



Hochschule  
Bonn-Rhein-Sieg  
University of Applied Sciences



R&D Project Proposal

# Multimodal Machine Learning for Robotics

*Priteshkumar Gohil*

Supervised by

Prof. Dr. Paul G. Plöger  
Santosh Thoduka

April 2019

## 1. Introduction

- The environment that we experience is multimodal because we use visual, audio, touch, taste and smell sensory information. A very good example could be a person talking to someone will combine audio (voice) with the visual (lips movement) information to understand them better
- The Autonomous mobile robot is equipped with multiple sensors with which they measures different aspect of its environment and perform various task.
- Multiple sensors on autonomous mobile robot measures the different information of same environment. For example, occurrence of an event is captured by multiple sensors. An event can be robot accidentally dropping an object while performing the task and reading from tactile, vision, audio sensors gives some information about the same event. These observations of same event from the sensors can be combined together which is called multimodality.
- According to Merriam Webster, vision is the prime part of sensation in modality.
- We focus to combine, process and relate multiple sensory modalities using machine learning techniques to solve a certain task in the field of robotics. Our primary focus is on combining visual information with other sensory data.
- Developing such a model can help to correlate visual information with multiple other modalities from robot sensors to understand the robot's environment better and increase the performance.
- Usually, different tasks in robotics are solved with single modality. A computer vision techniques are applied on visual sensory information. Signal processing and filtering is used for audio and tactile sensor data. However, information available from one sensor is processed independent of other sensor data. Where we are missing correlation between different sensor information while processing.

- The approach can help robot to understand it's environment better or accurately given a particular task.
- It can be applied to follow ing task performed by the robot,
  - Pick and place task
  - Fault detection and diagnosis
  - Audio-Visual Speech recognition
  - Event recognition
- Other than Robotics, multimodal machine learning can also be used for,
  - In the medical field such as neuroimaging to fuse data from EEG, MEG, MRI and increase the performance of diagnosis [? ].
  - Media captioning

### 1.1. Problem Statement

- What are you going to solve?
- How are you evaluating?

## **2. Related Work**

- What have other people done?
- Why is it not sufficient?

### **2.1. Section 1**

### **2.2. Section 2**

## **3. Project Plan**

### **3.1. Work Packages**

The bare minimum will include the following packages:

WP1 Literature Search

WP2 Experiments

WP3 Project Report

Keep in mind that depending on your project, you will probably need to add work packages that are more suited to your projects.

### **3.2. Milestones**

M1 Literature search

M2 Experimental setup

M3 Experimental Analysis

M4 Report submission

### **3.3. Project Schedule**

Include a gantt chart here. It doesn't have to be detailed, but it should include the milestones you mentioned above. Make sure to include the writing of your report throughout the whole project, not just at the end.

# Multimodal Machine Learning for Robotics

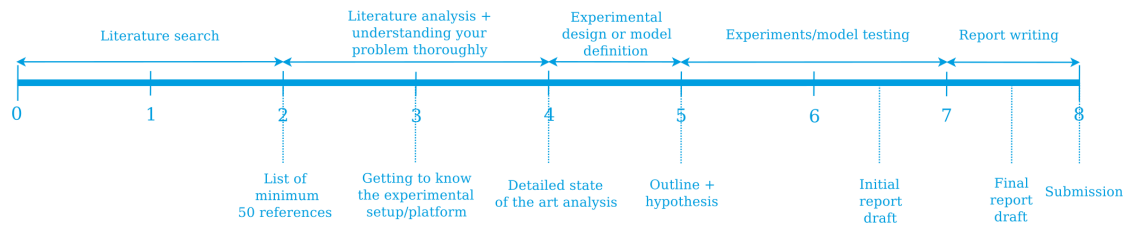


Figure 3.1:

## 3.4. Deliverables

### 3.4.1. Minimum Viable

- Survey
- Analysis of state of the art
- Simple simulated use case
- Demo on youBot or Jenny

### 3.4.2. Expected

- Comparison of approaches in the robot

### 3.4.3. Desired

- Integration to scenario