

Report on the project "development of software for capturing human movement and facial expressions"

Summary

At the time of the creation of the project, there were no ready-made video processing solutions for creating a mask for superimposing animations, and therefore our project was born.

The "Client-Server-Computing Machine" architecture was invented, which is quite comfortable and understandable to use for an ordinary person who does not know how to program, as well as who does not have a powerful computer that could perform all calculations. The sequence of actions of the client should have been as follows: the client shoots a video, it is sent to the server along with instructions on exactly what the client wants, the server sends the video to a computer that does the process of applying the mask, after which, in reverse sequence, the video is sent to the client with the mask applied, in addition, it sends a special file with parameters for overlays any animation he wants.

The tasks were distributed according to three teams, in three directions. This number of commands is determined by the architecture. I was on the team of the Computing Unit, so next we will talk exclusively about it.

The results met expectations and it turned out to realize almost everything planned.

Project Overview

The idea of creating this project appeared due to the lack of a full-fledged service that supports easy connection for any client, since all existing ones implied a good knowledge of programming in different languages.

Initially, it was thought that the client downloads a file to his computer, according to the instructions in the project repository, after which he can connect his phone's camera and receive a 3D model online, while there are no requirements for his computer.

The proposed solution

In general, our project contains three directions: a visual, a server and a computer, I worked on the latter, that is, I took part in creating a program that receives video input, which should subsequently be cut into frames, then a mask is applied to each frame, judging by the key points, and again everything is assembled into a video.

The tasks that were assigned to me:

1. Study of libraries and created projects necessary for the realization of our task
2. The development of this program at the initial stages, namely the selection of models for processing simple videos with the participation of one person
3. Testing the first version
4. Creating a modification for processing complex videos involving multiple people
5. Testing the operation of the final version

Now for each task:

1. Libraries

When it comes to implementing a new project, of course, first of all it is necessary to study existing programs, which is where I started.

The EasyMocap project was found. This project contained a lot of different modes, many of which were not fully implemented, for example, there were modes where individual body parts were taken, or searching for a person in a photo and applying a mask, and all that, but there was no full-fledged mode for processing the entire human body. So I had to select the desired mode and study all this code.

2. Initial stages

At this stage, it was necessary to develop a mode that would process at least one person in average quality, which was achieved, as a result, this mode was called bad.

3. Testing the first version

The bad mode performed well on simple videos where only one person was present and at the same time his behavior was predictable. Testing has shown that if you increase the number of people, then the superimposed mask begins to jump between people in the video and behaves very unstable, and if a person begins to behave unstable, quickly and abruptly move, for example, his hands, then the mask may not be in time and there will be a slight delay.

4. Creating complicated modes

Subsequently, the following modes were created:

- best mode for processing complex videos. Complex videos are videos containing a large number of people and unpredictable movements.
- Good - the average mode. Videos of medium difficulty

Also the bad mode, created initially.

5. Final testing

The final testing showed that the modes are well chosen, the best behaves stably even on videos where the number of people exceeds 20 people and correctly monitors the selected one, even if everyone is dancing, there was such a video. Good tracks the character well if there are 3 people. Bad only if there is one person.

In general, we can say that there is a suitable video for any video client.

Conclusions

As a result of the work, it turned out to realize almost everything planned, except for the online shooting mode, at the first stages it turned out that this was not possible. As a result, it was done so that the client first needs to take a video, send it, and then he receives a video with an overlay mask in response. In addition, everything is implemented as planned!

The team and I also decided that we will continue this project and in the future we will make a full-fledged application for this program and rent a computer with more powerful parameters that will assemble the model many times faster than it is now. This is not available on the market and I think it will be very useful and in demand. This report described my participation and my tasks, but the project was quite large and included many other things.