Tracking using object localizers:
To track the car or to track a car?

Agni Delvinioti
Intelligent Systems Lab UvA, Netherlands
17 February 2015, Amsterdam

Outline

CVPR submission

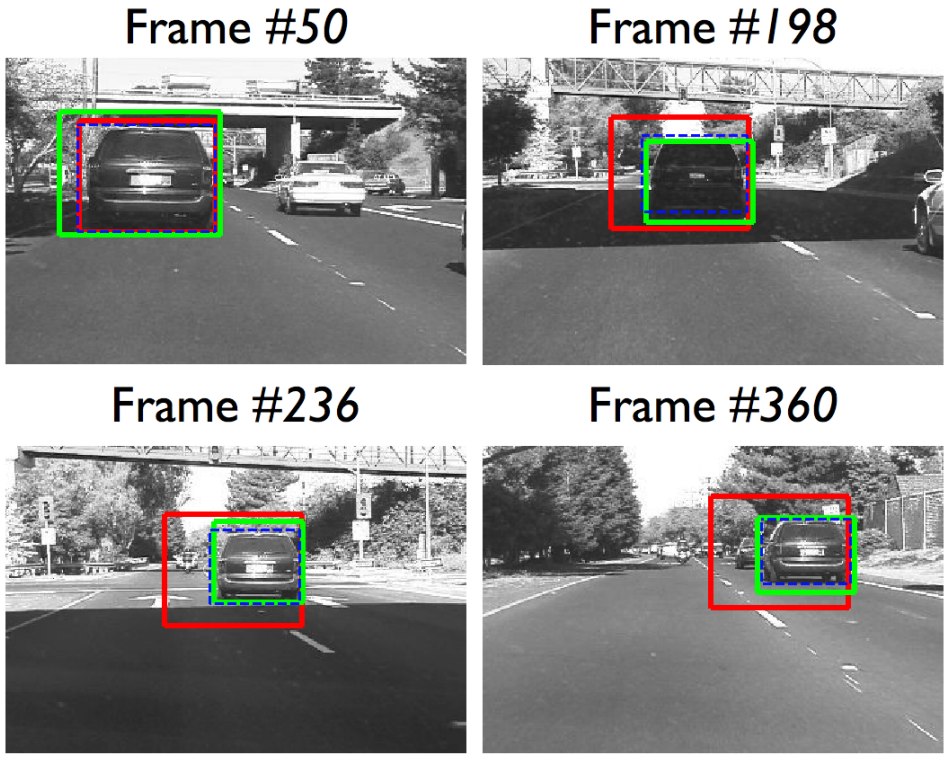
tracking map

future direction

approach

the problem

a base tracker: Struck (Hare *et al.* ICCV 2011)

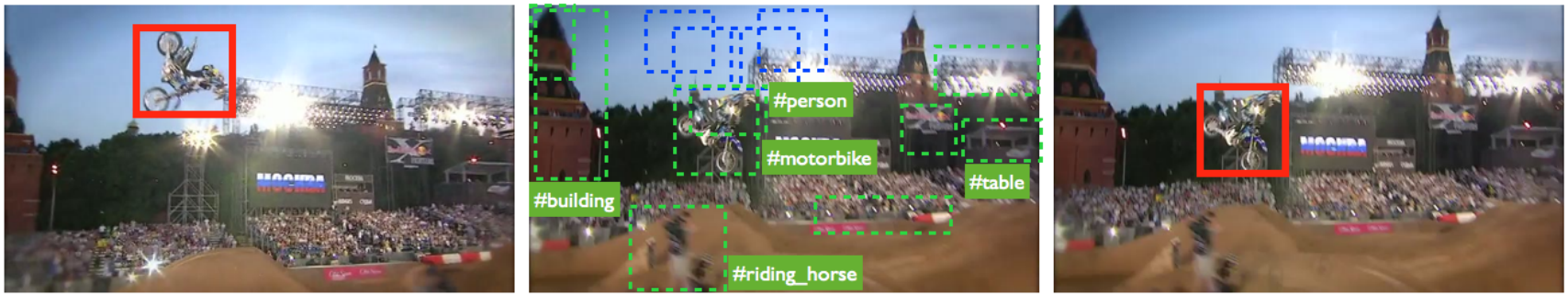


approach

the solution

a detector: Fisher and VLAD with FLAIR (van de Sande et al. CVPR 2014)

Detection at frame *i* | Category localization at frame *i+l* | Tracker decision at frame *i+l*

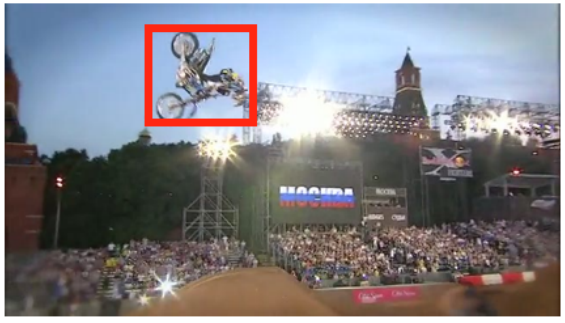


approach

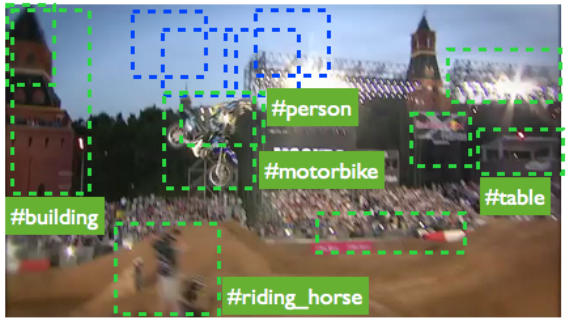
the solution

a detector: Fisher and VLAD with FLAIR (van de Sande et al. CVPR 2014)

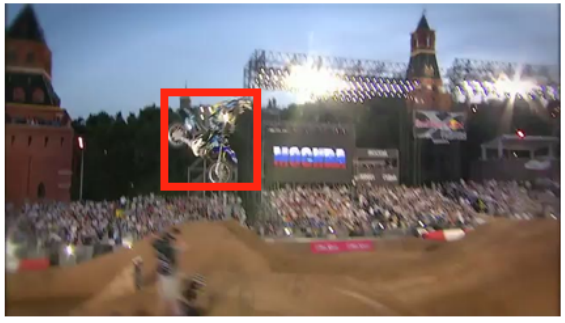
Detection at frame *i*



Category localization at frame *i+1*

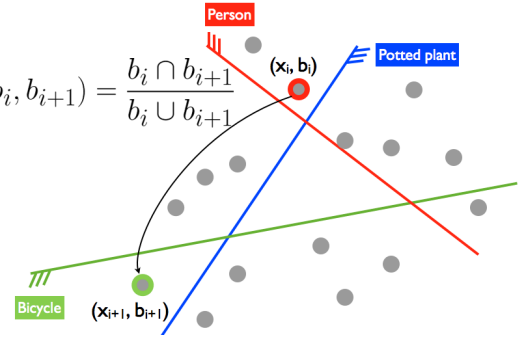


Tracker decision at frame *i+1*



Tracking the motorbike

$$\psi(b_i, b_{i+1}) = \frac{b_i \cap b_{i+1}}{b_i \cup b_{i+1}}$$



results on Tracking Benchmark (CVPR 2013)

Dataset	Object categories
PASCAL VOC 2007	aeroplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, sofa, train, tv monitor
TRECVID	potted plant, sheep, sofa, train, tv monitor, boat/ship, motorbike
Web	face

Initialization	Object categories
Manual (51)	face (23), person (19), car (4) non-defined (3), bottle (1), motorbike (1)
Automatic (51)	person (23), face (11), motorbike (4) non-defined (8), Boat_Ship (2), car (2), bottle (1)

results on Tracking Benchmark (CVPR 2013)

Dataset	Object categories
PASCAL VOC 2007	aeroplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, sofa, train, tv monitor
TRECVID	potted plant, sheep, sofa, train, tv monitor, boat/ship, motorbike
Web	face

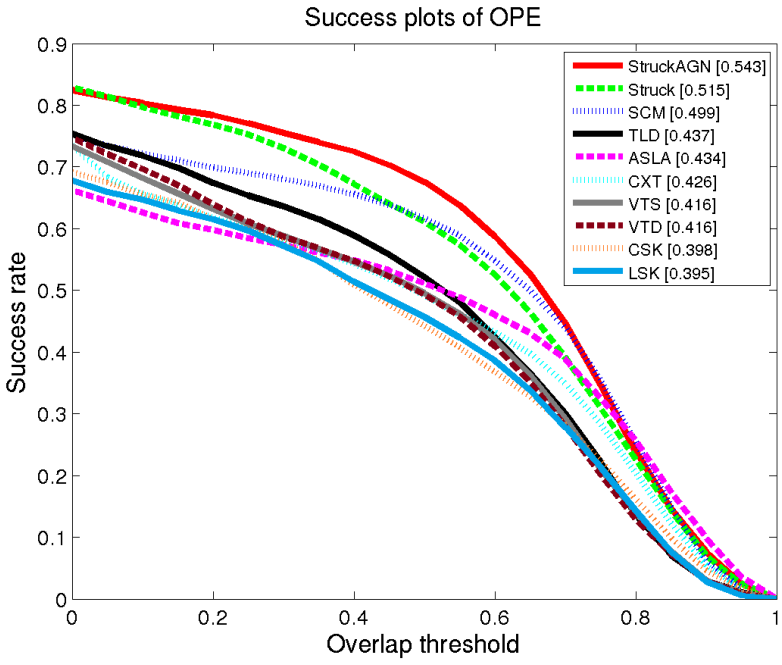
Initialization	Object categories
Manual (51)	face (23), person (19), car (4) non-defined (3), bottle (1), motorbike (1)
Automatic (51)	person (23), face (11), motorbike (4) non-defined (8), Boat_Ship (2), car (2), bottle (1)

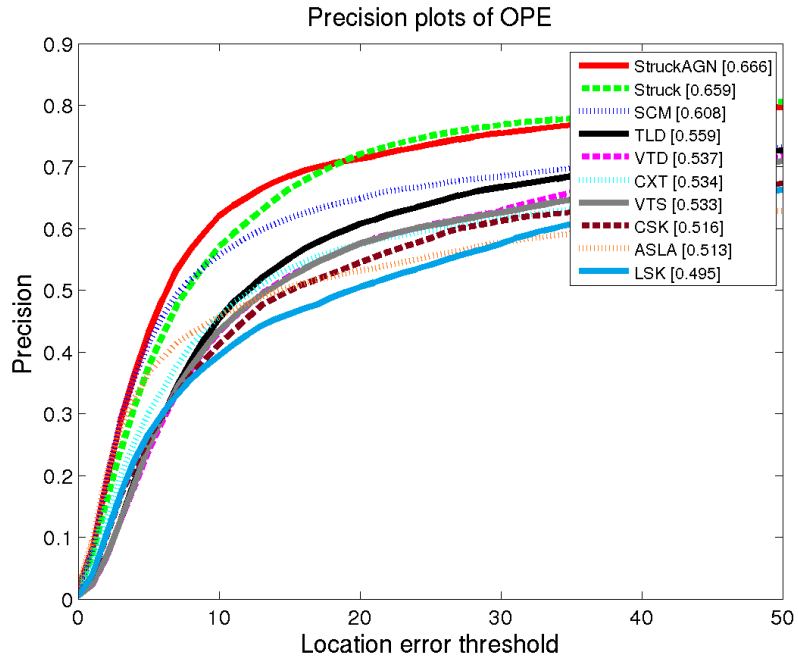
	Struck	Struck+GT	Struck+GT+Refine	StruckCL	StruckACL	StruckAGN
success rate	**0.61**	0.74	0.77	0.68	0.65	0.68
precision	**0.72**	0.80	0.83	0.73	0.72	0.71

success rate: 50% PASCAL overlap with GT
precision: 20 pixels mean center error

Struck: baseline tracker
Struck+GT: theoretical upper bound approximation
Struck+GT+Refine: theoretical upper bound approximation + refined search
StruckCL: initialization (box, category)
StruckACL: initialization (box) + automatic category selection
StruckAGN: initialization (box) + reverse localization

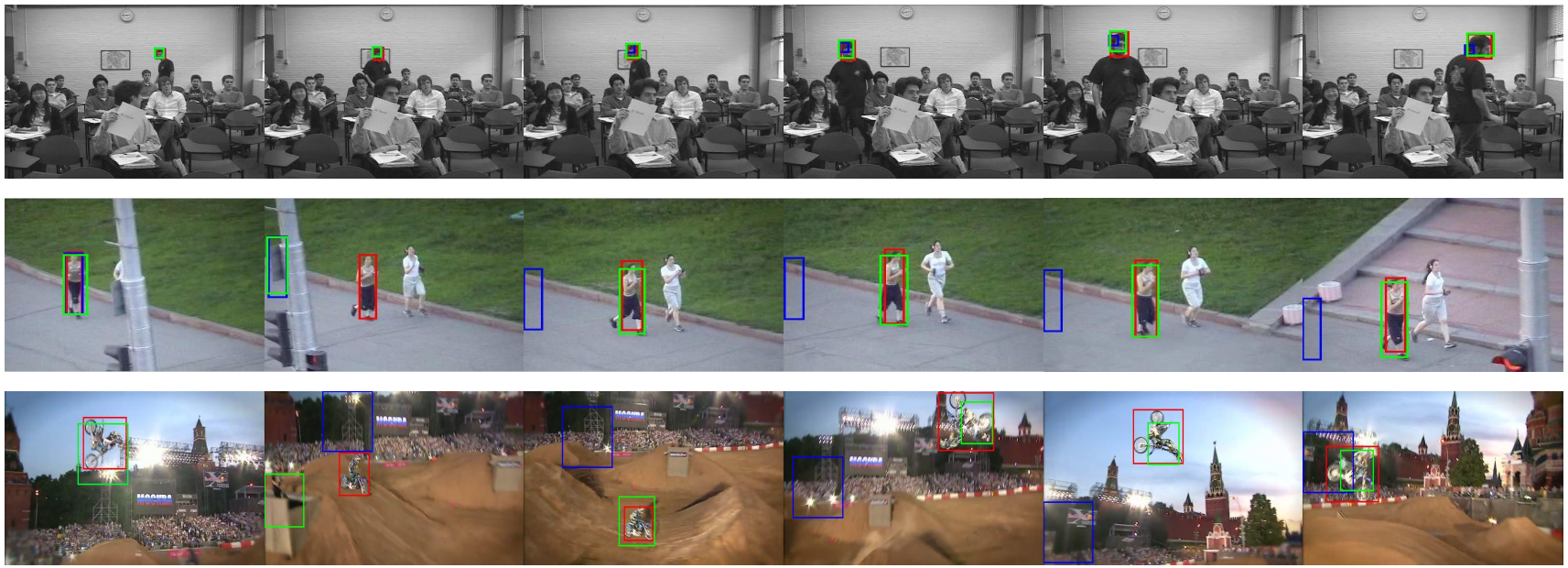
comparison with state-of-the-art





	AOG	StruckAGN	Struck	SCM	TLD	CXT	VTD	VTS	CSK	ASLA	LSK
success rate	0.75	0.68	**0.61**	0.62	0.52	0.49	0.49	0.50	0.44	0.51	0.46
precision	0.85	0.71	**0.72**	0.64	0.59	0.56	0.56	0.56	0.53	0.52	0.49

success & failure



success & failure

Outline

CVPR submission

tracking map

future direction

current state

What is the room for improvement on Tracking Benchmark (CVPR 2013)?

"0.84 the best current state-of-the-art (MEEM) success rate."

"0.93 success rate if we consider the best tracker per video among trackers available in the Benchmark."

evaluation

VOT challenge 2013-2014: systematic evaluation & unique framework for comparison

more datasets

- Tracking Benchmark (CVPR 2013): 50 videos - 11 attributes
- ALOV (PAMI 2014): 315 videos - 14 difficulty aspects
- Princeton Tracking Benchmark (ICCV 2013): 100 videos

modern trends

in general

- hierarchical appearance models: AOG, GOT, MQT
- memory management for long-term tracking: MEEM, SPLTT
- spatial-temporal constraints for robustness: SCT
- signal processing techniques for rich memory models: CSK, KCF

modern trends

in general

- hierarchical appearance models: AOG, GOT, MQT
- memory management for long-term tracking: MEEM, SPLTT
- spatial-temporal constraints for robustness: SCT
- signal processing techniques for rich memory models: CSK, KCF

component-oriented

- robustifying FoT (Matas & Vojir Computer Vision Winter Workshop 2011)
- occlusion & motion reasoning (Hua *et al.* ECCV 2014)
- SPLTT-self paced learning (Supancic & Ramanan CVPR 2013)
- DSST-accurate scale estimation (Danelljan *et al.* BMVC 2014)
- MEEM-entropy minimization (Zhang *et al.* ECCV 2014)
- Adaptive Objectness for Object Tracking (Liang *et al.* CVPR 2015 submission)

remarkable observations

"Appearance model is more important than motion model."
"Large number of non-overlapping negatives is crucial."
"Memory management/allignment over time appears profitable."
"Scale and occlusion handling are mandatory."
"Self awareness tools are required to avoid "accidental" matching."

Outline

CVPR submission

tracking map

future direction

motivation

few video examples: Struck (Hare *et al.* ICCV 2011)
best known tracker (Tracking Benchmark (CVPR 2013)+experimental survey (PAMI 2014))

TABLE III: The list of outstanding cases resulted from the Grubbs' outlier test and with $F \geq 0.5$.

Sequence	Tracker	Sequence	Tracker	Sequence	Tracker	Sequence	Tracker
0112	TLD	0411	ACT	1102	TLD	1203	MIT
0115	STR	0510	L1T	1103	HBT	1206	STR
0116	KAT	0512	STR	1104	TLD	1210	TLD
0122	TLD	0601	STR	1107	HBT	1217	TLD
0203	FBT	0611	MST	1112	STR	1218	TLD
0301	L1T	0705	TLD	1116	TLD	1221	TLD
0305	L1T	0901	HBT	1119	TLD	1303	TLD
0312	L1T	0916	STR	1128	TLD	1402	TLD
0314	KAT	0925	STR	1129	FBT	1409	STR
0404	FBT	1020	FBT	1134	FRT		

motivation

few video examples: Struck (Hare *et al.* ICCV 2011)
best known tracker (Tracking Benchmark (CVPR 2013)+experimental survey (PAMI 2014))

TABLE III: The list of outstanding cases resulted from the Grubbs' outlier test and with $F \geq 0.5$.

Sequence	Tracker	Sequence	Tracker	Sequence	Tracker	Sequence	Tracker
0112	TLD	0411	ACT	1102	TLD	1203	MIT
0115	STR	0510	L1T	1103	HBT	1206	STR
0116	KAT	0512	STR	1104	TLD	1210	TLD
0122	TLD	0601	STR	1107	HBT	1217	TLD
0203	FBT	0611	MST	1112	STR	1218	TLD
0301	L1T	0705	TLD	1116	TLD	1221	TLD
0305	L1T	0901	HBT	1119	TLD	1303	TLD
0312	L1T	0916	STR	1128	TLD	1402	TLD
0314	KAT	0925	STR	1129	FBT	1409	STR
0404	FBT	1020	FBT	1134	FRT		

* analysis of failures
* broad complexity $\rightarrow$ needed hierarchy
* initialization dependency & strong assumptions in update
* explicit state model different from observed model $\rightarrow$ self awareness metrics

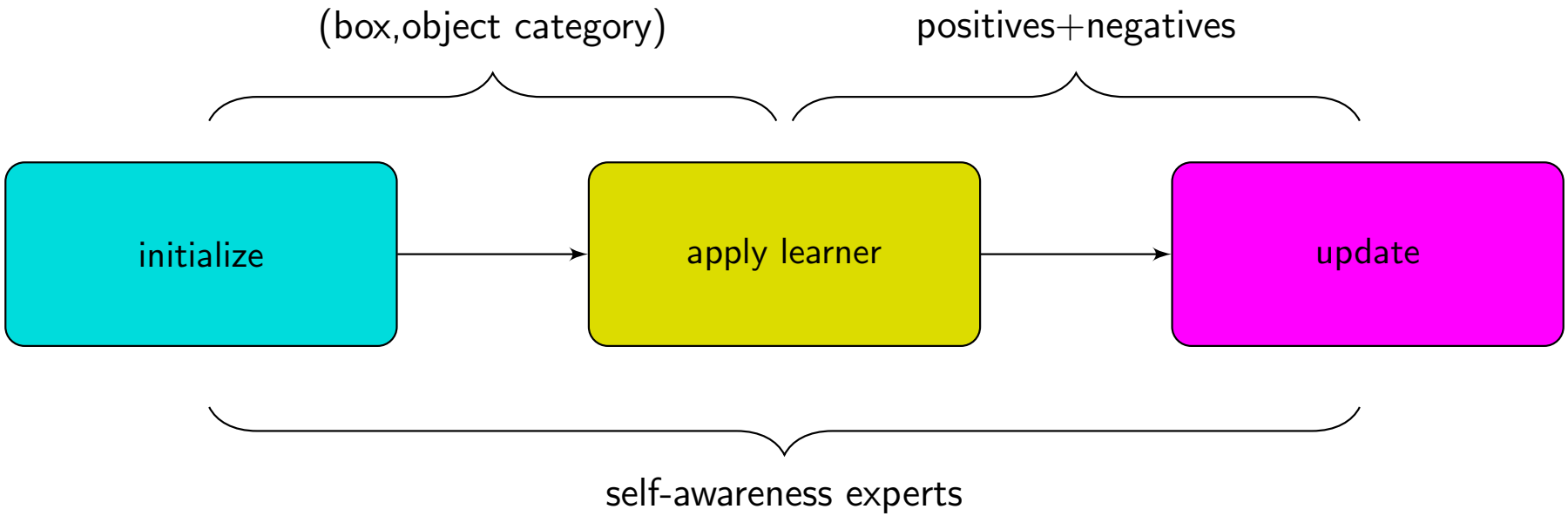
our paradigm

redefinition of tracking problem

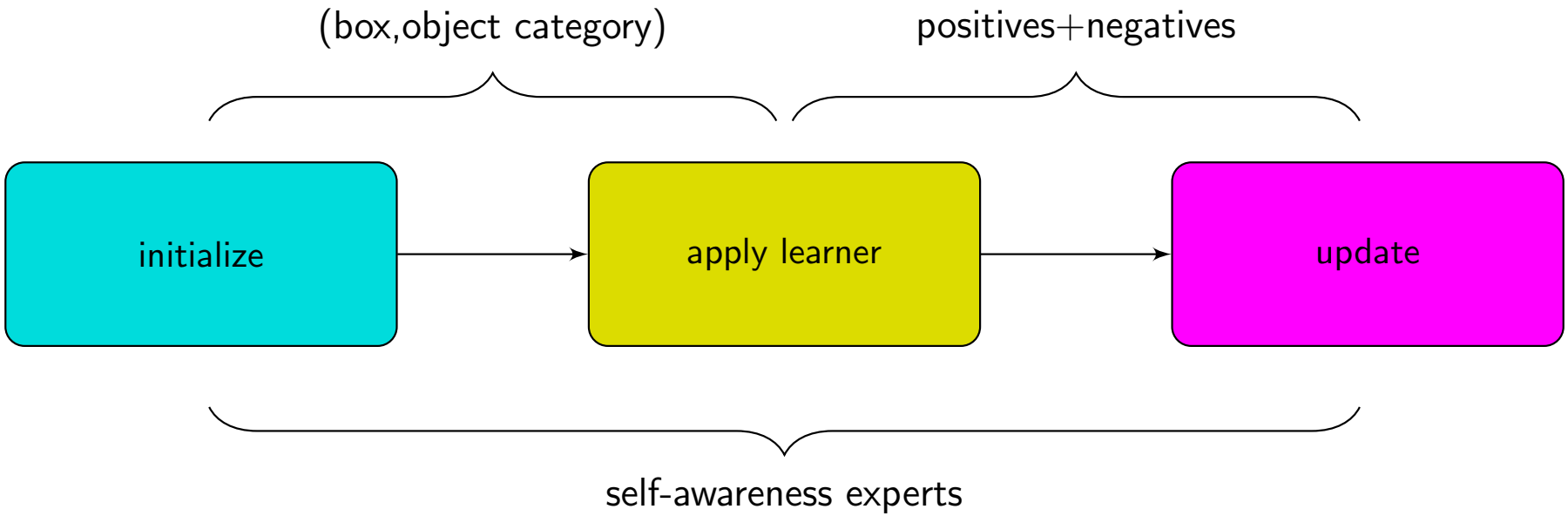
"Give the answer to the question **which object** to track by the inclusion **what and where** is the object to track?"

- solve a "better defined" problem when hard (box + object category information)
- use hierarchy to reduce to the traditional one $\rightarrow$ bottom-up fashion
- switch $\rightarrow$ self awareness

our paradigm



our paradigm



proposed solution

(I) study:

(a) initialization $\rightarrow$ detector

(b) update $\rightarrow$ detector

(II) compare state model with observed model:

(a) self aware of internal state

(b) self aware of external condition $\rightarrow$ detector

experiments

tools

- **3 trackers:**
  (1) Struck (CVPR submission)
  (2) MEEM
  (3) SPLTT
- **detector:** Fisher and VLAD with FLAIR (CVPR submission)

datasets

1. Tracking Benchmark: 50 videos
2. MEEM dataset: 10 challenging videos

evaluation

- OPE, SPE, TRE (precision+success rate)
- analysis of failures

experiments

1. oracle experiment: theoretical upper bound approximation
2. initialization & object category selection:
   - 2.1 box (baseline)
   - 2.2 object category (new!/automatic annotation)
   - 2.3 box+object category (CVPR submission)
3. integration scheme (extending CVPR submission):
   - preprocessing & filtering of the detections
   - reverse localization & detector-tracker regulator
4. update (new!):
   - self-awareness experts (scale, appearance & motion smoothness criteria)
   - hard negatives indicated by detector

Outline

CVPR submission

tracking map

future direction

Thank you for your attention!

Questions - Ideas ?