Pre-project report

Bachelorthesis in software engineering at OsloMet
Faculty of Technology, Art and Design

Group 15: Eskil Gaare Høstad, Jonathan Kjølstad

January 22, 2021

Contents

1 Presentation

Client: Huddly AS
Project title: Gesture Recognition for Video Meetings
Group number: 15

Group members:

Name	Student number	E-mail
Eskil Gaare Høstad	s331380	eskilgh@gmail.com
Jonathan Kjølstad	s321399	jonathan.kjolstad@gmail.com

We have teamed up with Huddly for our bachelors project. Huddly's goal is to enhance the experience of video communication by bringing intelligent cameras into the meeting room. Huddly does this by carefully combining hardware, software and artificial intelligence solutions to produce conferences cameras with intelligent features.

Huddly AS
Karenslyst Allé 51, 0279 Oslo
`https://www.huddly.com`

Contacts:

Name	Position	E-Mail
Company supervisors		
Elena You	Data Scientist	elena@huddly.com
Espen Wøien Olsen	Data Scientist	espen@huddly.com
Supervisor at OsloMet		
Jianhua Zhang	Professor	jianhua.zhang@oslomet.no

2 Summary

We will research and develop software that uses deep learning to identify and classify gestures, for use in communication with others and/or camera features. We intend that this will grant Huddly valuable insights into today's possibilities with new interactive gesture based functions, mostly directed at video communication. We intend to deliver an MVP (Minimum Viable Product) within the first 9 weeks based on the requirements specification.

3 Current situation

Contrary to regular cameras, Huddly's can not only provide a video feed, but by utilizing the onboard neural engine is able to see, understand, process and respond to what they are seeing. This means that it can automate tasks like controlling the camera itself and will also be able to generate and report high-quality analytics data about what it sees. The camera's capabilities are virtually endless, and Huddly is therefore always on the lookout for new intelligent features to implement.

At this point in time Huddly wants to gain new insights into hand gestures as a means of interacting with a machine. By moving simple functions in applications away from traditional means such as keyboard and mouse, to relying simply on gestures from camera input, one could provide a lot more flexibility of controls to the user. This is especially true for video communication applications where a video feed already would be present and user functions usually are quite limited and minimal. By instead utilizing gesture recognition to control these functions, video communication could be streamlined into a convenient and completely hands-off experience.

4 Goals and conditions

The aim of the project is to develop a proof of concept demonstration of gesture recognition on live video, using machine learning/deep learning techniques. The demonstration is only required to work locally.

Development of the gesture recognition model should rely on already existing solutions, as thesis a well-researched area. Furthermore it should be noted that the students are relatively new to the field of image related machine learning. Therefore it is not expected that the end report will contain a detailed analysis of the construction and specification of the model. Rather the focus of this project is on the implementation of available models. Nevertheless basic reports on aspects such as accuracy, precision, etc should be considered throughout the process and included in the final report.

For development the students are in principle free to use what OS and platform they deem fit. The recommended software tools are Python 3.6+ and well-documented libraries connected to Python 3.6+. This is to express that the task has been solved in a methodological and academic manner, and proving the students general insight of methods to solve such tasks.

A list of recommended libraries that fulfill these criteria are as followed:

- Keras
- Tensorflow
- Numpy
- Opencv

Other libraries may of course also be used, as long as they are regarded as established and well-documented.

5 Solution

The exact specifications of which gestures and which use-case demonstrations are to be included in the final demonstration is not well-defined in this phase of the project. In short this is due to challenges on estimating both workload and timetables for the development process. As such the plan is to use an iterative workflow. For the initial stage in the project the target should therefore be a MVP (minimum viable product), which is specified further below. The plan is to deliver a MVP as early as possible, evaluate it, and then add functionality according to what is requested and deemed appropriate for the development schedule.

5.1 MVP-specification

The target demonstration should consist of a program run locally on a video feed. This program should demonstrate machine learning classification of static hand gestures. In principle this can be as simple as displaying a surrounding box on a given gesture, with the appropriate label. Functionality for showcasing gesture-recognition use-cases should also be included, an example of this can for instance be toggling mute/unmute through a designated gesture. The exact specifications of which gestures/functionality to include in the MVP are still subject to change, in accordance with pragmatic choices regarding available data / models or other limitations.

After a MVP has been developed, the product should be evaluated by both the students and company supervisors. Appropriate changes can then be specified for the next iteration, as in a typical iterative work process. The final product will therefore probably be more elaborate than what is specified in the MVP, but should in essence consist of the same thing: Demonstrating efficacy of machine learning based gesture recognition and its applications during video meetings.

The intended demonstration should, as mentioned, utilize machine-learning to detect hand gestures. Specifications for this follows.

5.2 Deep learning model specification

Building a deep learning based model for classifying hand gestures is the main focus of this project. Through using established machine learning techniques and technologies, the students should develop, train and evaluate a model that fulfills this purpose. The end goal for this process is both the model and documentation on how it has been developed, how it performs, what technologies it uses, etc.

5.3 Alternatives

There are many alternatives for getting to the end goal in the process, such as different model architectures, datasets, gestures, ML-libraries, etc. Identifying which of these to use should be considered part of the process and developing the optimal product.

I has been decided that the students should use more general models as a basis for further development. This is in contrast to models specialized for areas such as hand-tracking, which are readily available. If we were to use a specialized model from reputable sources such as google, no doubt the finished product would be more accurate and optimized then we could dream of in the scope of this project. However, since an integral part of the project is to document the development of the model, using general models as a basis is more appropriate.

6 Development schedule

Rough schedule:

Week	**Main focus**
1 - 4	Research regarding gesture recognition, existing solutions and datasets. Getting familiar with appropriate tools and libraries.
5 - 9	Development of an MVP. We estimate here that the first version of the product will take longer to develop, and have therefore allotted more time for this then succeeding cycles.
10 - 15	Improving and adding to product through iterative process
16 - 20	Documentation and project report

Obviously the main focus is here meant as a general guideline. It should also be noted that we take aim on documenting the development process throughout all of the project. Generating and writing tests will also be regarded as a natural part of the development process.