**Topic 1: Multimodal sensor fusion in urban environments (All Speakers)**

We humans perceive environments through smell, vision, taste or touch. With our ‚god given‘ sensors we are able to receive information from our surroundings, allowing us to interact with it. But how do we train an autonomous system to understand and interact with its environments? The key components here are essentially sensory data, data interpretation and acting. By combining sensory data from different modalities, the overall perception capabilities of the system can be greatly improved. The autonomous system must be self-aware, aware of the situation and then act accordingly.

In general, there are three types of sensor fusion: **Competitive fusion** describes the integration of data from independent measurements of type identical sensors. **Complementary fusion** describes the data fusion of multiple independent sensors to close information gaps by observing different properties of the environment. **Cooperative Fusion** describes a technique to extract new information from two or more independent sensors which a single sensor cannot measure.

However, not all sensors are equally reliable and the correspondence between information from different sources is usually unknown. Therefore all sensors must be time synchronized and calibrated with each other.

This is where the Kalman filter comes in. The key idea here is to stack all the observation matrices together into one measurement and then solving the estimation with a classic Kalman filter. This can be used to quickly estimate the position, velocity, etc. of the object being measured. However, in an urban environment, where we also want to determine the location of other objects, multimodal sensor fusion can be best achieved by neural networks. When using NN’s, different fusion strategies such as early or late fusion can be applied. Since Robotics is going to be ubiquitous and impact society as a whole, sensor fusion becomes more important as it increases the effectiveness of the state estimation.

**Topic 2: Towards Autonomous Mobile Manipulations for Robots: A Compositional Approach** (**Dr. Pascal Meißner)**

Dr. Meißner presented his work on how to make robots work in unstructured / urban environments. For that, he introduced his research of a multi-modal service robot. The robot observes human interactions like speech, object positions and human body movement. With these observations, the robot creates a model to derive its actions from it and react to human interactions. The robot was also able to find objects. For this it had to analyze the relations between objects. After the analysis, it had a prediction and could search for other objects.

As his second work he showed an industrial robot, where the robot learns to pick and place objects in the right location. For this he trained three neural networks. One as a precondition: The neural network should estimate the success of grasping the object. Another one as a postcondition: This neural network should estimate the success of placing the demonstrated object. And the last neural network combines the predictions of the two others. The networks should learn in a self-supervised manner, so they learned the pick and place actions on their own with 25.000 real world experiments.

The last work was about a robot that could adapt its trajectories based on effects of the environment on it. Given sensory feedback and future waypoints of the original trajectory, a neural network was trained to predict joint accelerations at regular intervals. And with these predictions the robot could balance different objects on a plate.

At the end he mentioned some challenges in the field of autonomous robots and his personal research:

* How to select appropriate features for reinforcement learning in general?
* How to learn pre- and post-conditions of elementary actions like grasping?
* How to train robots for active perception of changing environments?