Overview of our researches on Machine-Learning and DataMining for Self-driving cars and Intelligent Transport Systems

Fabien Moutarde
Centre de Robotique (CAOR)
Mines ParisTech
PSL Research University

Fabien.Moutarde@mines-paristech.fr

http://people.mines-paristech.fr/fabien.moutarde

Automated driving (or smart functions) require *semantic* interpretation of car environment:

- **Locally around vehicle:**
  - automated detection and understanding of road signaling
  - categorization of objects around (cars, pedestrians, etc…)
  - forecasting of moving « objects » trajectories/behaviors
- **On broader space-time horizon:**
  - precise ego-localization (cf. GPS uncertainty/outage)
  - predict traffic evolution on large area for optimal route choice/adaptation
- **Inside the car:**
  - driver identification (for automatic switch of settings)
  - Recognize activity/gestures of driver, for monitoring his attention and/or for gestual commands

Past and current research @CAOR/Mines_ParisTech

- Locally around vehicle:
  - Detect & recognize Traffic Signs, traffic lights, etc…
  - Localize objects of important categories (cars, pedestrians, motorbikes, bicycles, etc…)
- On broader space-time horizon:
  - Visual precise ego-localization
  - Predict traffic evolution on large area for optimal route choice/adaptation
- Inside the car:
  - Recognize gestures of driver for gestual commands

Outline

- <u>**Visual detection & recognition of traffic signs and object categories (cars, pedestrians, etc…)**</u>
- **In-car Human Gestures & Activities recognition**
- **Road traffic mining and forecasting**
- **Onroad precise visual car ego-localization**

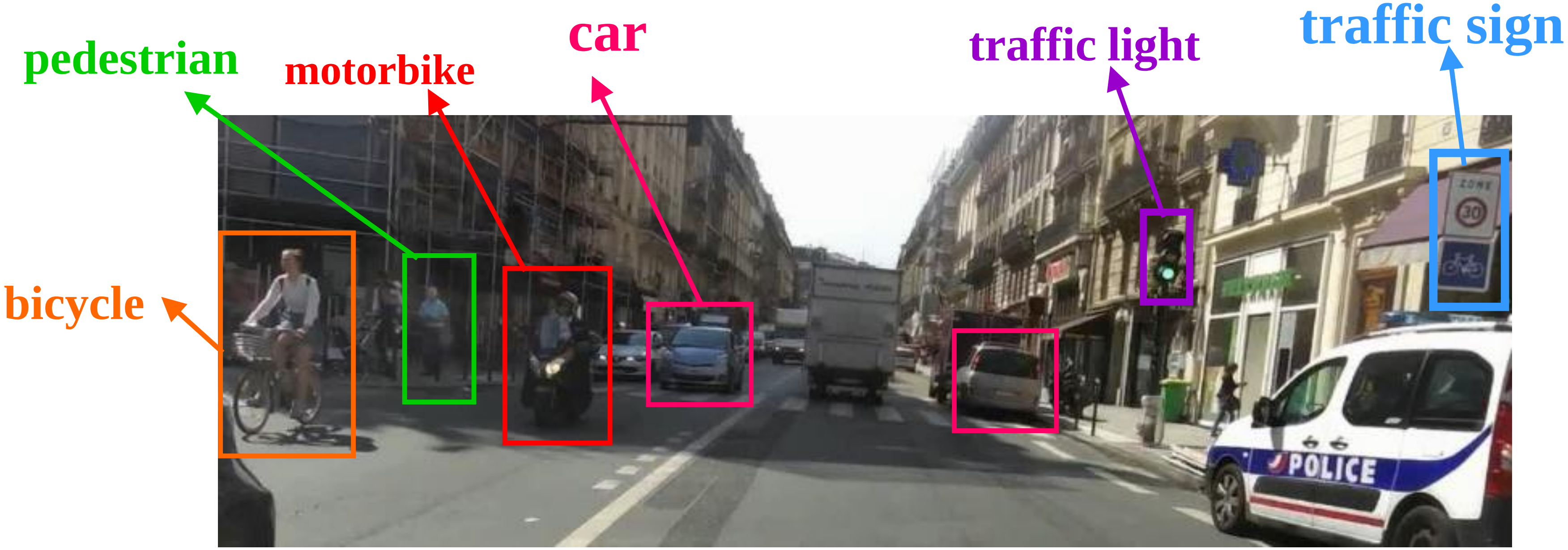
Traffic Sign Recognition (TSR)

- **Very little intrinsic variation of object**
  → main recognition challenge = *robustness to illumination & contrast changes + small 3D rotations*
- **Large number of classes (~100)**
- **Input feature for classification ?**
  - **Vector of pixel values**
  - **HoG (Histogram of Orientations of Gradients)**
  - **…**
- **ML algo used @CAOR:**

  Histogram of Oriented Gradient (HoG) features
  + Support Vector Machine (SVM) for detection
  + Random Decision Forest (RDF) for recognition

TSR démo

Real-time scene understanding for ADAS and self-driving cars



- **Key componant for Advanced Driving Assistance Systems (ADAS) and self-driving cars**

- **<u>Very large intra-class variability</u>: person or car model, shape, colors → challenge = find sth common to all instances AND discriminant v.s. other categories**
- **Strong real-time constraint: process at least ~10 frames/second**

Machine-learning for visual recognition of object categories

Classifier training, e.g. using « boosting », applied to image examples extracted from videos

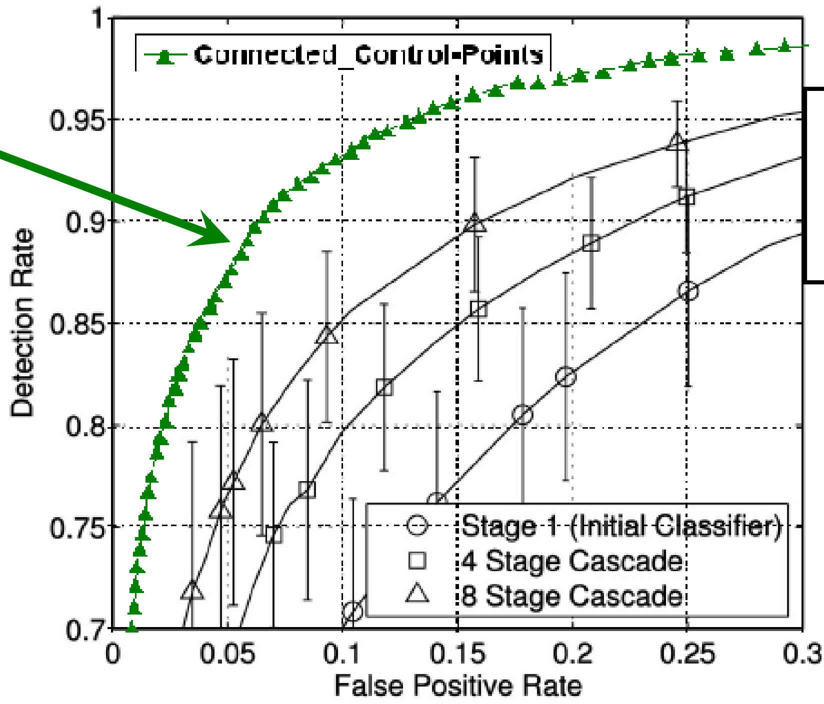
(boosting principle: assembling and weighting many elementary « weak » classifiers into one « strong « classifier)

$$H(x) = \mathrm{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

Excellent results for pedestrians recognition with adaBoost + our features

Public pedestrian examples database collected by Daimler, with 4800 positive image examples and 5000 negative (all of size 18x36 pixels)

Performance of CAOR's classifier (with connected Control-Points features)



« standard » Haar-based features (openCV)

Object category detection démo

<u>Cars (backviewed)</u>: ~ 95% detection with less than 1 false alarm per frame

<u>Pedestrians (daytime)</u>: ~80% detection with less than 2 false alarms per frame

Outline

- **Visual detection & recognition of traffic signs and object categories (cars, pedestrians, etc…)**
- **<u>In-car Human Gestures & Activities recognition</u>**
- **Road traffic mining and forecasting**
- **Onroad precise visual car ego-localization**

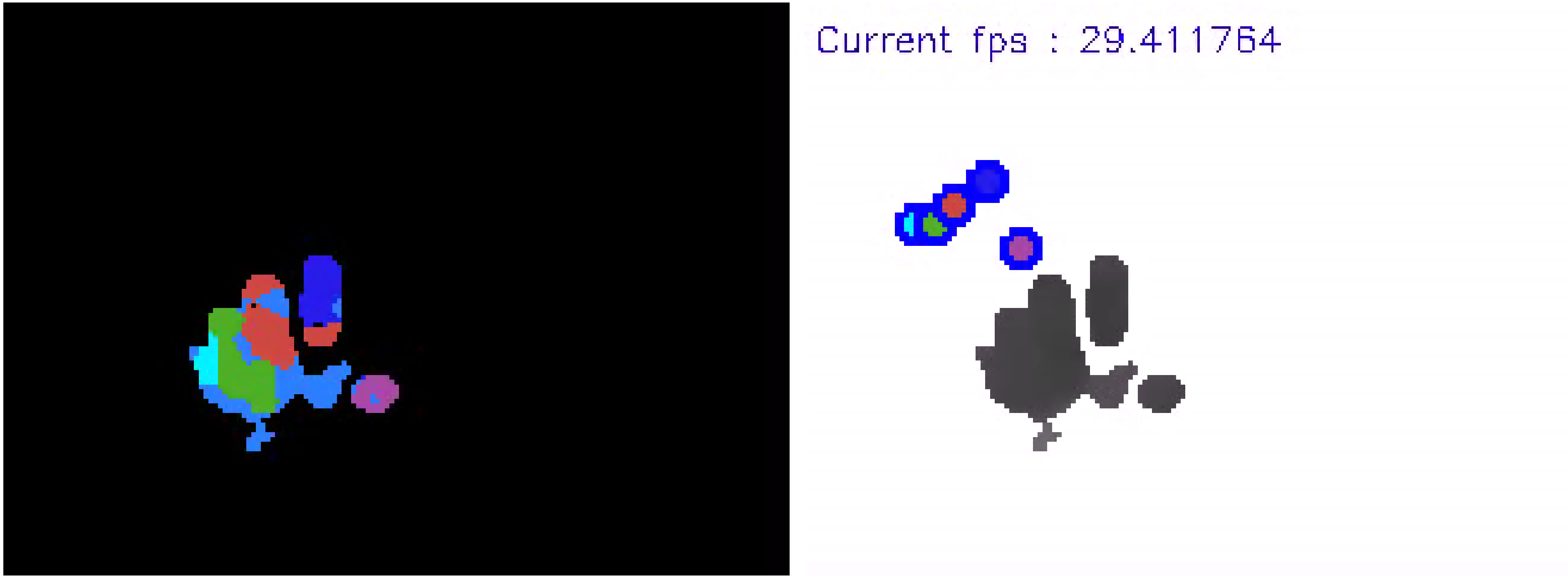
Gestures recognition *inside* car

Goal = *touchless HMI (for infotainment, etc…)*
that avoids perturbing attention/driving
(e.g. fingers gestures while holding the driving wheel)

Micro-gestures (fingers) areas

Macro-gestures area (swipes, etc…)

ML for Gestures recognition *inside* car



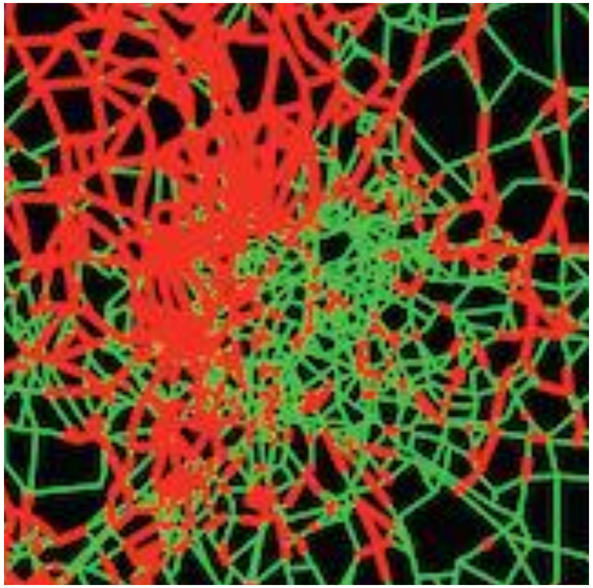
3D « time-of-flight » camera (PMD camBoard Nano)
+ segmentation/labelling of fingers by RandomForest
+ gestures recognition by HMM or/and DTW

Outline

- **Visual detection & recognition of traffic signs and object categories (cars, pedestrians, etc…)**
- **In-car Human Gestures & Activities recognition**
- **<u>Road traffic mining and forecasting</u>**
- **Onroad precise visual car ego-localization**

Traffic mining and forecasting

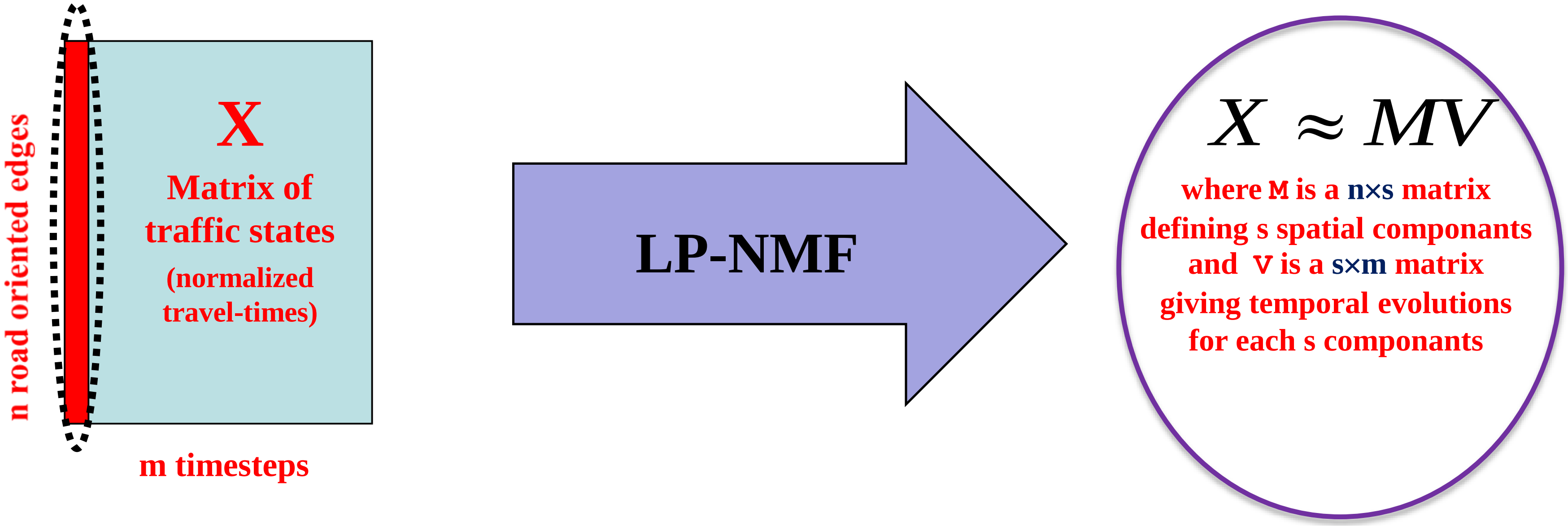
- Goal: forecast large-scale (~70km) long-term (~1h-2h) evolution of traffic, for re-planning of fastest itinerary



- Input data:
  - current traffic state + evolution since beginning of day
  - history of travel-times on hundreds of days

Dimension reduction for traffic data

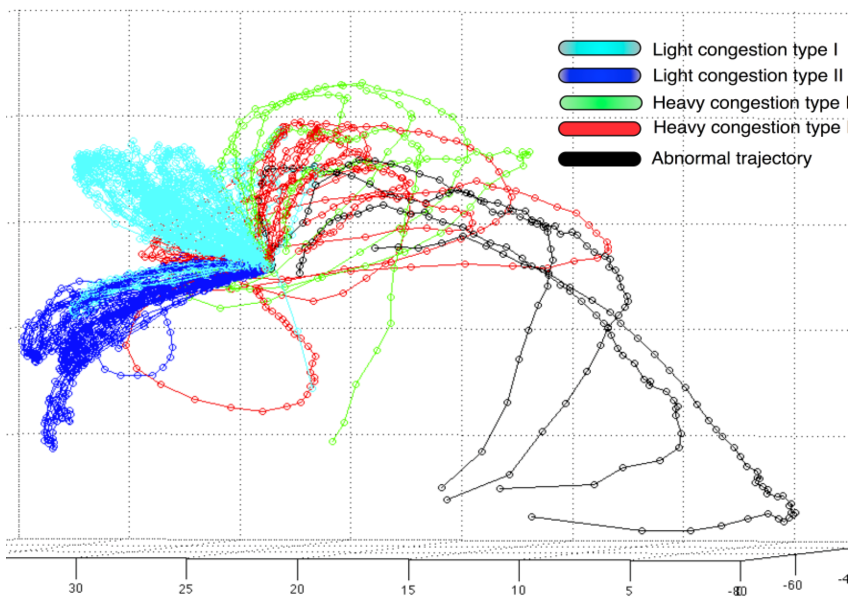
- **with Locality-Preserving Non-negative Matrix Factorization (LP-NMF)**



→ **Each traffic state of full area (vector in $]0;1]^n$ with n~5000) mapped onto a compact representation in $\Re^s$ (with s~15)**

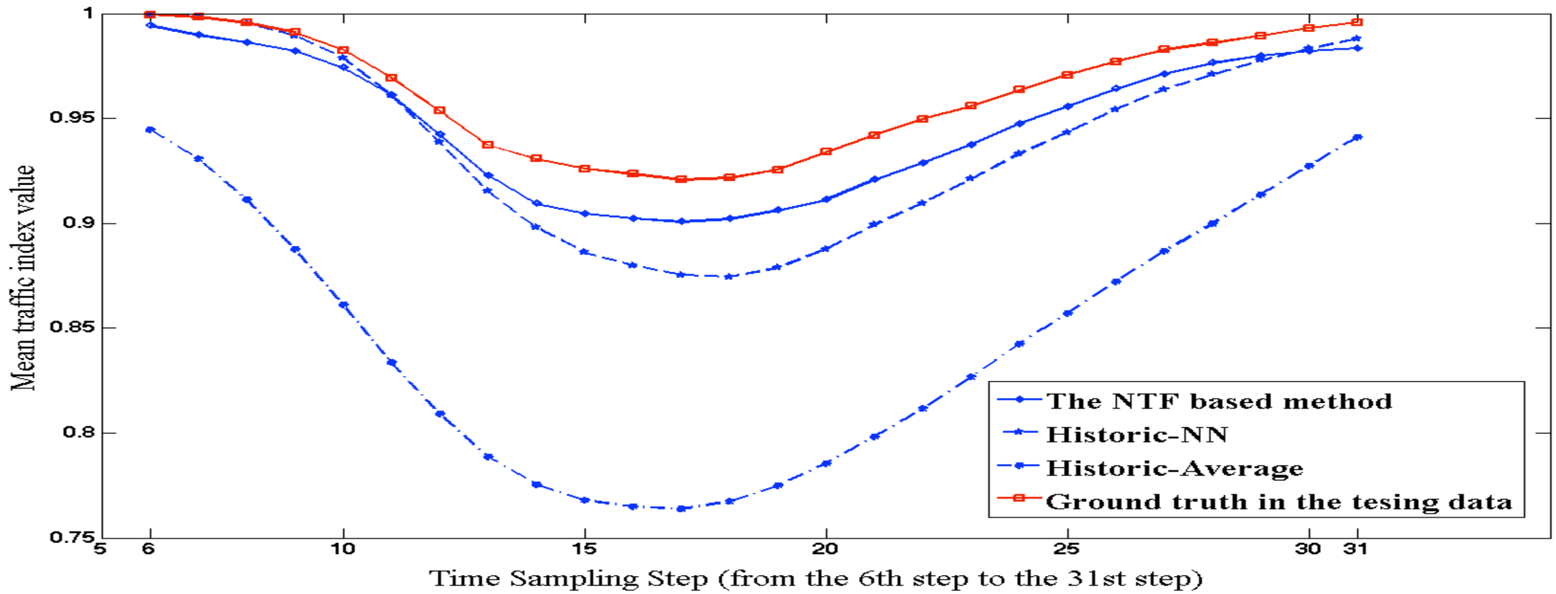
Partitioning of days into several typical temporal evolutions

- **Each day in historic mapped onto a *<u>trajectory of d successive points in $\Re^s$</u>***
- **Apply clustering (e.g. K-means) on the set of trajectories → partition days of history into several big types of daily evolutions**



Traffic forecasting

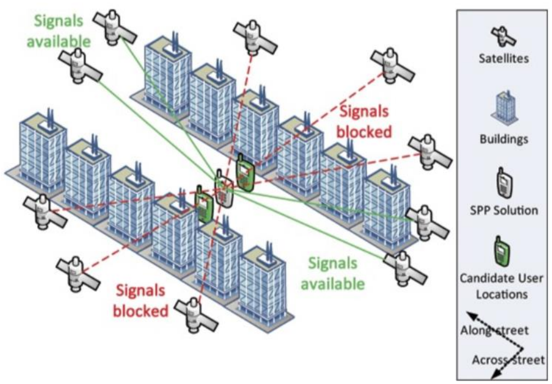
1. Given beginning of day (b vectors in $]0;1]^n$), estimate start-of-trajectory as b vectors (p_j) in $\Re^s$
2. Find in history the K most similar start-of-days in $\Re^s$ (efficient search as s<<n)
3. Future assumed to be linear combination of those K rest-of-days (with ponderations = similarities)

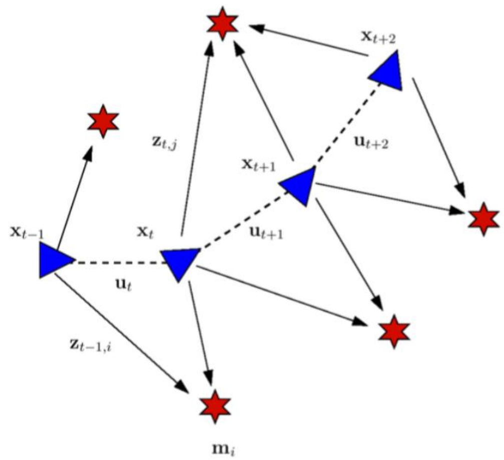


Outline

- **Visual detection & recognition of traffic signs and object categories (cars, pedestrians, etc…)**
- **In-car Human Gestures & Activities recognition**
- **Road traffic mining and forecasting**
- **<u>Onroad precise visual car ego-localization</u>**

Work in collaboration with VeDeCom (phD thesis of Li YU)



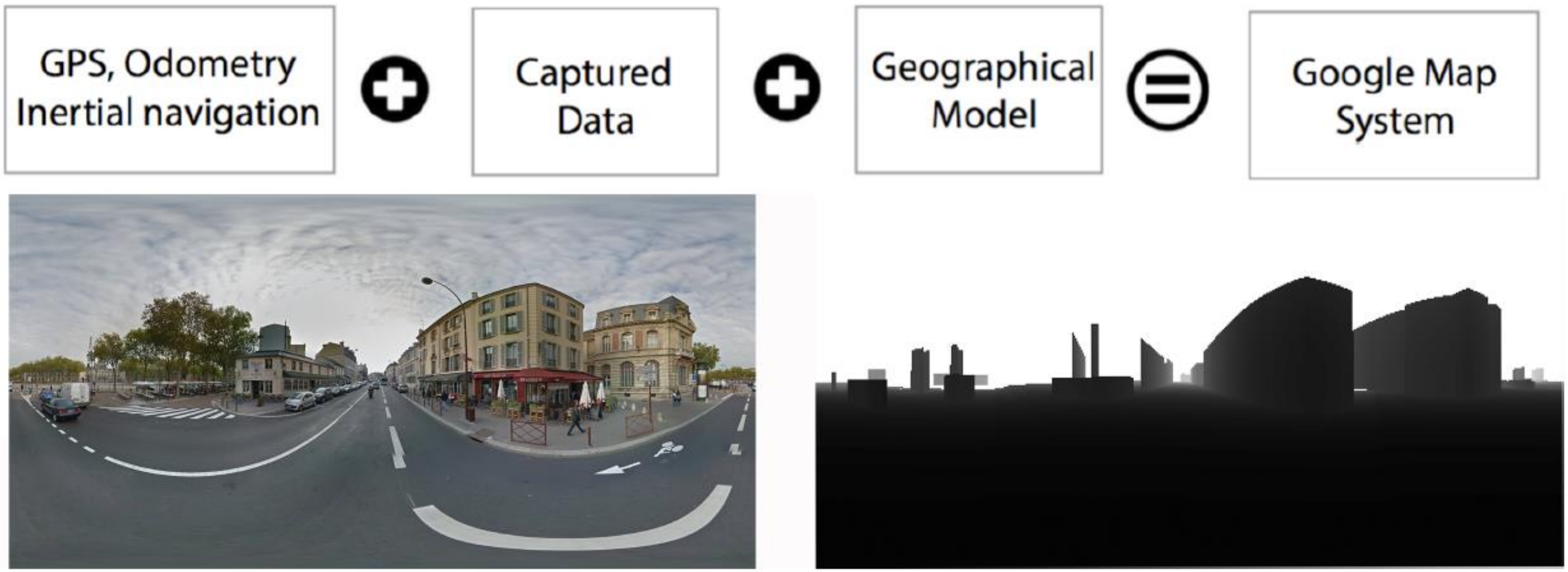


GPS	SLAM	Data Fusion
Signal Degradation	Incremental Drift Relative positioning	Uncertainty of sources

Our goal: <u>camera-based localization</u> technique <u>leveraging data from GIS</u> (Geographical Information Systems) such as Google StreetView

Image data in Google StreetView

Google Street View



An example of Street View panorama (13312*6656) and its depth map(<200m) at the localization [48.801631, 2.131509] with a yaw degree(158.39°) w.r.t the north direction.

Our proposed approach

Real-time mining of nearest reference images from GIS by approximate matching of current image from car

+

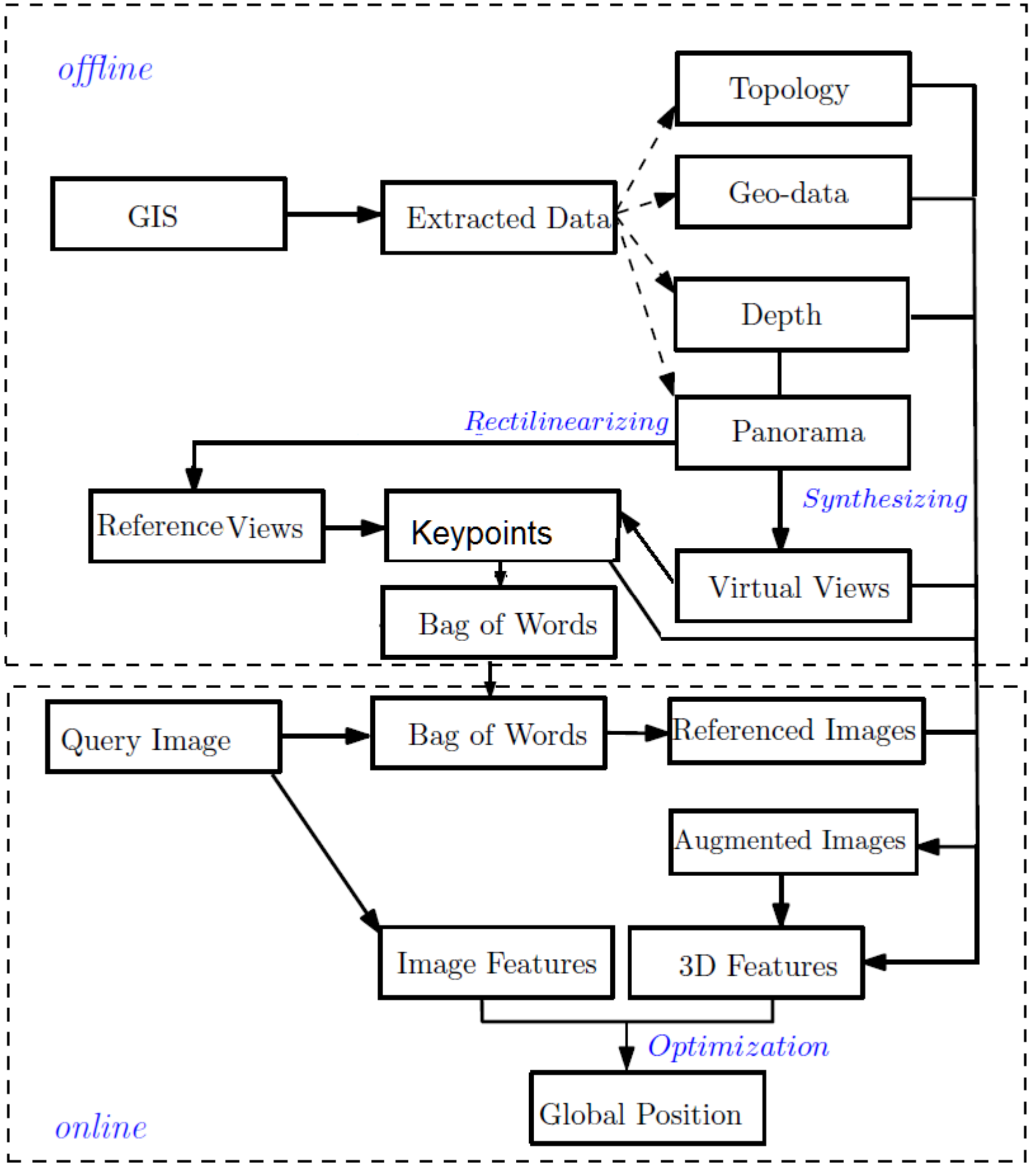
Estimation of precise pose (position+heading) by comparison current_image/reference_images

~ Place recognition

+

~ SLAM-like visual pose estimation (by geometric image comparison)

Pipeline of our algorithm

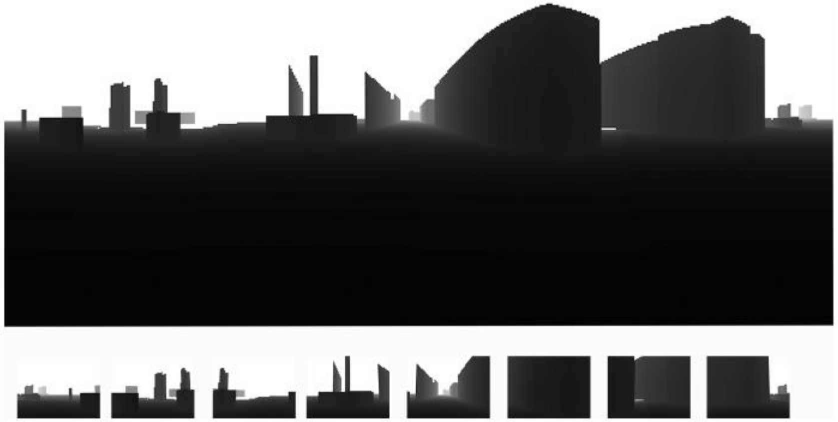


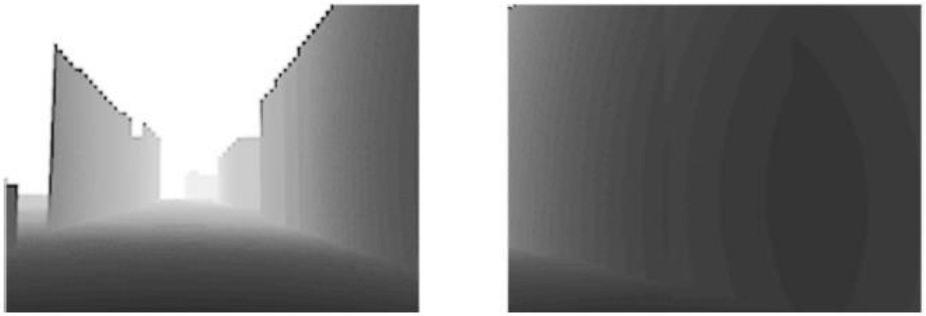
Offline preparation of GIS data

- **Generate rectified images from panoramas**
- **Generate synthetized intermediate images *between* panoramas**
- **Compute keypoints on obtained reference images**
- **Build bag-of-words descriptors for them**

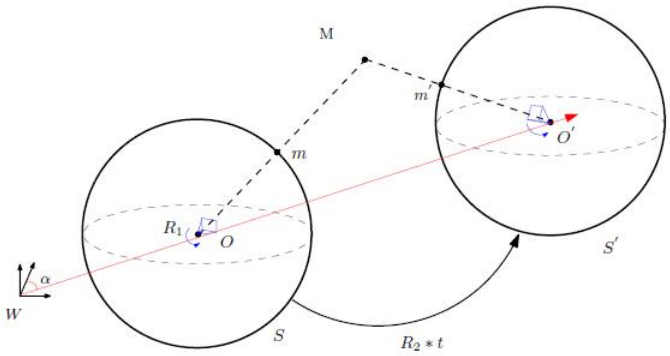
Rectifying StreetView images

Example of rectified images:





Synthetizing intermediate views



Translation distance	$2m$	$4m$	$6m$	$8m$
Invalid camera position	0	3	11	27
Uniform distribution	N	Y	Y	N
Ratio of virtual views with null pixels	0	0.125	0.5	1

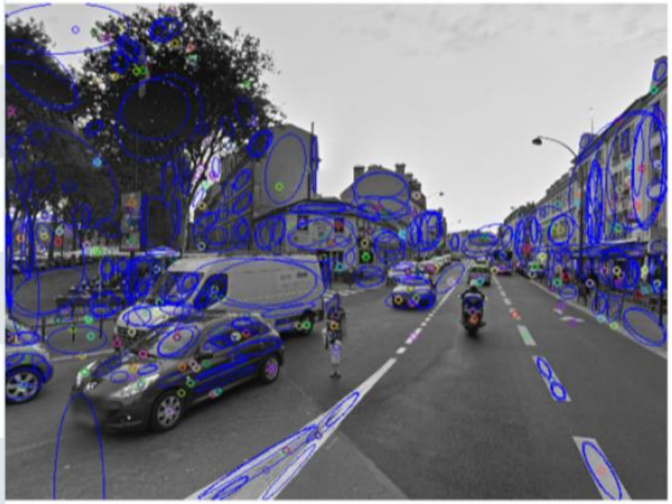
4-meter forward/backward virtual panoramas are constructed from the original ones

Database construction

t=0

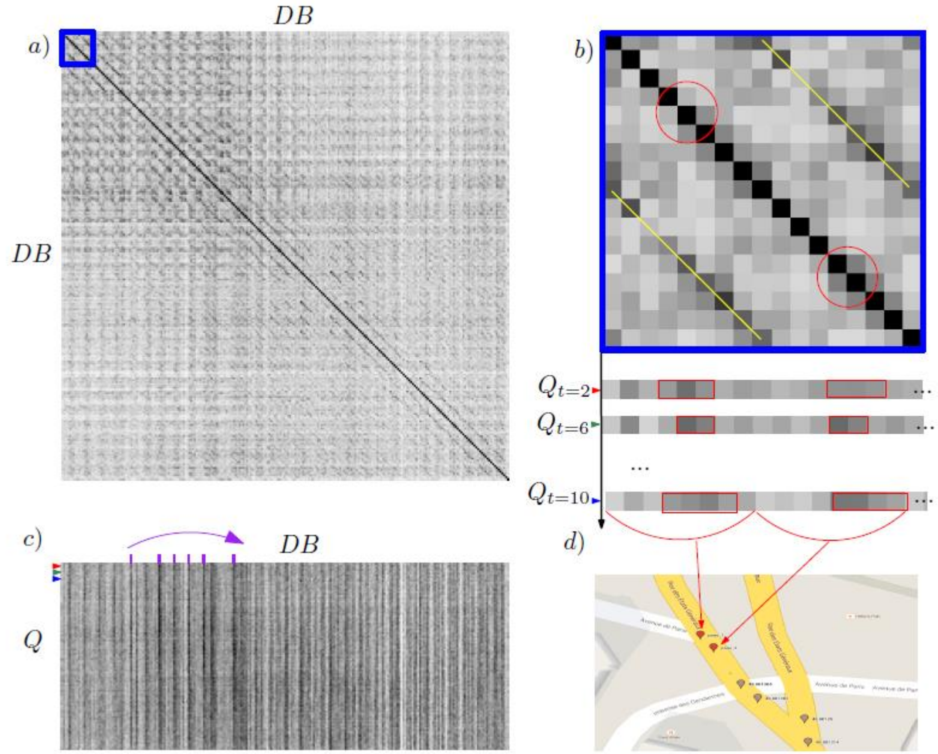
Construct 2 independent bags of words		
No.	1	2
Detectors	SIFT	MSER
Descriptors	484202	91026
Parameterization		
Size of bags	5000	2000
IF-ITF weighing		
Combination of two bags		
Search by cosine similarity		

....



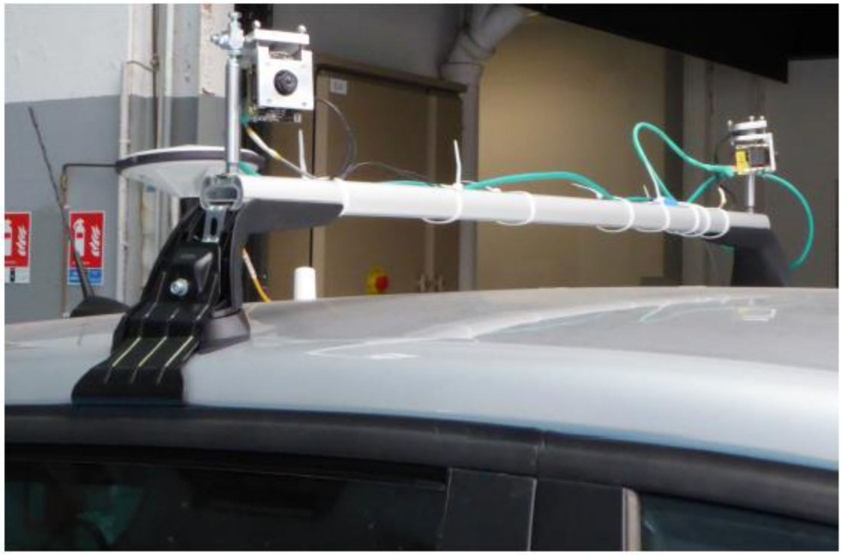
SIFT - local point
MSER - local region

Search efficiency optimization



Topologic relationship co-similarity matrix helps to reduce 89.7% impotent searching.

Acquisition system used

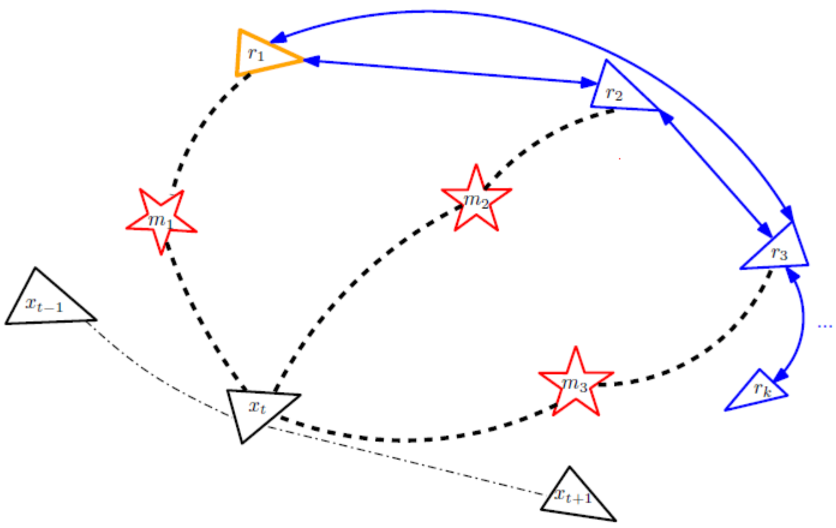


Left: Capturing system by MIPSee Camera(57.6°FoV, 20fps);
A Real Time Kinematic GPS as ground Truth;

Right: A sample of current image (640*480).

Real-time localization method

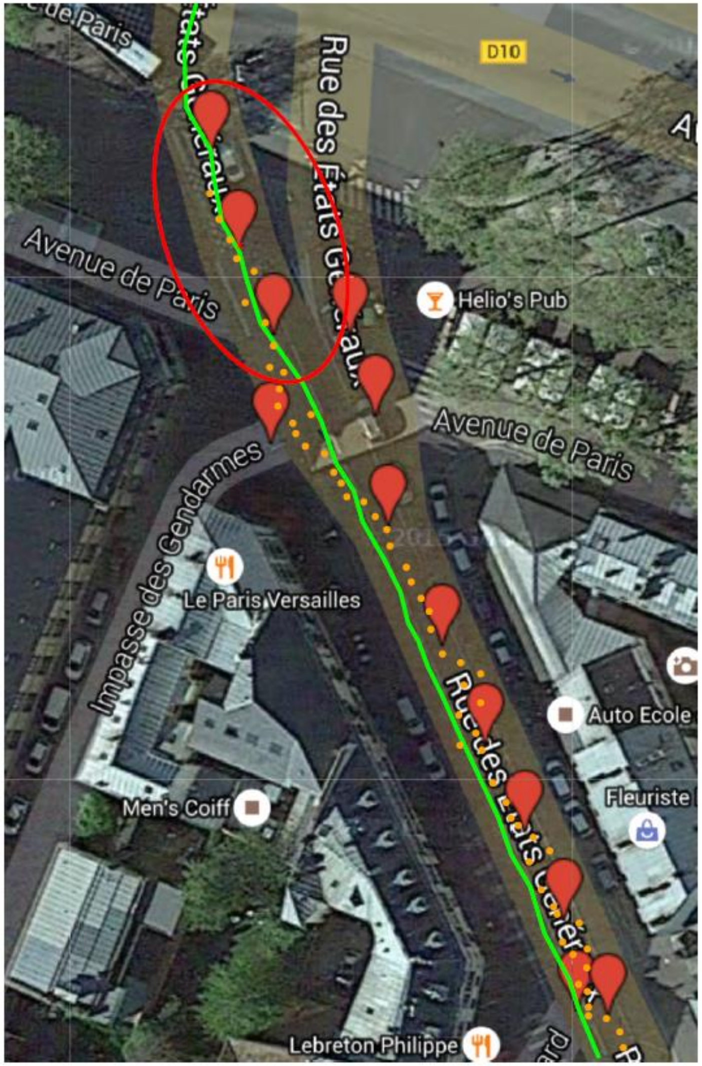
Coarse to fine Localization:
- Topologic localization: Bag of words => Referenced images
- Metric localization: RANSAC PnP => pose



Red: observations
Blue: camera position of Google
Black: camera position of vehicle

$$\Theta^\star = \arg\min_{\Theta} \sum_i \pi\left(\|\mathbf{m}_i - \mathbf{P}(\mathbf{M}_i, \Theta)\|\right)$$

Results with only original panoramas



- -13 panoramas in a 287m street
- -Ground truth in green
- -Average error <6.5m, 58.6% <2m
- -Standard GPS <8m
- -58/423 images improved with metric localization

Results with « augmented » StreetView

	Original Street View	Augmented Street View
Continuity	137/1046	281/1046
Average Error	$3.82m$	$3.19m$
Ratio in $[0m, 1m[$	21.89%	41.28%
Ratio in $[1m, 2m[$	28.47%	27.40%
Ratio in $[2m, 3m[$	44.53%	19.22%
Ratio in $[3m, 4m[$	5.11%	12.10%

- 1046 query images
- 498m trajectory
- 28 existing panoramas
- 53 virtual panoramas synthesized

with augmented Street View:
More query images are localized
68.68% positions lie within the error [0m, 2m[

Conclusions on visual ego-localization

- **Feasibility of meter-level real-time urban localization with just monocular camera**
- **Interest of leveraging images in GIS such as Google StreetView**

Ongoing and future work:

- **compare our method with deep-learning (PoseNet)**
- **automatic update of GIS database when structural change detected ?**

General conclusions and ongoing/future research

CAOR has addressed many domains of Machine-Learning / Datamining applications for intelligent vehicles & ITS:

- **computer-vision / pattern recognition for visual scene semantic understanding, visual ego-localization of car, driver gestures recognition**
- **traffic-mining and forecasting**

<u>Ongoing and future work:</u>

- **unified recognition for driver fingers_micro-gestures & hand_macro-gestures**
- **deep-learning for localization (~PoseNet),**
  + possibly other Self-driving cars functions
- **Categorization of 3D points clouds from LIDAR ?**
- **deep-learning for analysis & recognition of gestural time-series (+possibly other types, eg trajectories?)**

Questions ?