A System for Real-time Detection and Tracking of Vehicles from a Single Car-mounted Camera

Claudio Caraffi, Tomas Vojir, Jiri Trefny, Jan Sochman, Jiri Matas

TOYOTA

cmp

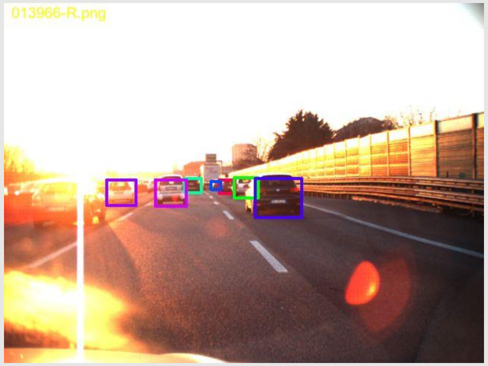
Toyota Motor Europe **Center for Machine Perception, University of Prague**

ITS Conference, Anchorage, 18 Sep 2012

An “easy” problem

Vehicle detection & tracking on motorways:

- Only vehicle rear, rigid object
- Limited street furniture, no pedestrian
- Single camera
- Variable lighting condition



Quantitative evaluation:

- Extended dataset → ~~expensive annotation~~
- Laser-scanner for semi-automatic annotation of ~30000 frames. **Dataset released**.

TOYOTA

Algorithm architecture

- Target initialized by vehicle detector (Wald Boost)
- KLT features tracking (Flock of Trackers)
- LrD: Learn specific target and re-detect to correct tracker drift (Randomized Forest)

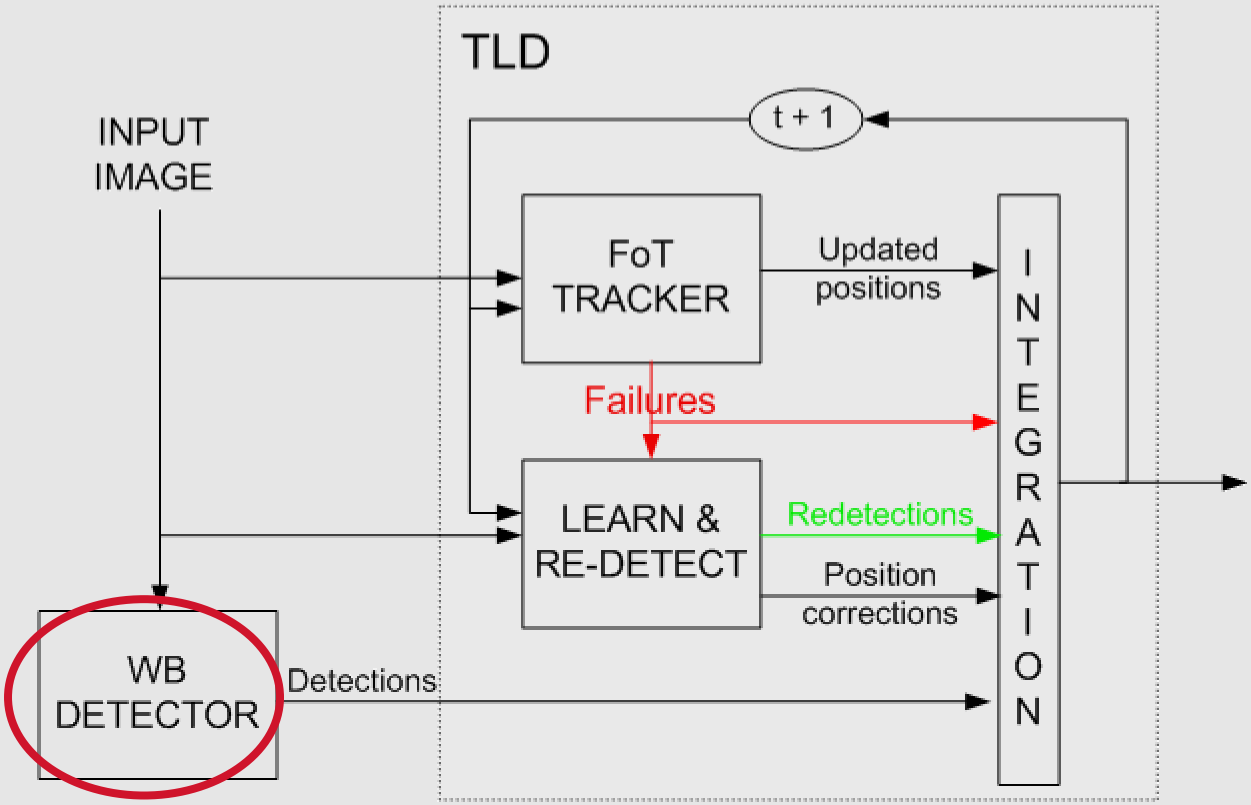
- Wald Boost detections are also used to correct known targets → Precise bounding box
- In case of tracker failure, WB & LrD try to recover the target in a Kalman-filter predicted position → long-term vehicle identity maintenance

Vehicle Detection & Tracking

TOYOTA

m p

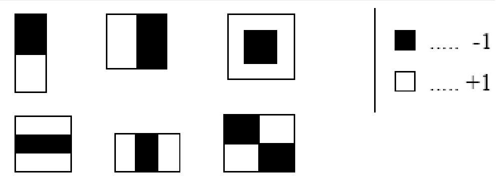
Vehicle detection and tracking

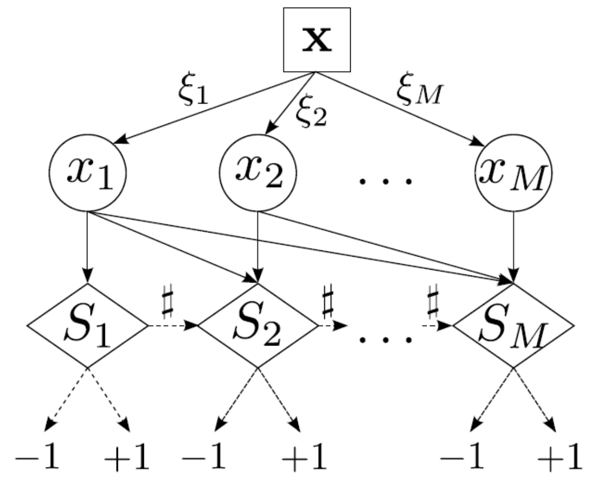


TOYOTA

WaldBoost Detector 1

- WaldBoost:
  - Sliding window algorithm
  - Sequence of weak classifier (AdaBoost)
  - Focus on **time to decision**
  - After each classifier, verify if we can take a decision “not object” (Asymmetric)

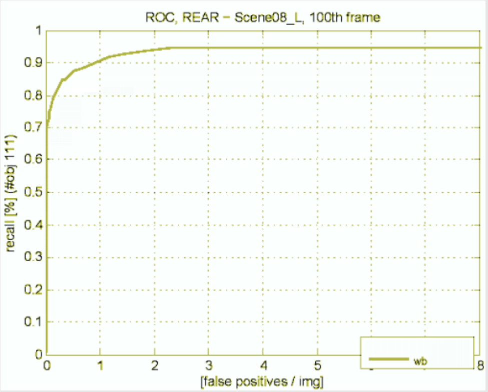




TOYOTA

WaldBoost Detector 2

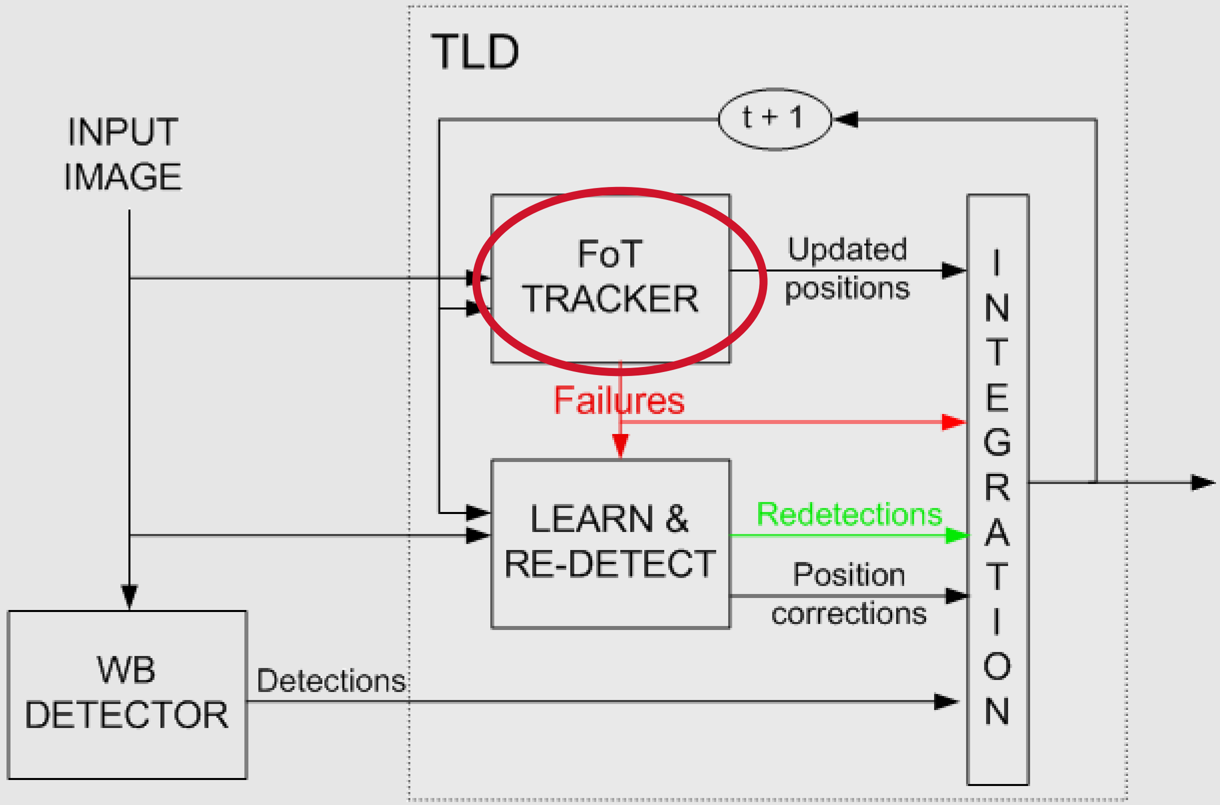
- 5000 car images (and a few trucks…) training dataset
- One billion negative samples
- Performance on test sequence:



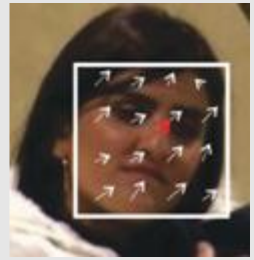
- 1.74 average classifiers evaluated per window for rear vehicle detection
- Limits: longer time to decision in crowded traffic scenes.

J. Sochman and J. Matas, "WaldBoost - Learning for Time Constrained Sequential Detection," in CVPR 2005.
J. Trefny and J. Matas, "Extended Set of Local Binary Patterns for Rapid Object Detection," in CVWW 10: Proceedings of the Computer Vision Winter Workshop 2010.

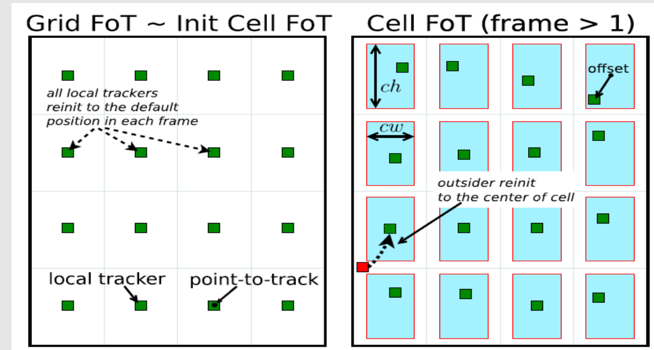
Vehicle detection and tracking

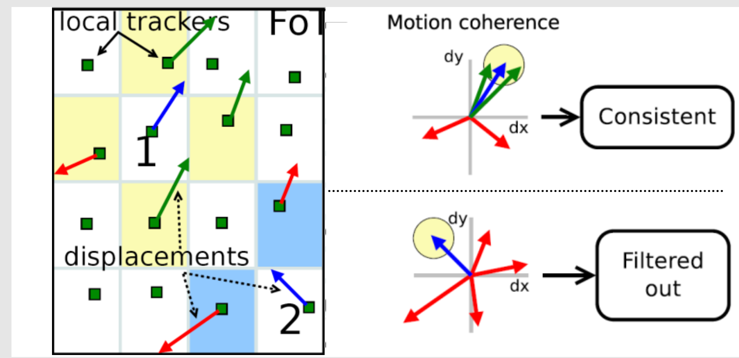


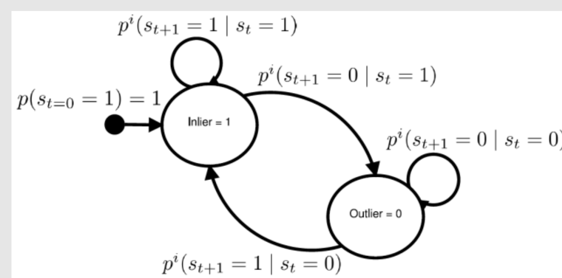
Flock of trackers (FOT) 1



- Target divided in cells
- One KLT feature tracker per region
- KLT feature pre-extraction skipped
- Tracker evenly placed, **“naturally” converge to a good feature**
- Trackers are evaluated (in/outlier):
  - ~~Previous: Forward-backward KLT check, cost is twice~~
  - Neighbourhood consistency (Nh)
  - Markov model predictor (Mp)
  - Norm. cross correlation (NCC)
  - Combination: 10% KLT time
- Estimation of translation & scaling







TOYOTA

Flock of trackers (FOT) 2

- Fast computation (4.8 ms)

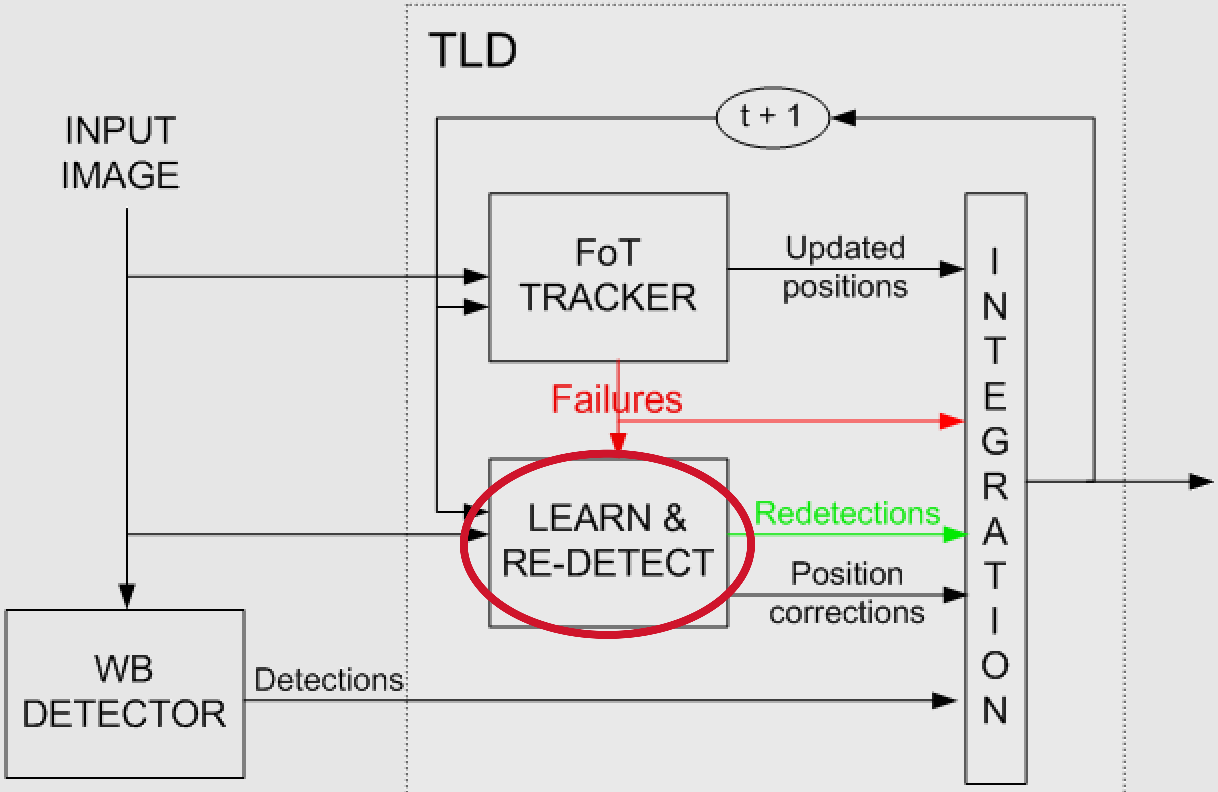
sequence	[10]	[4]	[2]	[3]	[5]	T_Σ
1	17	n/a	94	135	**761**	**761**
2	75	**313**	44	**313**	170	76
3	11	6	22	101	**140**	**140**
4	33	8	118	37	97	**264**
5	50	5	**53**	49	52	52
6	163	n/a	10	45	**510**	**510**
best	0	1	1	1	3	**4**

Comparison with recently published methods on public sequences

- Effective for rigid objects
- Failure in case of strong illumination changes

T. Vojir and J. Matas, "Robustifying the Flock of Trackers," in Computer Vision Winter Workshop, 2011.

Vehicle detection and tracking



Learn & re-Detect

- At target confirmation (after 3 detections) create a specific random forest classifier (RF): Insert as positive samples the current target window and its affine warps
- Keep updating the RF, collect:
  - Negative samples: RF detector firing outside the tracked object
  - Positive samples: current target window if similarity (NCC) > 0.75
- Limit: Similar vehicles in the scene

Z. Kalal, J. Matas, and K. Mikolajczyk, "P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints," in Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

TOYOTA

m p

Scheduling (10 Hz objective)

AVERAGE COMPUTATION TIME [ms]		
	Image resolution	
Process	640x480	1024x768
WaldBoost*	16.61	42.99
Warping + RF Learning	8.82	21.24
LD position correction	5.06	3.99
LD negative samples	2.65	2.74
FoT	3.12	6.29
WaldBoost verification	1.27	0.87
LD verification	3.47	1.60

Only one expensive operation per frame:

- WB detector (every 3rd frame)
- Random forest generation (frame after 1st det.)
- LrD (other frames): correct position for each target, collect negative samples for one target

Note: WB runs every 3 frames, 3 detections to confirm a target → delay up to 0.9 s

TOYOTA m p

Results 1

TOYOTA

m p

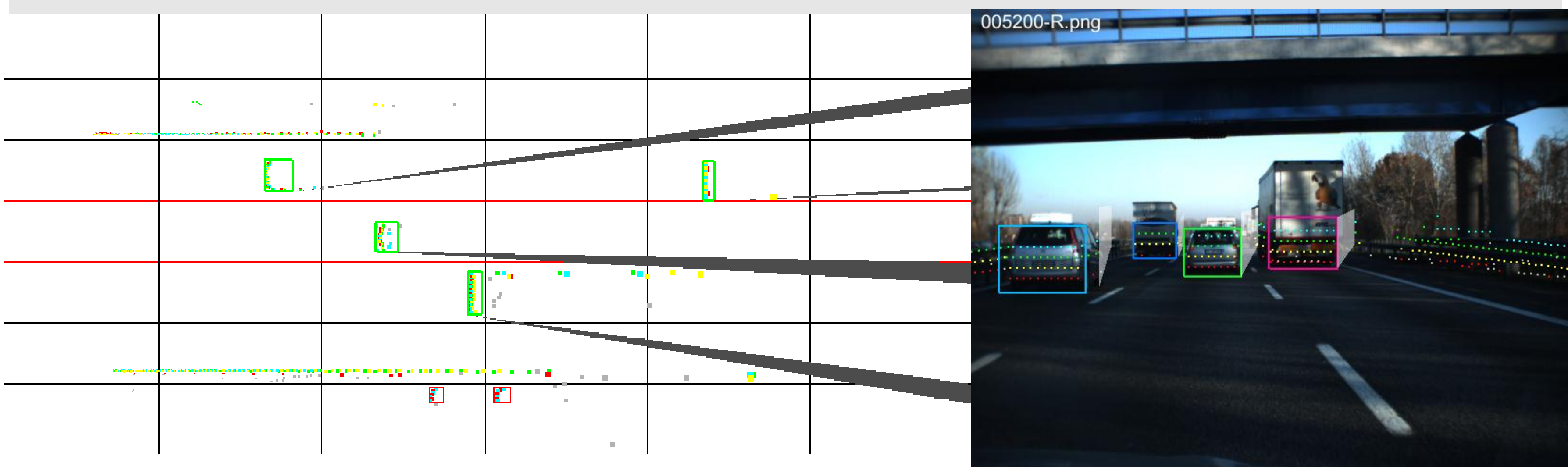
Quantitative evaluation: Dataset and annotation

- Acquired in December 2011 in cooperation with VisLab, University of Parma, Italy
- 28 clips for a total of ~27 minutes
- Two 1024x768 color cameras,
  IBEO 4-layer laser-scanner, ego-trajectory
- Idea: Use laser-scanner data to extract automatically vehicles
- Project vehicles into the image to create an approximated Ground Truth (GT')

Laser-scanner detections

Easy scenario:

- Extract surfaces perpendicular to the driving direction, discard static objects
- Project into the image using fixed calibration parameters and a flat-ground assumption



TOYOTA

Features of GT'

- Consistent ID available
- 3D position available
- Object width estimated over full time of observation
- Car – Truck classification based on width

Limits of GT'

- No motorbike
- Unreliable beyond 60-70 meters (< 3 laser reflections)
- Imprecise target side boundaries (quantization and noise)
- No vehicle length (although possible)
- No vehicle height (arbitrarily set)
- Static calibration insufficient for projection (oscillation and non-flat ground)

Oscillation: best pitch matching

Vision algorithm GT’ pitch correction GT’ fixed calibration

Match GT' ↔ Algorithm results

- Given the mentioned limitations, we introduce a custom overlap measurement

$$O = {O_w}^2 \cdot \underbrace{O_x \cdot \sqrt{O_y}} \quad (Overlap\ score)$$

AREA TERM ***POSITION TERM***

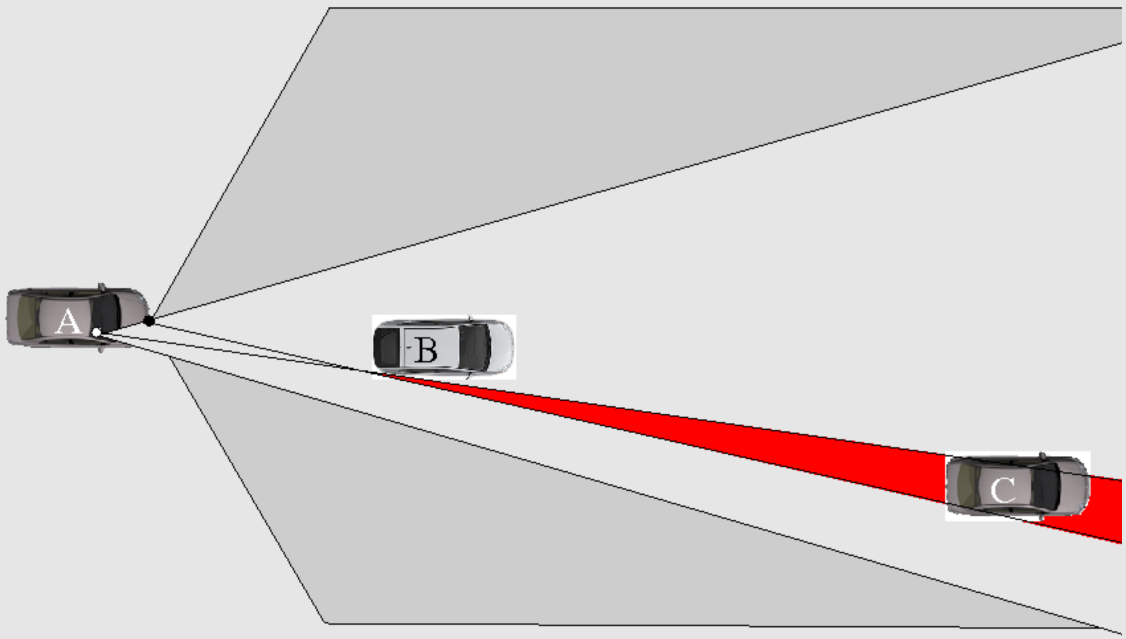
$$O_w = \frac{\min(w'_G, w_S)}{\max(w'_G, w_S)}$$

$$O_x = \frac{\left\| \cap\left(\left[{x_0}_G, {x_1}_G\right], \left[{x_0}_S, {x_1}_S\right]\right) \right\|}{\min(w_G, w_S)}$$

$$O_y = \frac{\left\| \cap\left(\left[{y_0}_G, {y_1}_G\right], \left[{y_0}_S, {y_1}_S\right]\right) \right\|}{\min(h_G, h_S)}$$

TOYOTA

Different sensor positions



- A correct detection can be classified false positive
- An object not visible in the image can be classified false negative

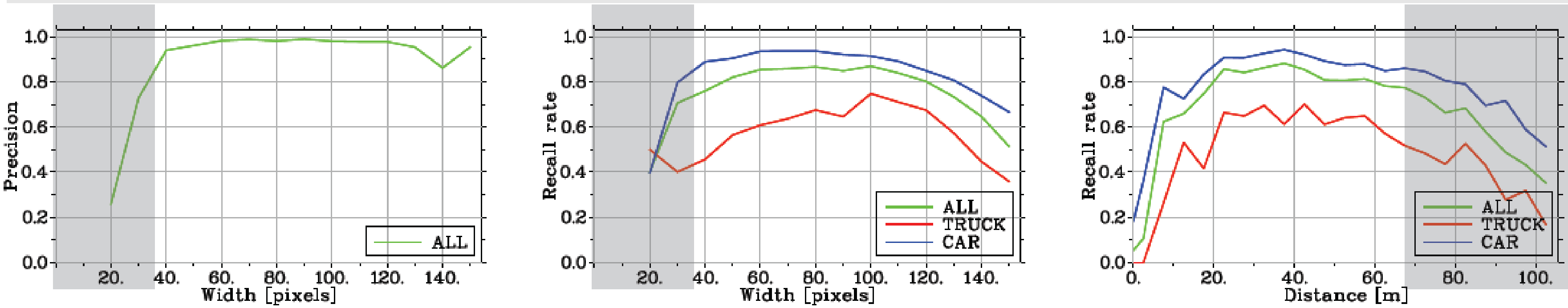
Different sensor positions

- The vehicle marked in yellow is not considered in the statistics

TOYOTA

Quantitative evaluation

Daylight:



(*a*) Precision in function of width

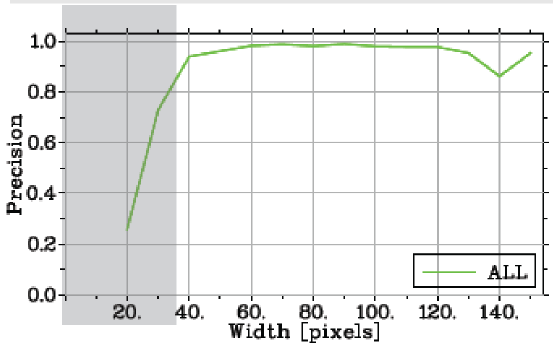
(*b*) Recall rate in function of width

(*c*) Recall rate in function of distance

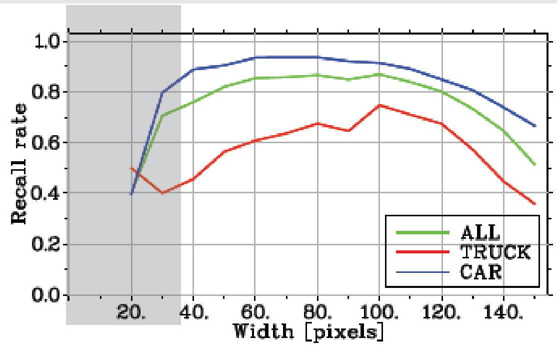
- Very few false positives
- Delay of detection reduces “recall rate”
- Poor results for trucks
- Statistics beyond 70 meters not significant

Quantitative evaluation

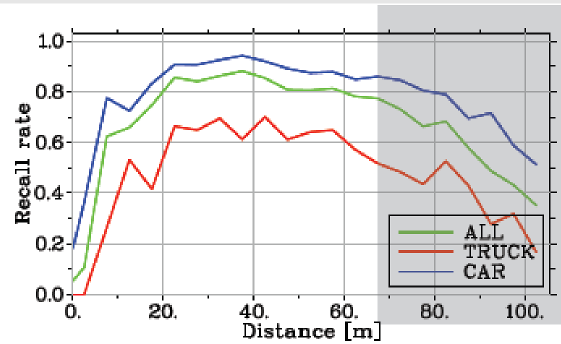
Daylight:



(a) Precision in function of width

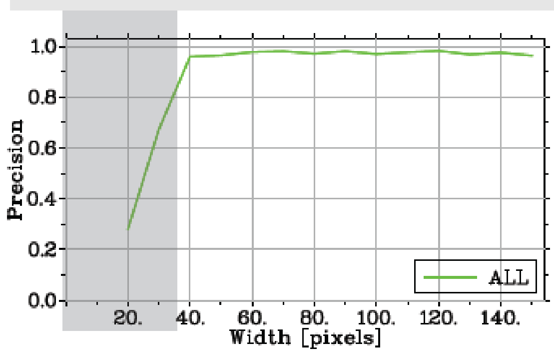


(b) Recall rate in function of width

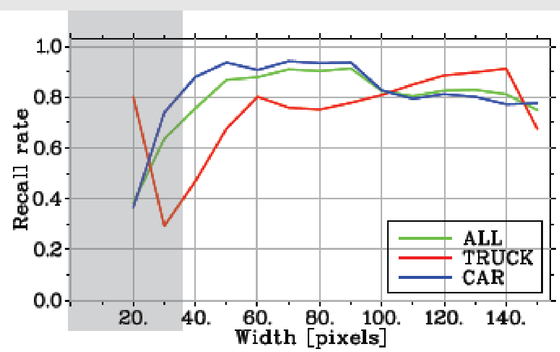


(c) Recall rate in function of distance

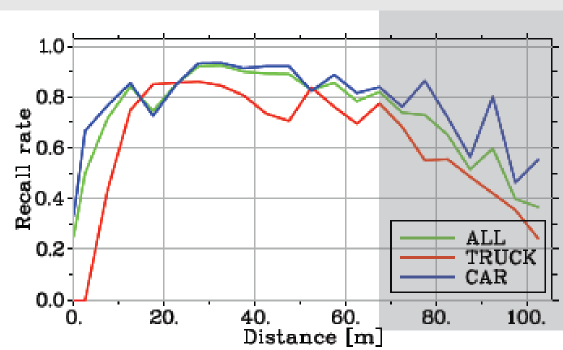
Sunset:



(a) Precision in function of width

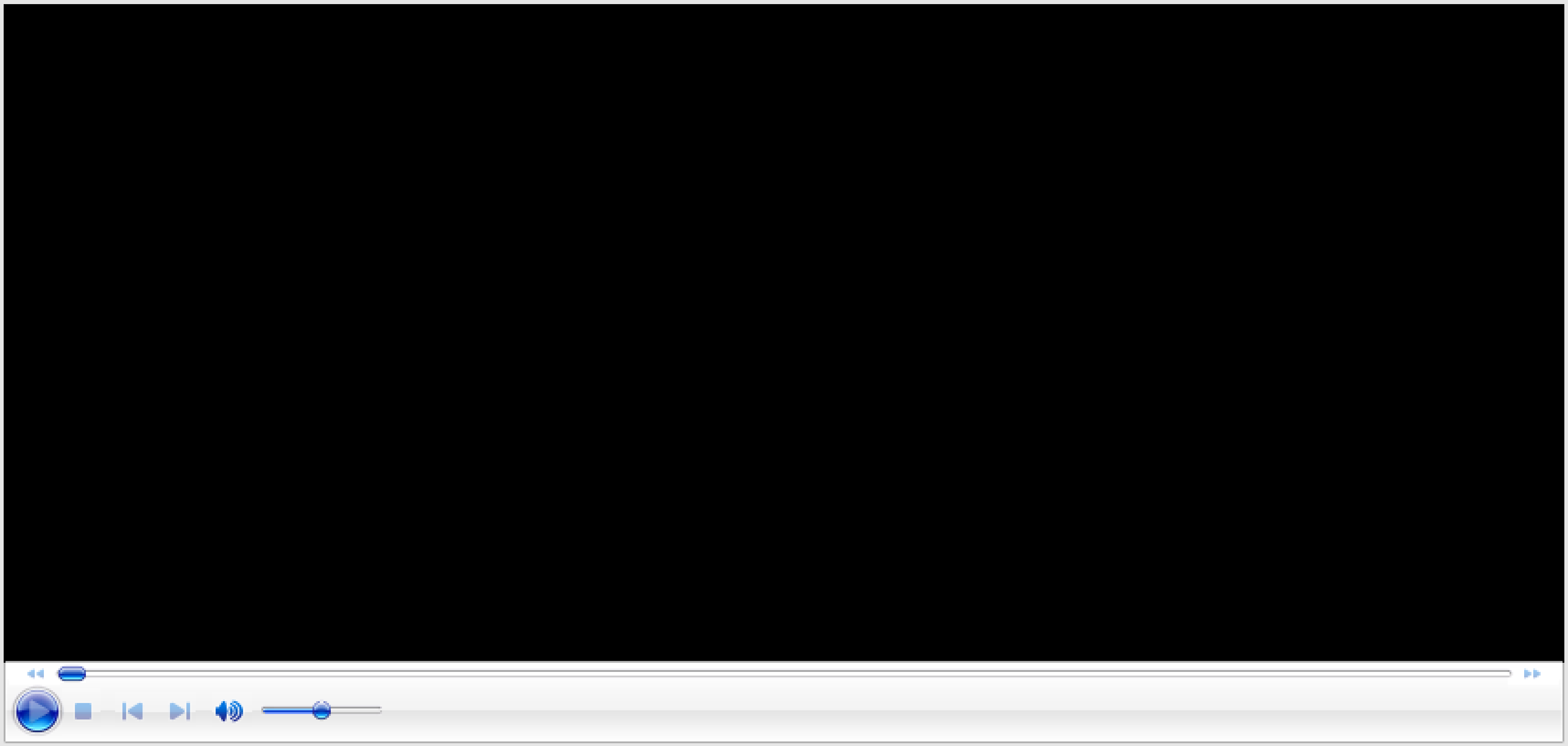


(b) Recall rate in function of width



(c) Recall rate in function of distance

Results



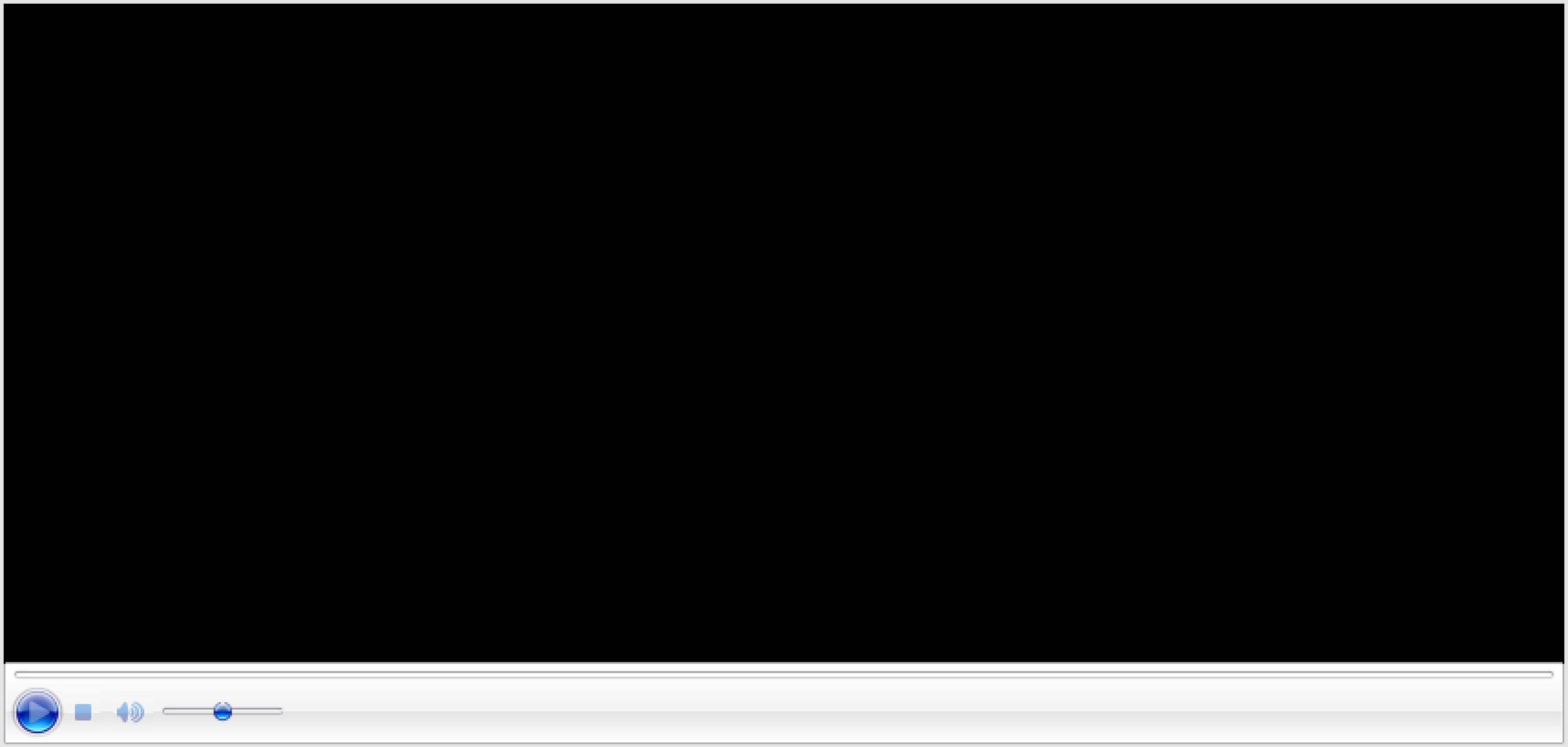
TOYOTA

Conclusions

- Real-time reliable detection and tracking of cars
- Next steps: measure & reduce confirmation time
  focus on trucks & other classes
  different scenarios
- Dataset is made public
- Images available, ground truth + software Oct ’12
- http://cmp.felk.cvut.cz/data/motorway/
- … or google: “motorway dataset”

TOYOTA m p

Thanks!



http://cmp.felk.cvut.cz/data/motorway/

TOYOTA

m p