Perception for autonomous vehicles: relevant applications using geometric and learning based models

T. CHATEAU, ISPR/ Institut Pascal
UMR 6602, UCA/CNRS/SIGMA,
Clermont Ferrand, France, 2019

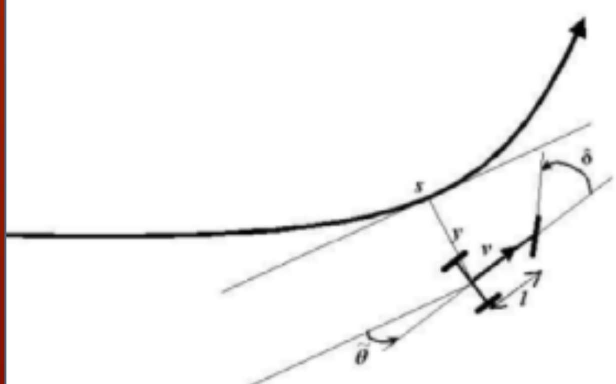
Content

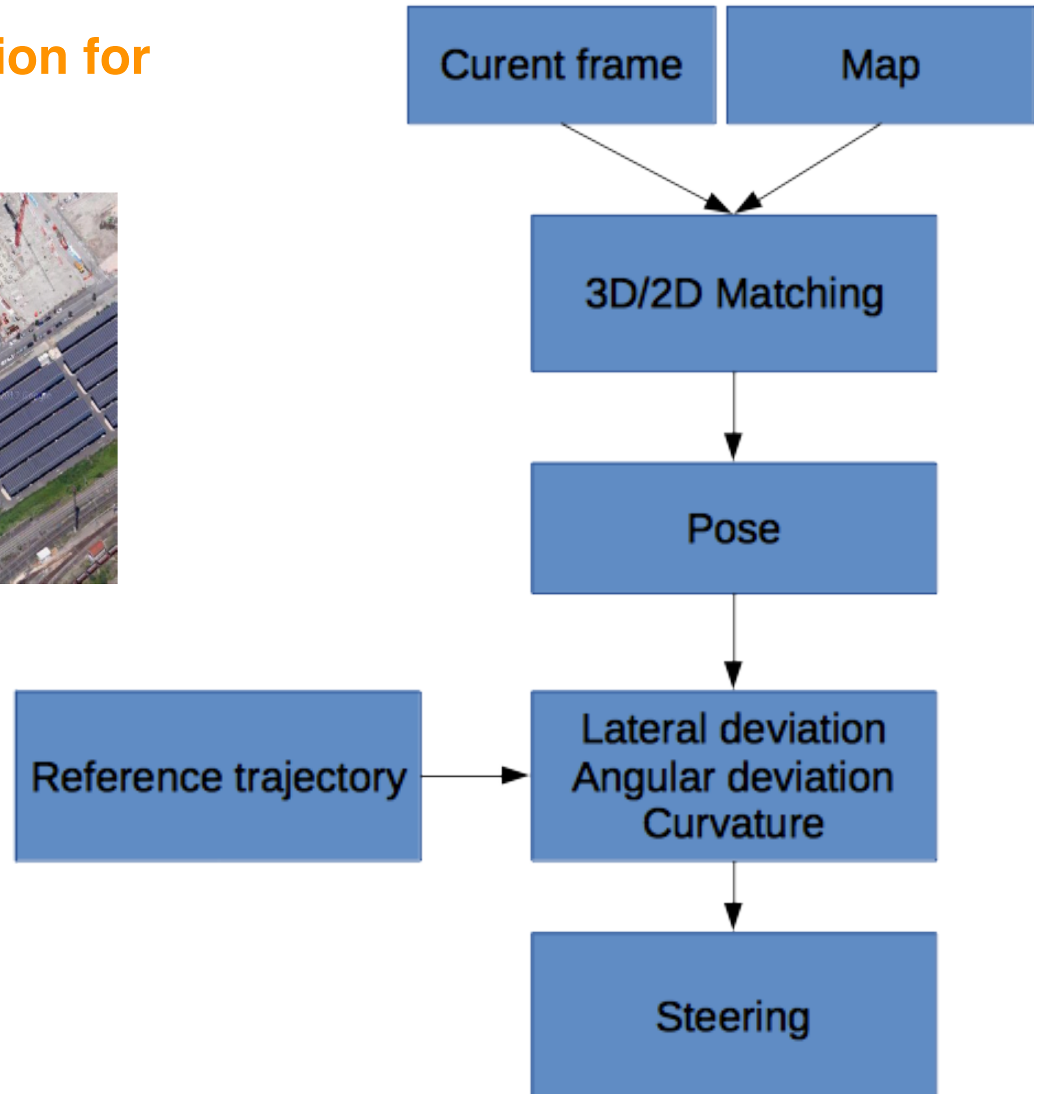
I) Monocular 3D Localisation for autonomous vehicles

II) Deep Learning: Object detection

Localization and real-time navigation

I) Monocular 3D Localisation for autonomous vehicles

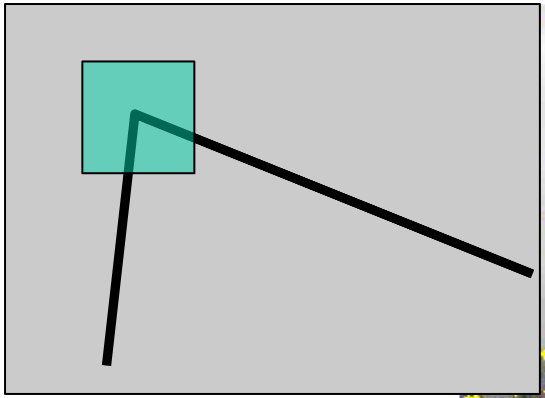


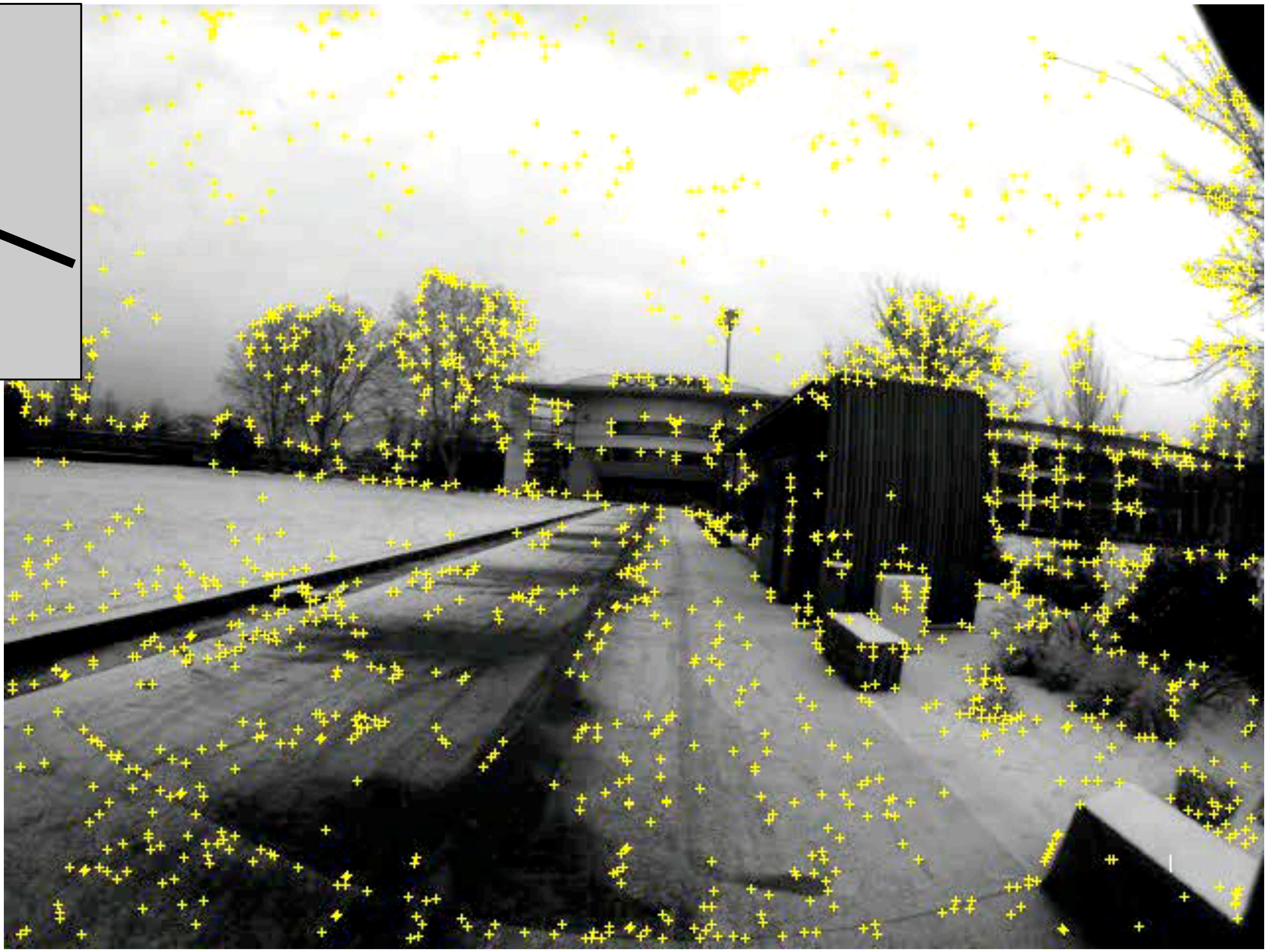


Monocular 3D Localisation for autonomous vehicles

Monocular based localisation for automatic guidance: step 1: Building a 3D map and reference path

Monocular based localisation for automatic guidance: step 1: Building a 3D map and reference path





Interest point detection

Step 2: 3D reconstruction

- Correlation ZNCC (11x11 pixels ROI)

$$zncc(\mathbf{z}_t, \mathbf{z}_{t+1})$$

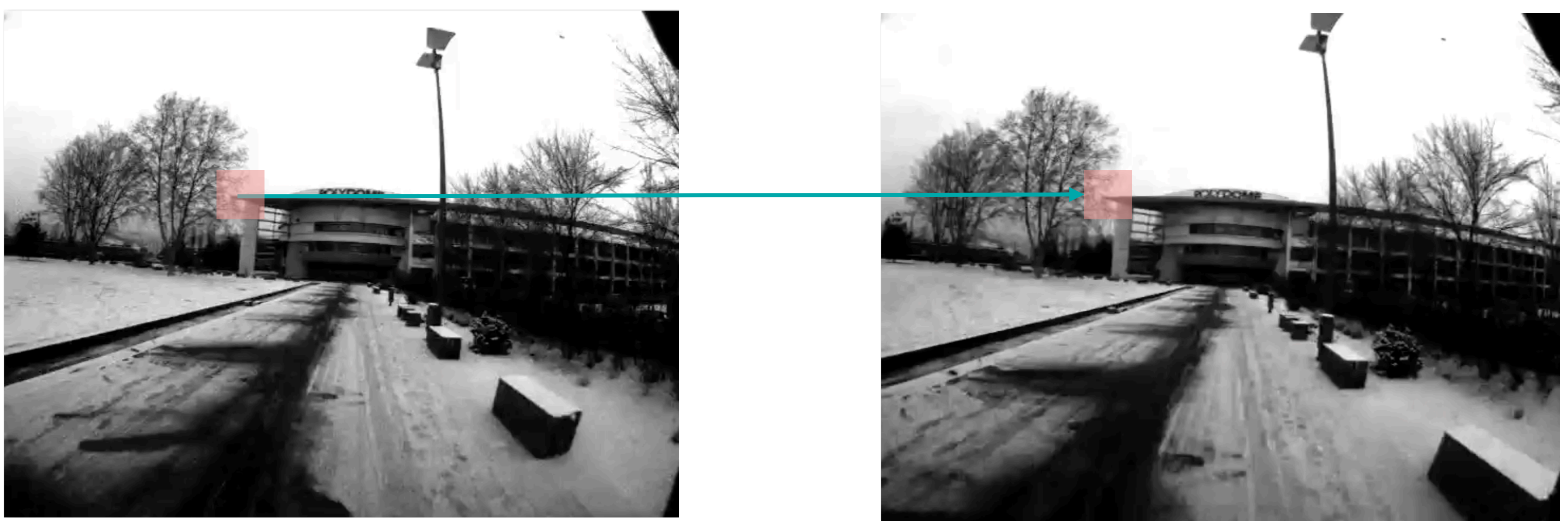
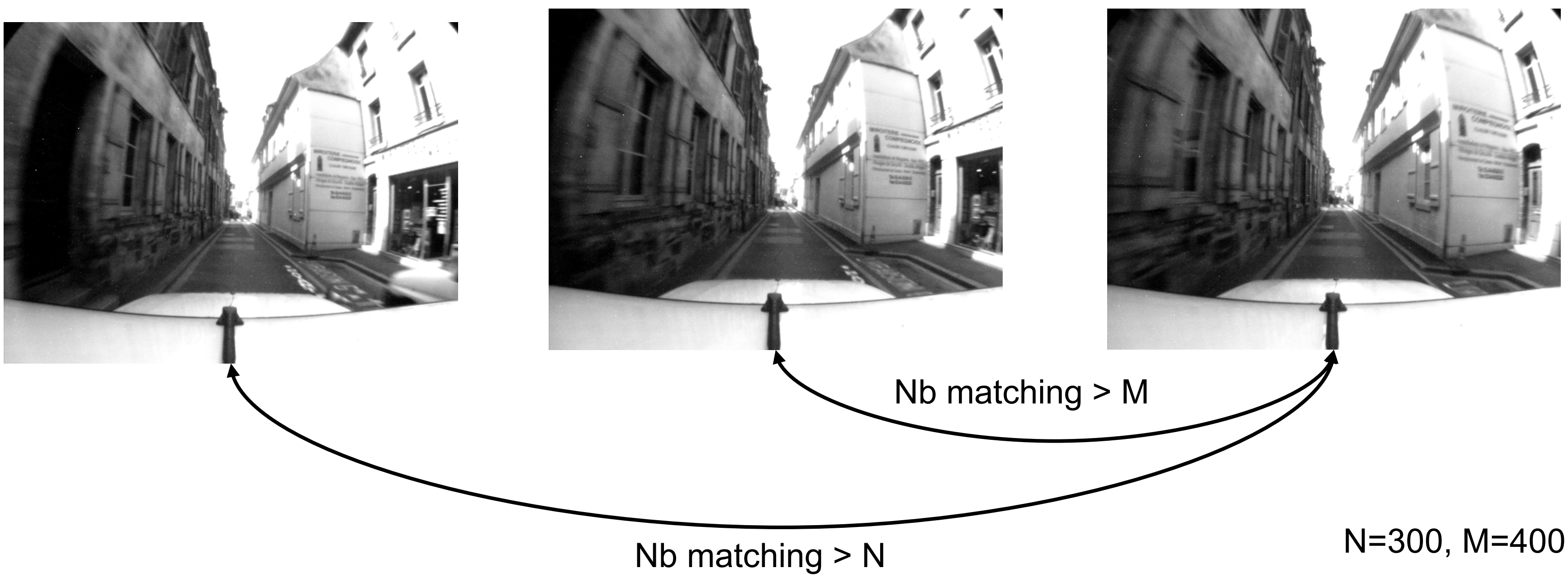


image t

image t+1

Step 2: 3D reconstruction

- Select key images:
  - far enough to produce a precise 3D reconstruction
  - close enough to keep matching points.



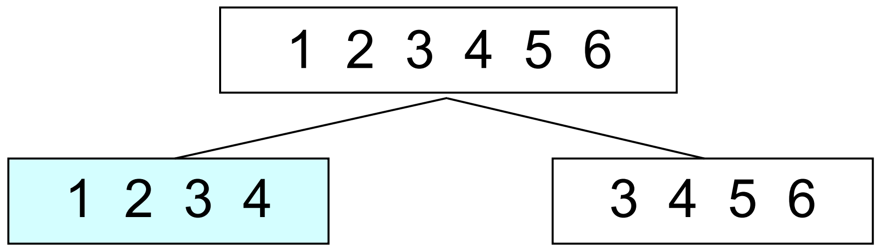
N=300, M=400

Step 2: 3D reconstruction

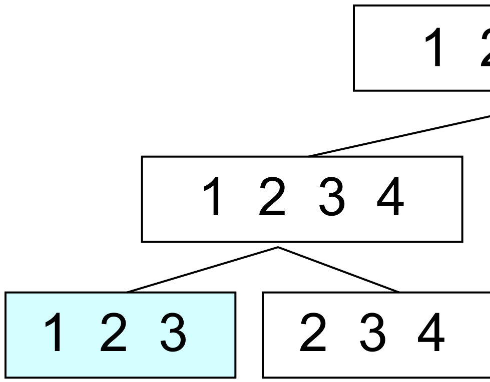
Key images

1 2 3 4 5 6

Step 2: 3D reconstruction

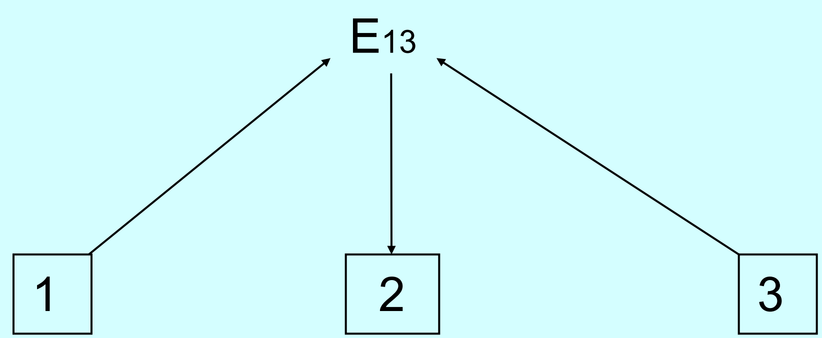


Step 2: 3D reconstruction

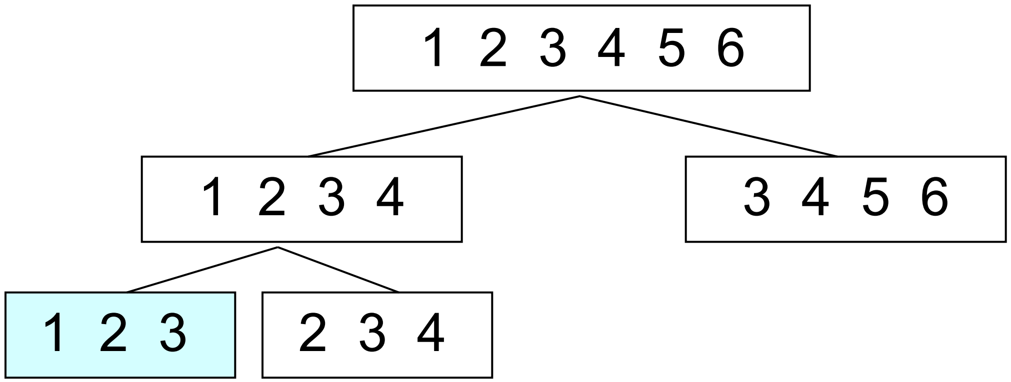


3D geometry estimation of the first 3 images

- Essential matrix (5 points algorithm)
- 3D points reconstruction
- Bundle adjustment.



Step 2: 3D reconstruction

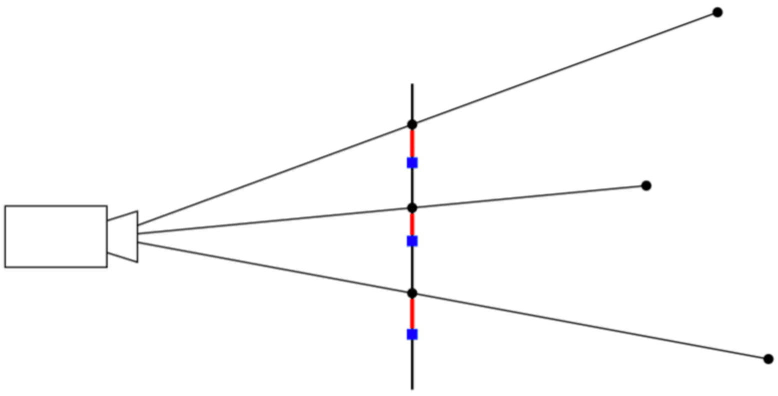


Bundle Adjustment :

estimate $(C_1,\ldots,C_n,Q_1,\ldots,Q_m)$ that minimize:

$$\sum_{i=1}^{N}\sum_{j=1}^{M}\left\| q_i^j - \pi\left(C_i, Q^j\right)\right\|^2$$

Step 2: 3D reconstruction



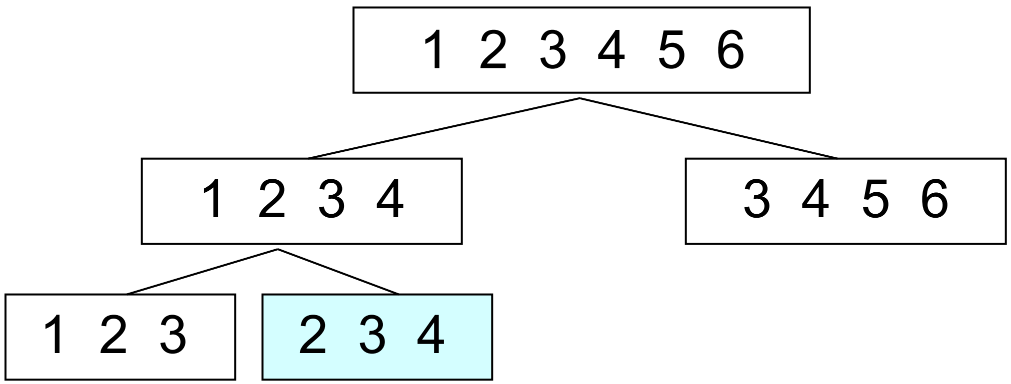
$$F(X) = Y - \epsilon$$

Bundle Adjustment :

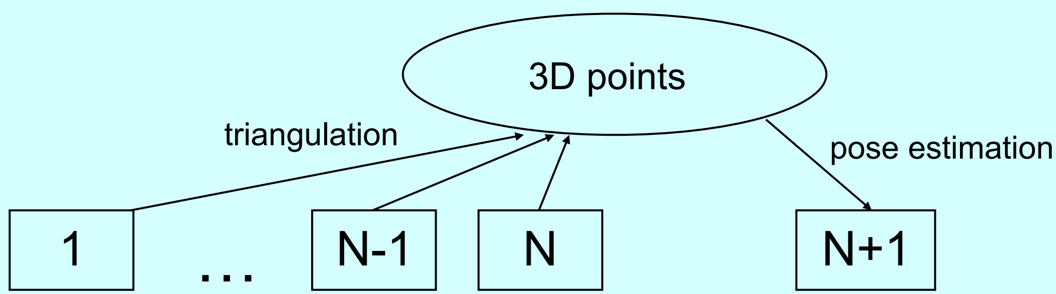
estimate $(C_1,\ldots,C_n,Q_1,\ldots,Q_m)$ that minimize:

$$\sum_{i=1}^{N}\sum_{j=1}^{M}\left\| q_i^j - \pi(C_i, Q^j)\right\|^2$$

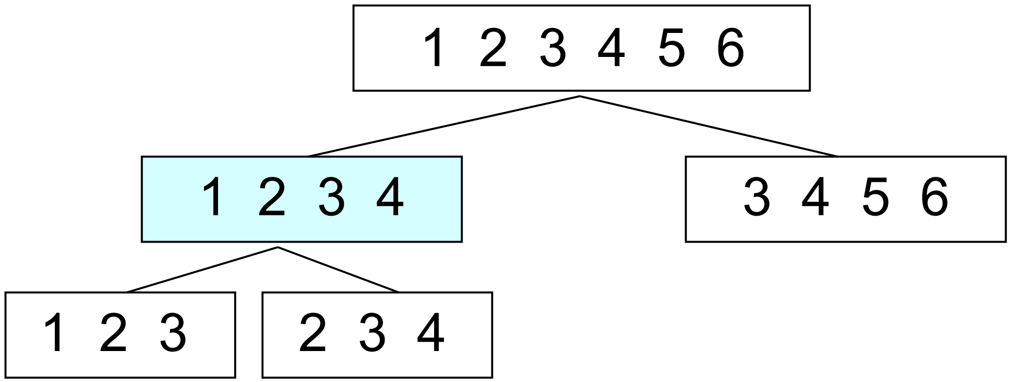
Step 2: 3D reconstruction



3D reconstruction is made for each subset of 3 key images

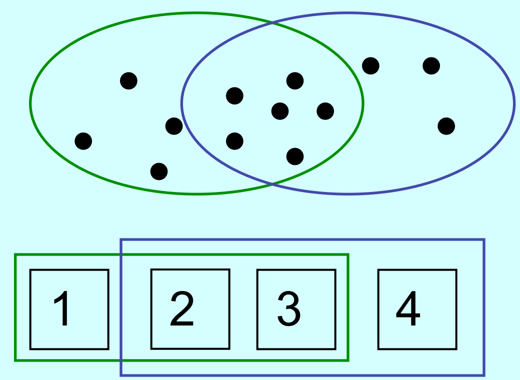


Step 2: 3D reconstruction

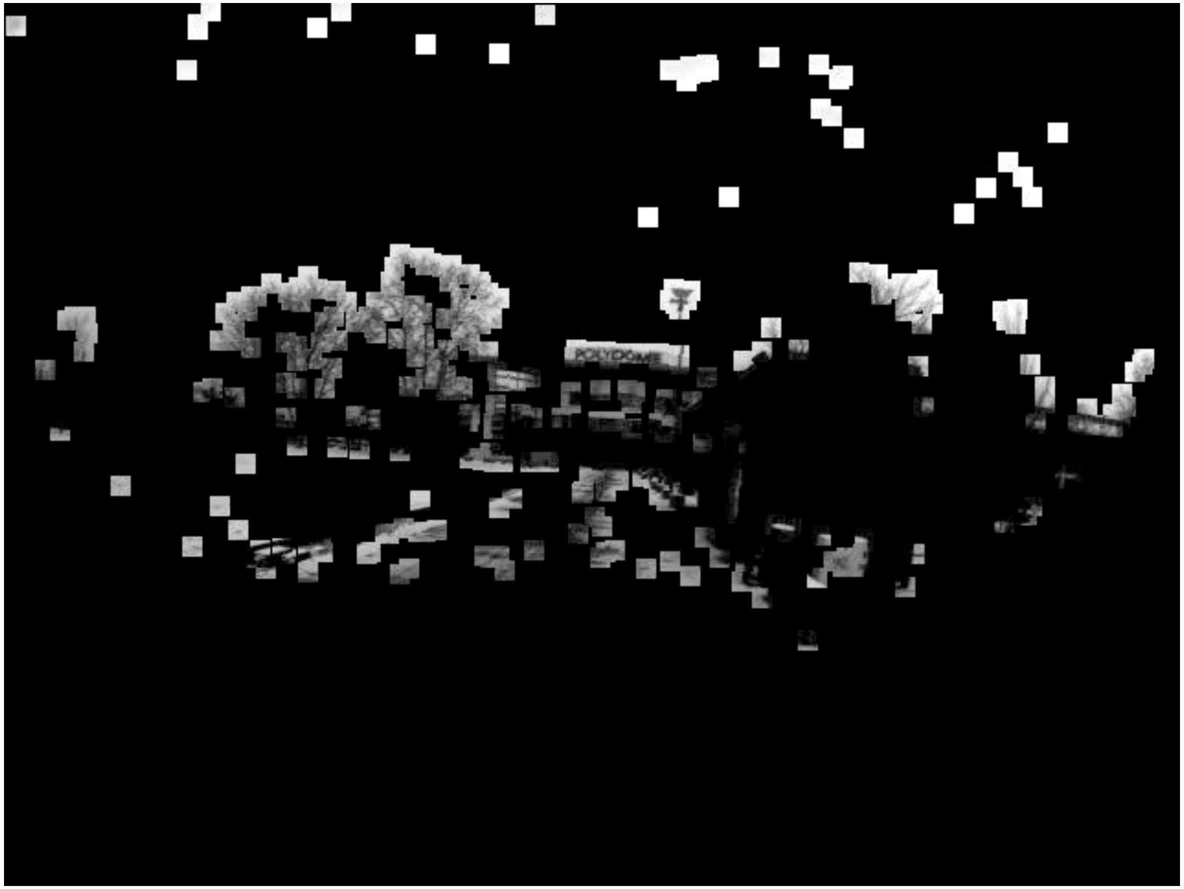


Fusion step:

- based on subsets with common images



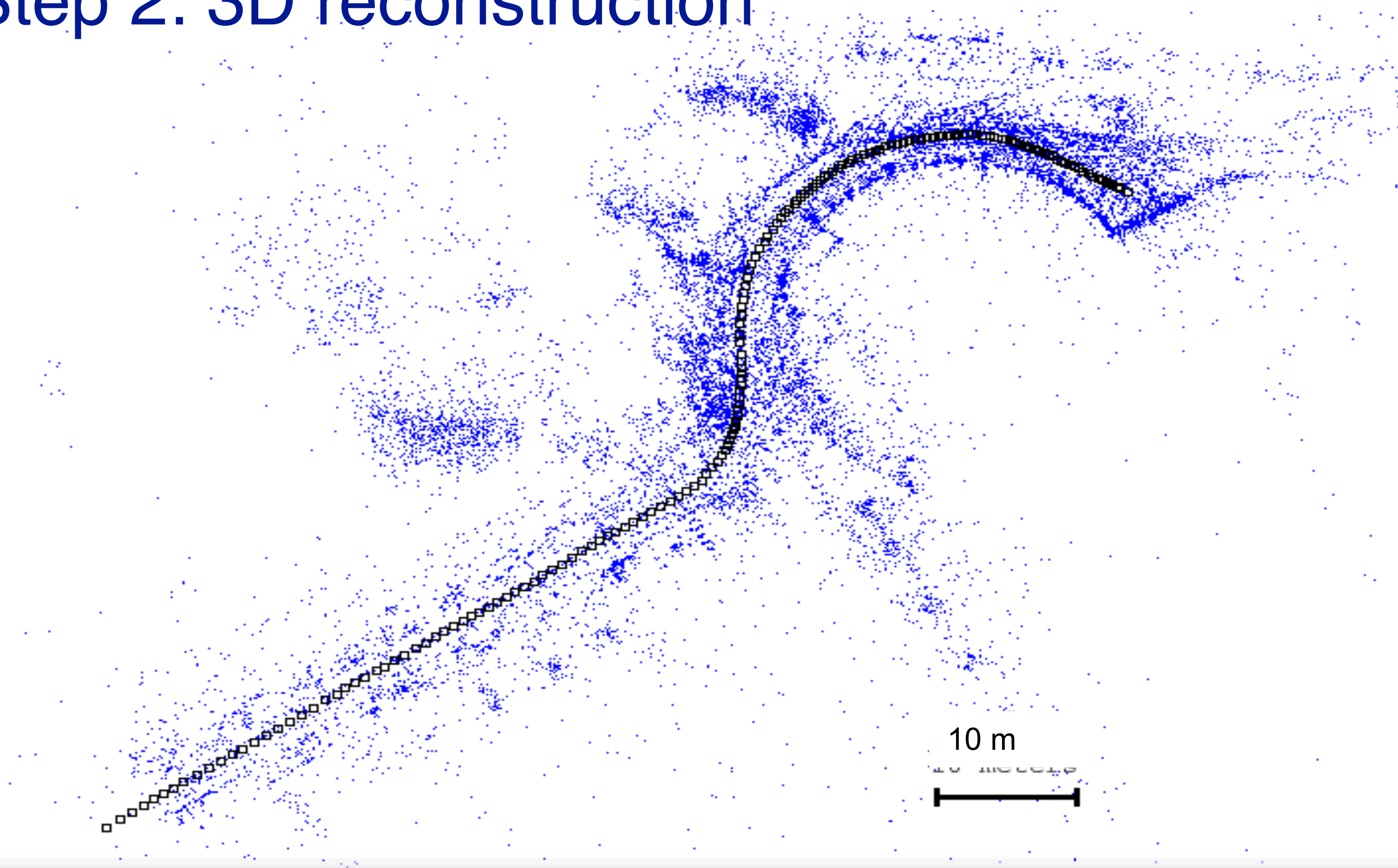
Step 2: 3D reconstruction



(125 m, 172 key images, 23000 3D points)

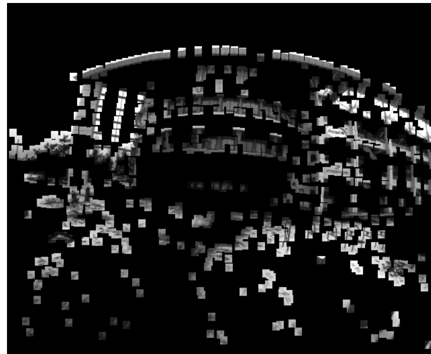
Visual Memory: 3D points

Step 2: 3D reconstruction



3D reconstructed points
(125 m, 172 key images, 23000 3D points)

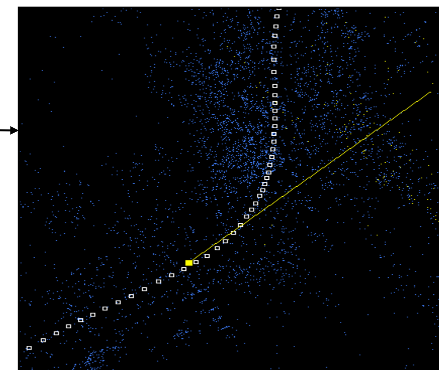
Step 3: realtime online localization



Visual memory
3D patchs

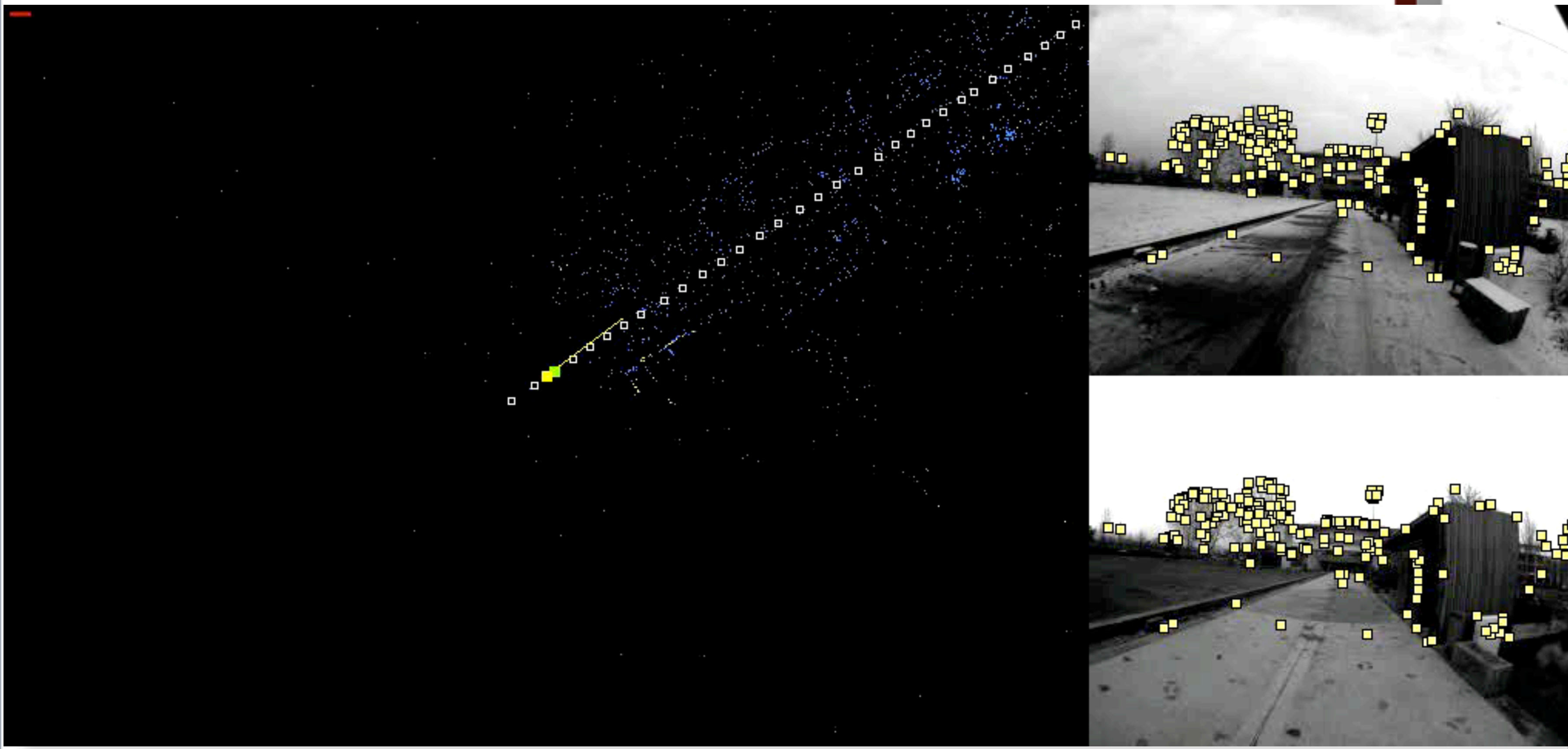
Current image

2D/3D matching
and localization



Camera
localization
6 dof

Real time Localisation and Control



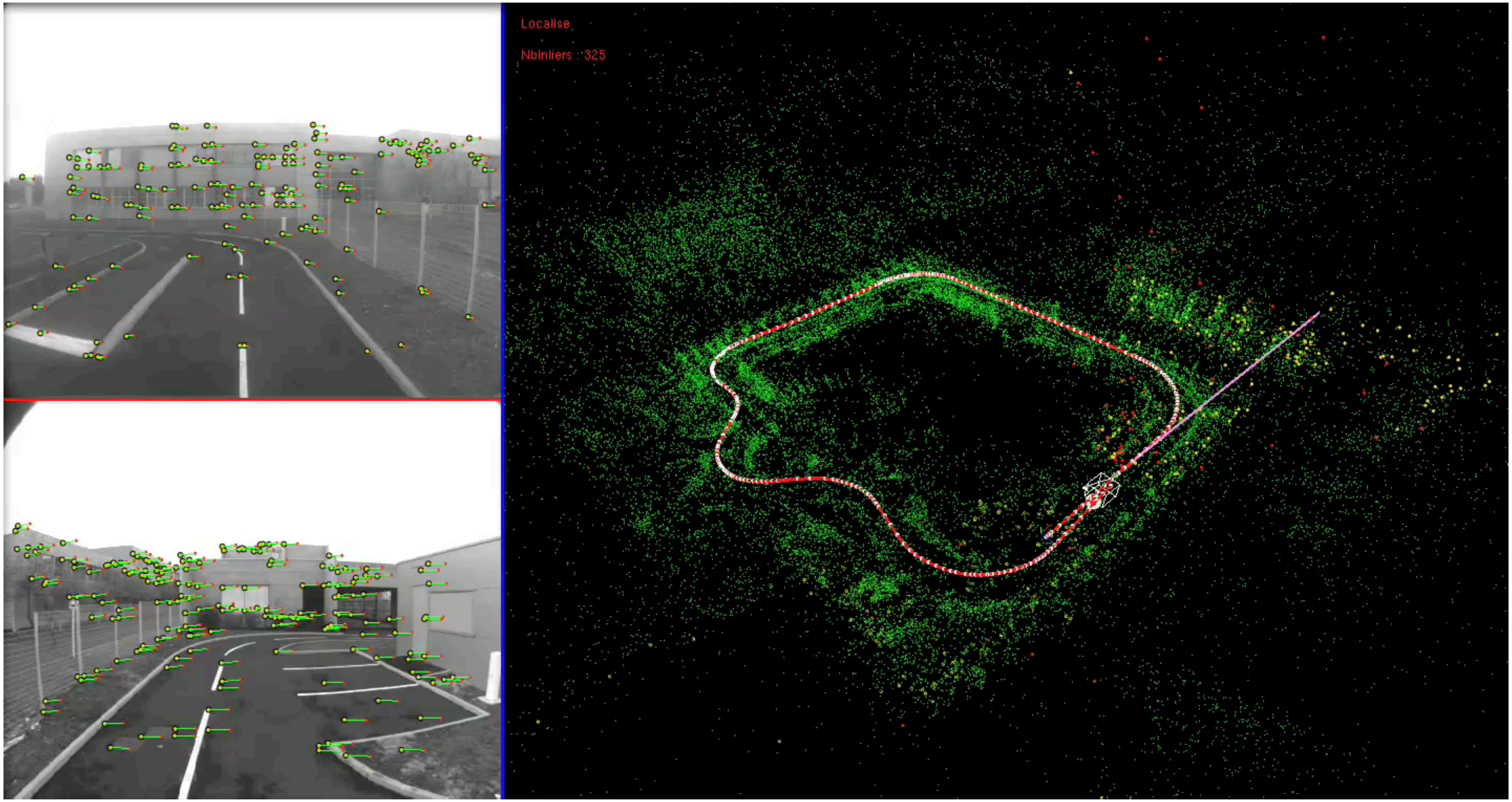
Realtime localization (15 fps) – precision: 10cm

Eric Royer, Maxime Lhuillier, Michel Dhome and Thierry Chateau, Localization in urban environments : monocular vision compared to a differential gps sensor. IEEE **CVPR2005**, Computer Vision and Pattern Recognition. San Diego, USA, June 2005

ISPR/ComSee: 3D-Localisation

Monocular based localisation for automatic guidance

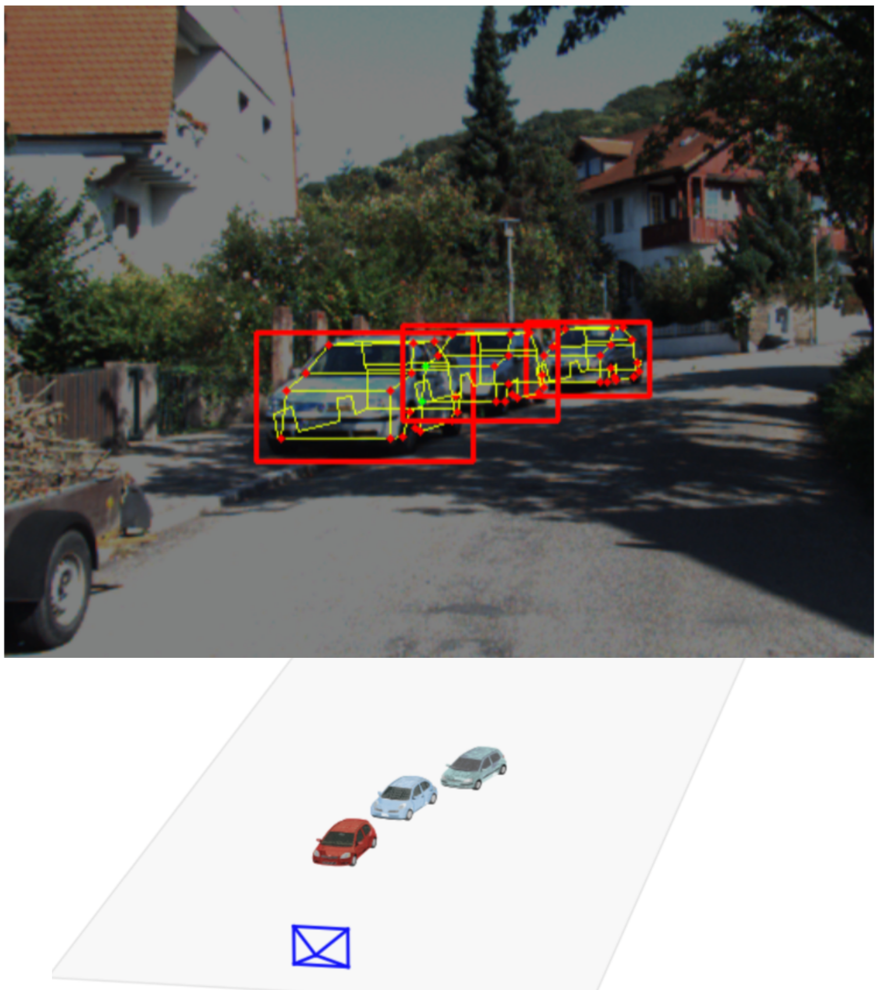
Toward the Vipa Project: using two cameras (rear-front)



INSTITUT PASCAL

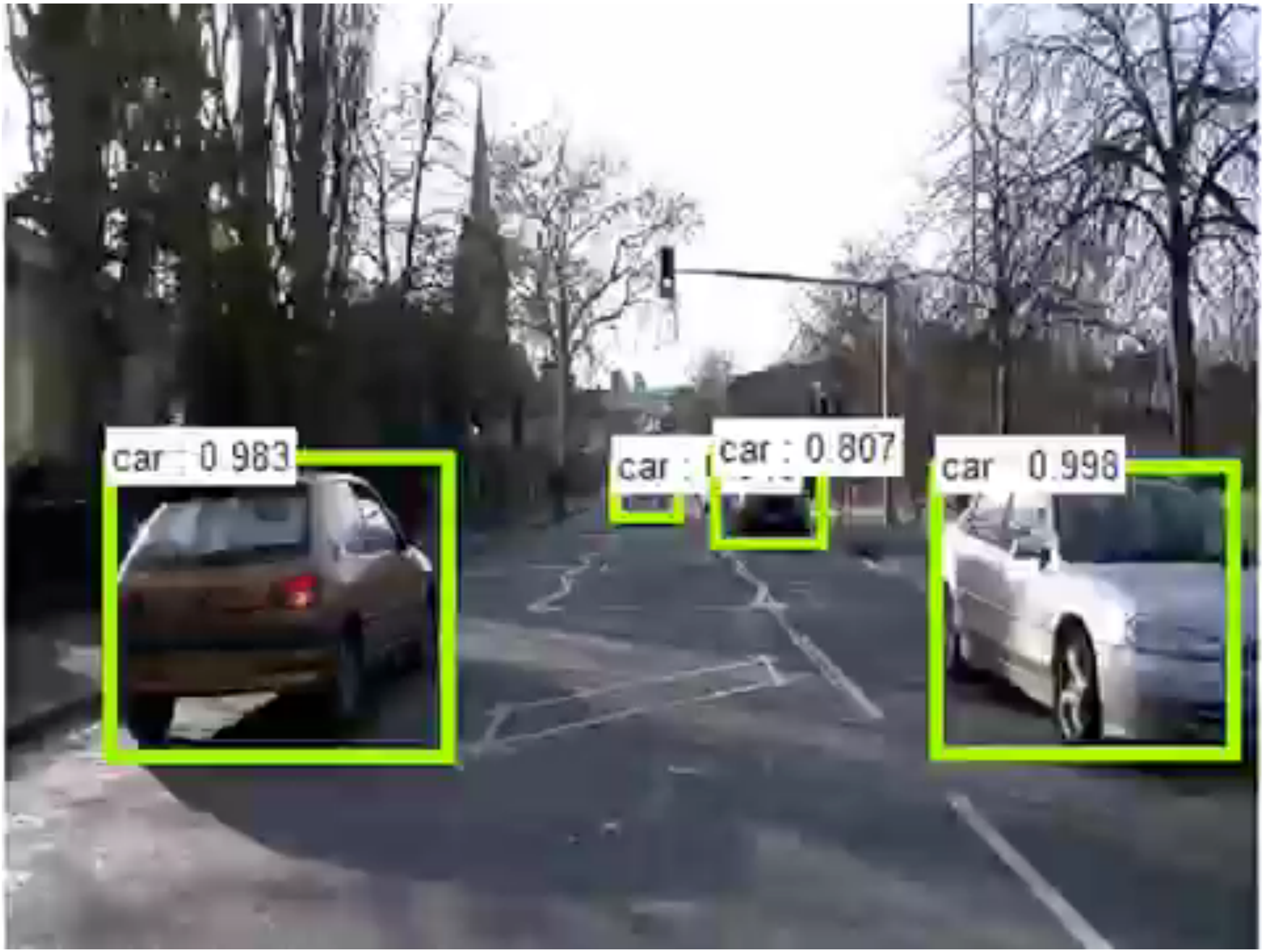
Content

II) Deep Learning: Object detection



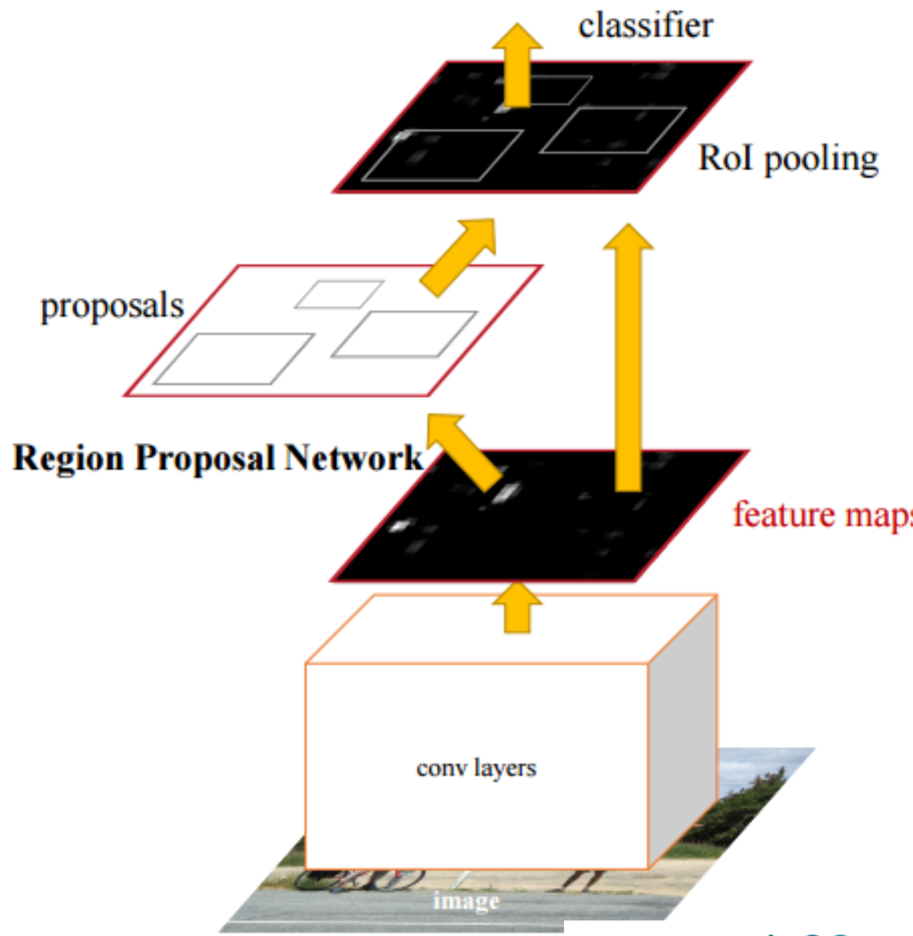
Object Localisation and Categorization (FasterRcnn)

Faster-Rcnn (realtime)



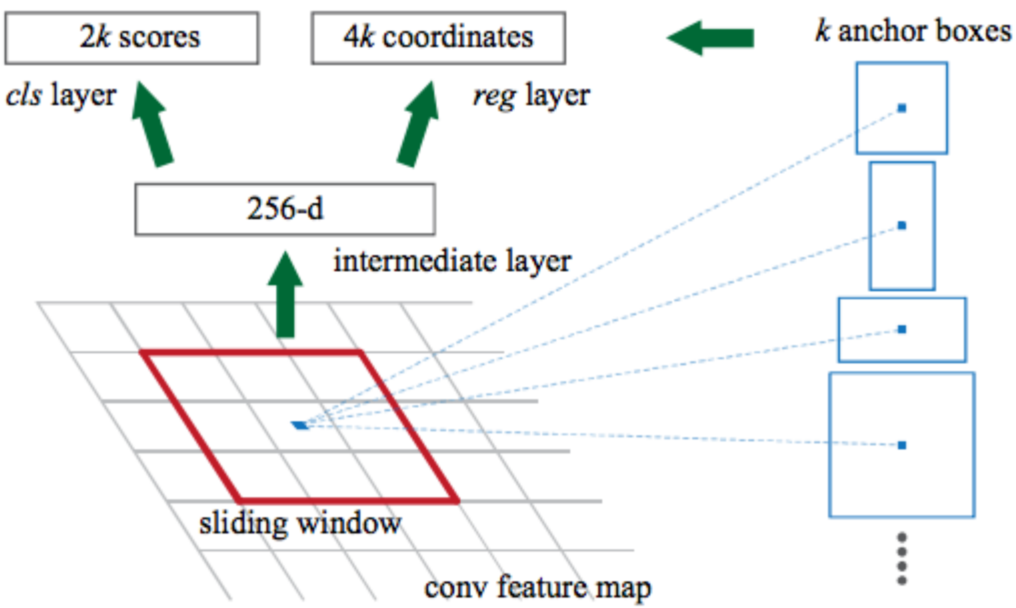
https://www.youtube.com/watch?v=WZmSMkK9VuA

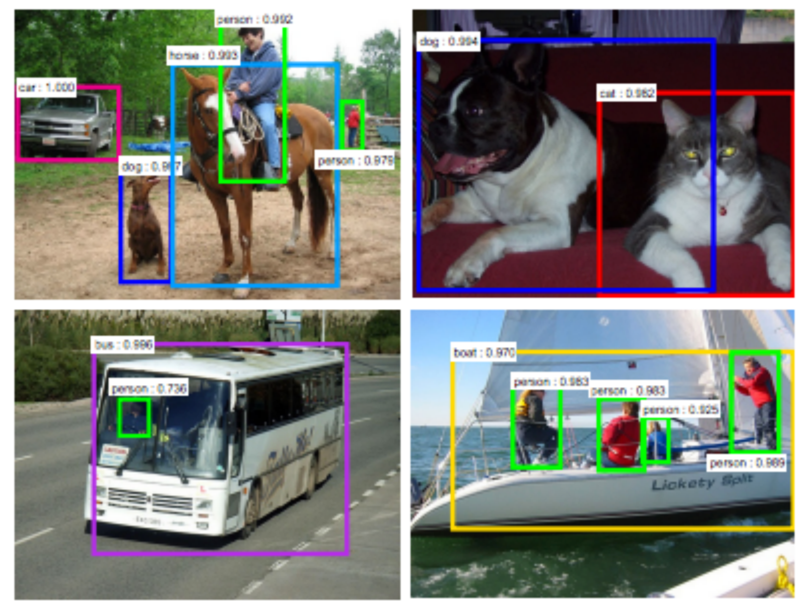
FasterRcnn (2016):
Region Proposal Network + Classification Network



Object Localisation and Categorization (FasterRcnn)

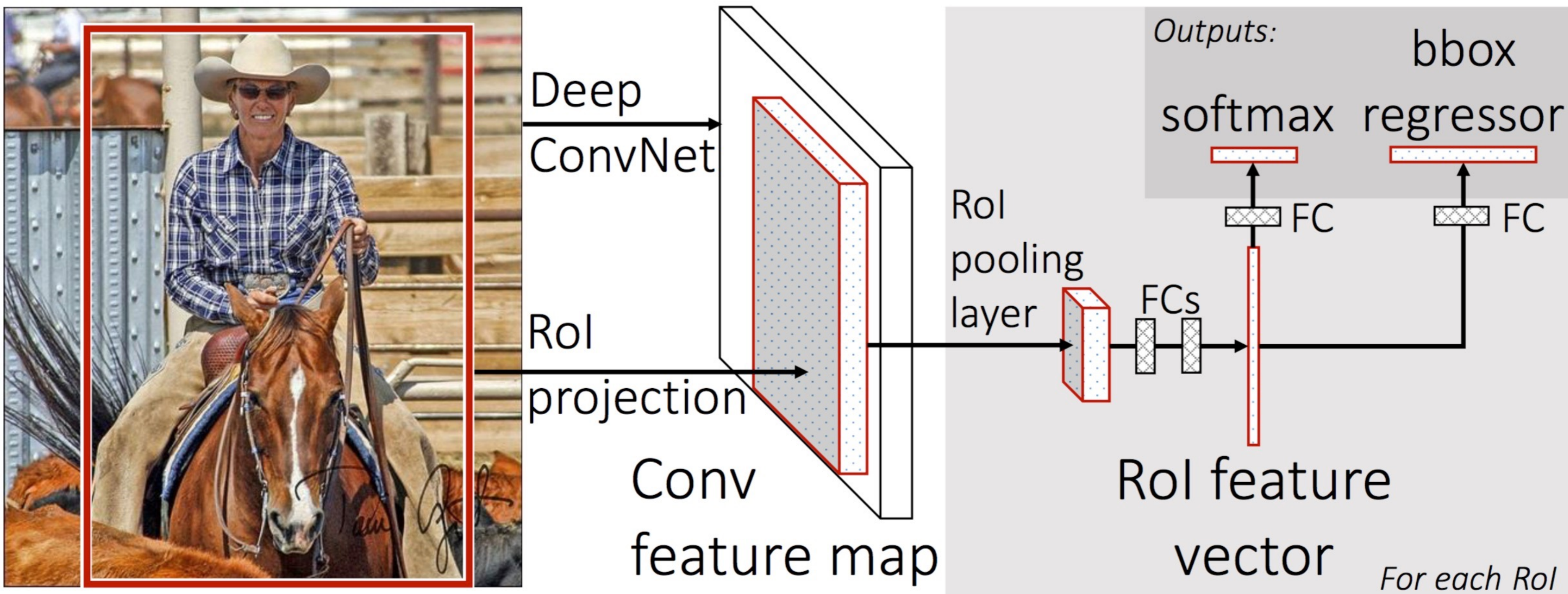
Region Proposal Network



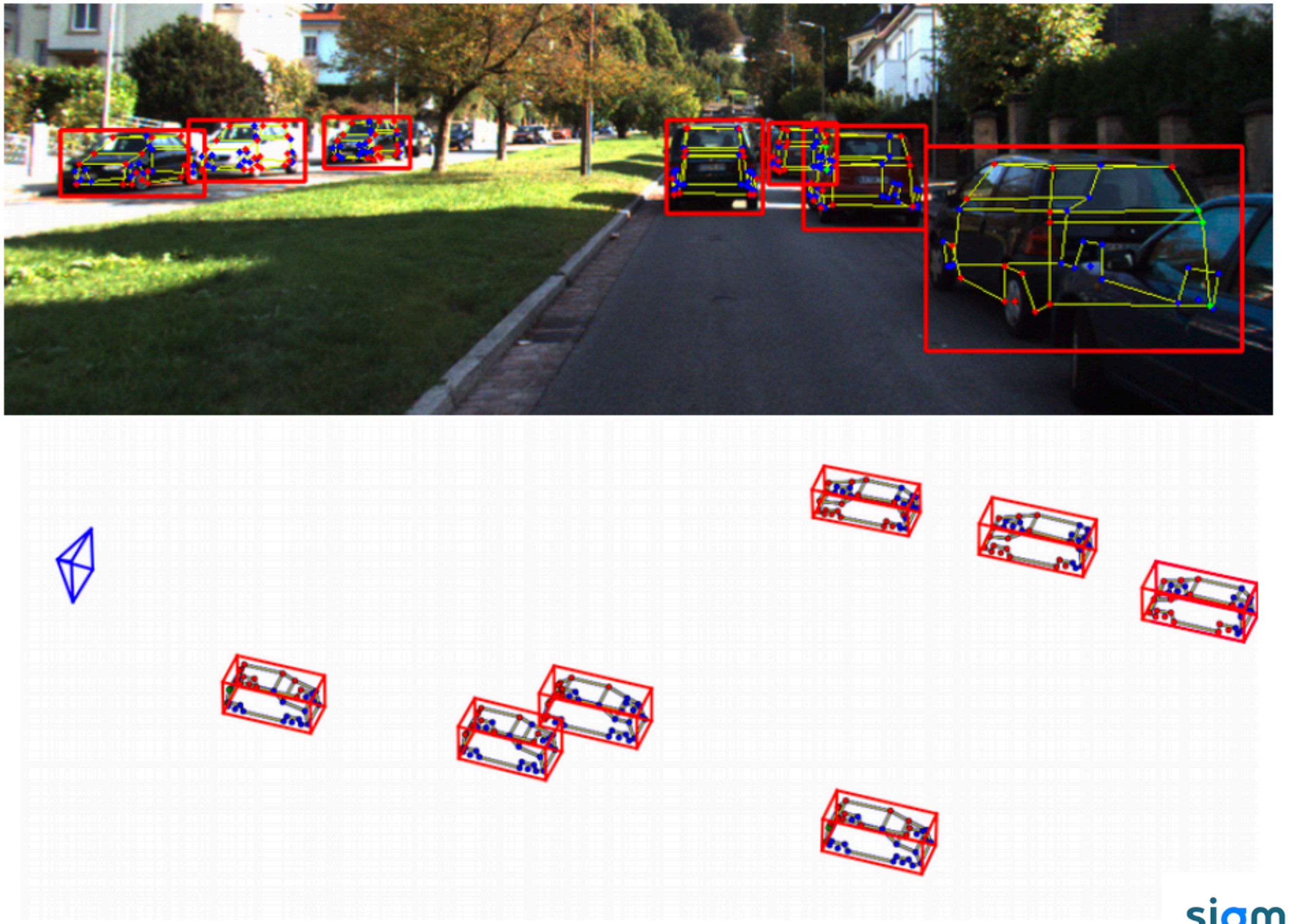


Object Localisation and Categorization (FasterRcnn)

RCNN

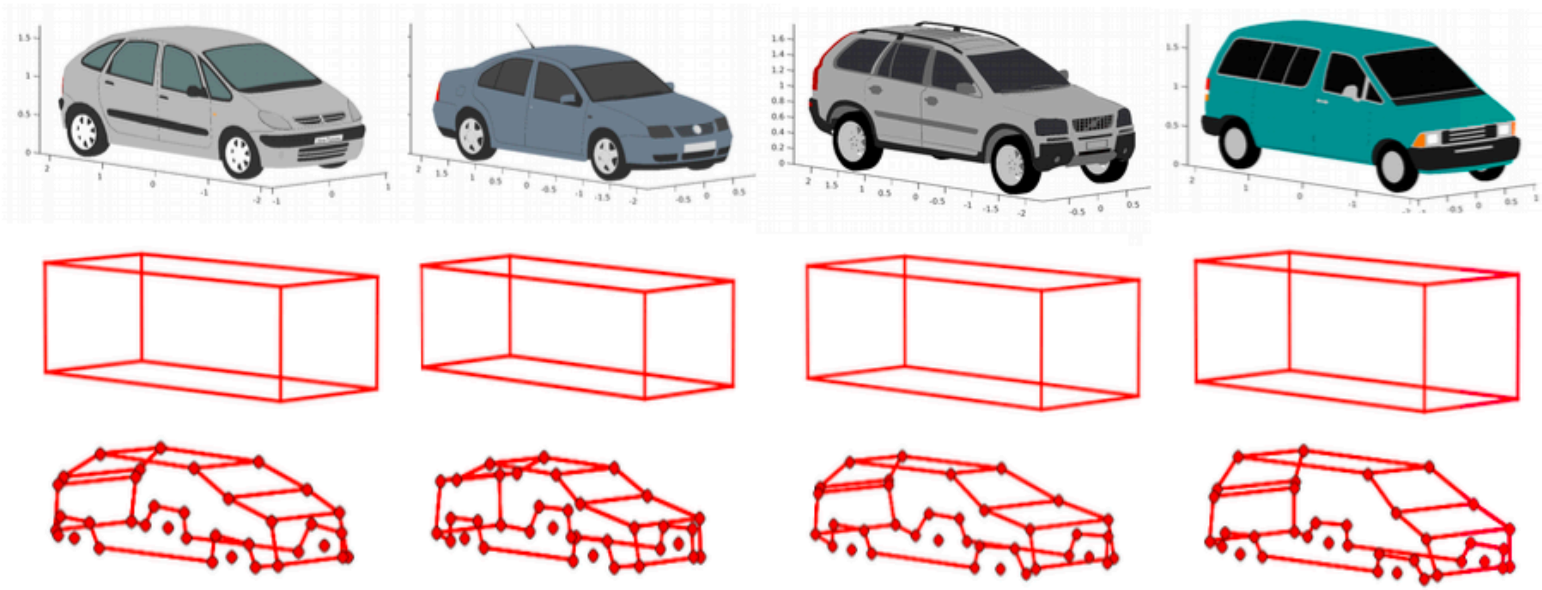


3D detection and localisation of vehicles Using a monocular camera



Deep Learning for 3D vehicle understanding from monocular images: toward many-task networks

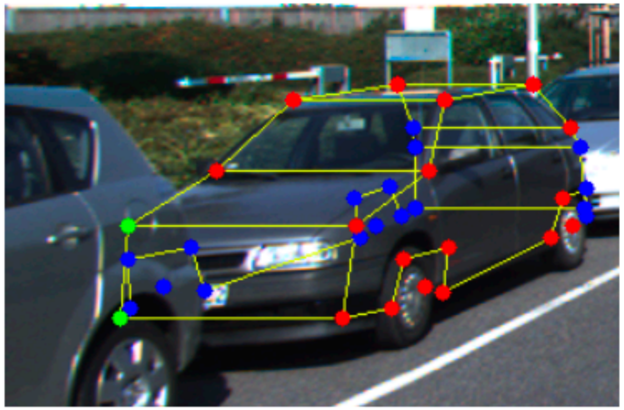
3D samples of shape and template dataset



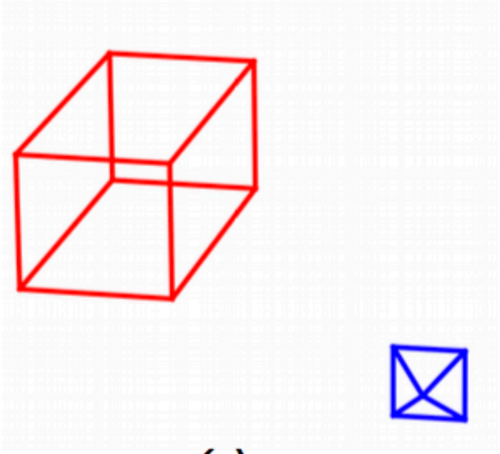
Deep Learning for 3D vehicle understanding from monocular images: toward many-task networks

Bounding box and part detection (with visibility estimation, green and blue)

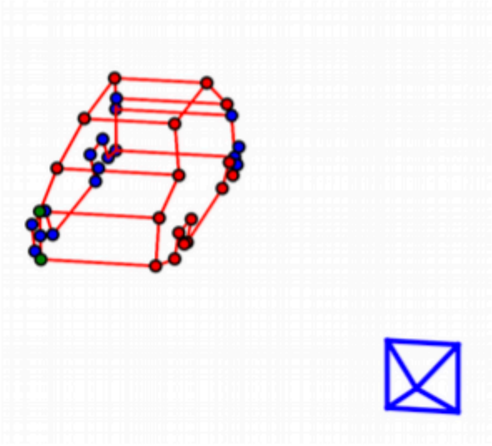
(a)



(b)

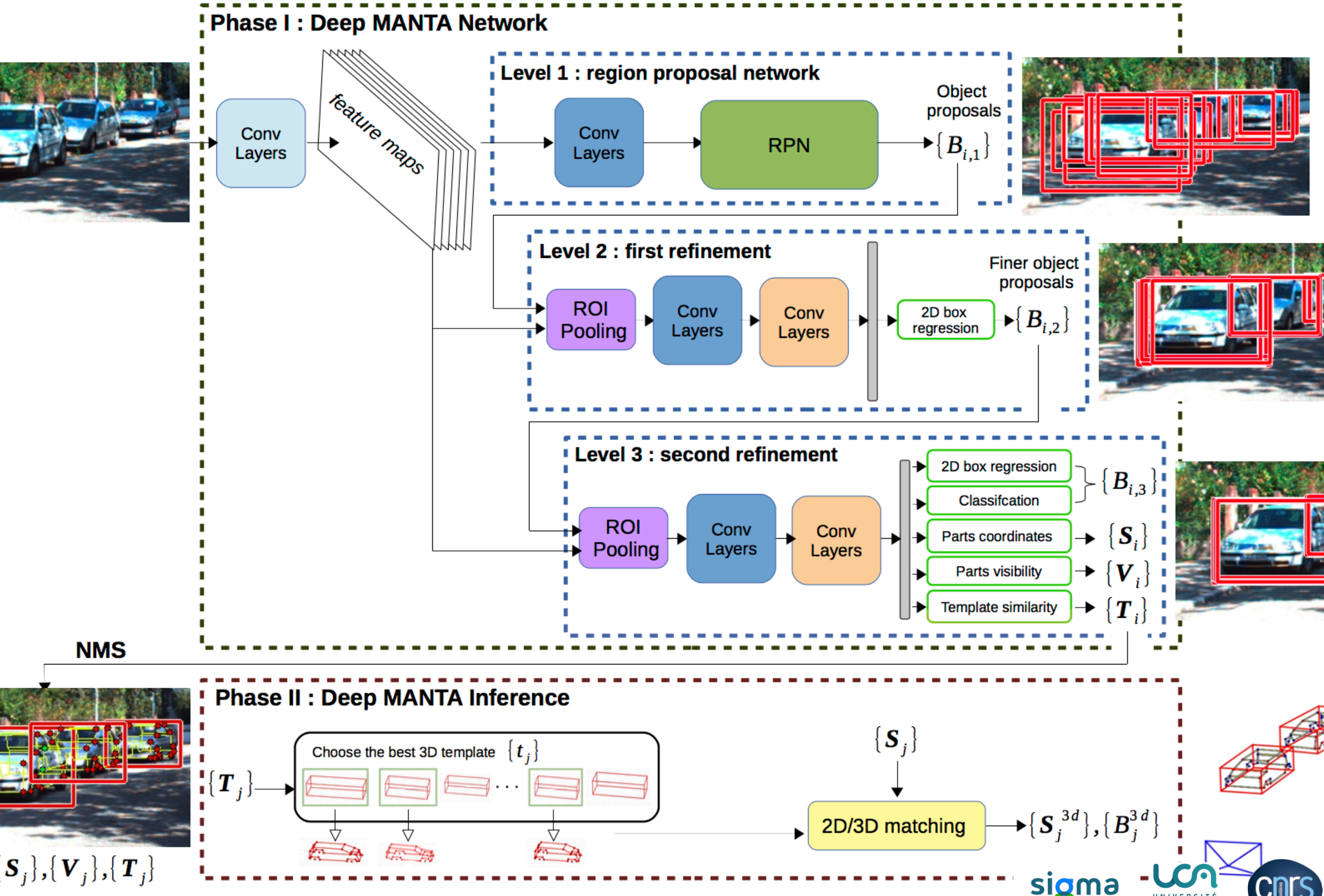


(c)



(d)

A coarse to fine strategy



Loss functions

RPN Loss

Detection loss

Parts Loss

Visibility Loss

Template similarity loss

$$\mathcal{L} = \mathcal{L}^1 + \mathcal{L}^2 + \mathcal{L}^3$$

with

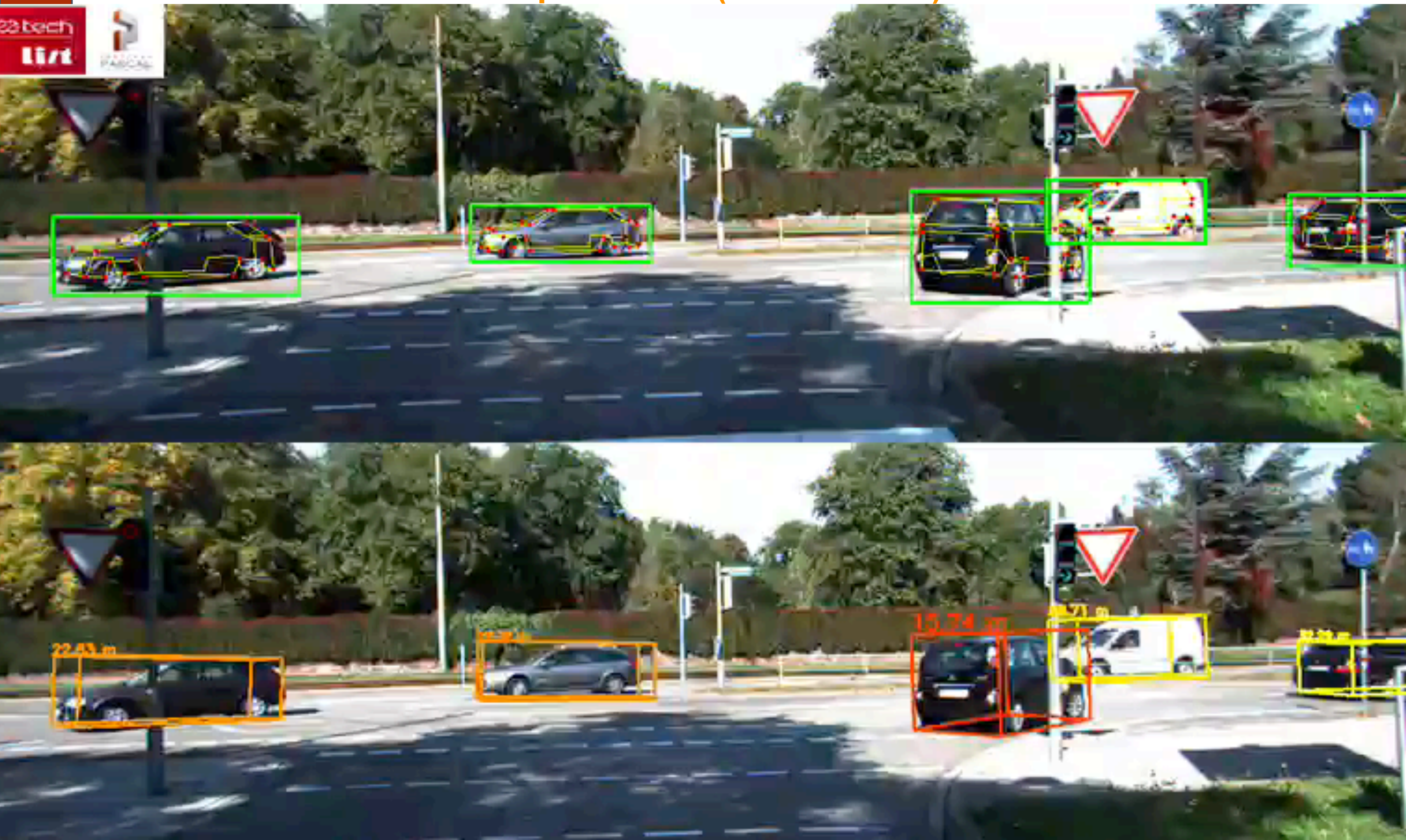
$$\mathcal{L}^1 = \mathcal{L}_{rpn},$$

$$\mathcal{L}^2 = \sum_i \mathcal{L}^2_{det}(i) + \mathcal{L}^2_{parts}(i),$$

$$\mathcal{L}^3 = \sum_i \mathcal{L}^3_{det}(i) + \mathcal{L}^3_{parts}(i) + \mathcal{L}_{vis}(i) + \mathcal{L}_{temp}(i),$$

Deep Learning for 3D vehicle understanding from monocular images

Experiments (Kitti Dataset)



Experiments (Kitti Dataset)

Detection and orientation (**ranked 1st during 20 months on orientation**)

2017

Method	Type	Time	val1 AP Easy	val1 AP Moderate	val1 AP Hard	val1 AOS Easy	val1 AOS Moderate	val1 AOS Hard
3DVP [31]	Mono	40 s	80.48	68.05	57.20	78.99	65.73	54.67
Faster-RCNN [27]	Mono	2 s	82.91	77.83	66.25	-	-	-
SubCNN [32]	Mono	2 s	95.77	86.64	74.07	94.55	85.03	72.2
Ours nms = 0.4	Mono	0.7 s	97.05	88.94	78.25	96.90	88.68	77.83
Ours nms = 0.5	Mono	0.7 s	96.98	89.58	79.77	96.83	89.31	79.31
Ours w vis	Mono	0.7 s	**97.90**	**91.01**	**83.14**	**97.60**	**90.66**	**82.66**

Today

	Method	Setting	Code	Moderate	Easy	Hard	Runtime	Environment
1	MVRA + I-FRCNN+			**89.93 %**	90.60 %	79.78 %	0.18 s	GPU @ 2.5 Ghz (Python)
2	Deep MANTA			89.86 %	**97.19 %**	80.39 %	0.7 s	GPU @ 2.5 Ghz (Python + C/C++)

F. Chabot, M. Chaouch, J. Rabarisoa, C. Teulière and T. Chateau: Deep MANTA: A Coarse-to-fine Many-Task Network for joint 2D and 3D vehicle analysis from monocular image. CVPR 2017.

The KITTI Vision Benchmark Suite

A project of Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago

AP: mean average precision

AOS: average orientation similarity

Florian Chabot, Mohamed Chaouch, Jaonary Rabarisoa, Céline Teulière, Thierry Chateau. *Deep MANTA: A Coarse-to-Fine Many-Task Network for Joint 2D and 3D Vehicle Analysis from Monocular Image. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, Honolulu, United States. ⟨hal-01653519⟩