Sekvenční deformace klasifikačních oken pro detekci nerigidních objektů

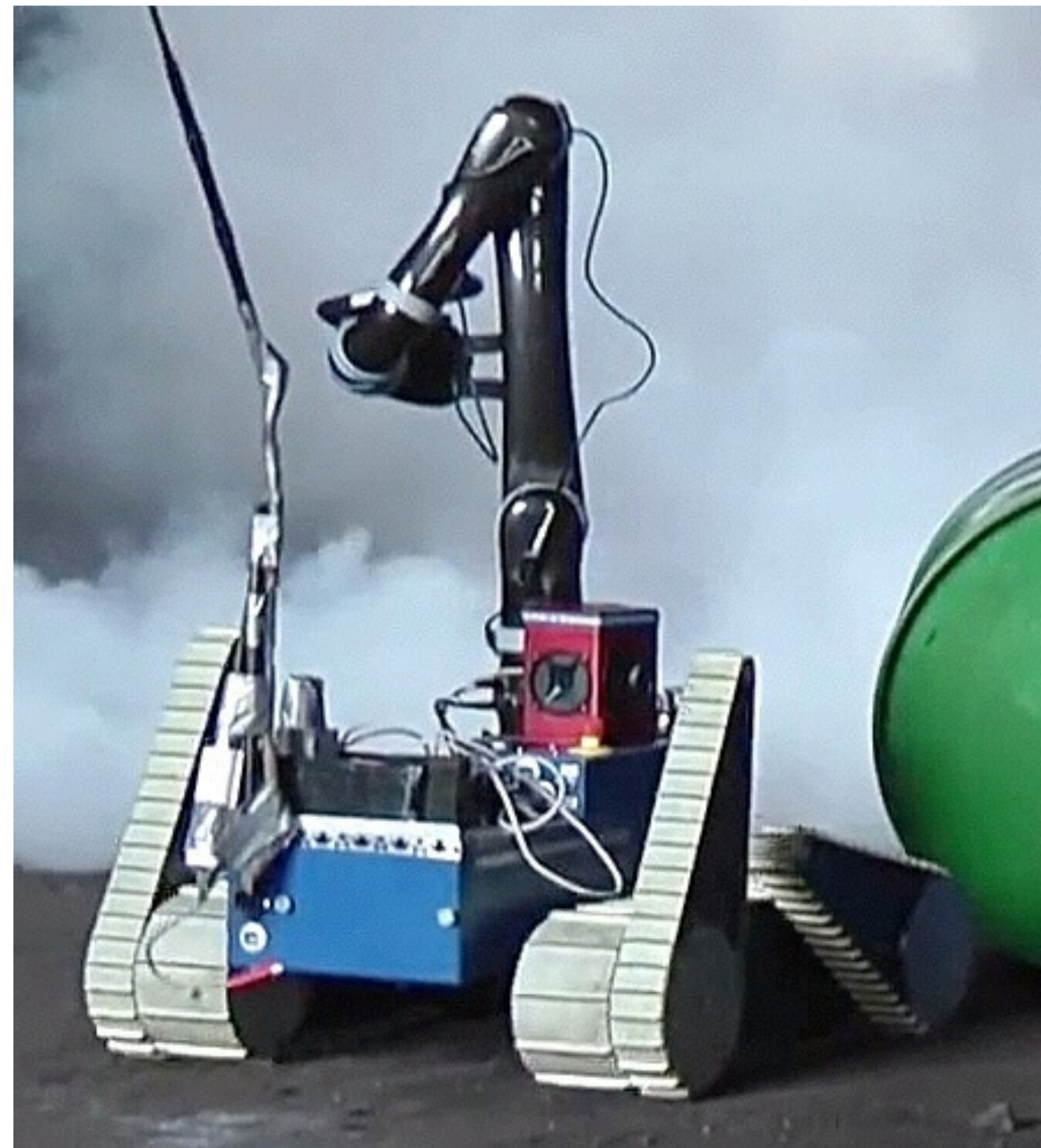
Tomáš Svoboda
http://cmp.felk.cvut.cz/~svoboda
svobodat@fel.cvut.cz
ČVUT, FEL, Katedra kybernetiky
Centrum strojového vnímání

společná práce: K. Zimmermann, D. Hurych, and J. Matas

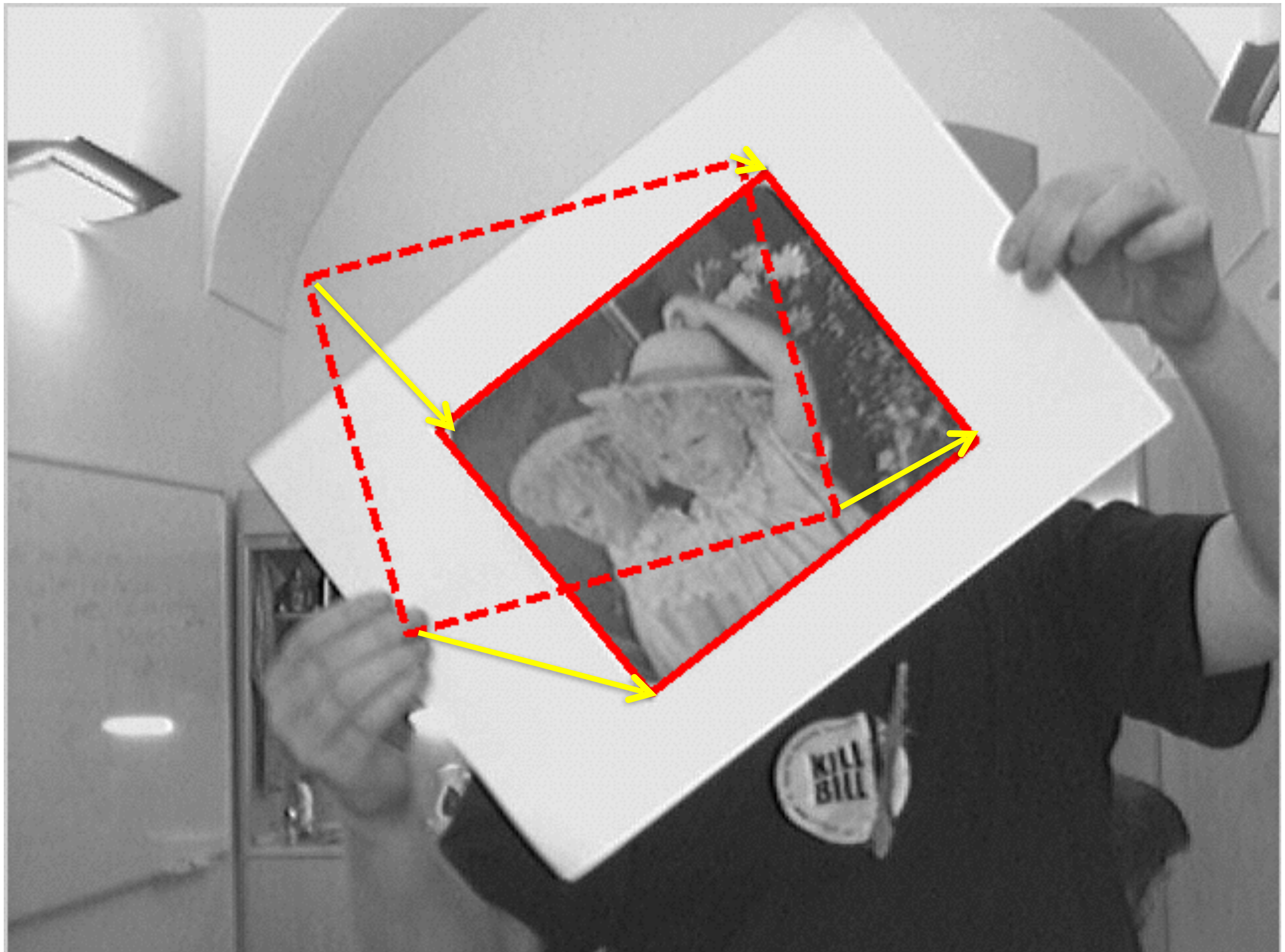
Robot perception

need for speed

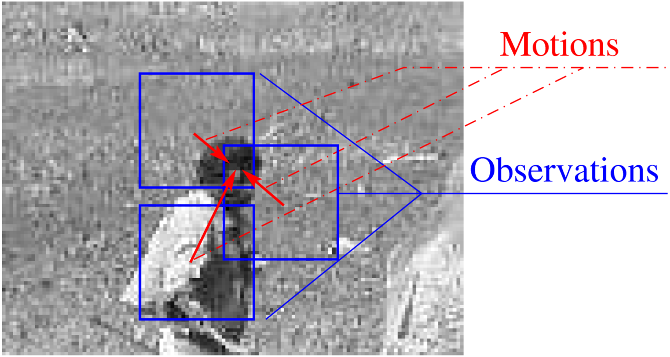
- on-board computation
- results any-time
- many concurrent processes



Learn to deform/align



Linear mapping for tracking

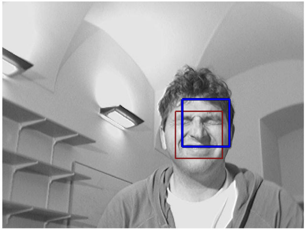


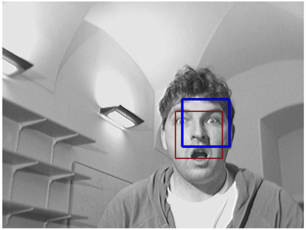
$\Phi($$) = (0,0)^T$ $\Phi($$) = (12,7)^T$

$\Phi($$) = (-14,2)^T$ $\Phi($$) = (-9,18)^T$

$\Phi($$) = (14,-14)^T$ $\Phi($$) = (-16,-12)^T$

Learning alignment for one predictor





- $\varphi($  $) = (0,0)^\top$
- $\varphi($  $) = (-25,0)^\top$
- $\varphi($  $) = (25,-15)^\top$

- $\varphi($  $) = (0,0)^\top$
- $\varphi($  $) = (-25,0)^\top$
- $\varphi($  $) = (25,-15)^\top$

Connection to KLT

$$\Delta \mathbf{p} = \mathtt{H}^{-1} \sum_{\mathbf{x}} \left[\nabla I^\top \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right]^\top [T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x}; \mathbf{p}))]$$

where:

$$\mathtt{H} = \sum_{\mathbf{x}} \left[\nabla I^\top \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right]^\top \left[\nabla I^\top \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right]$$

Reformulating *regression*:

Regression matrix

transformation
alignment

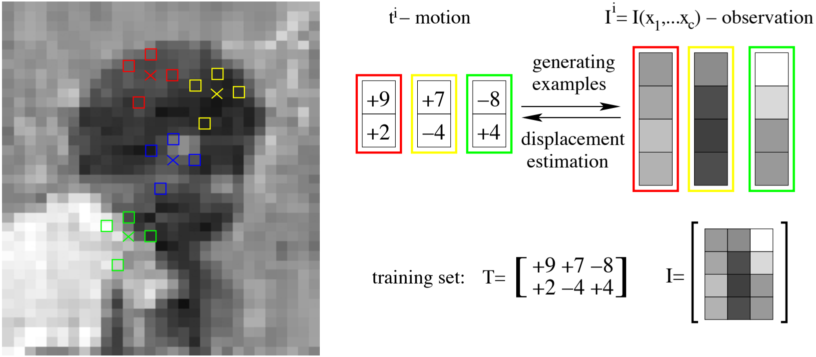
$$\mathbf{p} = \varphi\Big(I(\mathbf{x})\Big) = \mathtt{H}\Big(I(\mathbf{x}) - T(\mathbf{x})\Big)$$

Observation

Template, model

How to get $\mathtt{H}$?

Generating training examples



Training set: $(\mathtt{I}, \mathtt{P})$
$\mathtt{I} = [\mathbf{I}^1 - \mathbf{T}, \mathbf{I}^2 - \mathbf{T}, \ldots \mathbf{I}^d - \mathbf{T}]$ and $\mathtt{P} = [\mathbf{p}^1, \mathbf{p}^2, \ldots \mathbf{p}^d]$.

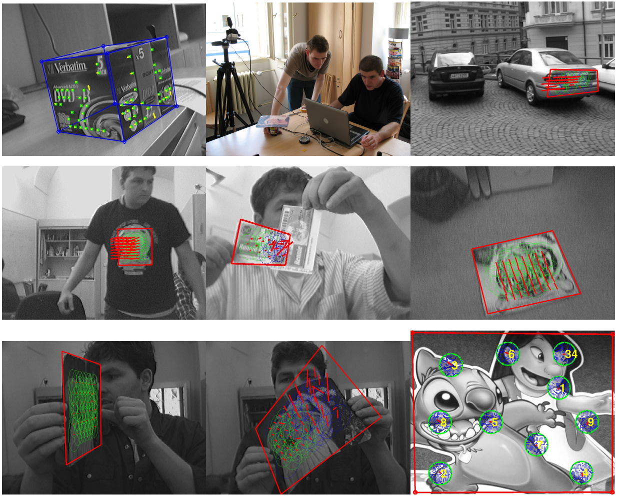
Object detection

Changes in view angles and object deformations cause problems to standard object detectors






Motion blur, fast motion, views from acute angles and other image distortions.



Sliding window detection

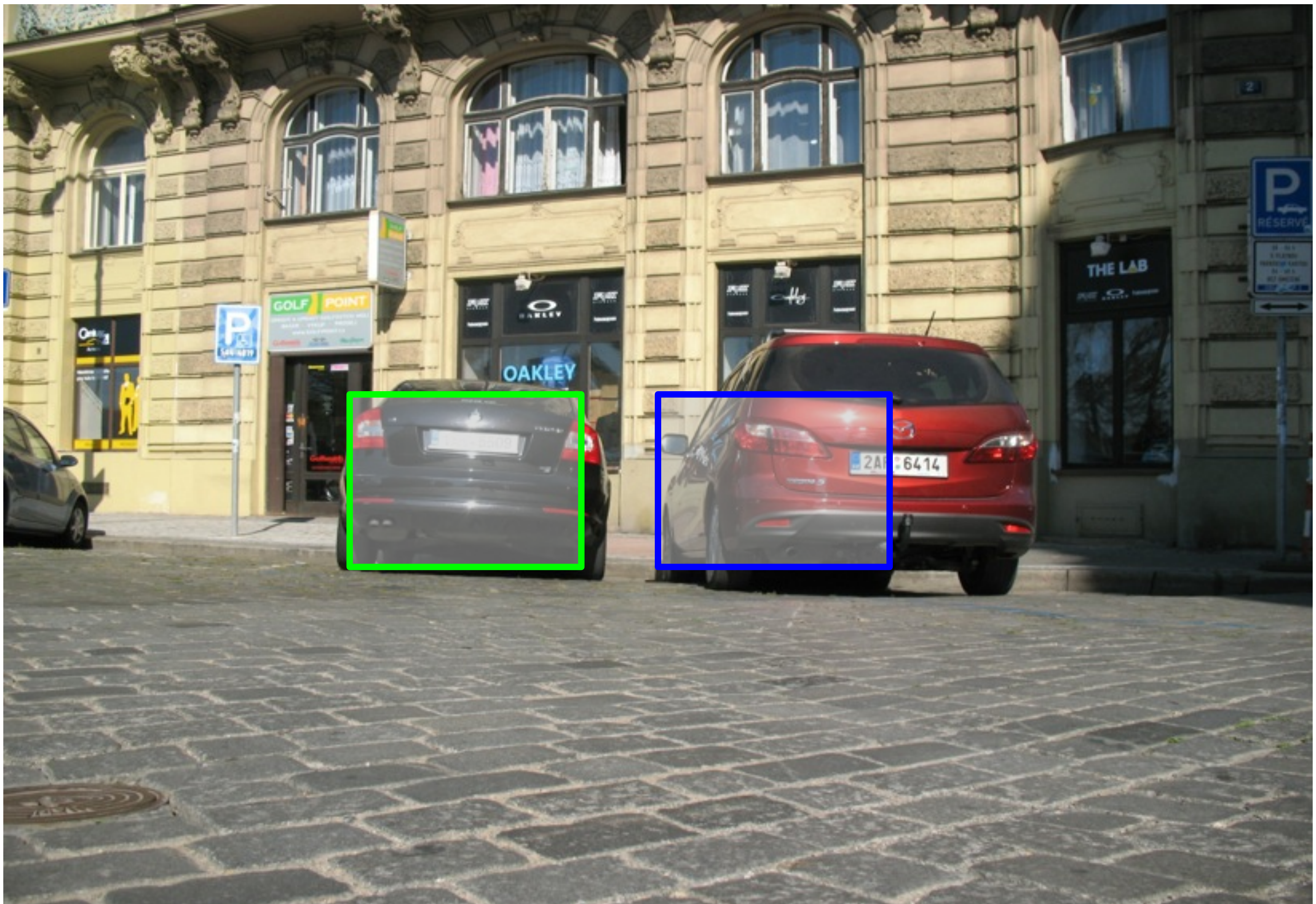
- sliding sparsely over
- aligning each candidate (by the learned regressor)
- repeat if necessary


GOLF POINT
OAKLEY
THE LAB
P
2AF 6414


GOLF
POINT
OAKLEY
THE LAB
P
2AF 6414

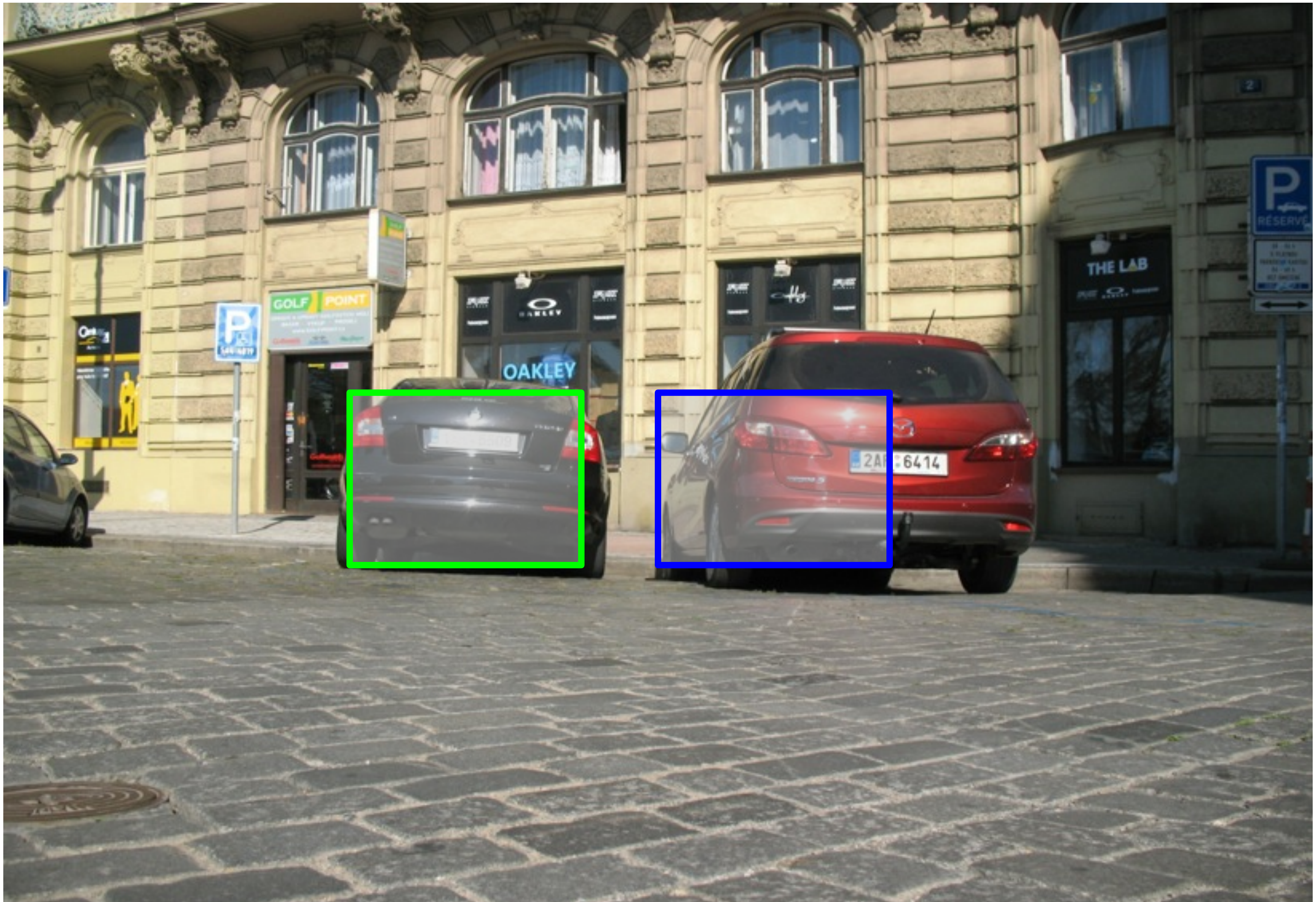

GOLF POINT
OAKLEY
THE LAB
2AF 6414

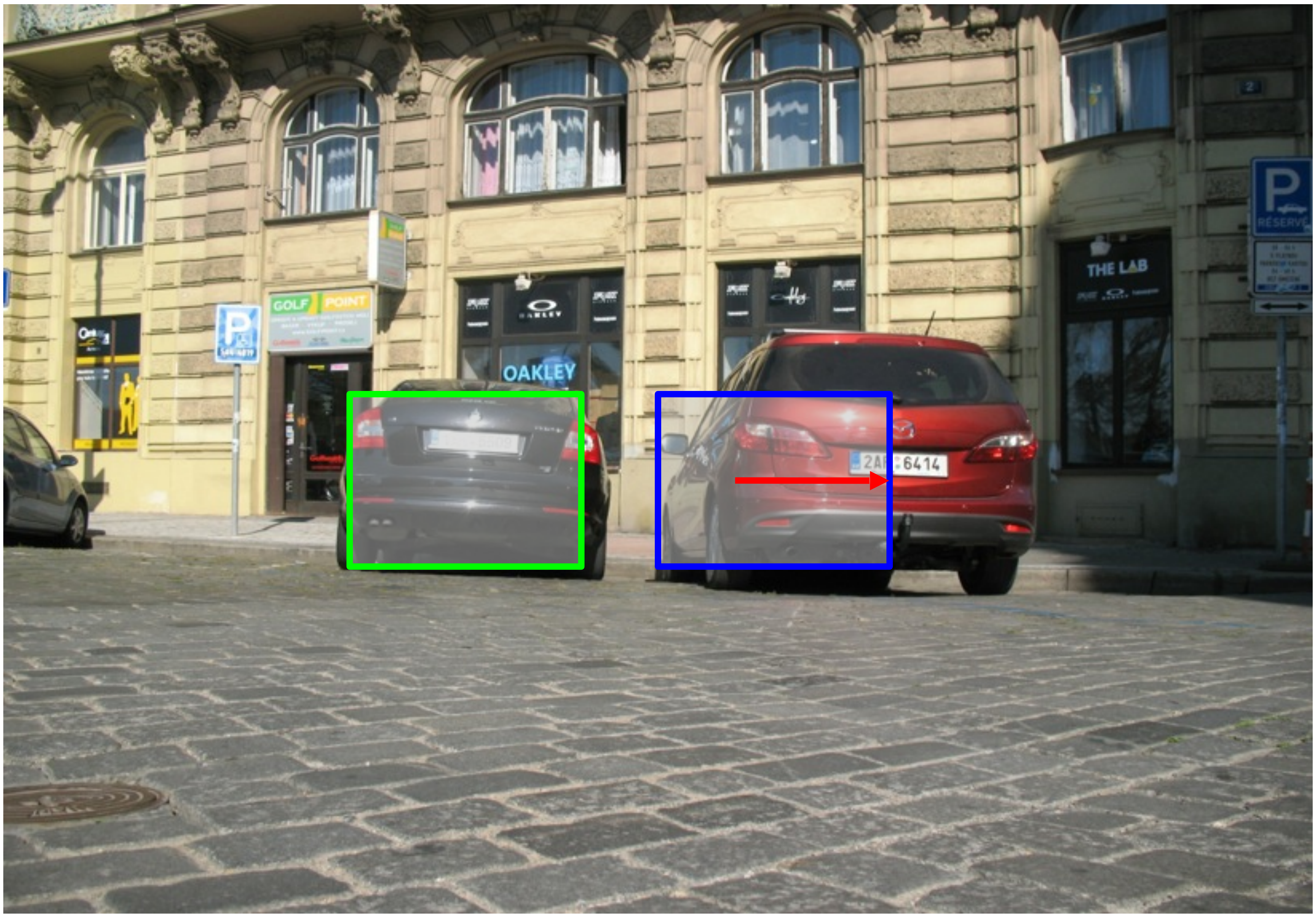

GOLF POINT
OAKLEY
THE LAB
P
2AF 6414

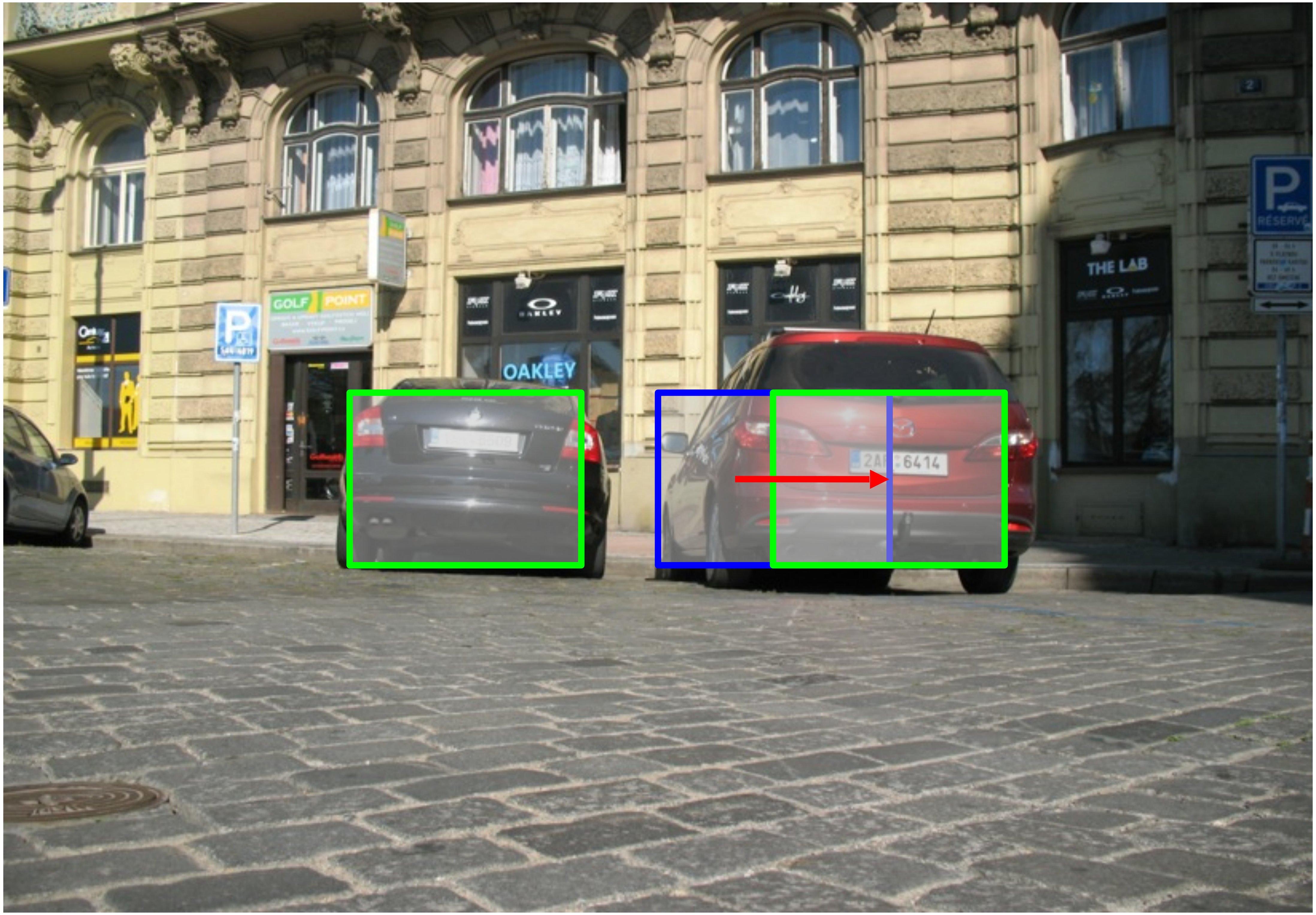

GOLF POINT
OAKLEY
THE LAB

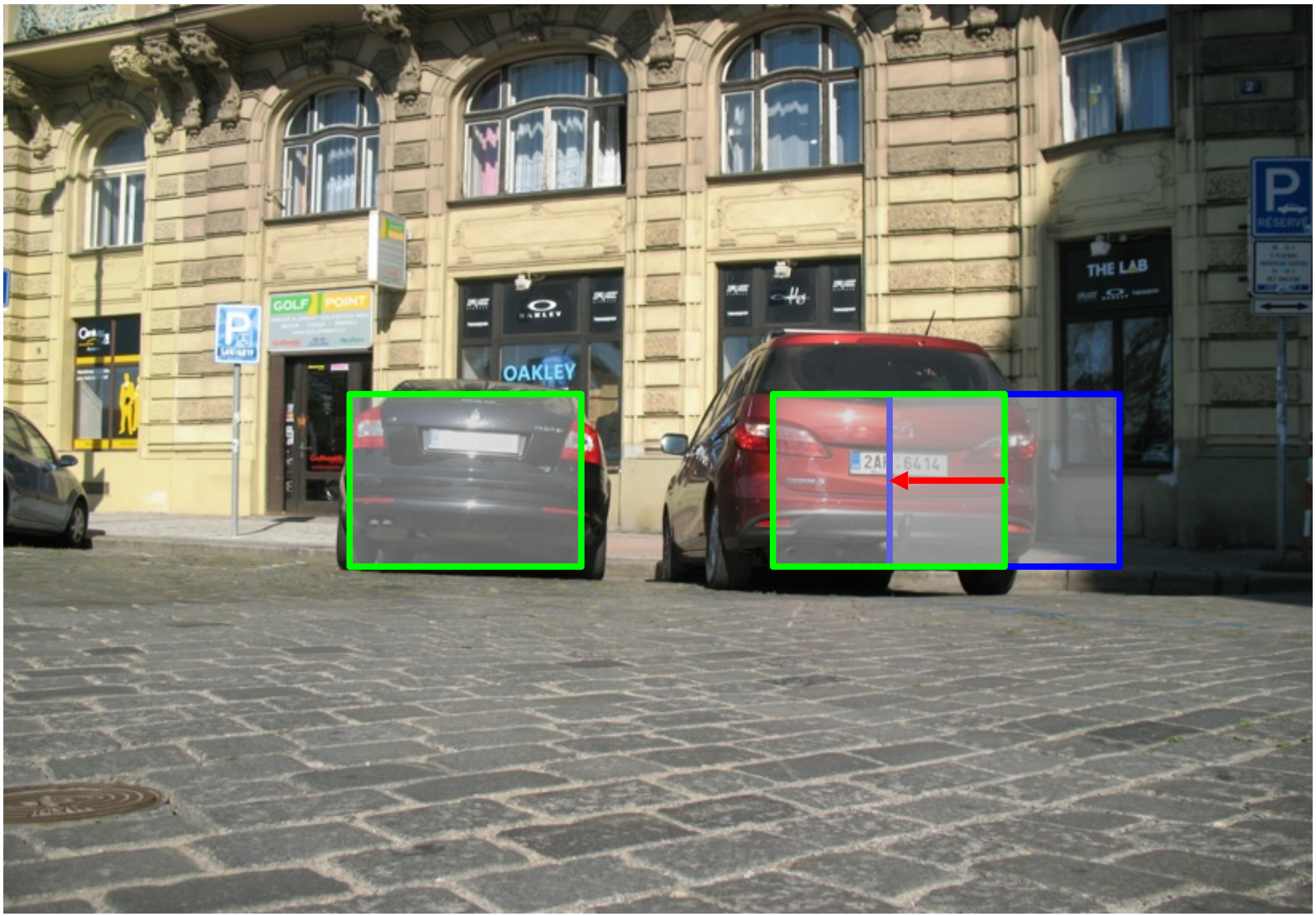
GOLF
POINT
OAKLEY
THE LAB
2AF 6414

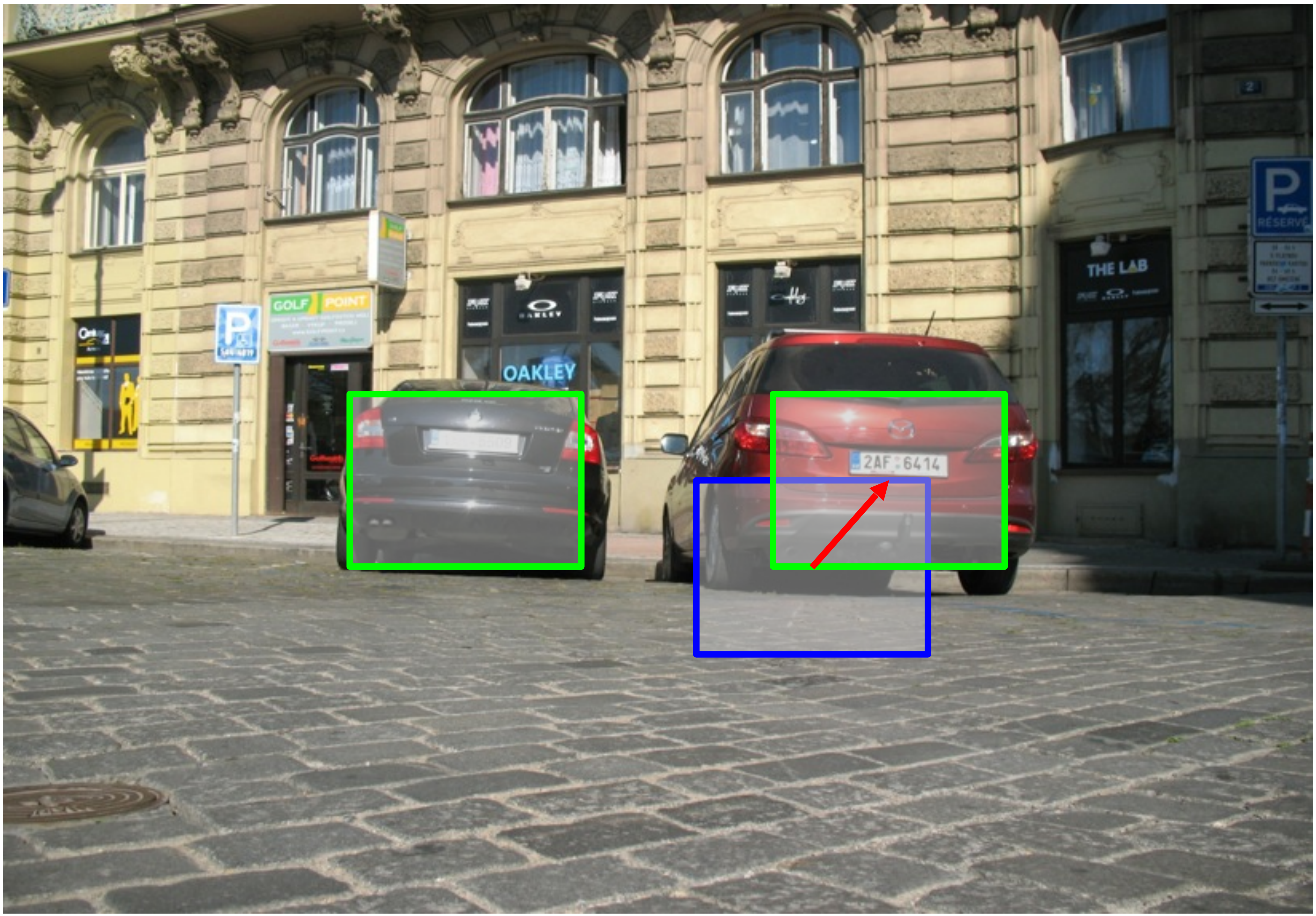

GOLF POINT
OAKLEY
THE LAB
P
2AF 6414


GOLF
POINT
OAKLEY
THE LAB
P

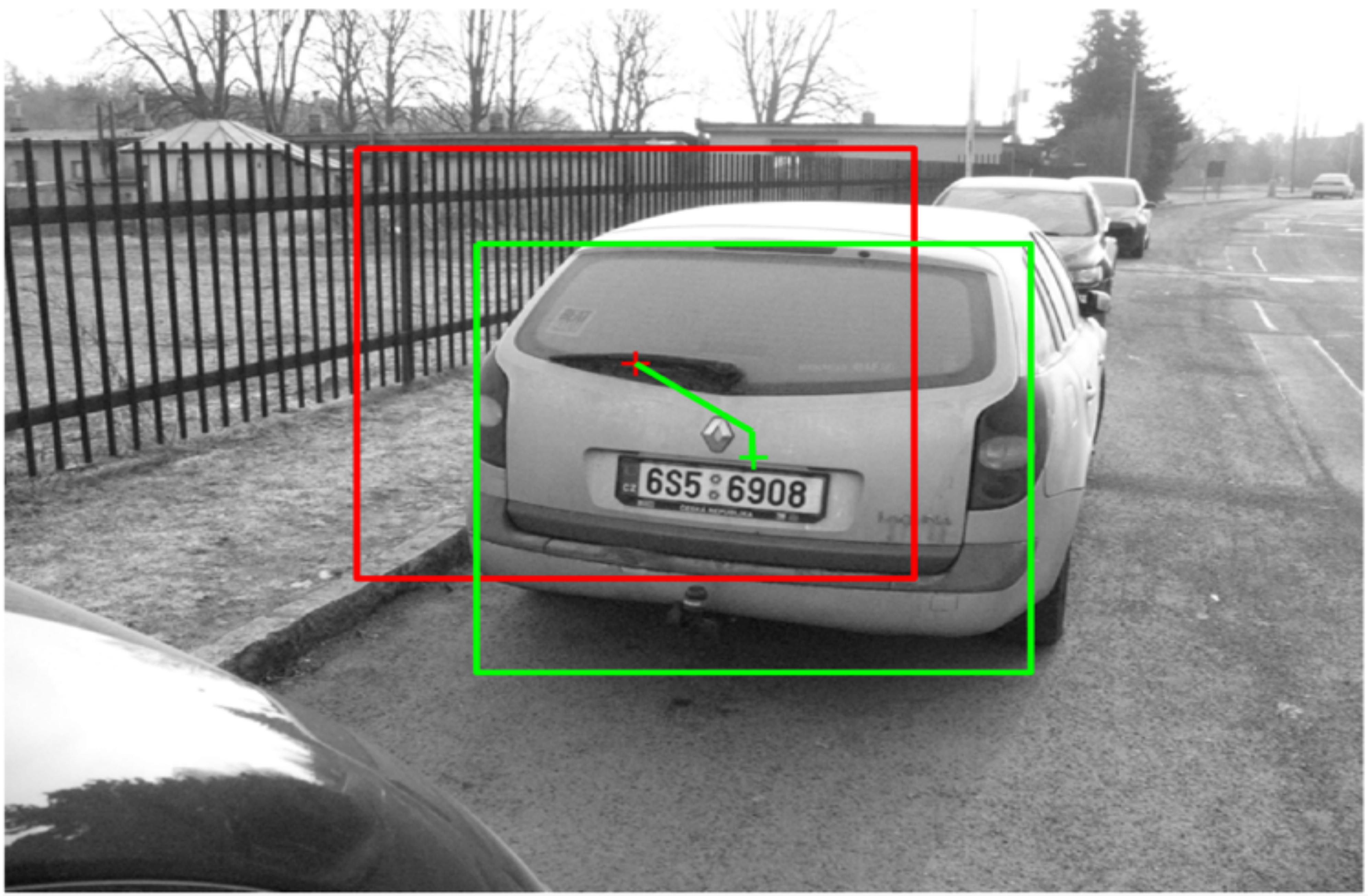

GOLF POINT
OAKLEY
THE LAB


GOLF POINT
OAKLEY
THE LAB
P
RESERVÉ

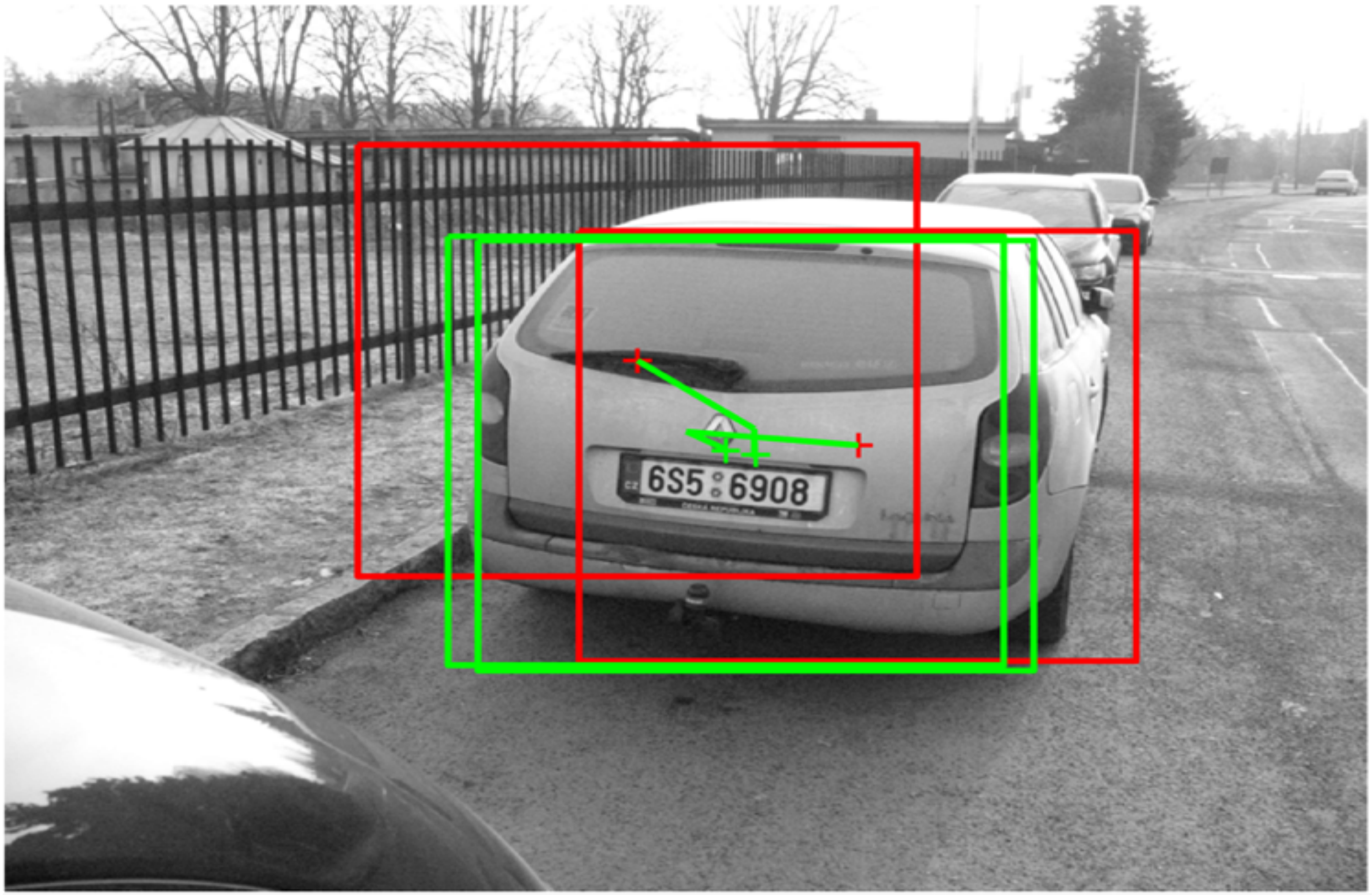

GOLF
POINT
OAKLEY
THE LAB
P


GOLF POINT
OAKLEY
THE LAB
2AF 6414

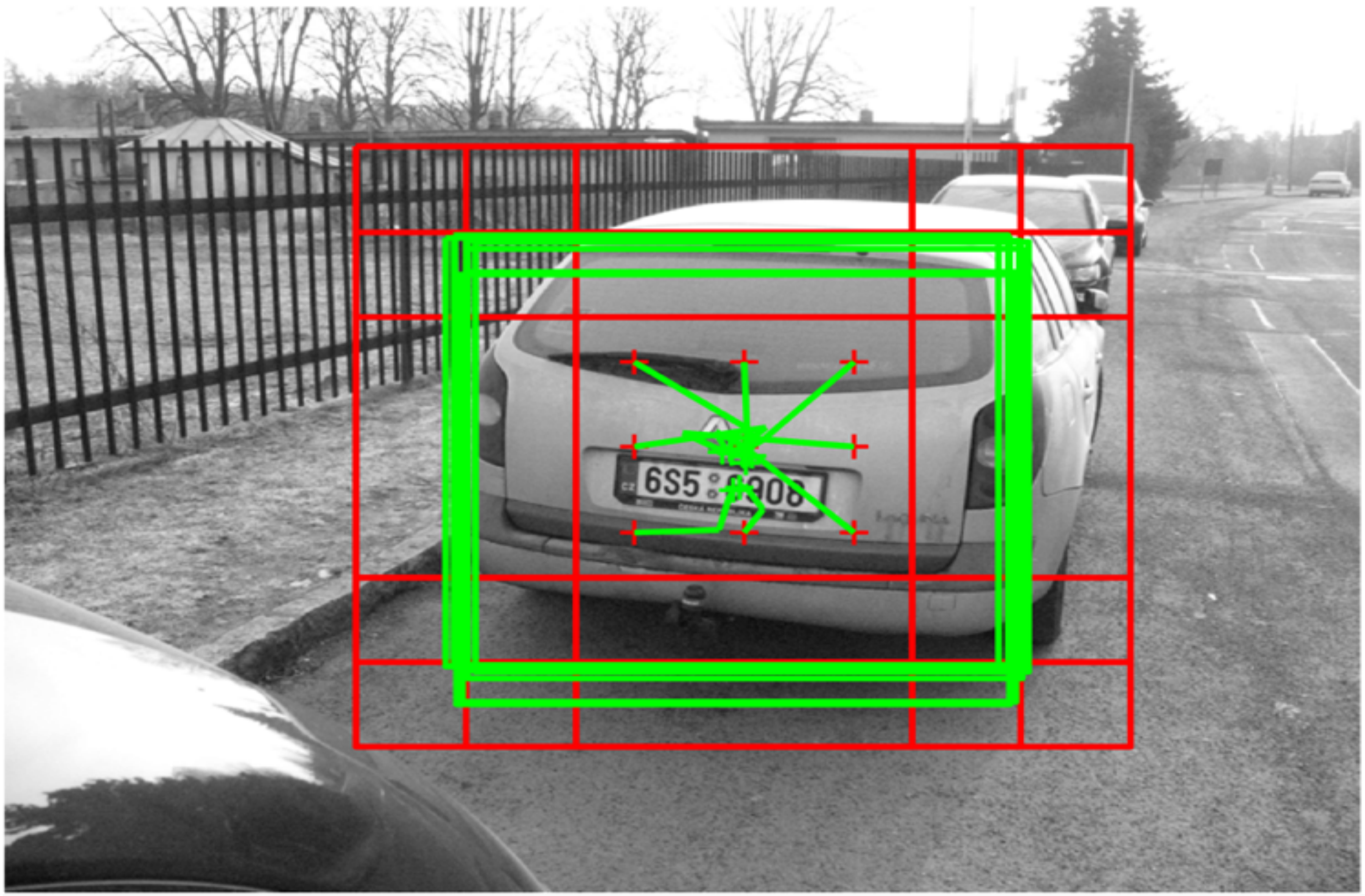
Detection with aligned cascade



Detection with aligned cascade

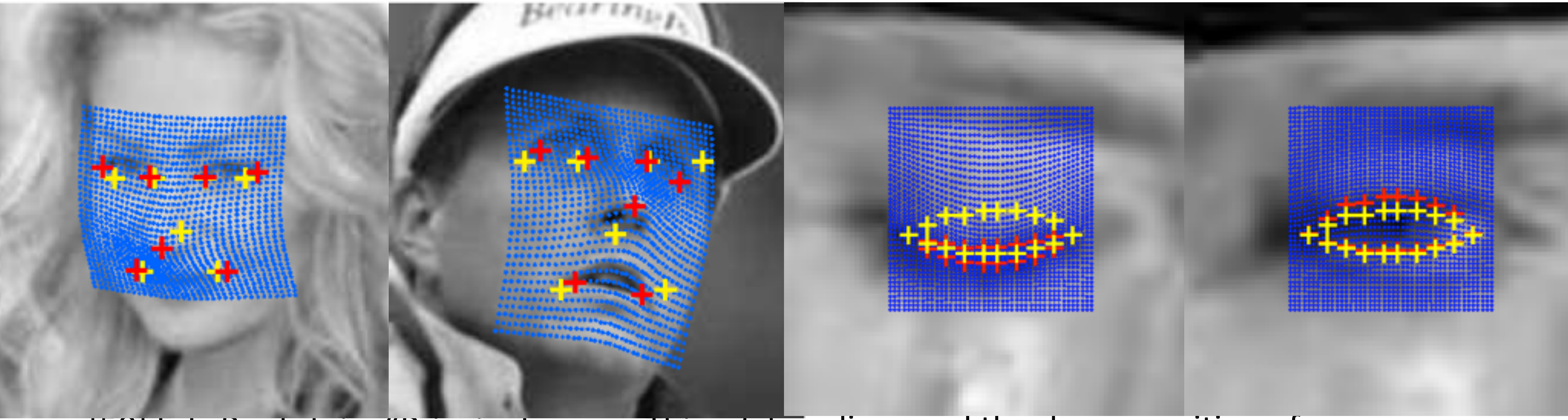


Detection with aligned cascade



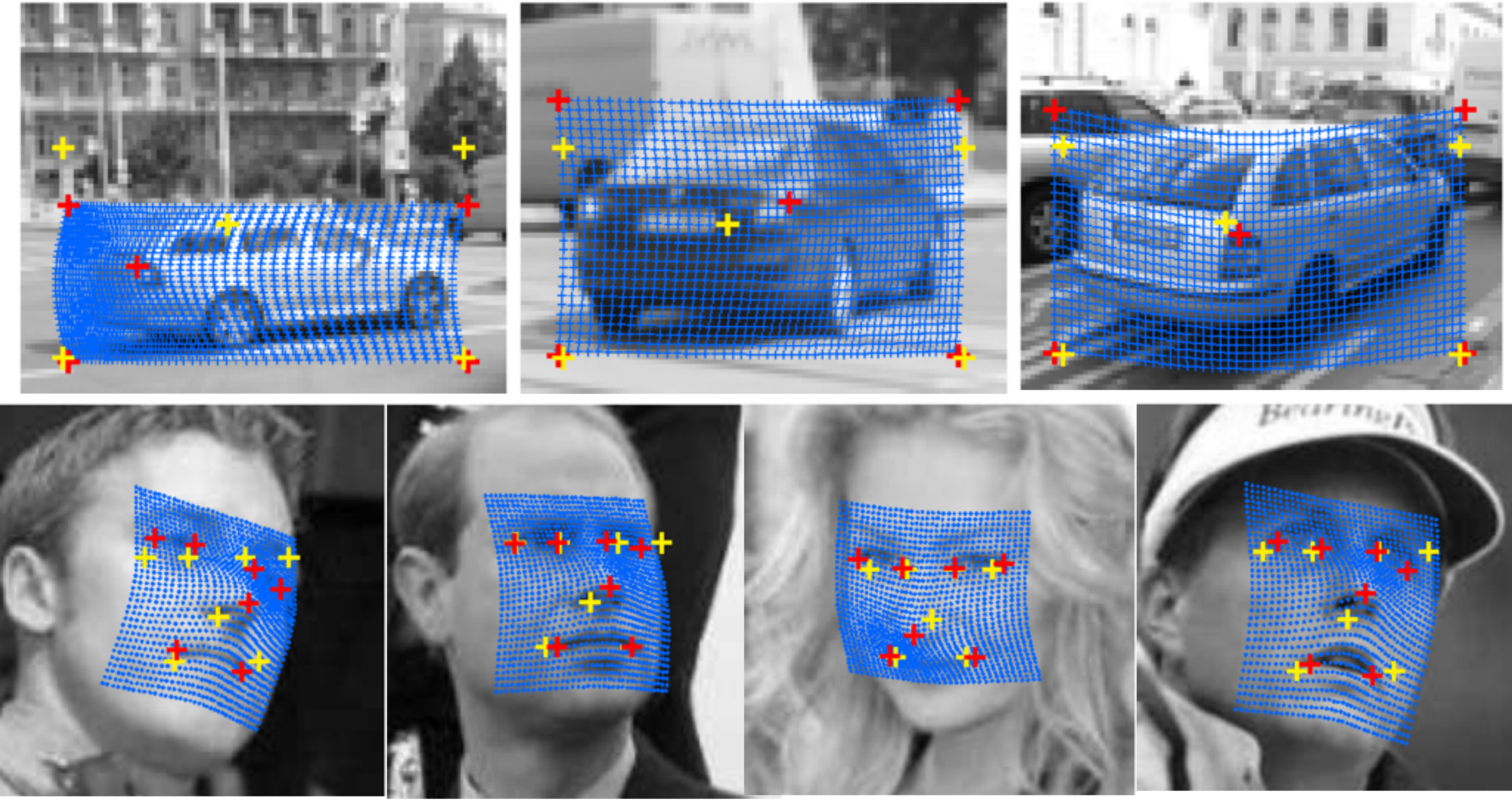
Deformable Object Tracking

- Ground truth annotation – red crosses
- Starting positions – yellow crosses
- Generate the deformed grids [50]

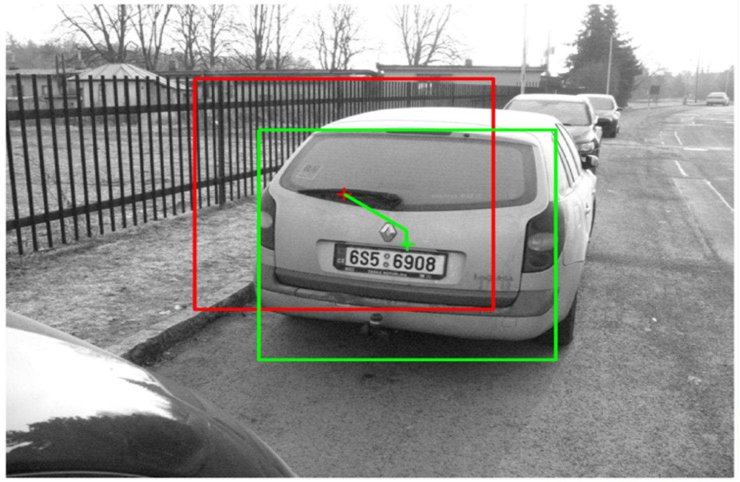


[50] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," TPAMI

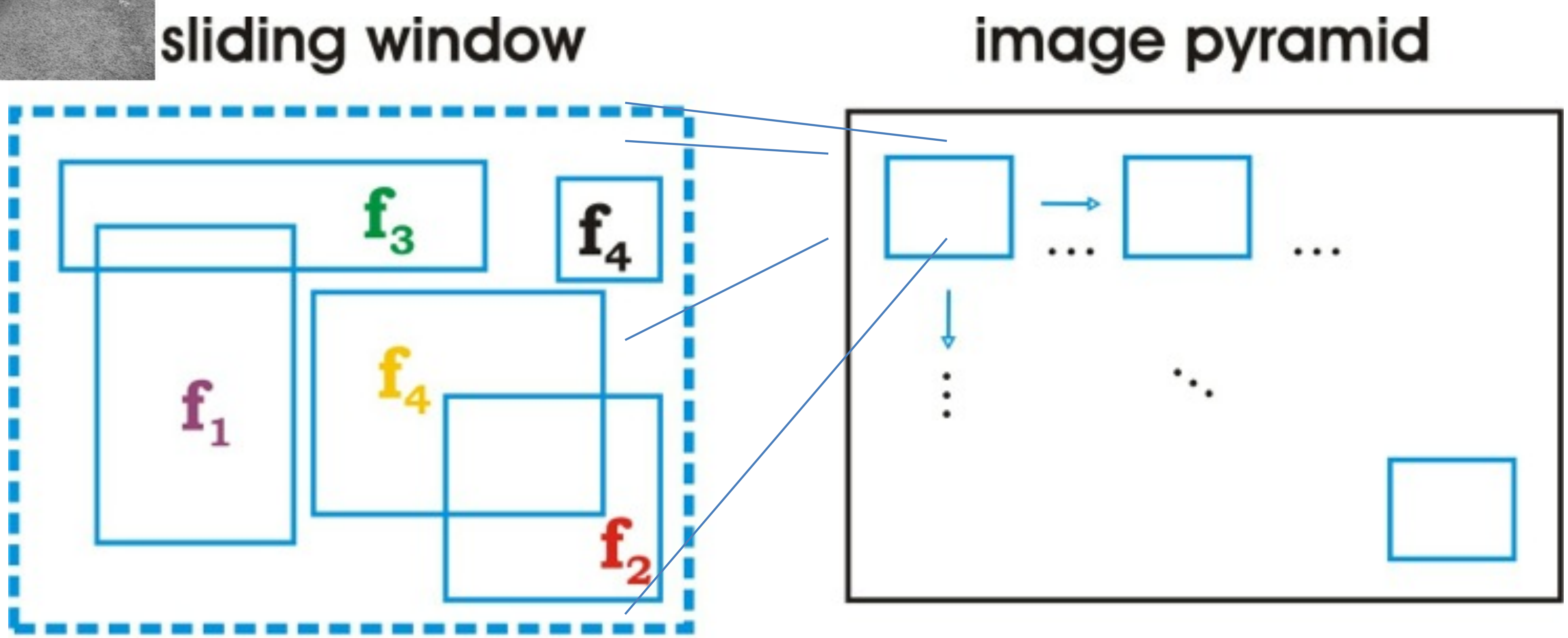
Detecting deformable objects

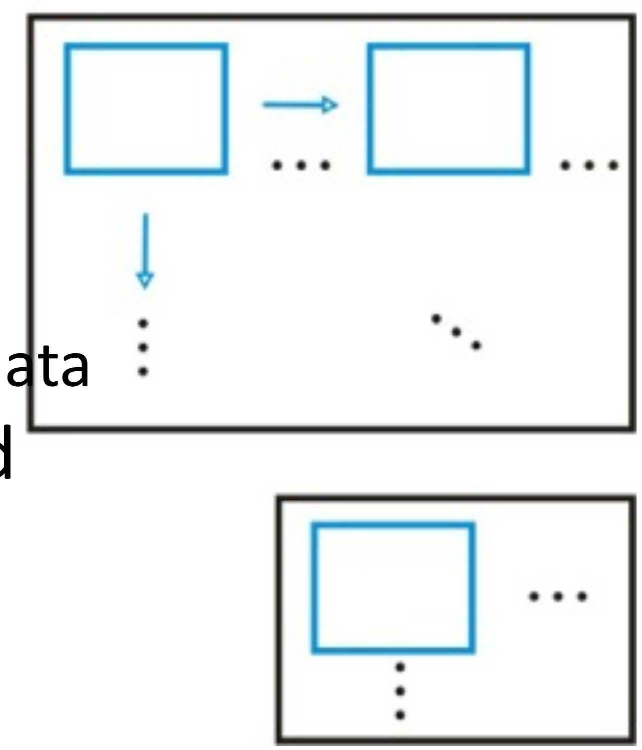


F. L. Bookstein, “Principal warps: Thin-plate splines and the decomposition of deformations,” TPAMI



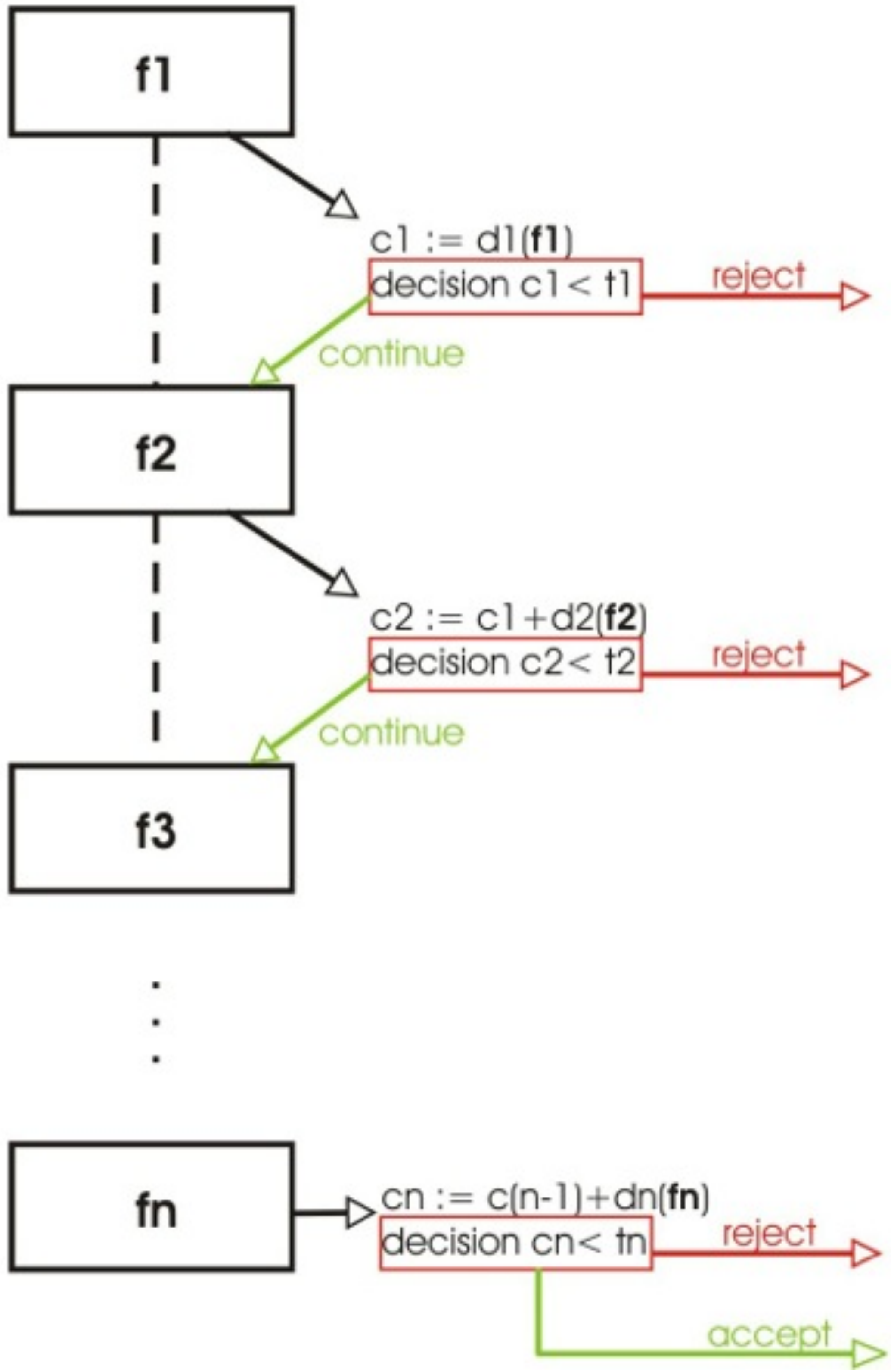
Sliding window detector





- fixed features positions in the sliding window
- search space reduction for speed-up
  - deteriorates detector’s performance - not well aligned data
- more complex deformations usually not considered

Sequential Decision Process

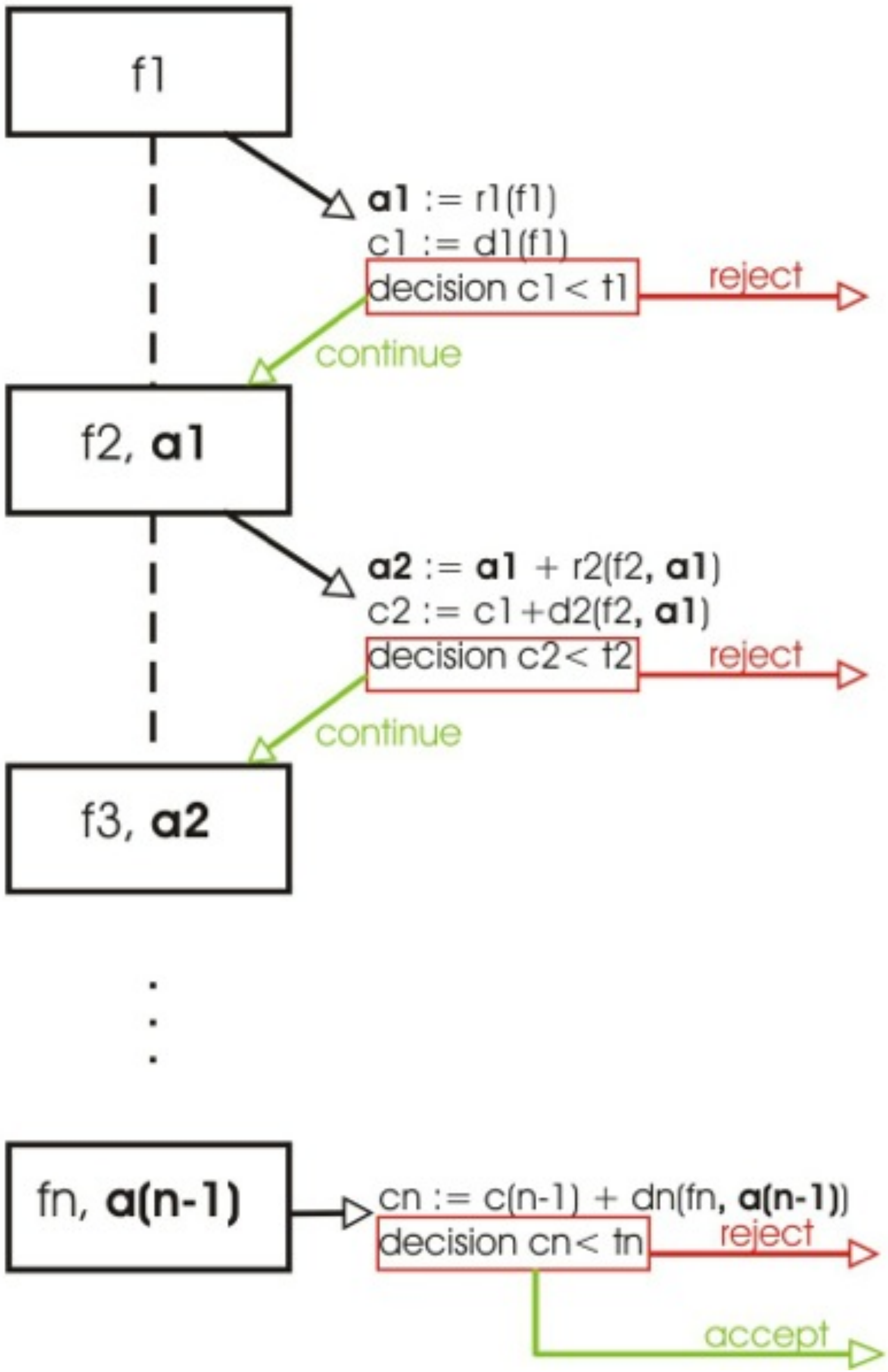


f_i – feature

c_i – confidence

d_i – weak classifier

Sequential Decision Process with Local Interleaved Sequential Alignment



f_i – feature

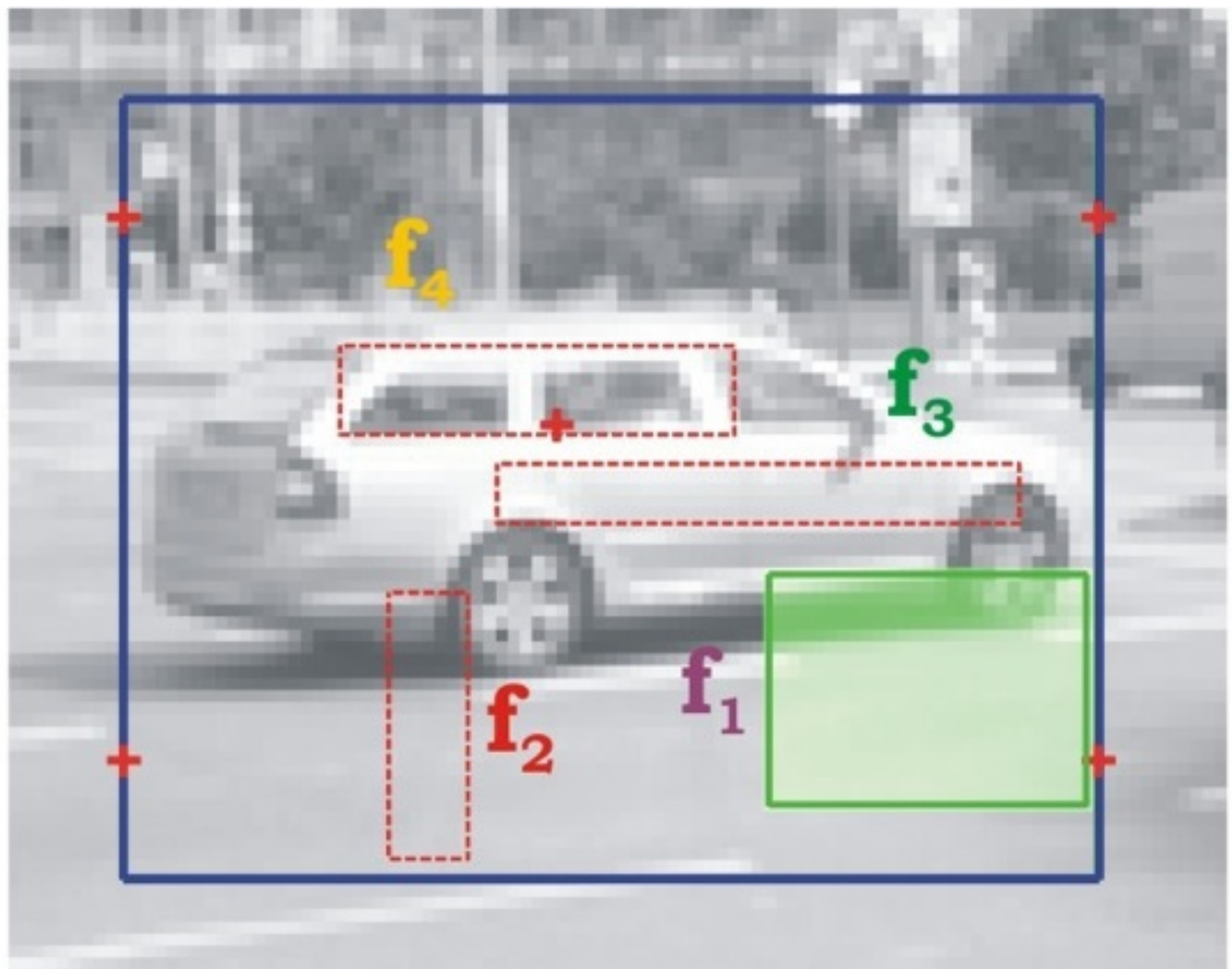
c_i – confidence

d_i – weak classifier

a_i – alignment param.

r_i – regressor

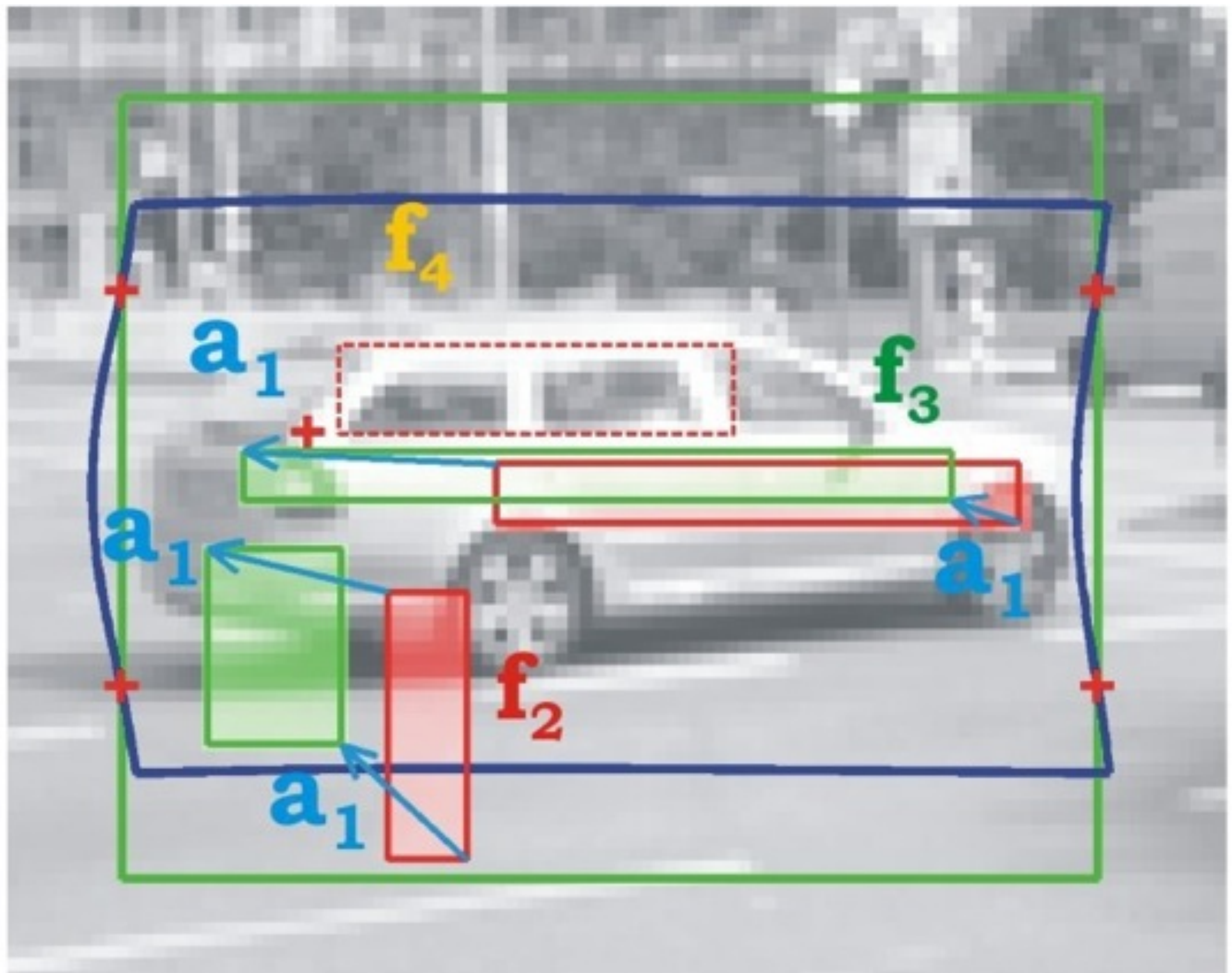
SDP with LISA classification



$$a_1 := r_1(f_1)$$

$$c_1 := d_1(f_1)$$

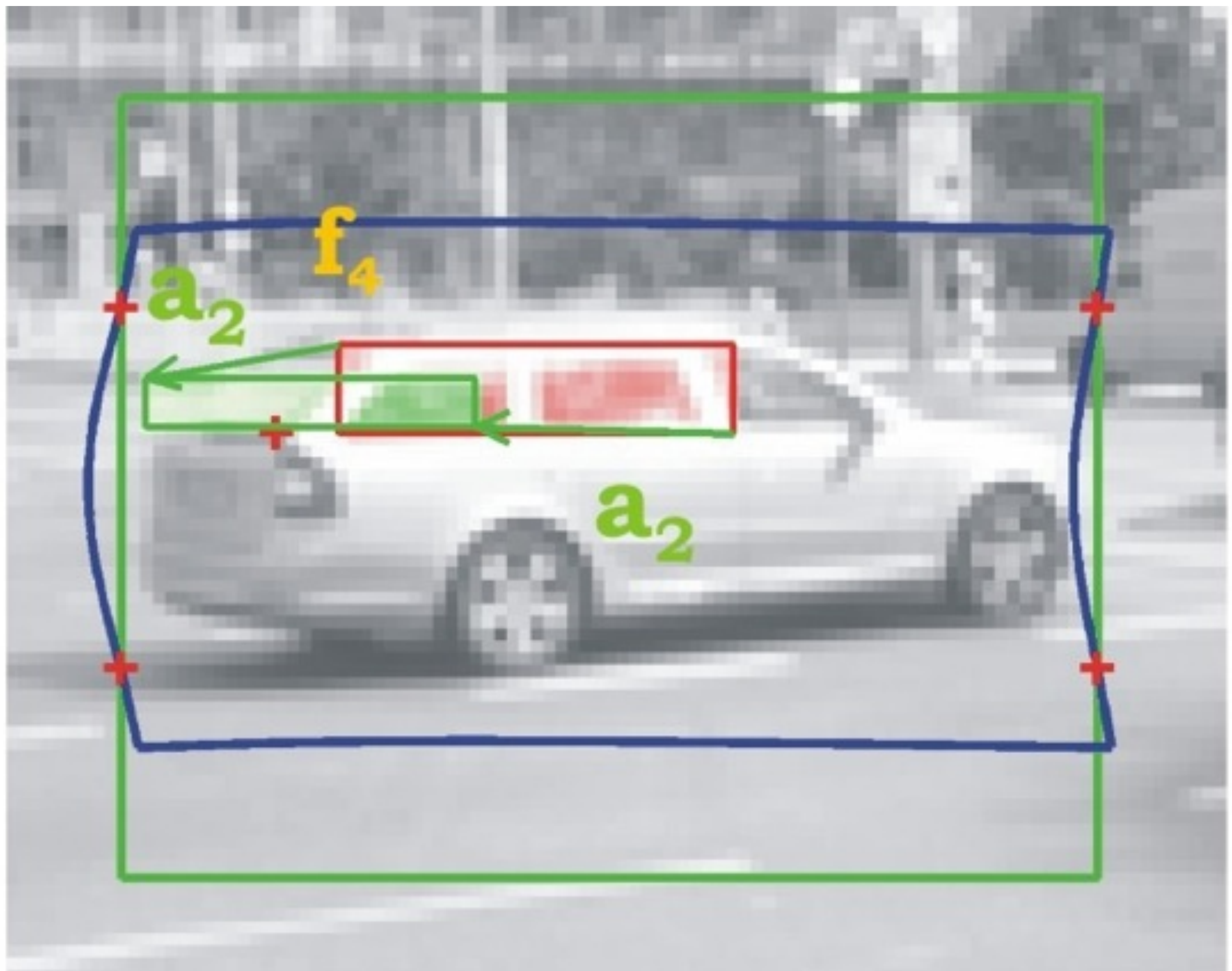
SDP with LISA classification



$$\mathbf{a}_2 := \mathbf{a}_1 + \mathbf{r}_2(\mathbf{f}_2) + \mathbf{r}_3(\mathbf{f}_3)$$

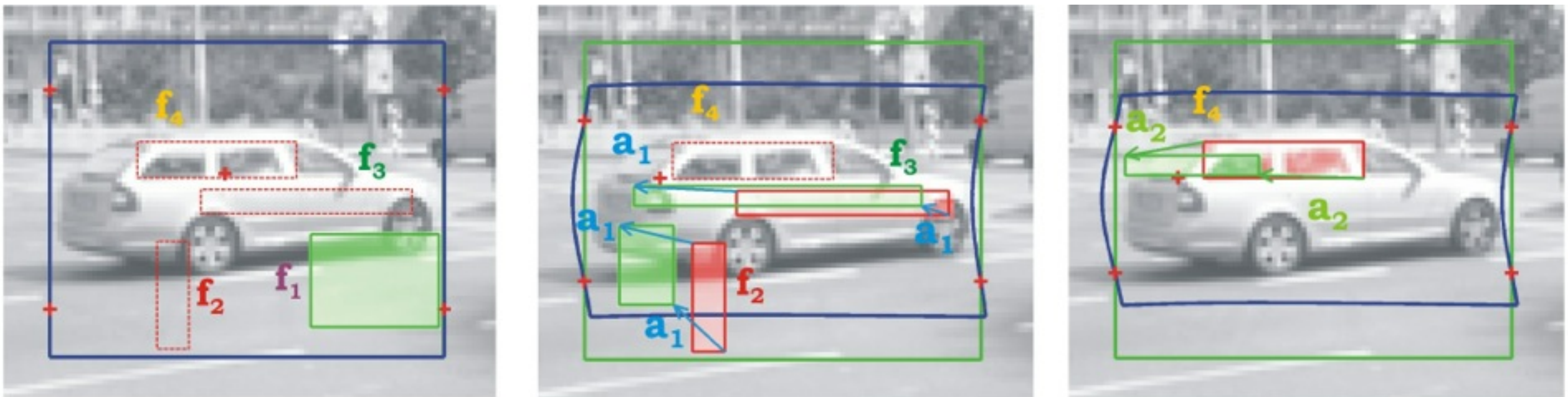
$$\mathbf{c}_2 := \mathbf{c}_1 + \mathbf{d}_2(\mathbf{f}_2) + \mathbf{d}_3(\mathbf{f}_3)$$

SDP with LISA classification



$$c_3 := c_2 + d_4(f_4)$$

SDP with LISA classification

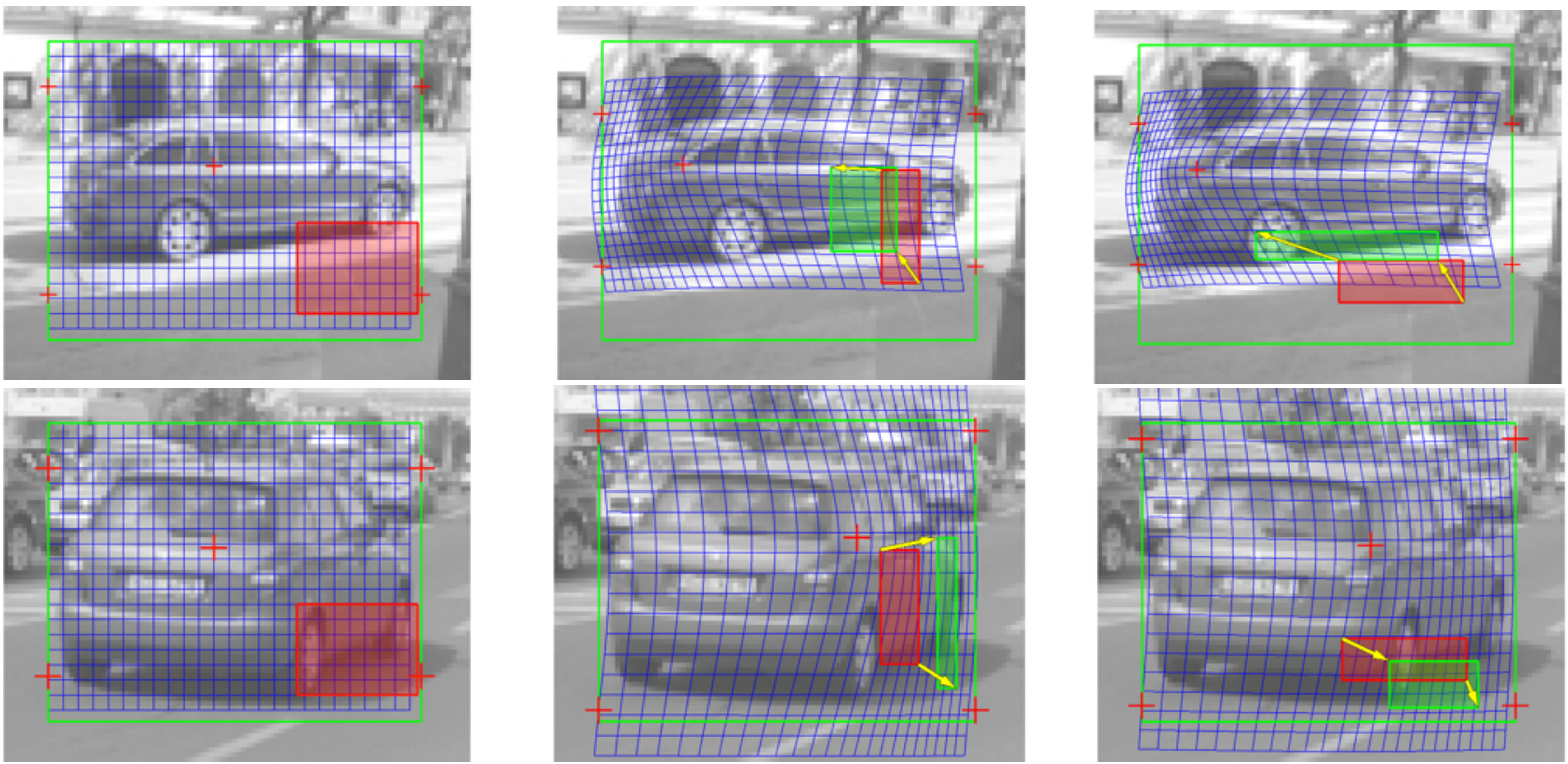


$\mathbf{a}_1 := \mathbf{r}_1(\mathbf{f}_1)$
$\mathbf{c}_1 := \mathbf{d}_1(\mathbf{f}_1)$

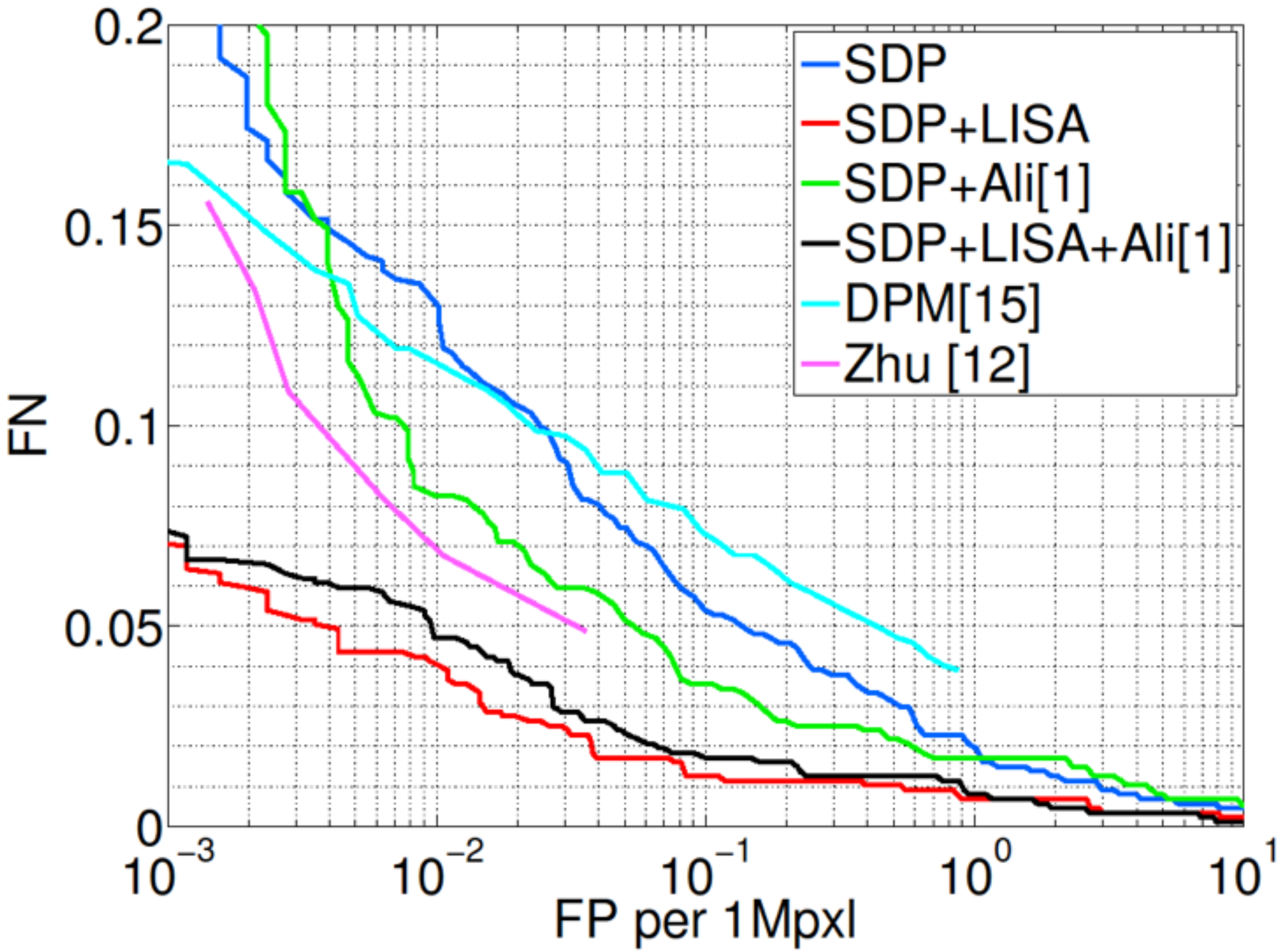
$\mathbf{a}_2 := \mathbf{a}_1 + \mathbf{r}_2(\mathbf{f}_2) + \mathbf{r}_3(\mathbf{f}_3)$
$\mathbf{c}_2 := \mathbf{c}_1 + \mathbf{d}_2(\mathbf{f}_2) + \mathbf{d}_3(\mathbf{f}_3)$

$\mathbf{c}_3 := \mathbf{c}_2 + \mathbf{d}_4(\mathbf{f}_4)$

	f_1	f_2	f_3	f_4	...
confidence update	$d_1(f_1)$	$d_2(f_2, \mathbf{a}_1)$	$d_3(f_3, \mathbf{a}_2)$	$d_4(f_4, \mathbf{a}_3)$	...
alignment update	$r_1(f_1)$	$r_2(f_2, \mathbf{a}_1)$	$r_3(f_3, \mathbf{a}_2)$	$r_4(f_4, \mathbf{a}_3)$	...



AFW faces dataset



[1] K. Ali et al., *A real-time deformable detector*, TPAMI, 2012

[12] X. Zhu et al., *Face detection, pose estimation, and landmark localization in the wild*, in CVPR, 2012

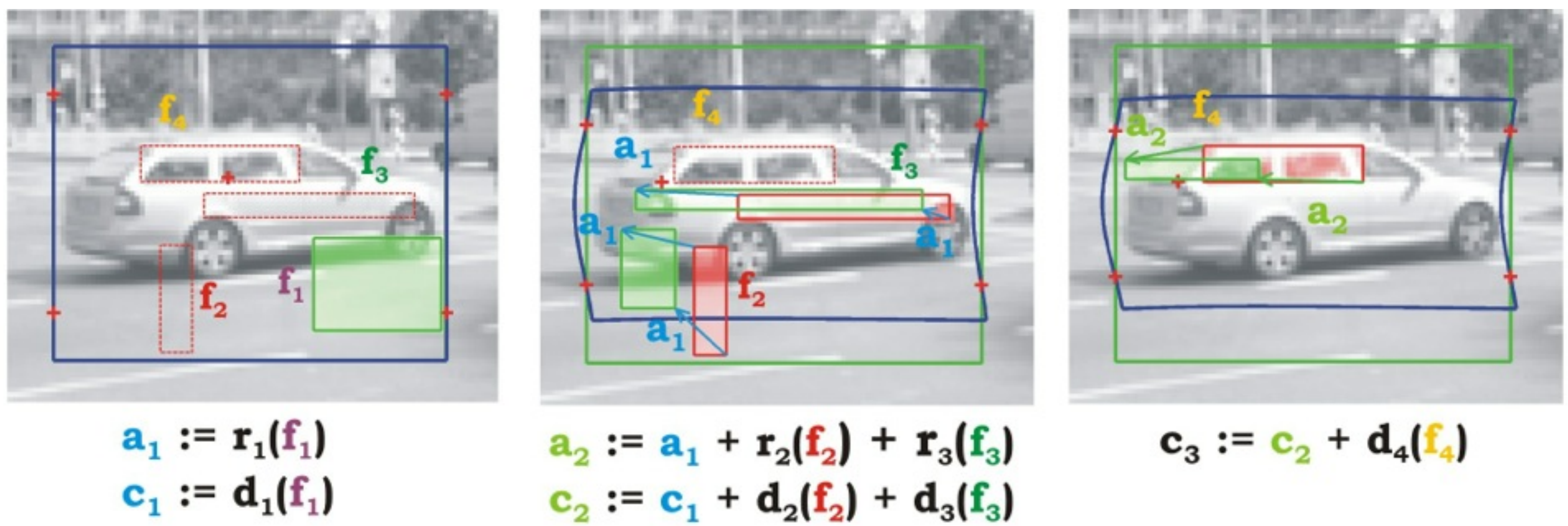
[15] P. Felzenszwalb et al., *Object detection with discriminatively trained part-based models*, TPAMI, 2010

Method	SDP+LISA	SDP+Ali [1]	Zhu [12]	DPM [15]
Running time on VGA	33ms	41ms	17.2s	10.5s


50.3
40.6
42.9
40.6

Conclusion

- exploiting features (computed once) pays off



- K. Zimmermann, D. Hurych, T. Svoboda. Non-Rigid Object Detection with Local Interleaved Sequential Alignment (LISA). *In IEEE Transactions on Pattern Analysis and Machine Intelligence.* 36(4), 2014
- K. Zimmermann, T. Svoboda, J. Matas. Anytime learning for the NoSLLiP tracker. *Image and Vision Computing.* 27(11), 2009
- K. Zimmermann, J. Matas, and T. Svoboda. Tracking by an Optimal Sequence of Linear Predictors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 31(4), 2009,