UPPSALA UNIVERSITET

Intelligent Interactive Systems
Projects

In the project you will develop an intelligent interactive system that can detect the six basic emotions (anger, disgust, fear, happiness, sadness, surprise) and respond with artificially synthesized emotions. The system consists of three components: (1) *The recognition of facial landmarks* (Computer Vision), (2) the *attribution of emotional state based on a given set of facial landmarks* (Machine Learning), and (3) the *expression of emotions in an artificial agent* (Emotion Synthesis).

You will work in **primary groups** (3 – 5 students) on one of the three parts of the system. Detailed descriptions of each sub-task are given below. In addition, each primary group is matched with two other groups working on the other parts of the system in a **secondary group.** The goal of the secondary group is to show the full pipeline of the interactive system as a live demo in the final presentation: The system should automatically detect the position of the facial landmarks from a RealSense video stream and feed the landmarks to the machine learning algorithm which outputs the emotion the user is most likely expressing. The embodiment of the agent then reacts to the input by showing an emotional expression itself.

Grading is only based on how the primary group's work performs against the goals set in the individual project proposals and thus the majority of your time should be allocated to this sub-task. However, in order to pass the project, you have to engage with the other members of the secondary group to get a pipeline for the full system and reflect about this joint work in the final project report. It is your responsibility to jointly work on defining and providing required interfaces, not to fix the other group’s bugs.

The final report shall also contain an elaboration about potential ethical concerns you identified with this emotional responsive agent.

1) Computer Vision (CV)

Your task is to automatically detect the position of facial landmarks in a human face using a RealSense RGB-D camera. Start by collecting a dataset of at least one person showing the six basic emotions in front of the camera and manually mark the facial landmarks in the frames. Look into the literature which set of facial landmarks is usually used to detect certain emotions. You do not need to detect all possible landmarks in the face, you can choose to use a meaningful subset for your task. However, remember to agree on the subset with the Machine Learning group because they need to train their models on the same subset that you are providing. Once you have your dataset collected, train a model that can automatically detect those landmarks and output the position in a given video frame. One approach to this problem is to start from an initial approximation of where the landmarks are in a neutral face. However, you can also suggest other methods in your project proposal. Once you have your algorithm optimized so the detection and location accuracy is reasonably good for a single user, you can extend your model to work for different faces.

You will be graded on how sophisticated the Computer Vision techniques are, the accuracy of your facial landmarks and the discussion and explanation of the obtained results and ethical implications of the emotional responsive agent. In addition, you have to provide a representation of the facial landmark positions that can be used by the Machine Learning group.

Milestone (second feedback session)

In the second feedback session you should have your dataset collected and a very simple version of your pipeline set up: Your system should be capable of generating a reasonable set of landmarks (they need not be accurate) that can be given to the machine learning group via the agreed interface. Further, you should be able to visualize these landmarks on a live stream captured by the RealSense camera.

2) Machine Learning (ML)

Your task is to assign one of the six basic emotions to a set of 3D facial landmark positions. You will be given a labeled dataset to train and test the system on, but you need to pre-process the data to extract important features from the dataset. Try different feature selection methods and compare at least four classifiers to solve this problem. First, replicate the two classifiers from assignment 1 part 1. As a second step, pick two other machine learning techniques (e.g. random forests, multi-layer perceptron based classifiers, etc) and try do understand how they work and what output measures they provide. Evaluate your results and choose the classifier to perform in the final live demo based on the obtained results.

You will be graded on how sophisticated the used machine learning techniques are, the performance of your classifiers on the given dataset and the discussion and explanation of the obtained results and ethical implications of the emotional responsive agent. In addition, you have to be able to read data from the interface of the Computer Vision group (the performance of those provided landmarks is **not** part of the grading) and provide an interface to the Emotion Synthesis group to pass the project. The interface you provide as an output should not only contain the selected label but also the associated probability scores computed by the classifier.

Milestone (second feedback session)

In the second feedback session you should be able to read in facial landmark positions as input, extract the features and have the two basic classifiers from assignment 1 part 1 working that selects the correct emotion with a higher than chance probability. Your classification result shall be published in a format that can be used by the Emotion Synthesis group.

3) Emotion Synthesis (ES)

Use the virtual 2D version of the NAO robot and the IrisTK virtual human to express the six basic emotions. You can use the body posture for the virtual NAO and facial expressions for the virtual human. In addition, you may use speech output and / or sound effects for both systems. In order to evaluate your ability to show emotions, make a small user study and evaluate how well participants can recognize which emotion the agent is expressing. Investigate if the recognition rate, perceived intensity and naturalness is the same in both embodiments for each emotion. Choose the embodiment for the final presentation based on the results in this first part of your project.

Your second task is to define a simple reaction schema to emotional input from a human user. Depending on which emotion the user is most likely showing and the confidence that this emotion is actually displayed, come up with an agent's response based on your synthesized emotion. It could be mimicking the same emotion, or another sensible schema, like reacting sad if the user is angry.

You will be graded on how sophisticated your synthesis approach, user evaluation and reaction schema is, as well as the analysis, discussion and explanation of the obtained results and ethical implications of the emotional responsive agent. In addition, you have to be able to read data from the interface of the Machine Learning group (the performance of the provided emotional input is **not** part of the grading).

Milestone (second feedback session)

In the second feedback session you should have at least two emotions synthesized in both embodiments. In addition, you should have the dialogue pipeline working that takes an emotion as an input and produces an emotional expression in response. Which response to choose can still be hard-coded, it's important that you can show that the pipeline is in place.