Deep Learning for gesture recognition based on novel hardware (Google project Soli)

Project description

We are currently witnessing several drastic shifts in the computing landscape. The traditional PC and hence the traditional mouse and keyboard-based UI are no longer the primary computing paradigm but are increasingly complemented by other paradigms such as direct-touch and gestural interaction. As our environments become smarter and the line between virtual and real worlds becomes blurred, it also becomes increasingly clear that traditional forms of user input are no longer adequate for mobile scenarios. With technologies such as wearable computing, Augmented and Virtual Reality (AR/VR) transitioning from the research labs into the mainstream market the need for a general purpose, contact-and wireless, and high precision, high bandwidth man-machine interface becomes increasingly urgent.

This project is aiming at building an intelligent algorithm that could recognize low-effort, low-energy and high-bandwidth interactions involving primarily small muscle groups. The interaction paradigm is based on UWB radar sensor (Google project Soli: https://www.youtube.com/watch?v=0QNzIfSsPc0) which is sensitive to tiny motions, immune to optical occlusion, works both indoor/outdoor, through paper/clothing/thin plastics at flexible range.

Recently convolutional deep neural networks (DNNs) have made a significant impact in computer vision. DNNs have outperformed state-of-the-art machine learning algorithms in very large-scale image recognition and hand-written digits recognition. In a recent competition on multimodal recognition of 20 dynamic gestures from the Italian sign language, an algorithm based on convolutional neural networks ranked first among 17 competing methods. Convolutional DNNs forgo handcrafting discriminatory features for classification, and instead learn them automatically from the training data. Range-Doppler Images (RDI), which depicts the amplitude distribution of the received signals for certain range (z) and Doppler (v) values, is the main raw features of the radar sensor. A rigid moving object appears as a single point in the RDI and a non-rigid object, e.g., a hand, appears as multiple points. Different dynamic hand gestures can be viewed as different sequences of those RDIs, so 3D CNN based algorithm is very promising for this special sensor.

Key words

Google Project soli, Radar Sensor, Deep Learning, Gesture recognition, Novel interaction interface

Context

The goal of this project is to explore state-of-art algorithms to enable robust gesture recognition based on this novel hardware.

Work packages

- Literature survey on existing State-of-Art gesture/action recognition based on deep learning
- Get familiar with existing Radar API and define gestures and collect relatively huge amount of Data
- Get familiar with existing Deep Learning Toolbox(Caffe/Torch/Theano)
- Explore and implement proper CNN pipeline for radar data
- Bonus: Incorporate the CNN pipeline into the Radar API and build some interesting interaction applications

Required skills

- Good C or python skills,
- Good knowledge in machine learning and computer vision
- Highly motivated and independent

Type SA/MA(can be 1 or 2 students)
Partner Google ATAP
Time period Autumn 2015
Student(s)
Internal supervisor(s) Jie Song, jsong@inf.ethz.ch, Otmar Hilliges, otmar.hilliges@inf.ethz.ch
Formal Requirements

Work schedule: Please provide your supervisor a work schedule within two weeks after the start of the project. Generally, Bachelor and Semester projects last 14 weeks starting at the first day of the semester and end at the end of the semester. Master projects last 6 months and the starting date is agreed on with the supervisor.

Intermediate presentation: An informal intermediate presentation (about 10 minutes presentation/discussion) about your work will take place around mid-term. The goal of the presentation is to give a brief summary of the work done, to propose a plan for the continuation of the project, and to discuss about the main directions of the project.

Final presentation: The final presentation will take place at the end of the project. A test run is presented to and discussed with the supervisors 2–5 days before the public final presentation. Exact dates and times for the intermediate and final presentations will be arranged by the lab administration.

Report: A report has to be handed in to the responsible supervisor. The report has to describe the full work performed during the whole project. A preliminary version has to be handed in one week after the final presentation or as determined with the supervisor. The preliminary version of the report is discussed with the supervisors. The final report has to be handed over to the responsible supervisor in 3 paper copies. All documents and files, including the report (original data and as a PDF-file) and the final presentation, have to be saved on a CD/DVD and handed in together with the final report.

Evaluation: The project is assessed according to the AIT evaluation sheet and the “Merkblatt für Studierende zum Thema Plagiate” of ETH Zurich. The responsible supervisor hands both of them out at project start. The final presentation is evaluated based on the public final presentation. The report is evaluated based on the preliminary version with the requirement that the remarks are incorporated in the final report.

Plagiarism: Every student has to make himself/herself familiar with ETH Zurich rules regarding plagiarism.
http://www.ethz.ch/faculty/exams/plagiarism/index_EN

Duration of Final Presentations:

- Bachelor and Semester projects: 15 minutes presentation, 5 minutes questions / discussion
- Semester/Bachelor Project with integrated SoM: 22 minutes presentation, 8 minutes questions / discussion
- Master projects: 20 minutes presentation, 10 minutes questions / discussion

<table>
<thead>
<tr>
<th>Responsible Professor</th>
<th>Responsible Supervisor</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature</td>
<td>Signature</td>
<td>Signature</td>
</tr>
</tbody>
</table>

Zurich, Date: ________________.

Notes
The internet provides detailed information on how to write a scientific report and how to make a presentation. The report and presentation should be done according to these common guidelines and the instructions from your supervisors. We recommend reading the following instructions before you start with your work:

- Writing Guidelines for Engineering and Science Students
  http://www.writing.eng.vt.edu/
- Prof. Bernstein’s Student Guides
  http://aerospace.engin.umich.edu/people/faculty/bernstein/guide.html

The style and format of your report should follow common practices and the instructions of your supervisors. However, the ETH template is preferred for the presentation. We recommend you using Microsoft Office, OpenOffice or LaTeX to create your report and presentation. Please ask your supervisor for the template files.

Report

- Oetiker et al., The Not So Short Introduction to LaTeX 2e, May 2006.
- LaTeX Editor
  http://www.toolscenter.org/ or http://www.lyx.org/

Presentation

- Tufte, PowerPoint is Evil, WIRED, September 2003.
  http://www.wired.com/wired/archive/11.09/ppt2.html
- The LaTeX Beamer Class
  http://latex-beamer.sourceforge.net/