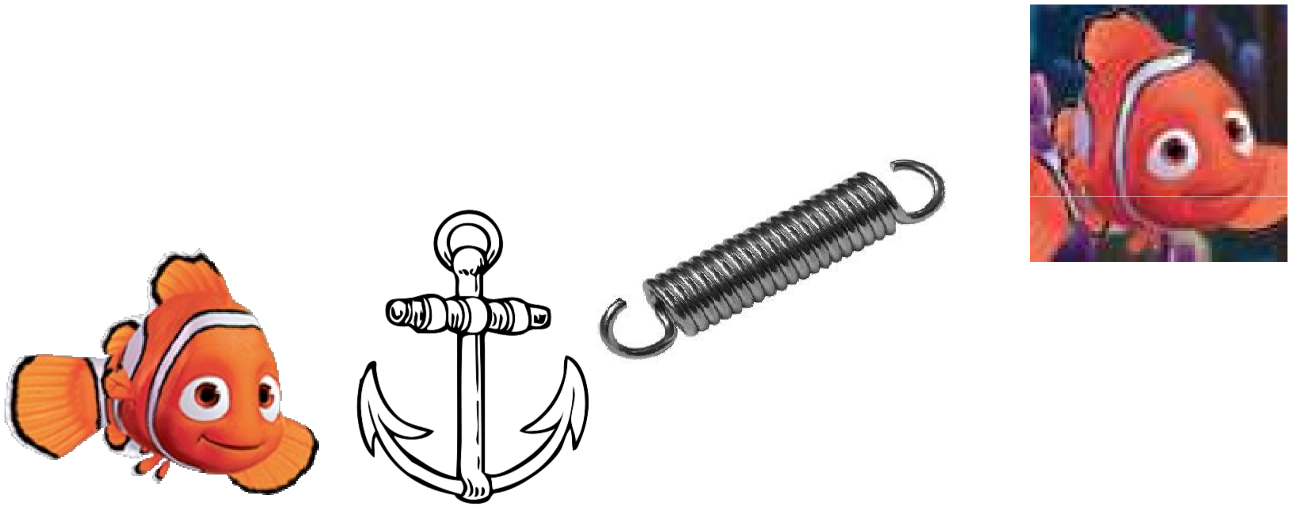


# Tracking is a semi-supervised problem!



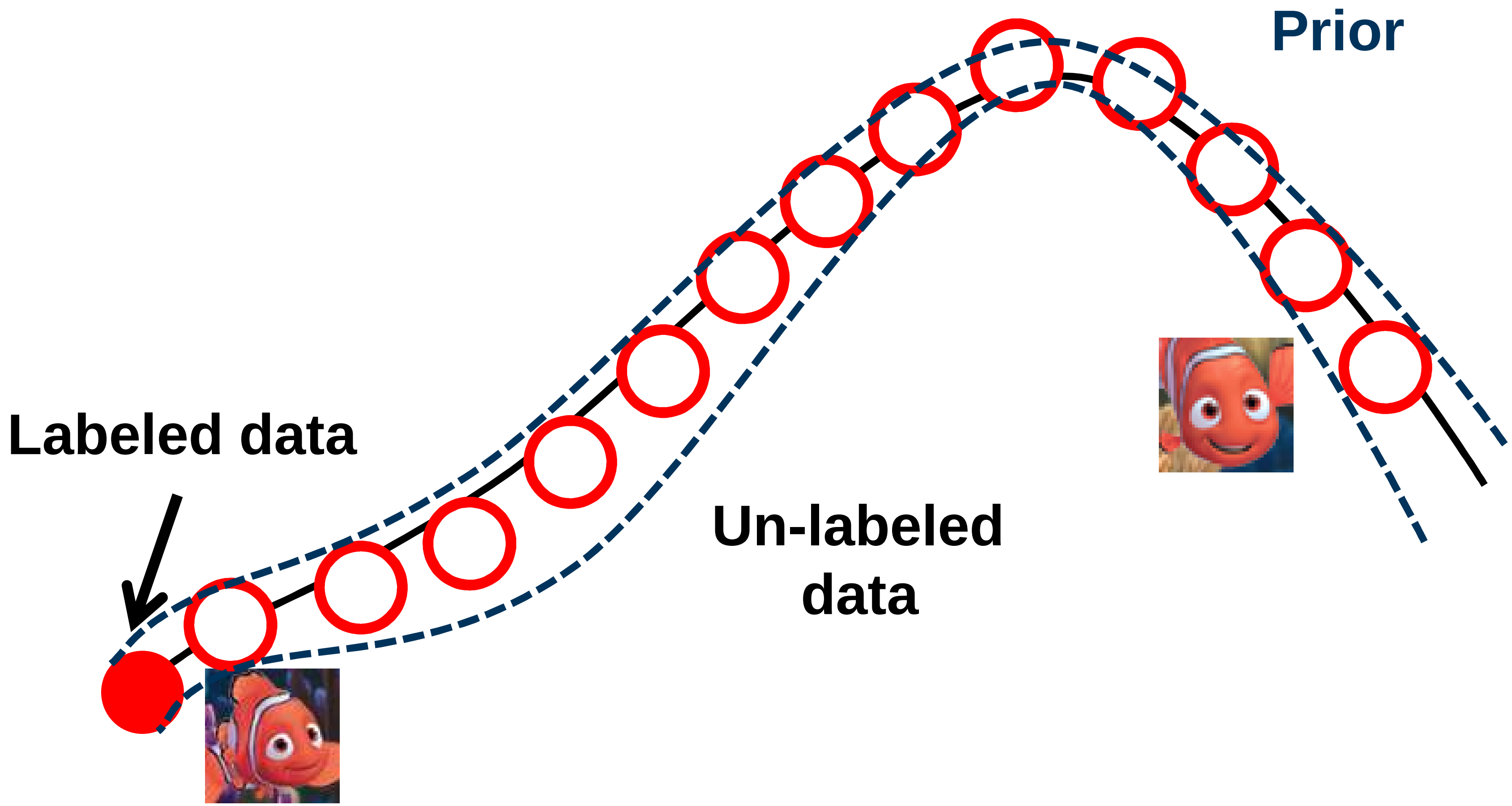
# Tracking is a semi-supervised problem!

**Current Model**

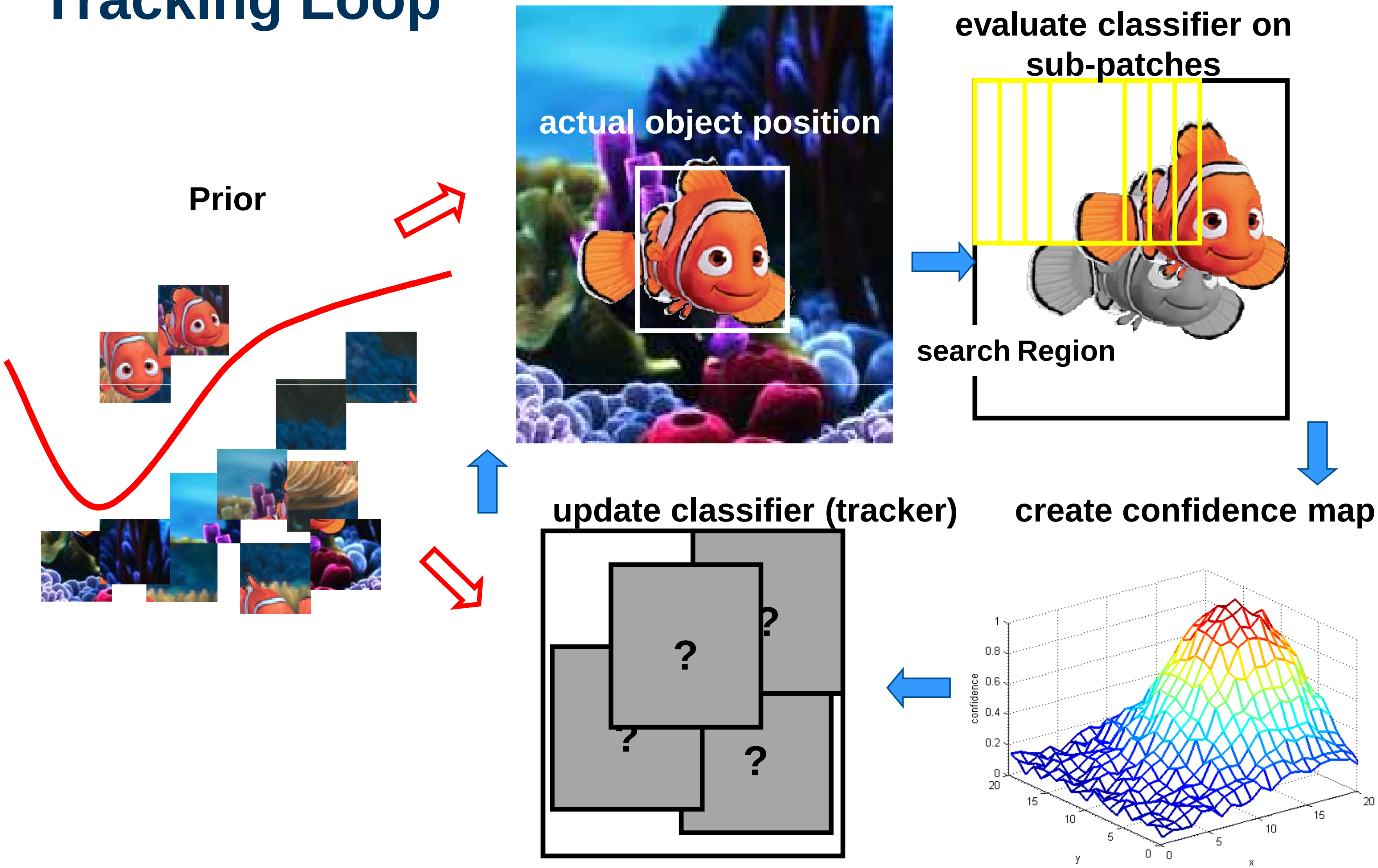


**Fix (initial) Model**
**Prior**

# Semi-Supervised Tracking



# Tracking Loop



# Robust Tacking

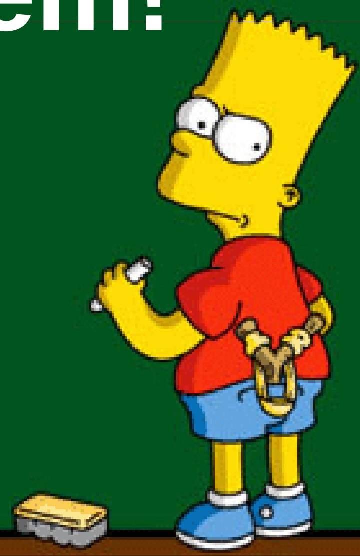
# Robust Tracking

# Drifting


CLICK HERE TO START

# LESSON LEARNED 6

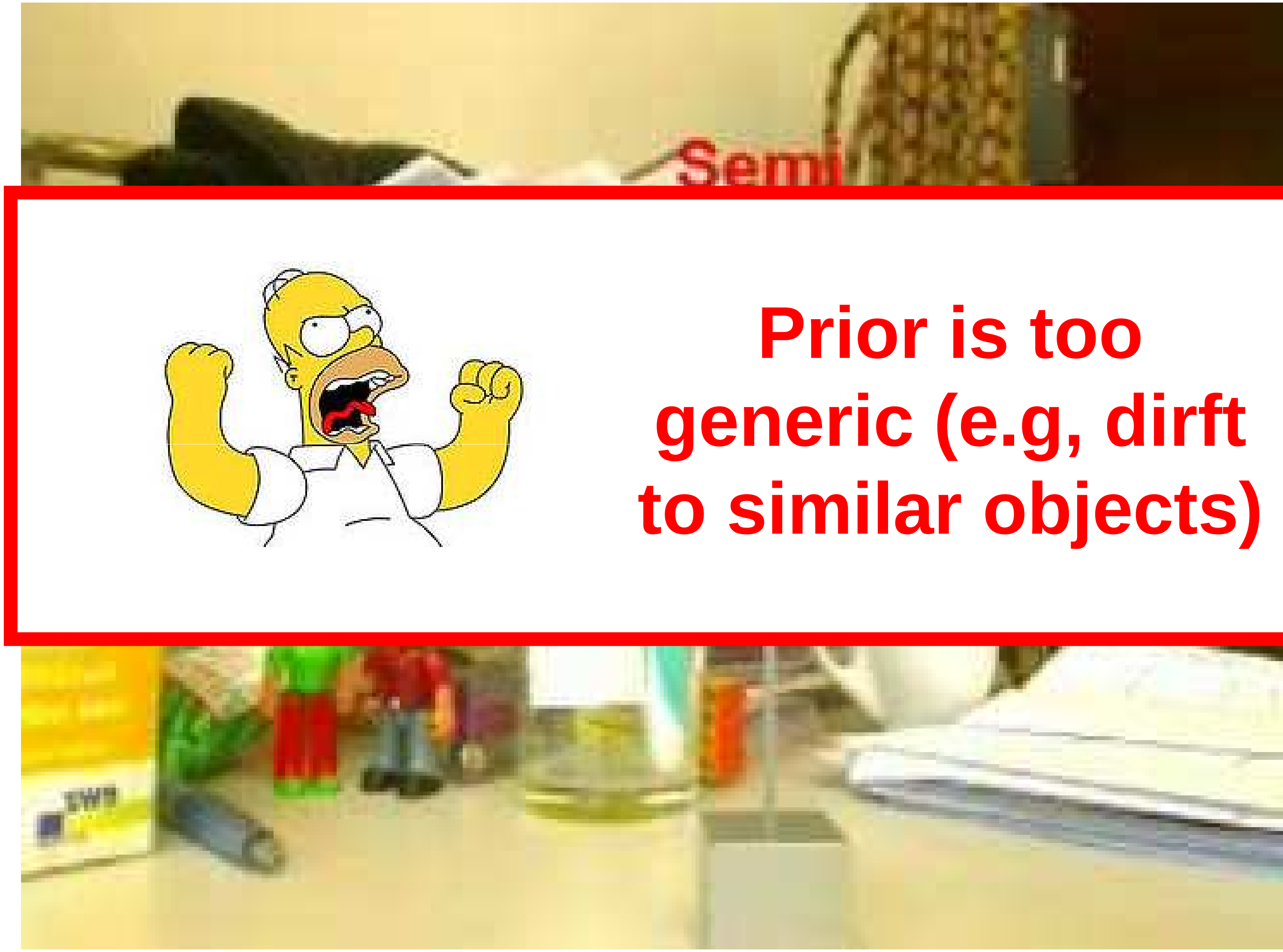
**Tracking is a semi-supervised problem!**



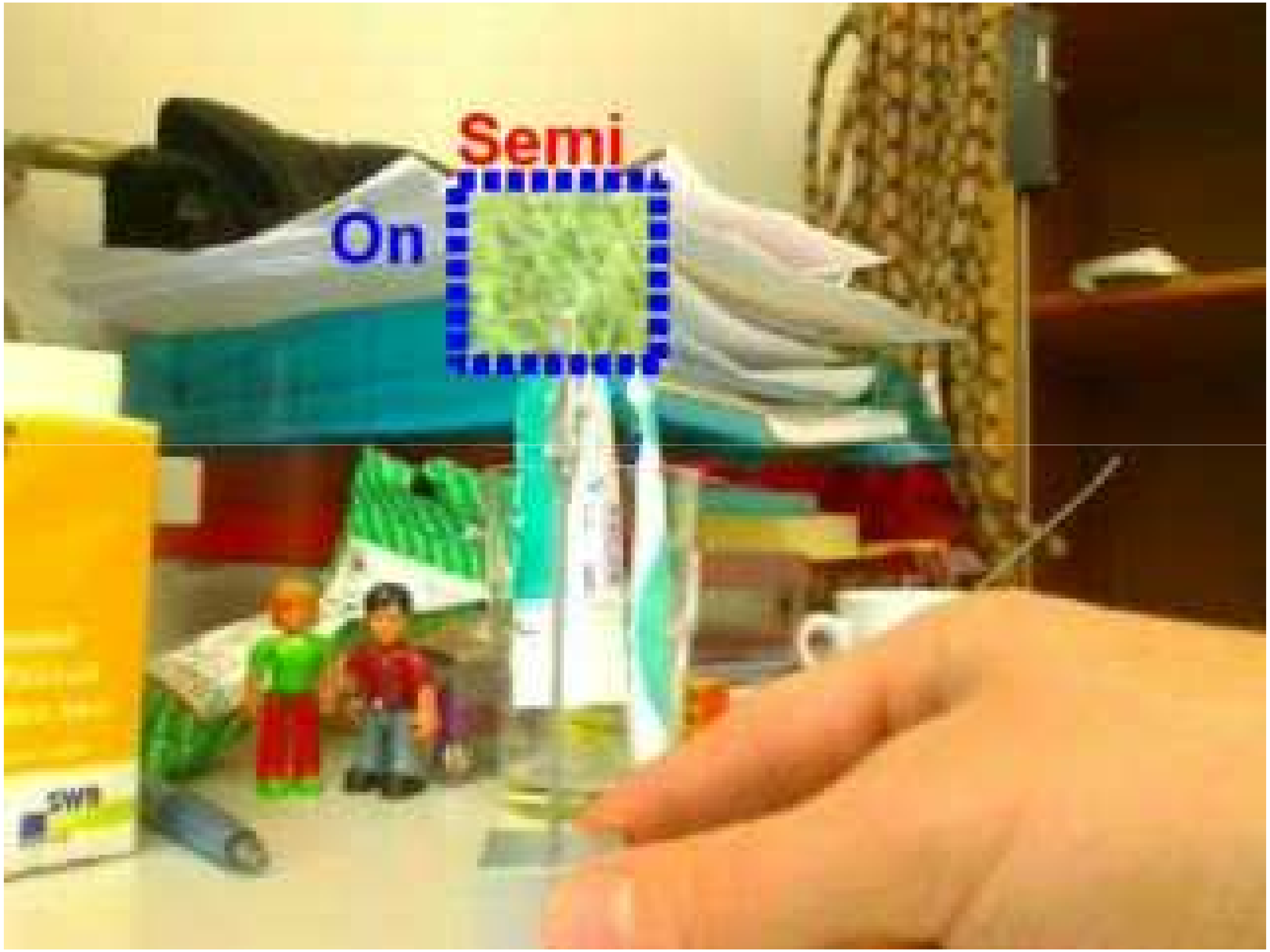
# Tracking Solved ☺


**Does it fail?**
**If yes, when?**


Semi
On




**Prior is too generic (e.g, dirft to similar objects)**

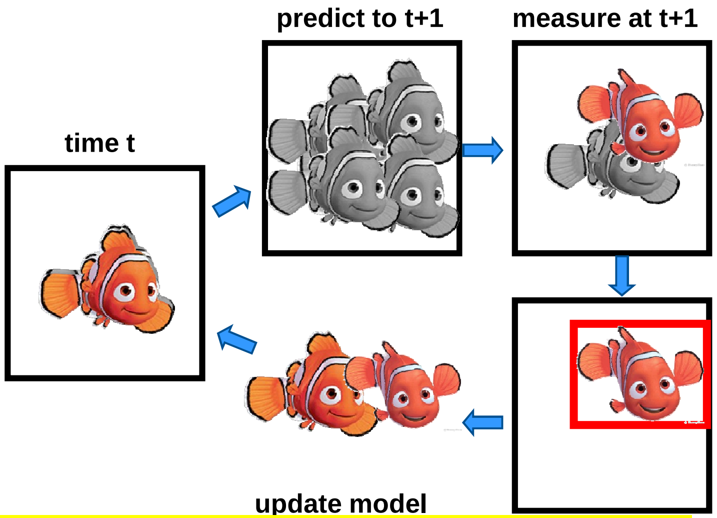

Semi
On


**Prior is too restrictive (e.g., partial occlusions)**

# LESSON LEARNED 7

**The prior (detector) is essential in semi-supervised machine learning.**

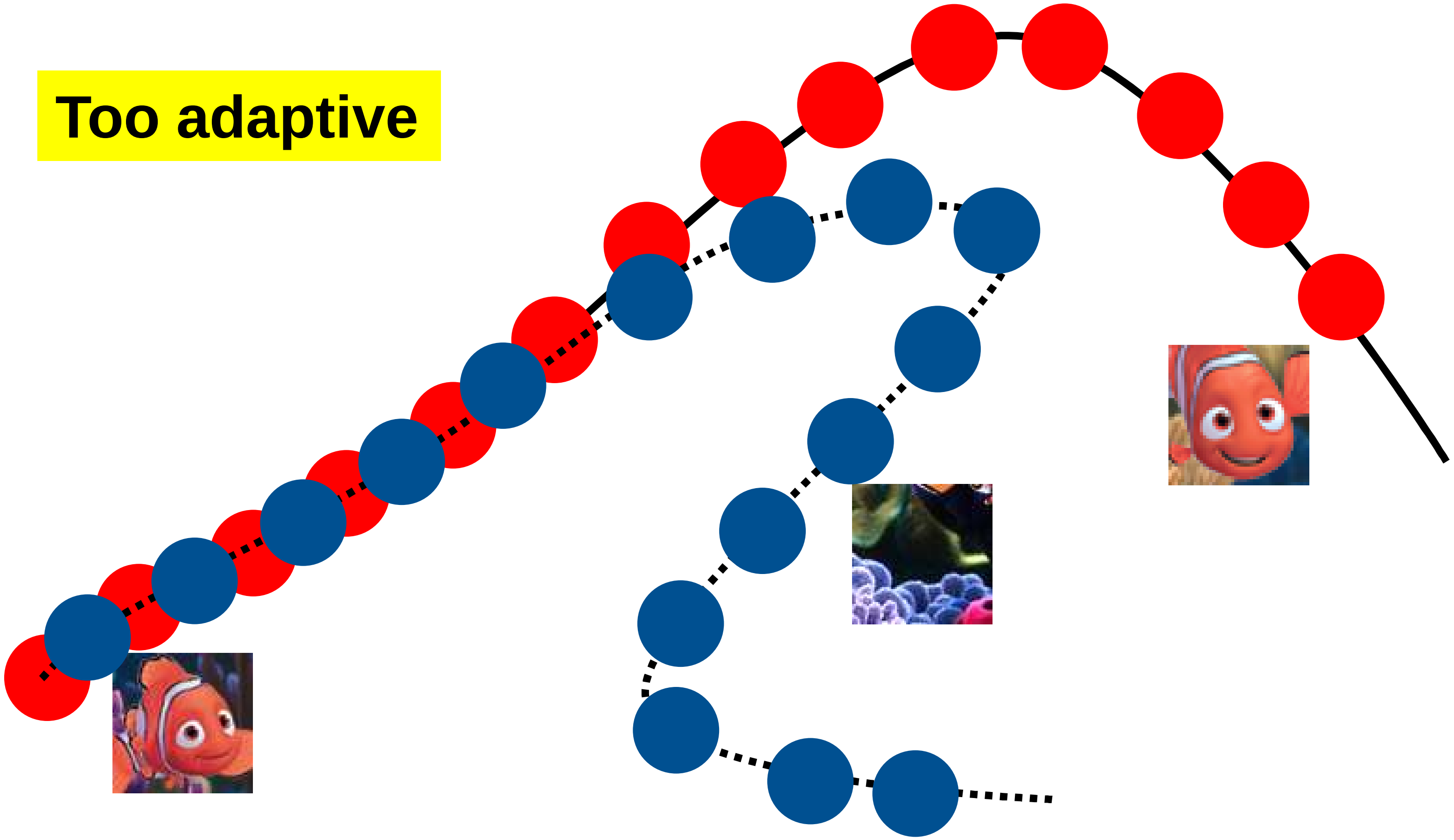
# SCENE CONTEXT



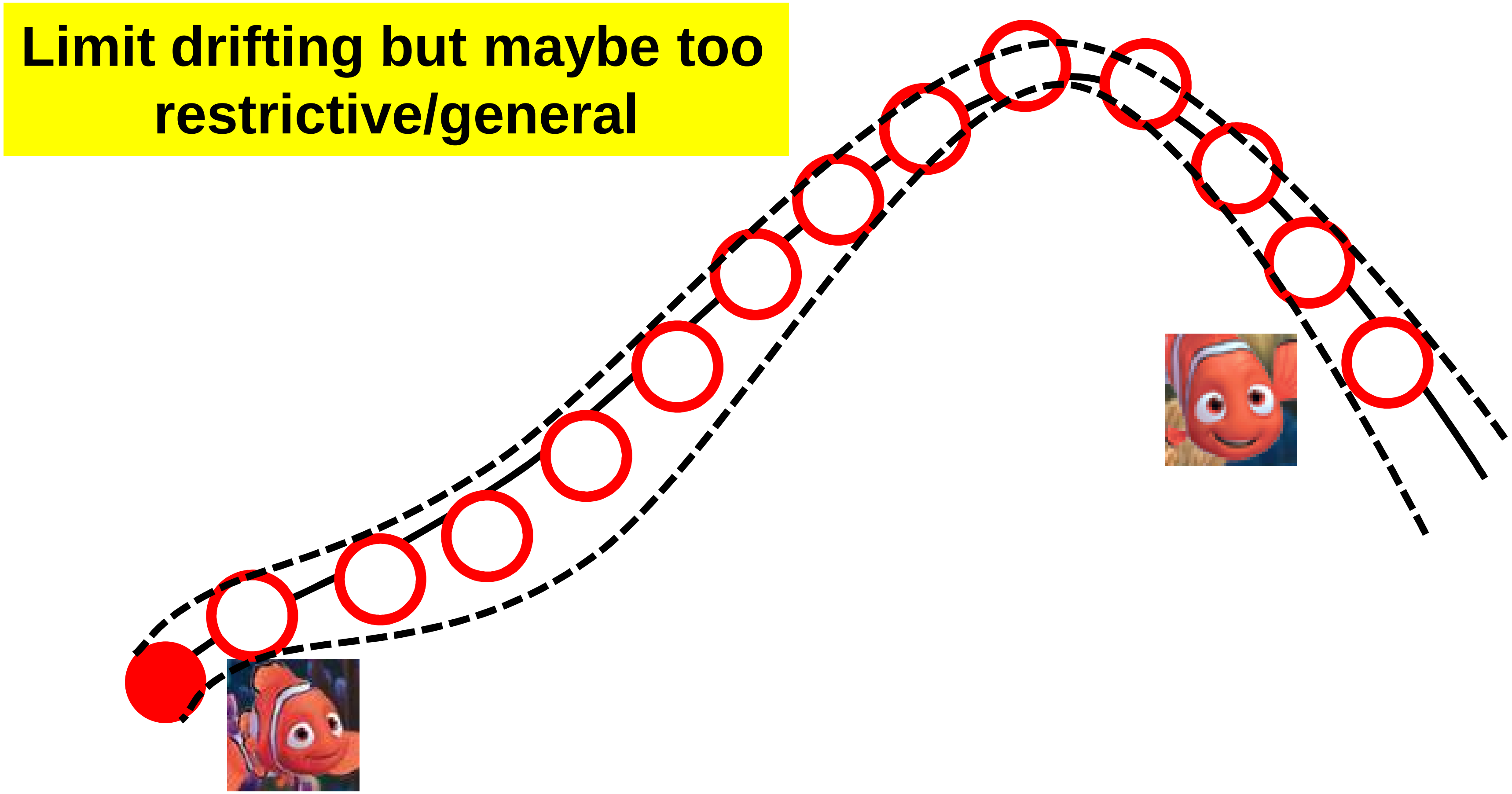

**learning (refining) a model during tracking II**

[Kalal et al. CVPR’10]
[Stalder et al. ICCV’09 WS on OLCV]

# Review: Self-Learning



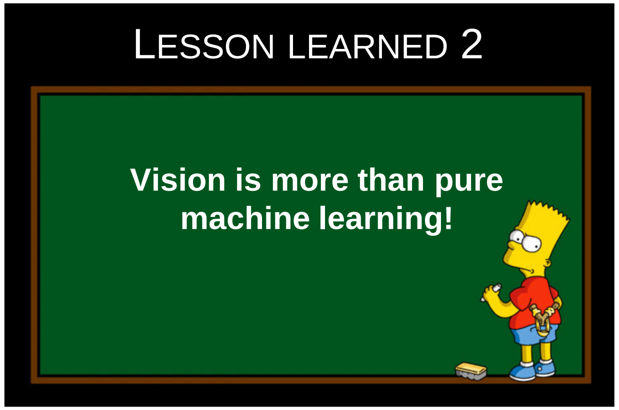
# Review: Semi-Supervised Tracking



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

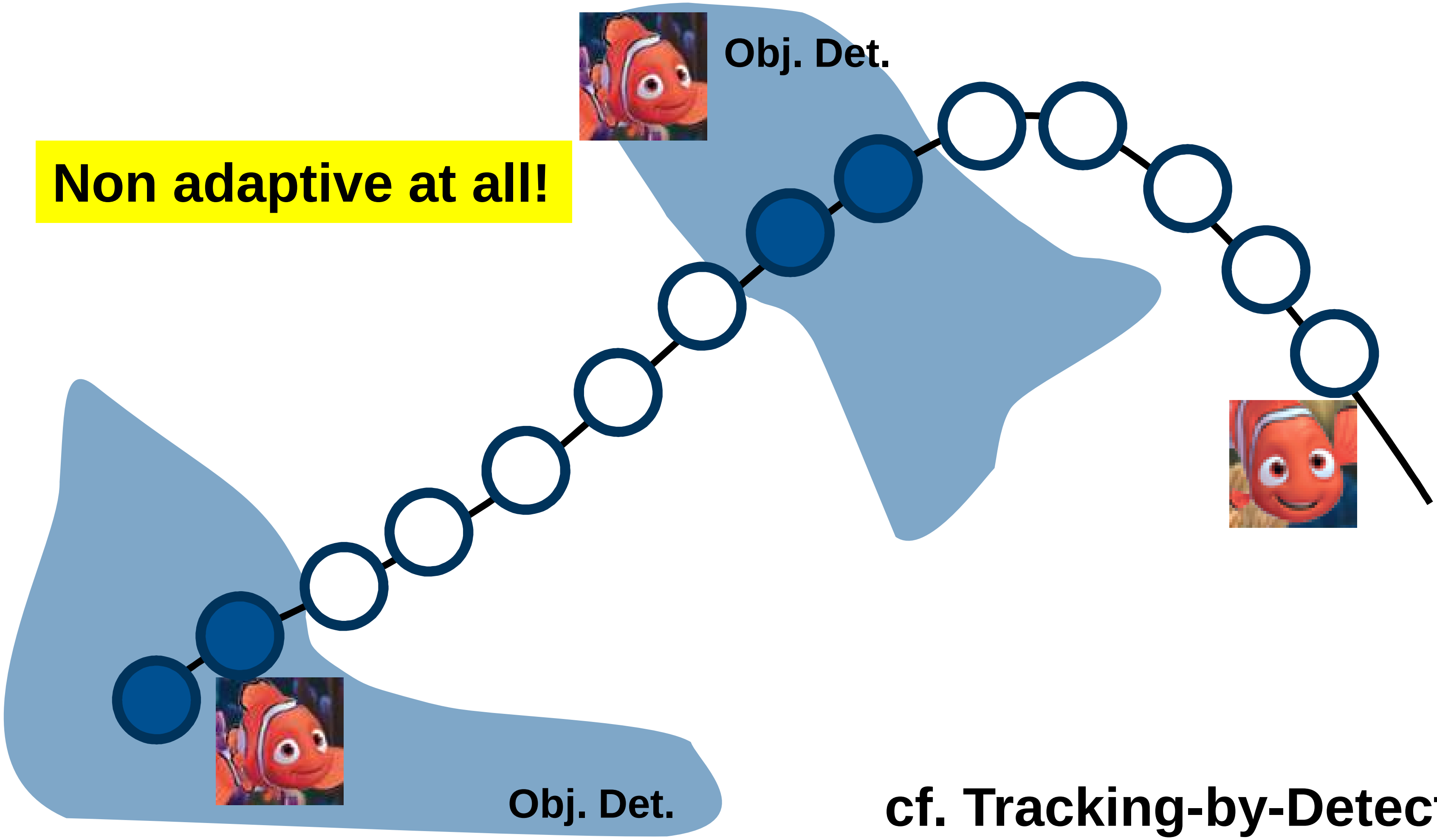
**Unlabeled data**

**Thomas Bayes**
**1702-1761**



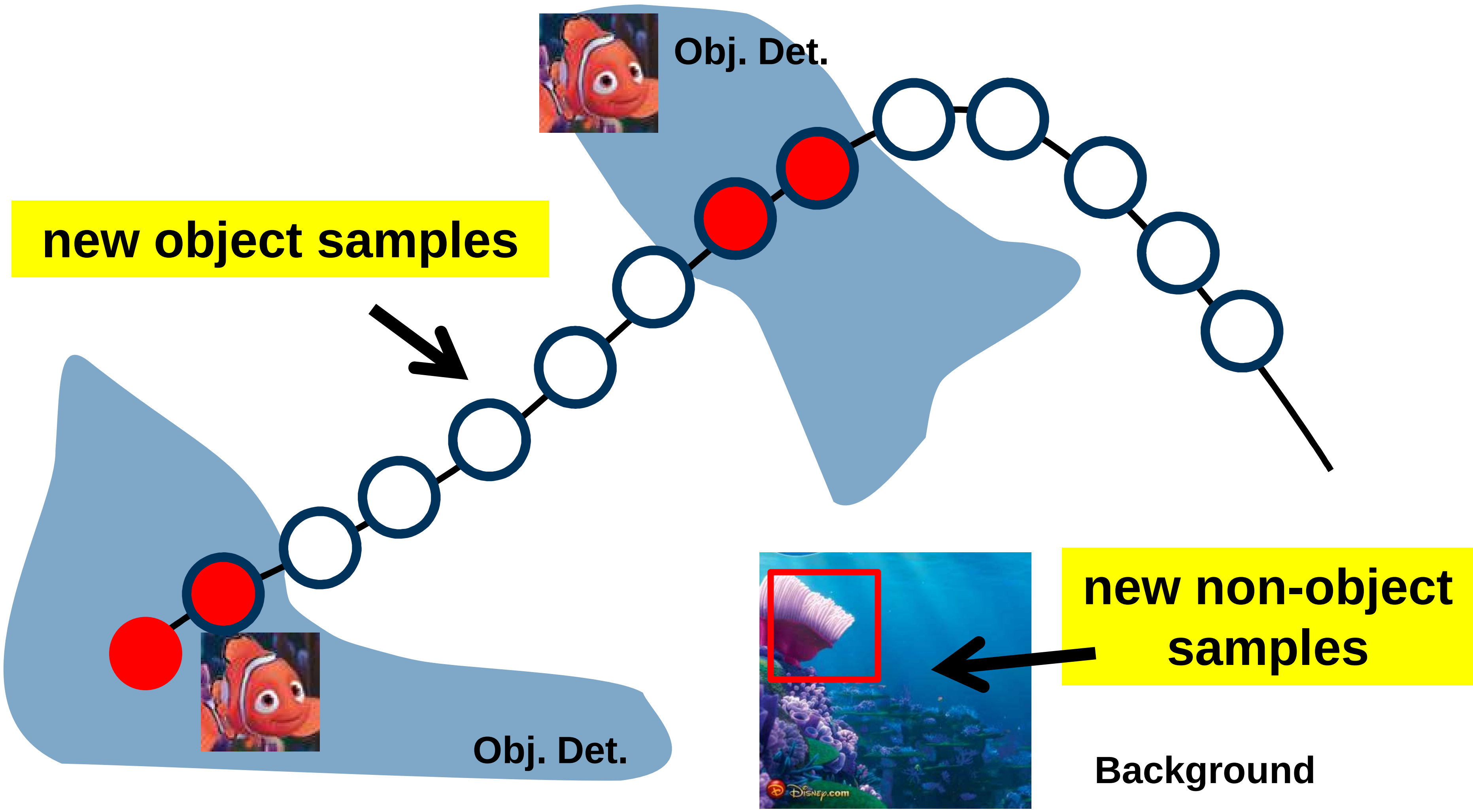
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

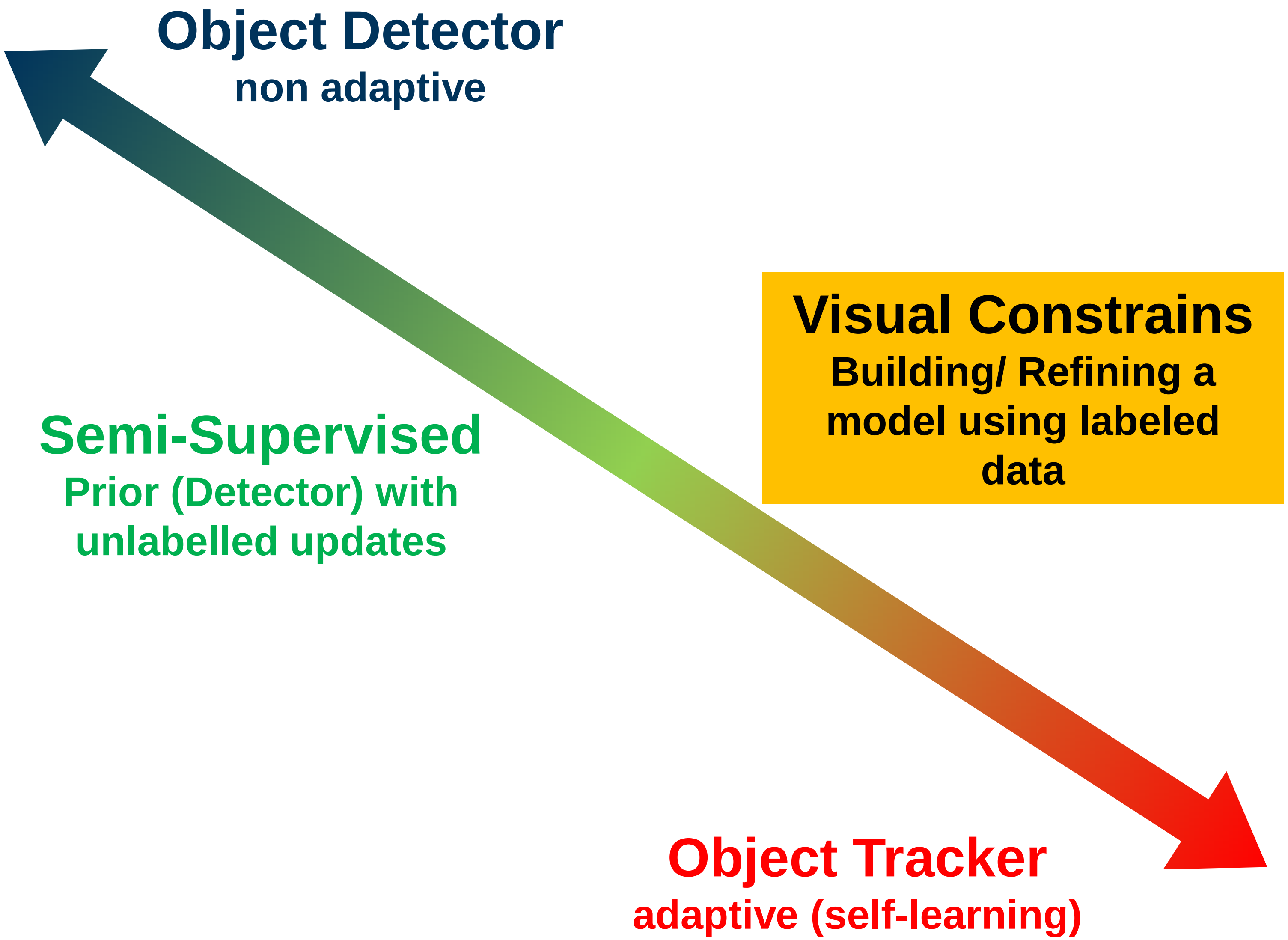
# Prior Model = Detector



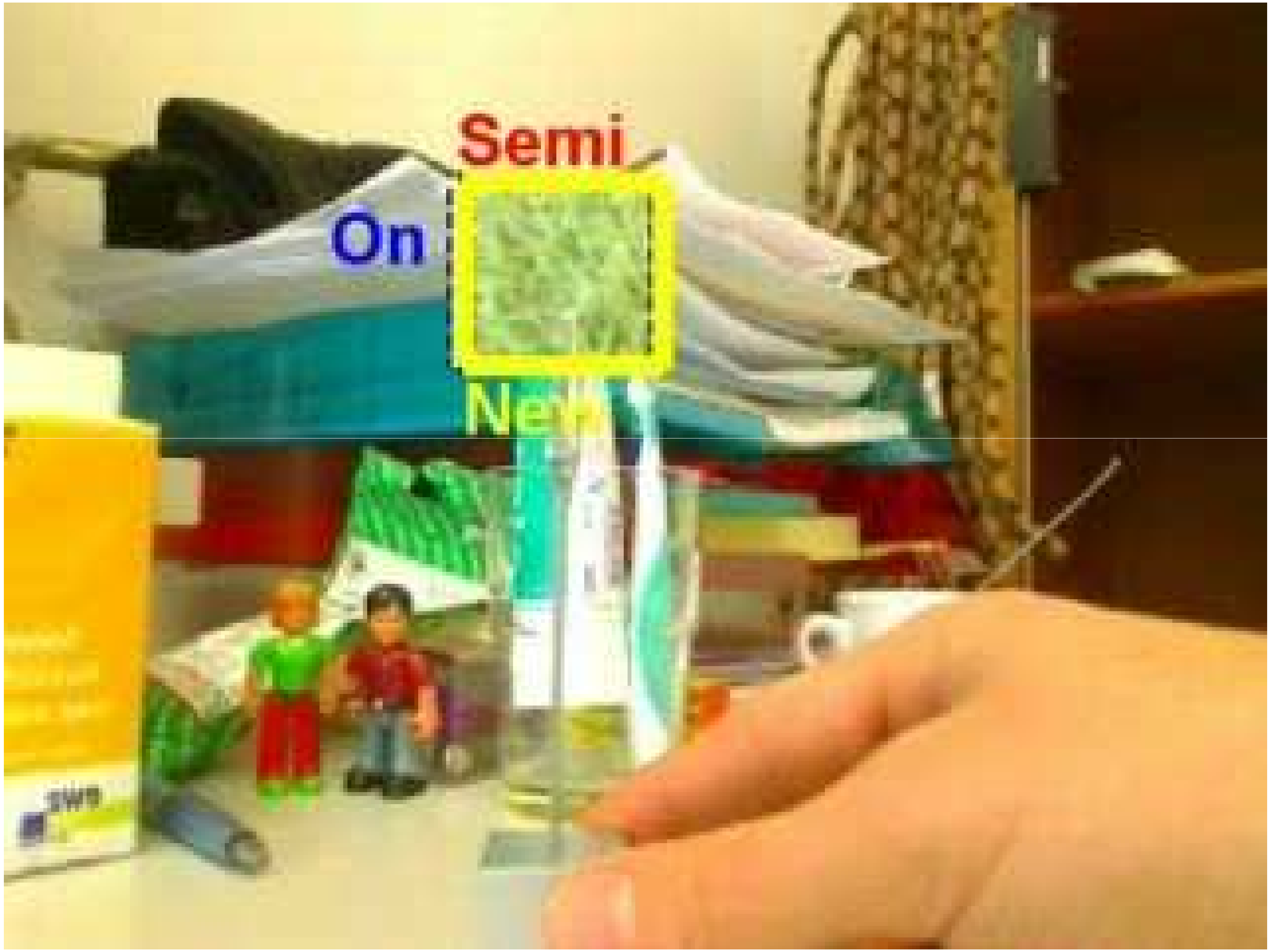

**cf. Tracking-by-Detection**

# Adapt the 'Prior' using Visual Constrains!




Object Detector
non adaptive
Semi-Supervised
Prior (Detector) with unlabelled updates
Visual Constrains
Building/ Refining a model using labeled data
Object Tracker
adaptive (self-learning)

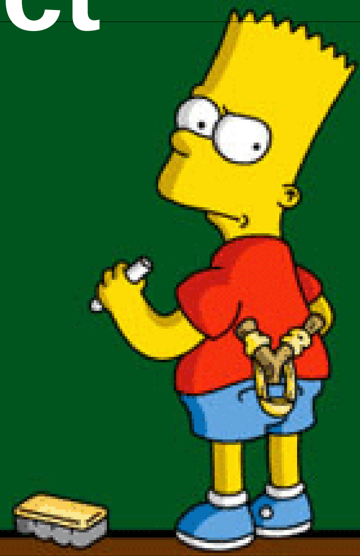

Semi
On
New


Semi
On
New

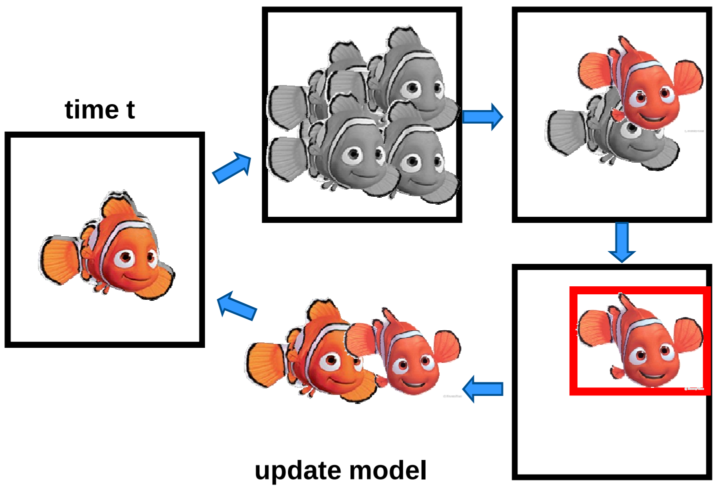

Learning
#2, fps:0.23, 1/28785

# LESSON LEARNED 8

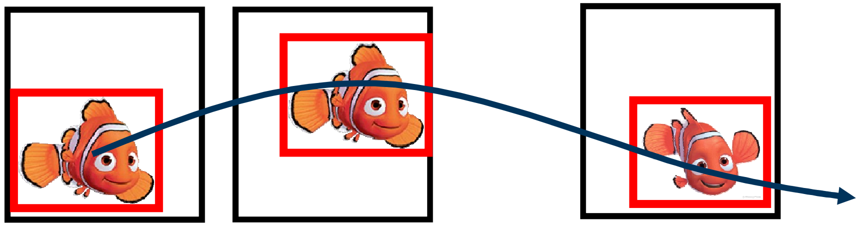
**Use visual constrains for learning an object detector.**



# OTHER OBJECTS




**exploit relations to other objects**

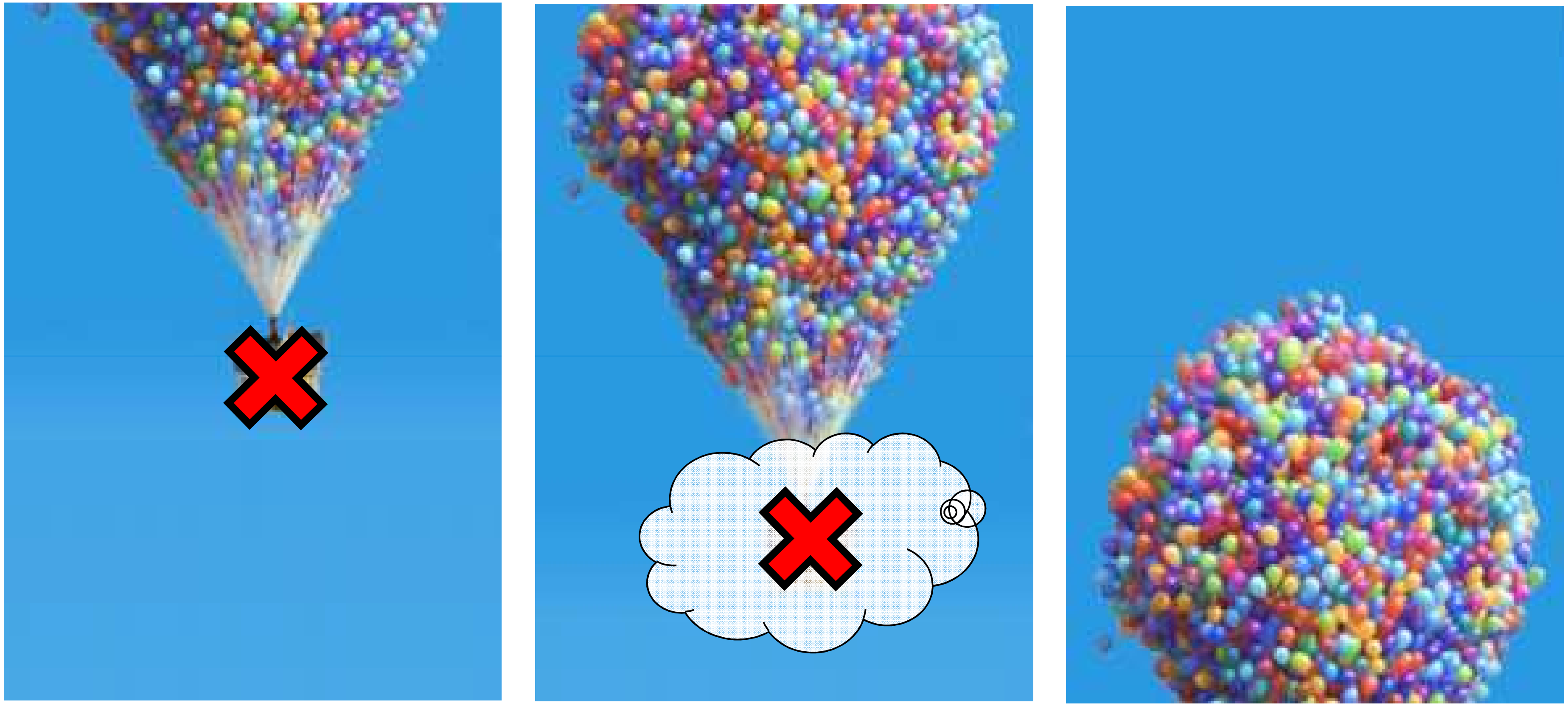



[Grabner et al., CVPR’11]

# I’m Carl – Track me…

# Tracking Carl

# Goal: Estimate the Position of an Object



Track the Object… … with occlusions … outside the image

**Many temporary, but potential very strong links exists between a tracked object and parts of the images.**

**We discover the dynamic elements – called SUPPORTERS – that predict the position of the target.**

# SUPPORTERS…

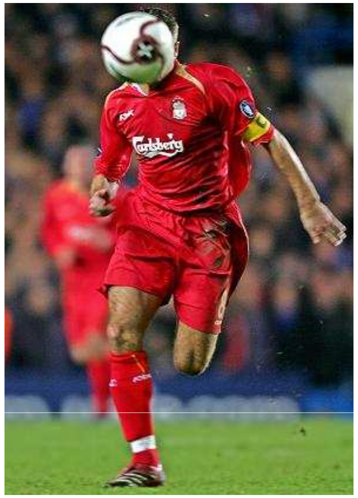
- … came with different strength.
- … change over time.

# SUPPORTERS…

- … came with different strength.
- … change over time.

# SUPPORTERS help Tracking of…

- … objects which change there appearance very quickly.
- … occluded objects or object outside the image.
- … small and/or low textured objects or even “virtual points”.

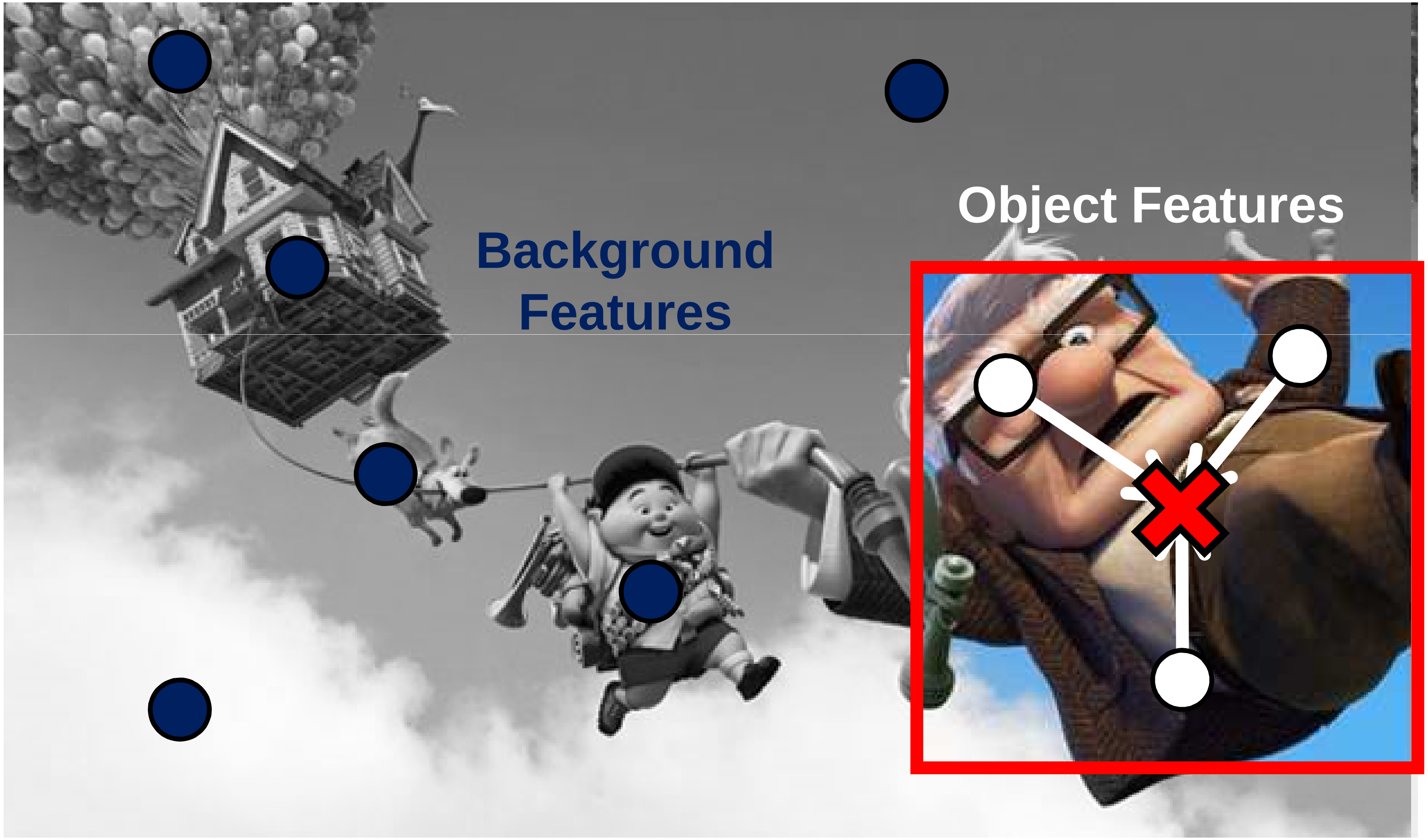



# Local Image Features as Supporters

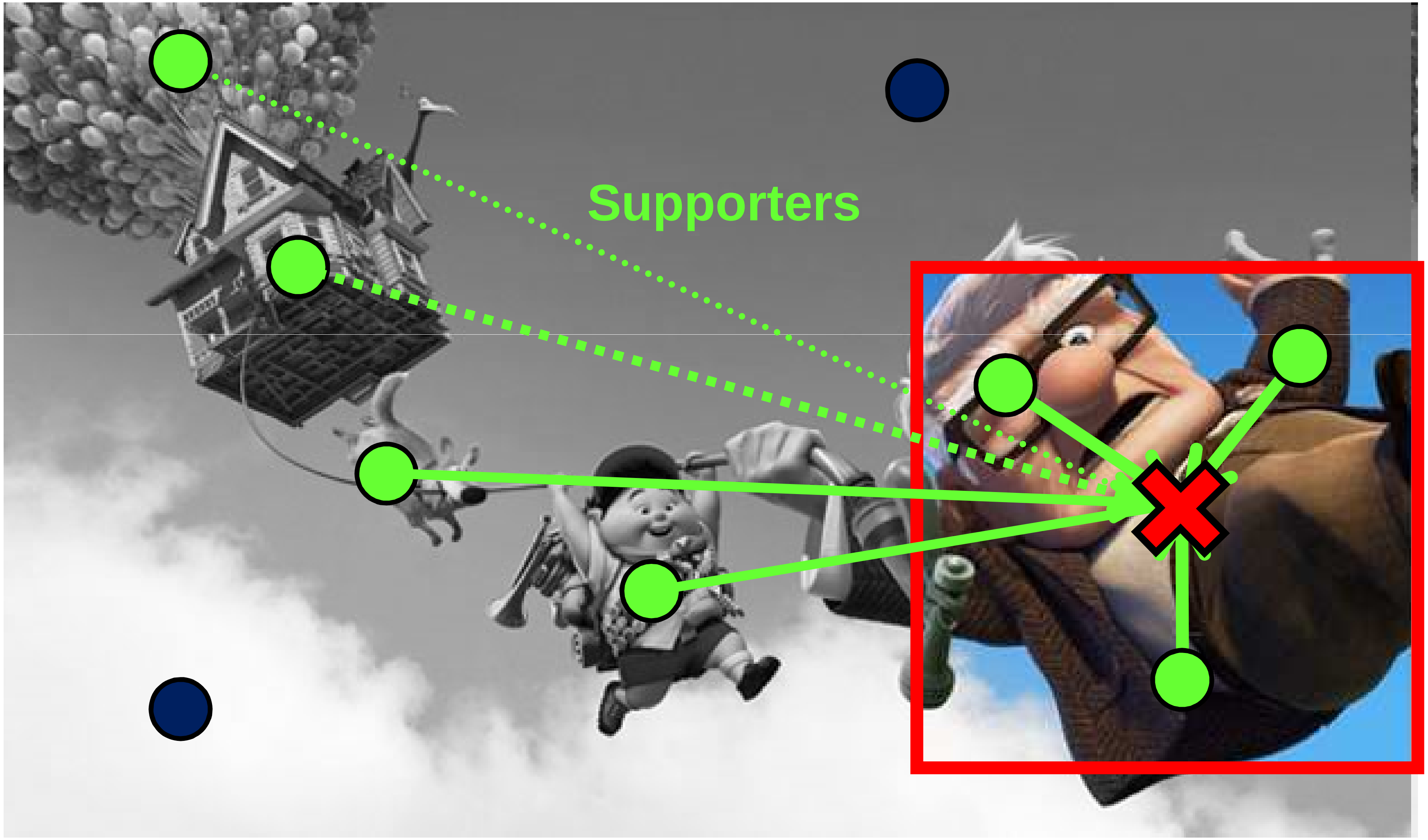
# Local Image Features as Supporters

# Local Image Features as Supporters

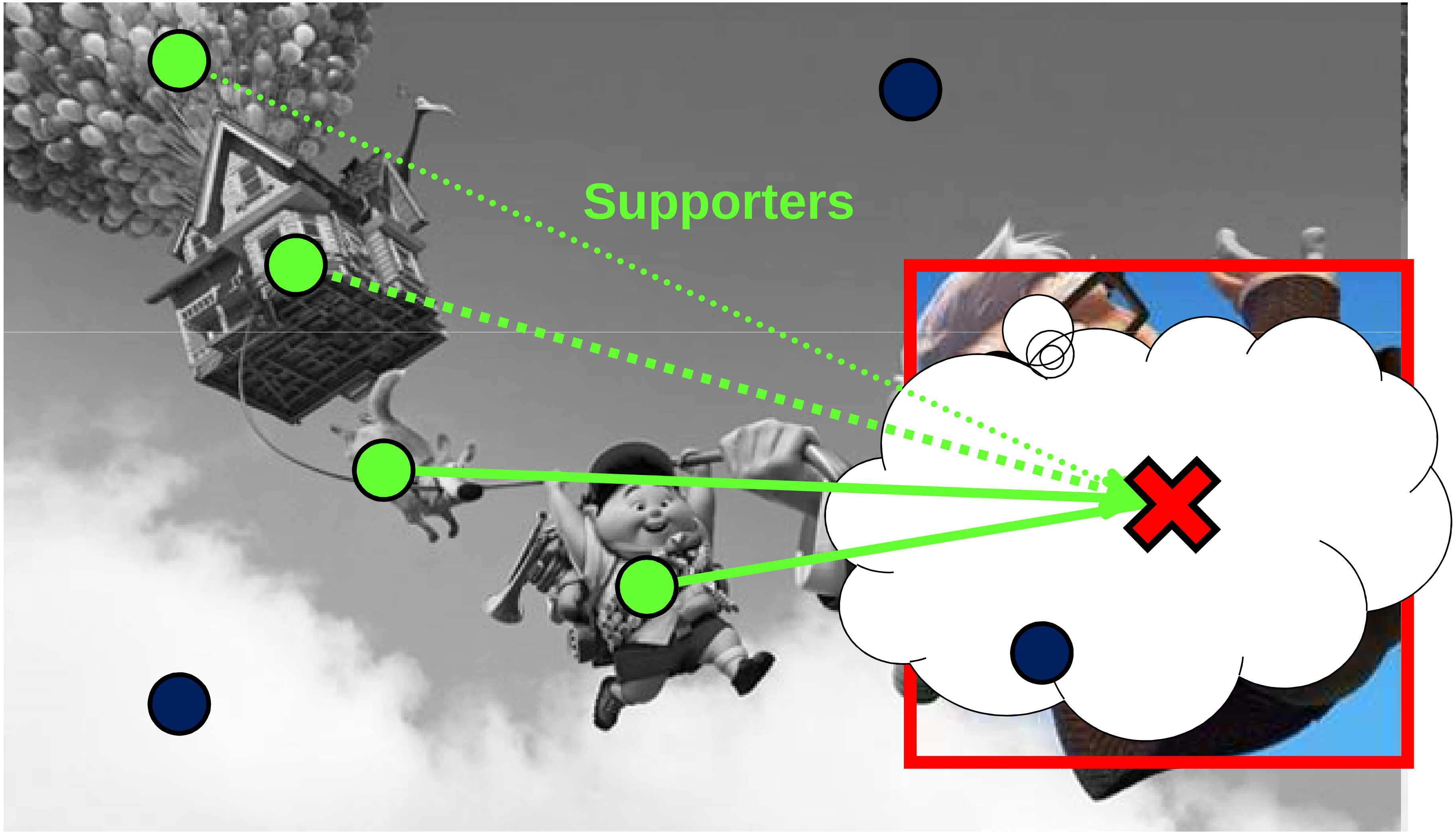
# Local Image Features as Supporters



# Local Image Features as Supporters



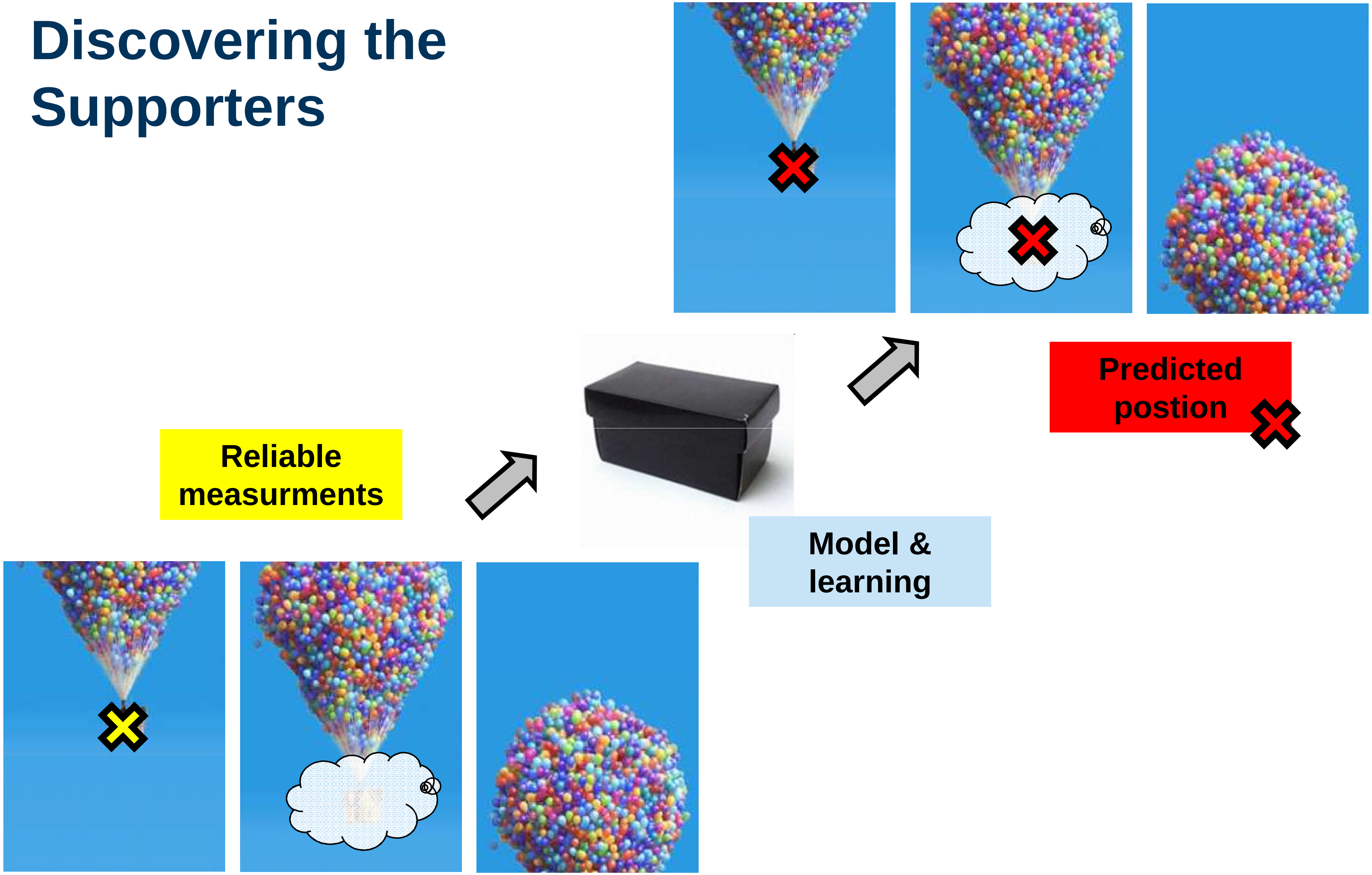
# Local Image Features as Supporters



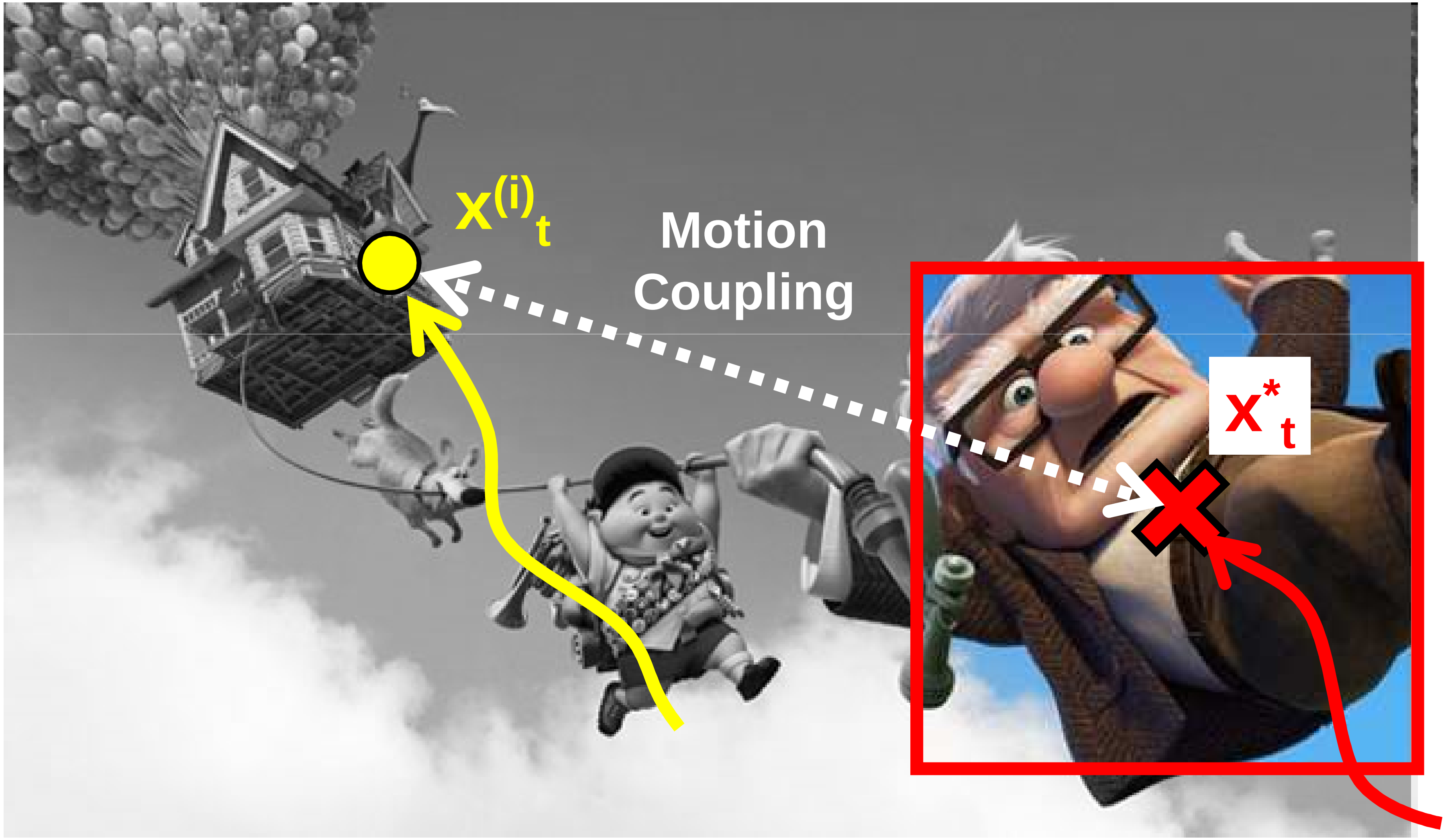
**SUPPORTERS are features which contribute to the prediction of the target position.**

**They at least temporarily move in a way which is statistically related to the motion of the object.**

# Discovering the Supporters



# Reliable Information & Motion Coupling



# Supporter

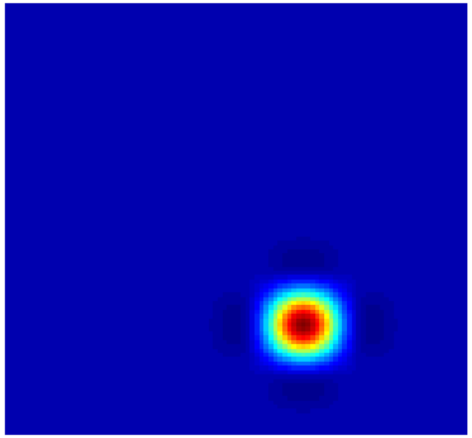
**Strong Supporter**

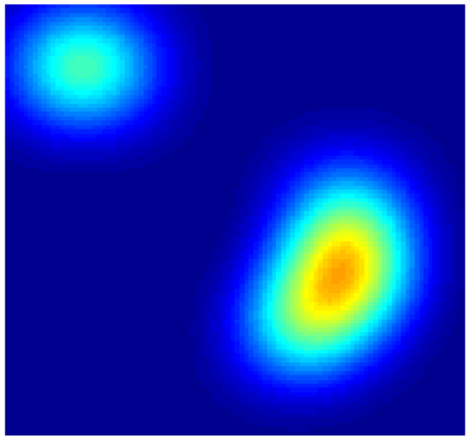
Strong motion coupling
Peaky vote

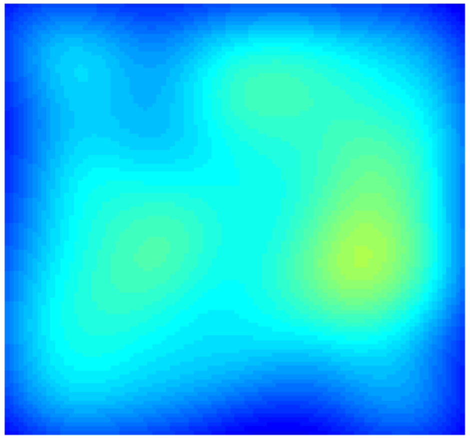
**Weak Supporter**

Weak motion coupling
Blurred vote

**Almost Unrelated Features**





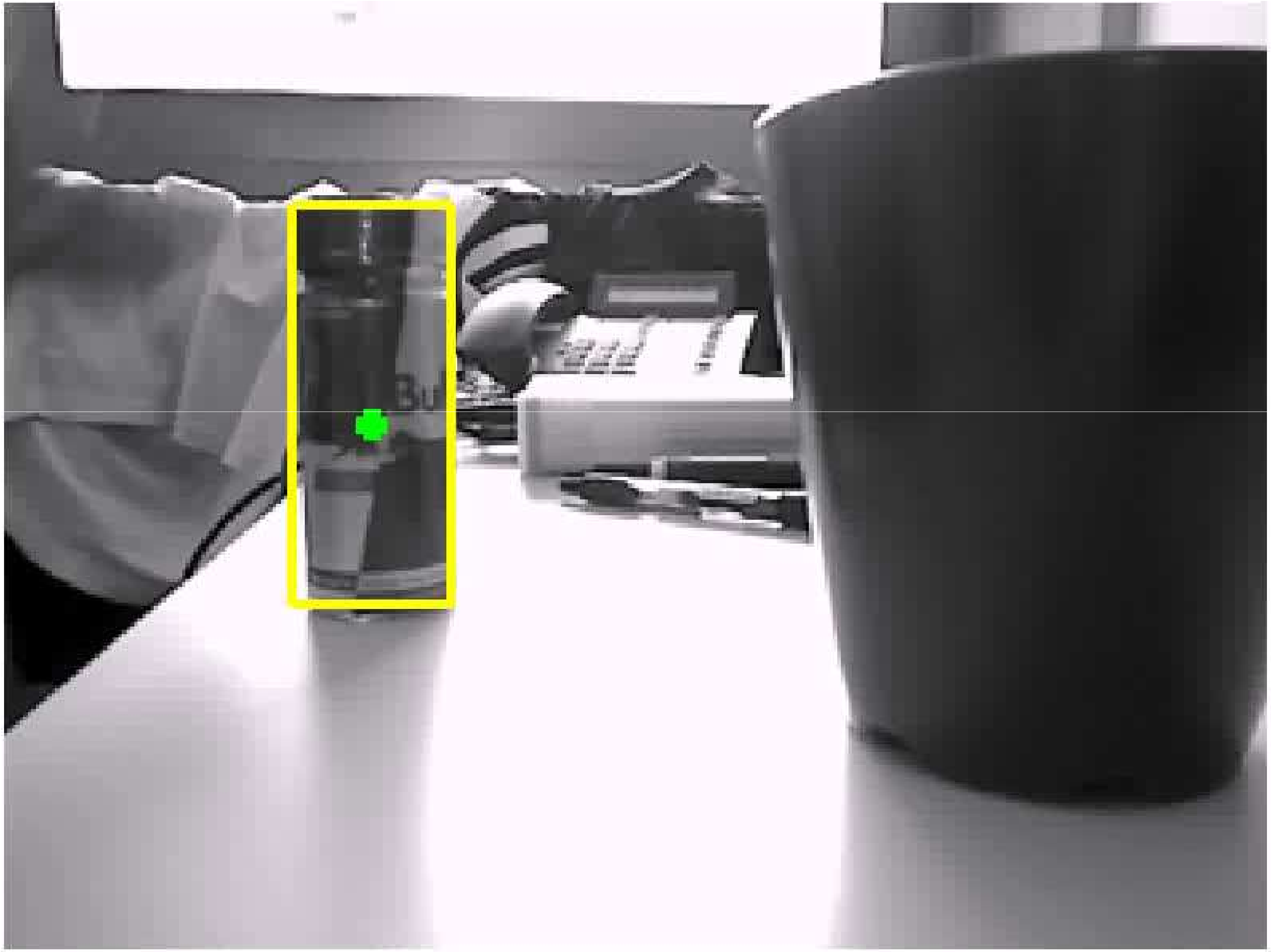


# Experimental Results: *ETH-Cup* Sequenze

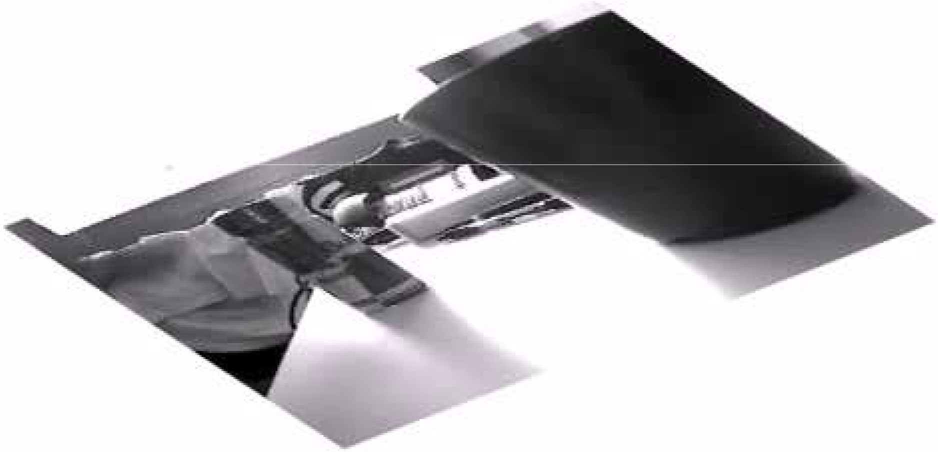
# *ETH-Cup*: Humans

www.vision.ee.ethz.ch/boostingTrackers

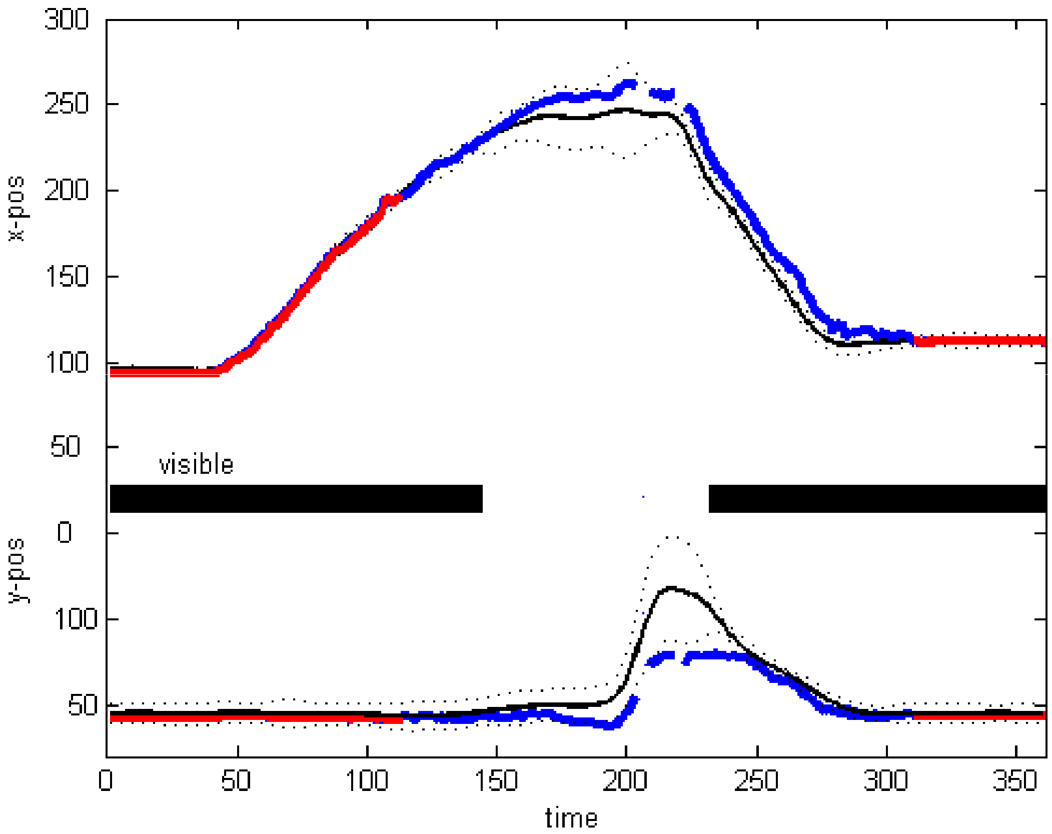
# *ETH-Cup*: Of the Web Tracker



# ETH-Cup: Our Result – Voting Space



# *ETH-Cup*: Improving Object Tracking

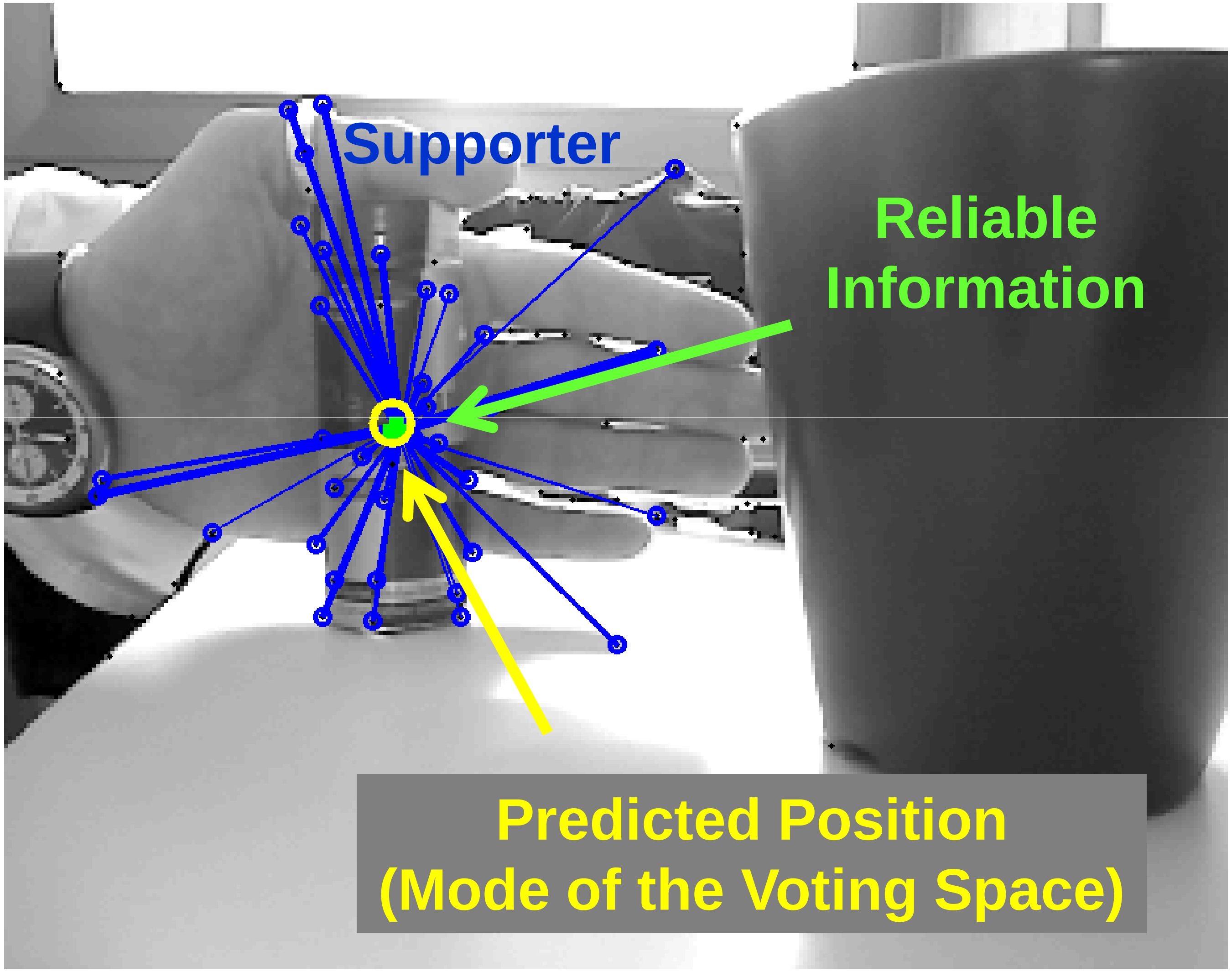



10 Humans
($\mu \pm 2\,\sigma$)

| | Rec. | Prec. |
|---|---|---|
| **Tracker** [Stalder et al, OLCV'09] | 45 % | 100 % |
| **+ supporter** | 89 % | 97 % |

**Please note, a simple interpolation would not work.**

# *ETH-Cup*: Supportes



# *ETH-Cup*: Supporters

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

BIWI Computer Vision

# Beyond the Image

**Voting Space**

**Supporters**

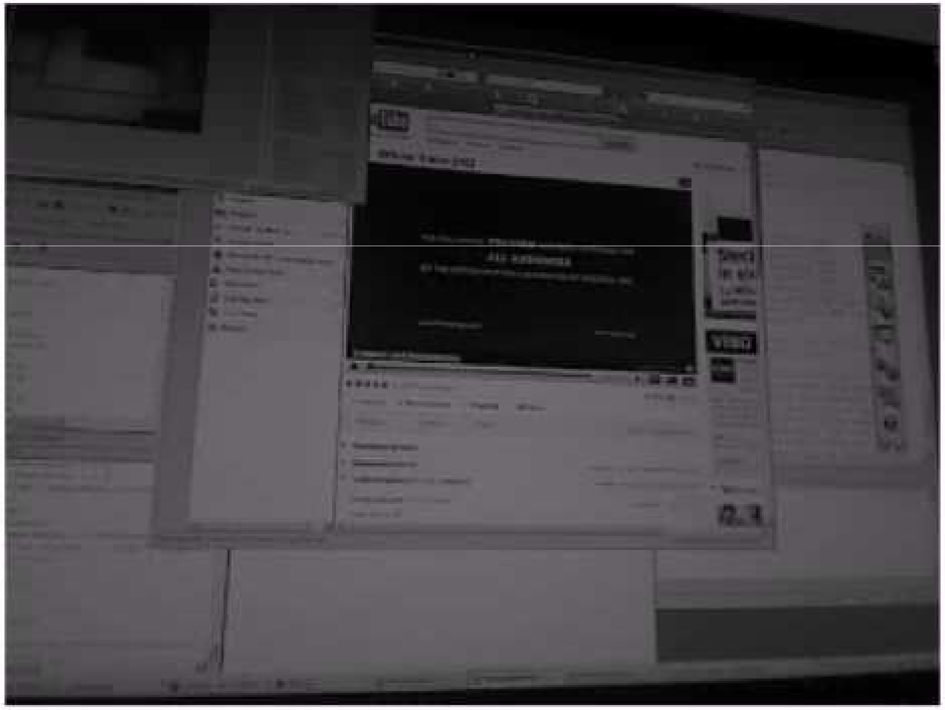
# Changing Supporter

**Voting Space**

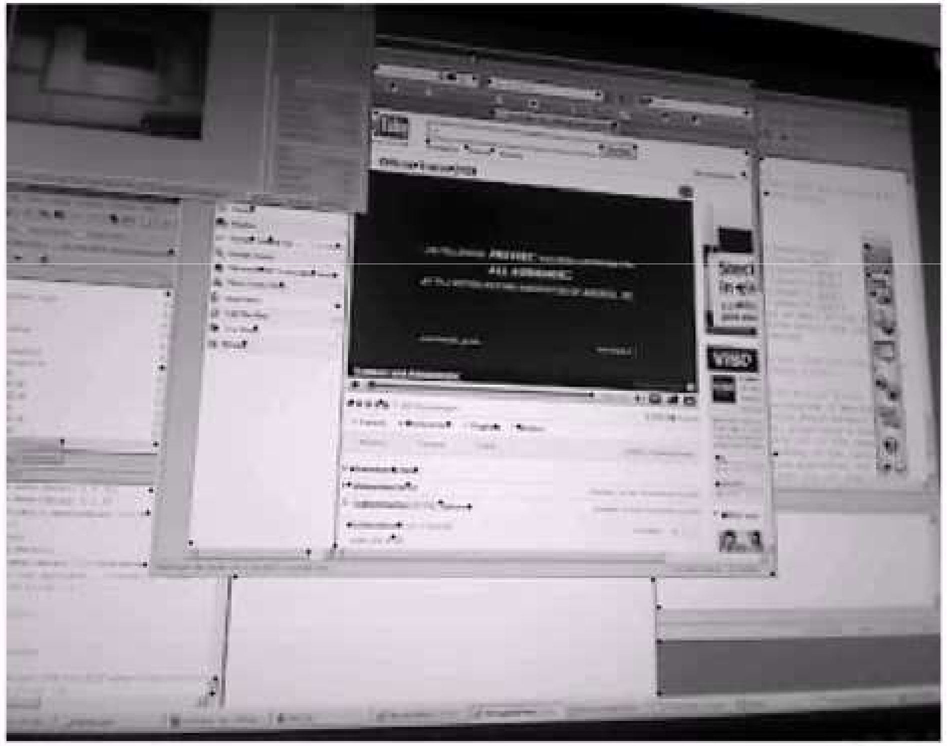
**Supporters**

# Appearance Change

**Voting Space**



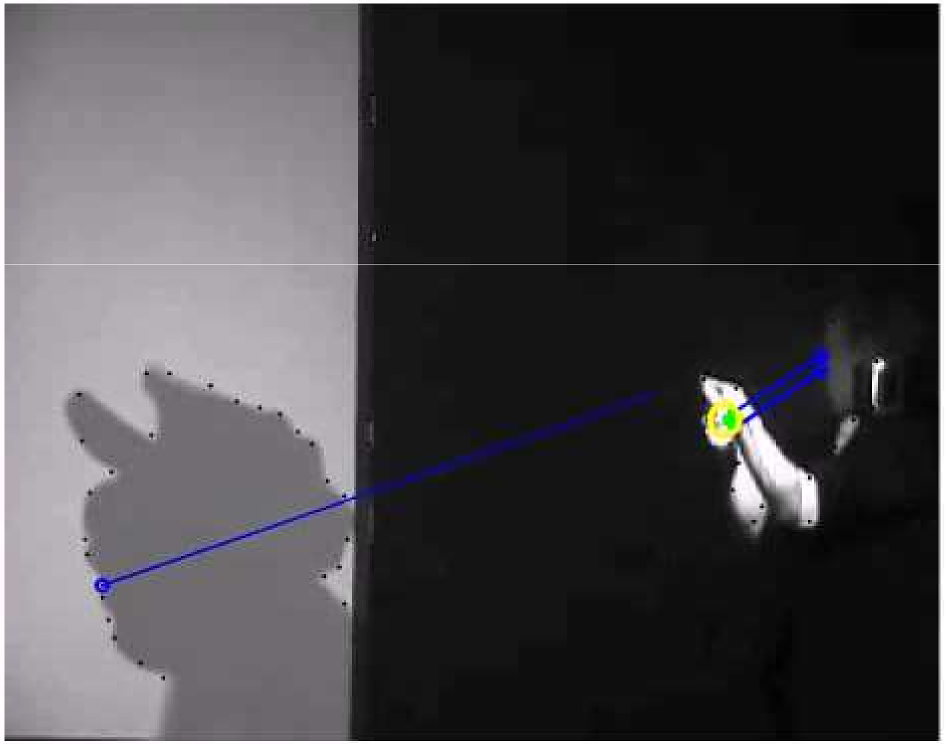
**Supporters**



# Coupled Motion

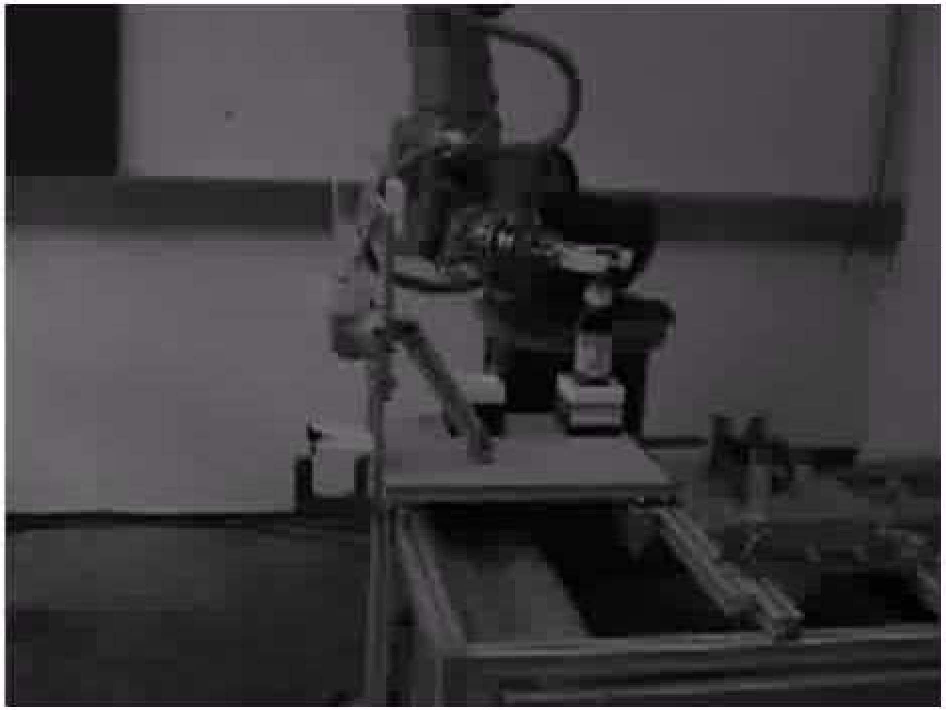
**Voting Space**

**Supporters**

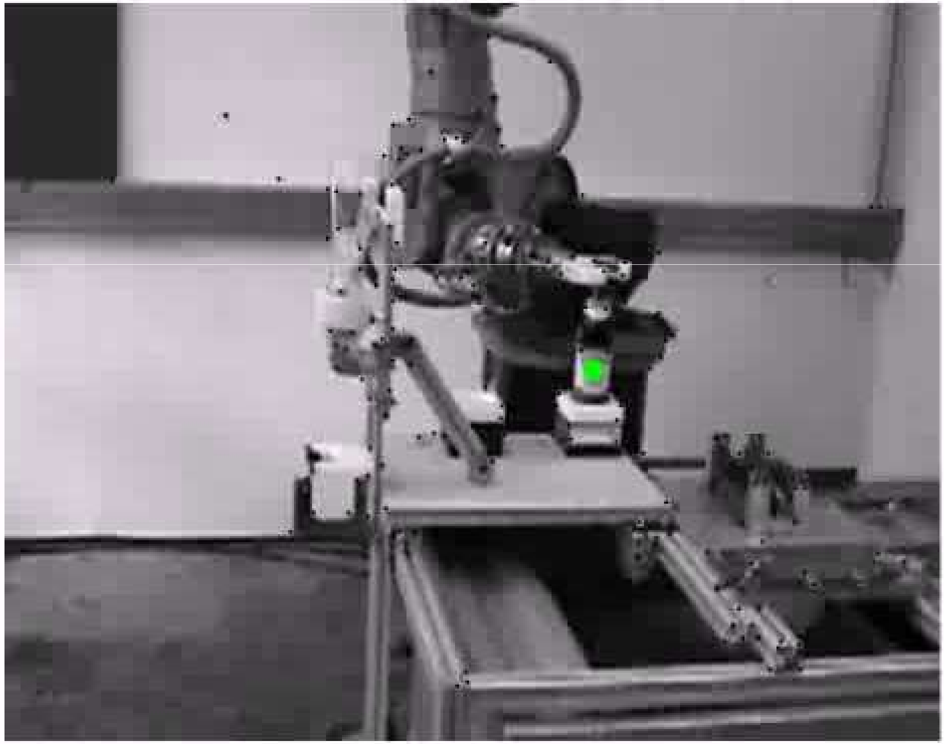


# Changing Supporters

Voting Space



Supporters



# Obviously, there are failure cases…
# …. and magician knows that.

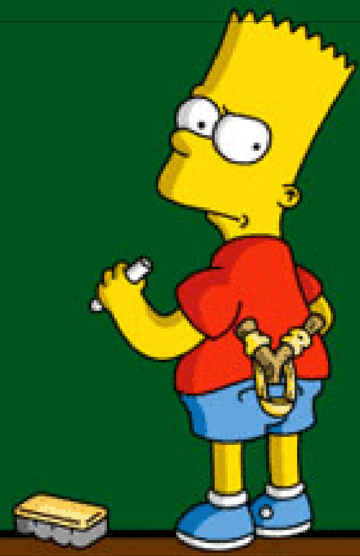
**Voting Space**

**Supporters**

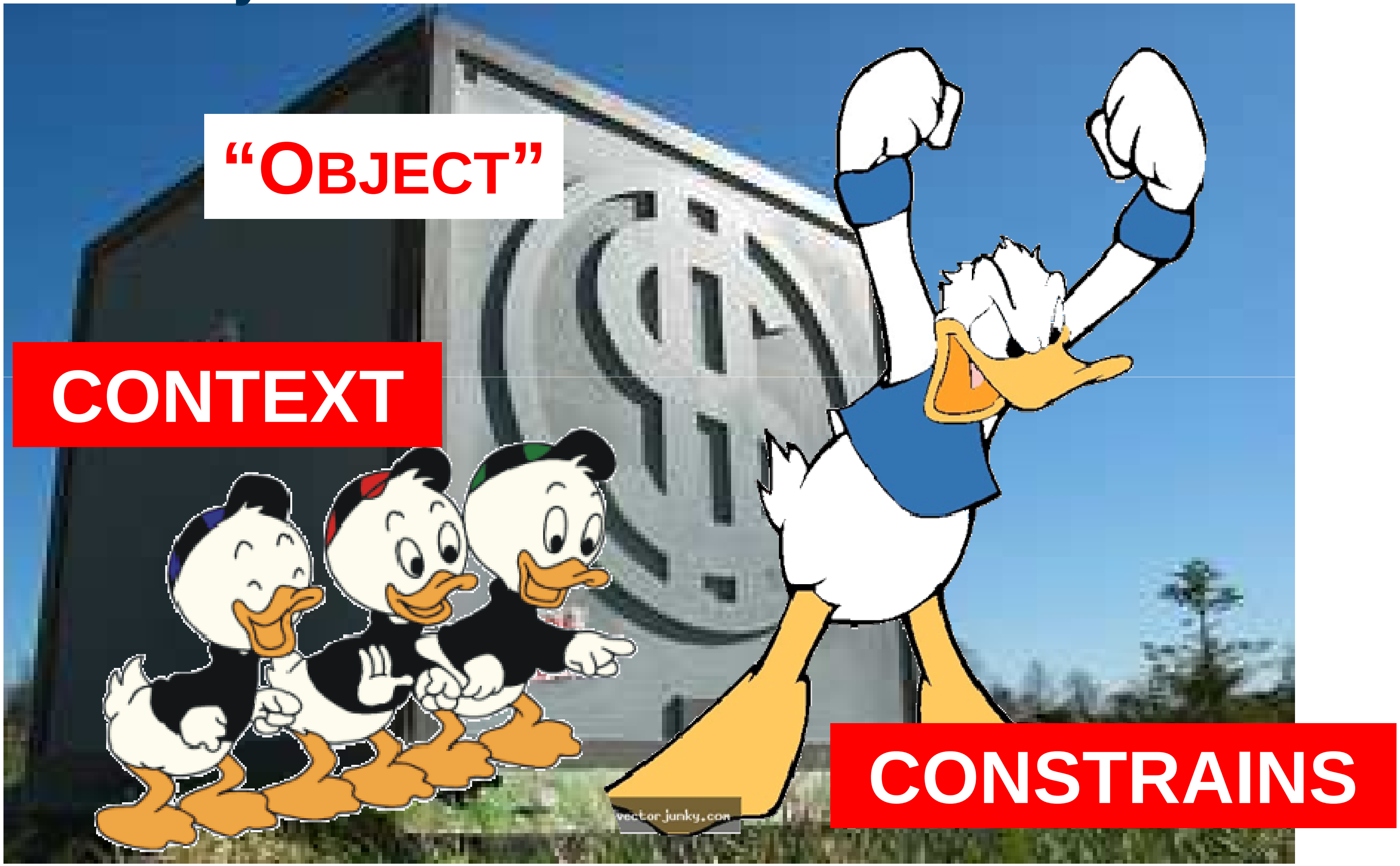
# LESSON LEARNED 9

## Look for supporters!

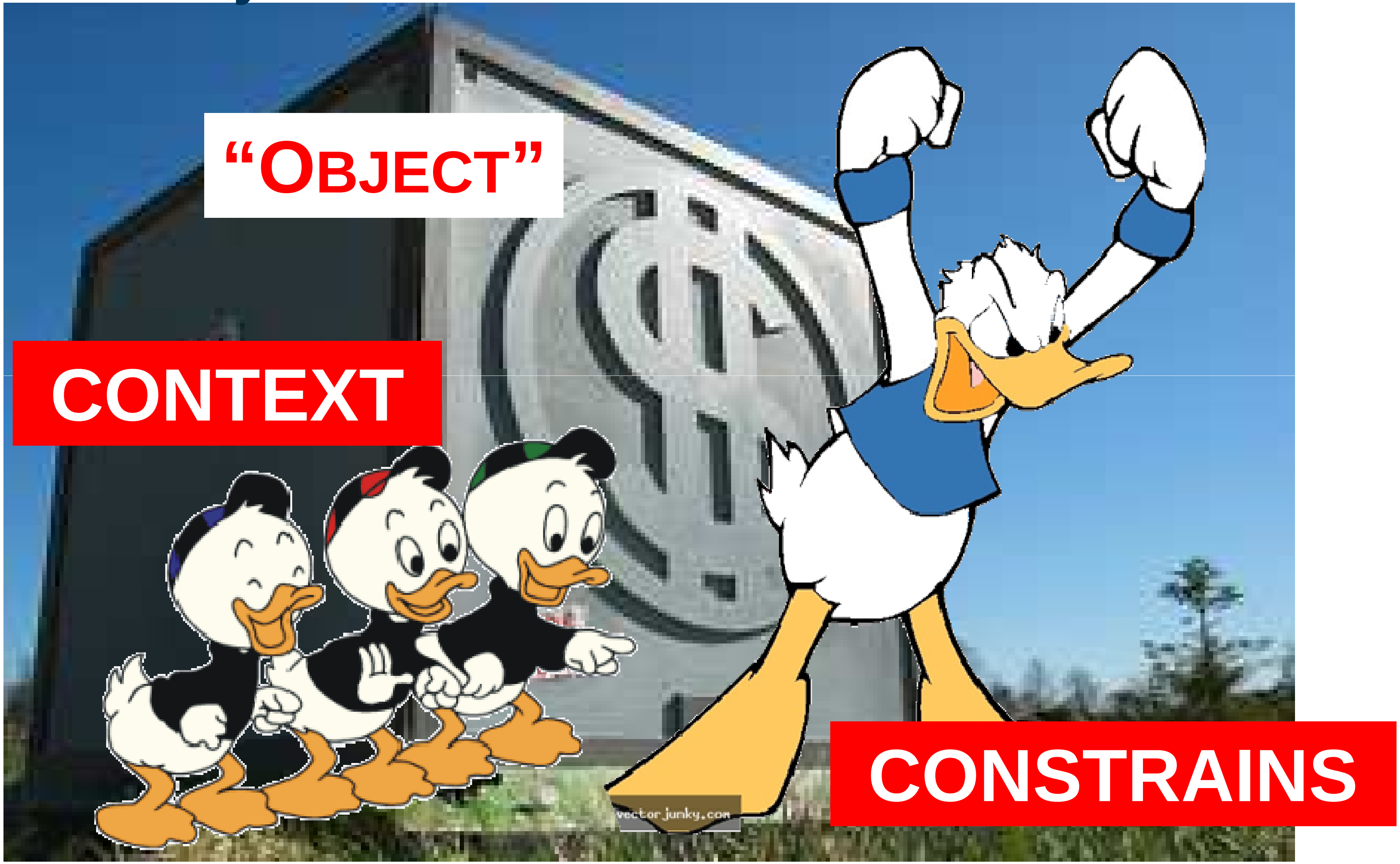
**They are common and take many forms.**



# Summary & Conclusion



# Summary & Conclusion



# Acknowledgments


Luc Van Gool

Severin Stalder

**Computer Vision Laboratory**
*ETH Zurich*

Jiri Matas

**Centere of Machine Preception**
*Czech Technical University*

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

Horst Bischof

Michael Grabner

Christian Leistner

Philippe Cattin

**Medical Image Analysis Center**
*University of Basel*