Task Description

Master Thesis

Markerless 3D Human Pose Estimation with Multimodal Deep Learning

The Chair of "Industry Grade Networks and Clouds" deals with applications and use cases of network technology and virtualization in industrial manufacturing. The main areas of application are industrial robotics, mobile industrial service robotics, human-robot-interaction, machine learning and Augmented Reality.

General scope of the thesis:

Pose Estimation or Skeleton Detection is a field of research with the goal of obtaining spatial information about the joints of an object. Information about the joints can be used in an industrial context to estimate the pose of a robot or a human worker. In the particular use-case of this thesis, the human pose estimation will be used for ergonomical assessment of factory workers. Given the pose of the joints of the human body, the ergonomics of working positions can be assessed with different methods (e.g. ‘Ergonomics Assessment Worksheet’ or ‘Rapid Entire Body Assessment’). Similar work has already been done for construction workers [2].

State of the art 3D pose estimation is currently done based on RGB images and deep learning with Convolutional Neural Networks (CNNs) [1, 3, 4]. Pre-trained networks can often be obtained from the paper’s authors. The hypothesis of this thesis is that state of the art 3D pose estimations based on RGB images can be improved by combining RGB images and depth images (RGB-D data).
Combining different sensor data with deep learning is also called Multimodal Deep Learning and can improve object recognition performance [5].


Objective of the thesis:
The goal of this thesis is to train a deep CNN to estimate the 3D human pose as a sequence of joint positions. To achieve this goal, the first step is a literature research of 3D human pose estimation with deep learning. The second step is to implement an existing state of the art neural network based on RGB-images as a baseline and then combine this network in a third step with additional depth data in a novel sensor fusion approach. The novel multimodal deep neural network has to be trained on a dataset (e.g. Human3.6M) and the 3D joint estimation outputs should be evaluated. The evaluation should show whether additional depth data can improve the 3D pose estimation on a specific dataset compared to the RGB-image baseline.

Finally, the trained neural network should be tested in an industrial environment with a use-case for ergonomical assessment of factory workers.

The chair IGNC is providing a GPU-based infrastructure for training CNNs. The student will have access to the laboratories and offices to work.

Requirements:
- Master student in computer science, computer engineering, electrical engineering or similar
- Knowledge of machine learning and especially deep learning
- Programming experience (Python)
- Experience with deep learning frameworks (Tensorflow, Keras, PyTorch, etc.)
Exemplary course of actions:

1. Literature research
   - Ergonomics assessment methods and the required joint information
   - Deep Learning state of the art approaches for 3D pose estimation
   - Sensor fusion of RGB and depth data in Deep Learning
   - Research of usable datasets for 3D pose estimation

2. Implementing an existing state of the art 3D pose estimation algorithm
   - Building a pipeline to train and test a given dataset
   - Getting the baseline model to work

3. Building a novel architecture to incorporate depth data
   - Early fusion vs late fusion
   - Network architecture
   - Preprocess depth data (e.g. depth image, point cloud, voxel)
   - ...

4. Training, testing and evaluation
   - Train the neural network on existing training data to output 3D human pose estimations
   - Compare the baseline model (step 2) against the new model with depth data (step 3)

5. Test the trained model in an industrial environment in the real world

Contact

Supervisor
M.Sc. Leon Eversberg
leon.eversberg@tu-berlin.de
www.ignc.tu-berlin.de

Professor
Prof. Dr.-Ing. Jens Lambrecht
lambrecht@tu-berlin.de
www.ignc.tu-berlin.de
Resources / Links

Paper

Dataset
http://vision.imar.ro/human3.6m/description.php

Others
https://heartbeat.fritz.ai/introduction-to-multimodal-deep-learning-630b259f9291
https://depositonce.tu-berlin.de/handle/11303/7612