Markuss Davis Skrastins

Formative draft outline of background and literature review

Direct interaction of a mobile phone and a PC display

Background

For multiple decades there have been propositions for a simple system, that would allow to operate a computer by manipulating a live video feed that is taken from a camera of another device. Even though such propositions date all the way to 1990s, no such systems have been observed to fully exist in either industrial, commercial, or private sectors. Such systems have been researched and implemented multiple times during 2010s, however there still doesn't seem to be any popular use for them. However, today it is almost a given, that everyone has a smartphone with a good quality camera unlike 20 or even just 10 years ago. Furthermore, more public (airports, public transportation, malls etc.) and private places have interactive displays than ever before.

A system that would allow to control a computer from any device by just pointing the camera at a computer's display would give a user a full control of any computer. This would increase users' ability to interact with any system without having a direct access to the physical unit easing the usability in such cases as presentations, interactive displays in public spaces and even a personal computer. Moreover, combining this with an ongoing pandemic, it would also make interactions with any public systems (ATMs, interactive navigation displays etc.) safer, as it would be possible to use them without a physical contact.

In this project, such a system and its implementation are researched. Although one of the challenges for such a system to be widespread is the connectivity between a display and a phone, the focus of this paper is the interaction between a display and a phone- matching video feed from a camera to the display and being able to control the computer from the phone. Throughout the project, the most accurate model will be chosen, and its strengths and weaknesses will be compared to different implementations of similar systems. This project will expand on existing ideas of pattern and feature matching; however, a convolution neural network (CNN) will also be implemented and the two methods will be compared.

Literature review

There have been multiple different attempts over the years for an easy and intuitive use case of interacting with any kind of display. Although the objective throughout multiple systems tends to vary, the main problem that is faced is feature matching.

Throughout the literature it can be observed that the approach for feature-matching is similar. In all approaches, the device, whose display is being captured, receives the incoming bit-stream from the camera and calculates the homography- the transformation between the camera and display image planes. However, since there are many different feature extraction techniques, a variety of them have been used and compared. For example, in the research paper [1], the technique used was Hough transformation, while in [2] two different scale-invariant technique were used and compared- SIFT and SURF. One of the main challenges that was observed was feature-matching when the camera is not completely parallel with the display. Pan, tilt, and roll can significantly influence the result, as they represent a real case use more accurately.

Most of the literature reviewed is about a decade old. This also can be noticed when some technical parameters are mentioned- the phone models, their base performance and quality/resolution of both camera and the display. This further suggests that some of the issues that were addressed in some of the literature could simply be eliminated simply because of technological advancements- phones are much faster with better resolution cameras. The same can be said about displays; although displays with resolution of 1080p are still the majority all round us, higher resolution displays, such as 4k, have become more prominent.

In conclusion, there are many different methods for doing a feature-matching based interactive system. However, given how much the average performance and quality of both cameras and displays have improved since the given literature has been published, it could pave a way for a more efficient feature-matching based system.

References

[1] S. Boring, D. Baur un A. Butz, «Touch Projector: Mobile Interaction through Video,» *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2287-2296, 2010.

[2] L. Herbert, N. Pears, D. Jackson un P. Olivier, «Mobile device and intelligent display interaction via scale-onvariant image feature matching,» *PECCS 2011 - Proceedings of the 1st International Conference on Pervasive and Embedded Computing and Communication Systems*, 2011.

[3] N. Pears, D. G. Jackson un P. Olivier, «Smart phone interaction with registered displays,» *IEEE Pervasive Computing*, sēj. 8, nr. 2, pp. 14-21, 2009.

[4] J. She, J. Crowcroft, H. Fu un P.-H. Ho, «Smart Signage: A Draggable Cyber-PhysicalBroadcast/Multicast Media System,» *IEEE Transactions on Emerging Topics in Computing*, pp. 232-243, 2013.