Deep-Learning for Intelligent Vehicles

Vehicle absolute ego-localization from vision, using only pre-existing geo-referenced panoramas

Pr. Fabien MOUTARDE
Center for Robotics
MINES ParisTech
PSL Université Paris

Fabien.Moutarde@mines-paristech.fr
http://people.mines-paristech.fr/fabien.moutarde

Some research results of the center for Robotics

Autonomous Vehicles, Intelligent Transport Systems, Mobile and/or Collaborative Robotics & Virtual Reality

International collaborations with: Berkeley, EPFL, Shanghai JiaoTong,...
Industrial contracts with: Valeo, PSA, Safran, Thales, SoftbankRobotics, etc...
Outline

• Introduction on AI for Intelligent Vehicles
• Visual ego-localization from pre-existing geo-referenced panoramas: classic approach vs. Deep-Learning inference
• Wider outlook on Deep-Learning for Intelligent Vehicles

From Intelligent Vehicles to Autonomous Vehicles

The 5 « automation levels » for vehicles defined by SAE

<table>
<thead>
<tr>
<th>NIVEAU 0</th>
<th>NIVEAU 1</th>
<th>NIVEAU 2</th>
<th>NIVEAU 3</th>
<th>NIVEAU 4</th>
<th>NIVEAU 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>conducteur seul</td>
<td>accompagné</td>
<td>assisté</td>
<td>guidé</td>
<td>passif</td>
<td>100% autonome</td>
</tr>
</tbody>
</table>

ADAS
HANDS OFF
EYES OFF
MIND OFF

Aucune aide
Aides primaires: APU, régulateur de vitesse adaptatif...
Aides intermédiaires: Alerte franchissement de ligne, park assist ...
Aides avancées: Dépassements dynamiques
Conduite autonome complète à la demande
Conduite exclusivement autonome

APPLICABILITY CAN BE CONDITIONAL (e.g. RESTRICTED TO ONLY MOTORWAYS, ...)

What are ADAS?

Acronym of Advanced Driving Assistance Systems = Intelligent functions for safer and/or easier driving

**Warning or Information**

- Lane Departure Warning (LDW)
- Forward Collision Warning (FCW)
- Pedestrian Collision Warning
- Blind Spot Monitoring
- Speed Limit Assistant
- etc...

**Active ADAS**

- Adaptive Cruise Control (ACC)
- Lane Keeping (LK)
- Autonomous Emergency Braking
- Automated Parking
- etc...

Artificial Intelligence (AI) for Autonomous Vehicles

Autonomous Vehicles are mobile robots!

- Perception
- Reasoning & decision
- ROBOTIC LOOP
  - Action
  - Reactive behavior (obstacle avoidance)
  - Trajectory planning

Real-time "understanding" of surroundings

Ego-Localization

Predicting movements/actions of others
Introduction on AI for Intelligent Vehicles

Visual ego-localization from pre-existing geo-referenced panoramas: classic approach vs. Deep-Learning inference

[work by former VeDeCom PhD student Li YU, co-supervised with G. Bresson and C. Joly]

Wider outlook on Deep-Learning for Intelligent Vehicles

Visual ego-localization motivations

• GPS not always available (indoor, tunnels, underground parkings, "urban canyons")
• GPS precision quite low (up to 10m error! [except for differential GPS]
• GPS directly provides position but NOT the orientation (only the local orientation of TRAJECTORY can be estimated over time)
• Odometry is quite imprecise (cf. wheel slip!), and subject to large rapid cumulative errors
• Inertial Measurement Unit (IMU) expansive if precise, and subject to cumulative errors
Outdoor visual ego-localization

Where am I? (position + bearing)

Google StreetView data

360° panoramas (RGB in UHD 13,312x6,656 pixels + coarse 360° depthMap)
~ every 10-50 m in ~3000 city centers worldwide
Using StreetView panorama for synthesizing rectified images

- Distortion of 360° images
  + unknown query viewpoint

⇒ Generate sustainable rectified views (with same focal length as on-board camera) in several orientations

Generating virtual views BETWEEN StreetView panoramas

Possible thanks to availability of (coarse) panoramic depth map in StreetView
Visual place recognition using geo-localized images

With enough (~8-12) rectified synthetic images generated for several viewpoints, coarse visual place recognition by standard Bag of VisualWords (BoVW) is possible.

Pre-compute 1 BoVW x ~10 views for each geo-tagged panorama.

Visual metric localization from geo-localized images

OFFLINE

Geo-tagged pre-existing panoramas

Geo-tagged synthetized views

Keypoints in synthetized views

BoVWs of synthetized views

ONLINE (onboard)

Query image (from on-board camera)

Query image

Keypoints

BoVWW MATCHING

Geo-tagged reference view

Translation & Rotation by Keypoints matching

Absolute position & heading
Visual metric localization from geo-tagged reference view

- Estimation of translation+rotation from reference view to query image by *Bundle Adjustment of keypoint descriptors matches (with outliers filtered by RANSAC)*

- Use geo-tag of reference view + estimated translation&rotation to estimate current absolute position and heading

---

Experiment: set-up

Techniques:
- MIPSee Cameras 57.6° Fov / 20 fps
- 640*480 resolution
- Real Time Kinematic(RTK) GPS as ground truth (<20cm)
Results of experiment with "augmented" StreetViews

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

with augmented Street View: More query images are localized

68.7% of estimated positions with error <2m

Deep-Learning of pose regression using only pre-existing images

Adapt PoseNet approach [Kendall et al. 2015]:

- Learn an only 3-DoF pose \((x, y, q)\) [instead of 6DoF]
- Start transfer learning from ResNet50 model [instead of InceptionV3] modified as follows:
  - final classifier replaced by a dropout layer
  - fully connected layer with 256 neurons added and connected to final 3-dimension pose regressor
- Train using ONLY images generated from PRE-EXISTING geo-referenced Google-StreetView panoramas [instead of many images from prior 1\textsuperscript{st} pass]
Results of DL visual localization trained on pre-existing geo-referenced images

<table>
<thead>
<tr>
<th>SeqID (length)</th>
<th>Nb of images</th>
<th>Nb of StView panoramas (nb of virtual ones)</th>
<th>Average localization errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Image features + geometry</td>
<td>Pose regression CNN</td>
</tr>
<tr>
<td>1 (234 m)</td>
<td>897</td>
<td>29 (1160)</td>
<td>2.85 m 7.62 m</td>
</tr>
<tr>
<td>2 (271 m)</td>
<td>898</td>
<td>29 (1160)</td>
<td>2.63 m 7.93 m</td>
</tr>
<tr>
<td>3 (222 m)</td>
<td>895</td>
<td>29 (1160)</td>
<td>Fail  Fail</td>
</tr>
<tr>
<td>4 (216 m)</td>
<td>901</td>
<td>34 (1360)</td>
<td>2.82 m 7.55 m</td>
</tr>
<tr>
<td>F (265 m)</td>
<td>554</td>
<td>29 (1160)</td>
<td>Fail 7.87 m</td>
</tr>
</tbody>
</table>

DL ego-localization errors (~ 7m) larger than with BoVW+geometry

BUT
Error comparable to GPS, and much faster inference (~75ms) than using BoVW+geometry (~3s !)

Conclusions on DL for Visual ego-localization of Intelligent Vehicles

- **Vision-based ego-localization** using BoW place recognition + keypoints matching, even if using as only references pre-existing geo-tagged panoramas from Google-StView, can provide in city centers positioning accuracy of ~3m, comparable to plain GPS.

- **Deep-Learning pose regression** = very interesting alternative to standard visual localization methods: currently still ~ 2 times less precise (positioning error~7m), BUT but much more real-time at inference (75ms/image vs. 3s/image).

⇒ Potential use as a fallback for mitigating GPS outages (urban canyons, tunnels, etc….)
Outline

• Introduction on Intelligent Vehicles
• Visual ego-localization from pre-existing geo-referenced panoramas: classic approach vs. Deep-Learning inference
• Wider outlook on Deep-Learning for Intelligent Vehicles

Deep-Learning for Perception by Intelligent Vehicles

Visual objects detection, semantic segmentation

From camera

Real-time estimation of Human poses [OpenPose, 2017]

From LIDAR (3D points cloud)
Image-based Deep-Learning

- Image classification
- Visual object detection and categorization
- Semantic segmentation of images

- Estimation of Human pose
- Inference of 3D (depth) from monocular vision
- Image-based ego-localization
- Realistic image synthesis

- Learning image-based behaviors
  - End-to-end driving from front camera
  - Learning behavior by *Imitation of Humans*, or with *Reinforcement Learning*

---

Unsupervised monocular depth estimation with left-right consistency
C Godard, O Mac Aodha, GJ Brostow - CVPR’2017 [UCL]
PoseNet: camera-pose regression with Deep-Learning

PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization
[ICCV’2015, pp. 2938-2946]

DL for realistic Image synthesis

"Video-to-Video Synthesis", NeurIPS’2018 [Nvidia+MIT]
Using Generative Adversarial Network (GAN)
Deep Reinforcement Learning for vision-based Autonomous Driving

Work by my Valeo/MINES_ParisTech PhD student Marin Toromanoff
1st prize at « CARLA Autonomous Driving challenge »!!

Reward shaping

Agent

Environment

Town02: Single Lane, EU
Weather: Heavy rain
Traffic Light: Red

Network input

Current Order: Left
Current Speed: 1.8 km/h
• Deep-Learning (DL) is now able to provide much more than just image analysis for visual objects detection

• For Intelligent Vehicles, DL is now even investigated \textit{beyond perception}, for machine-learning of \textit{reactive behavior} or even \textit{trajectory planning}