Pervasive Interaction – Using movement and speech to interact with computers

Thomas Riisgaard Hansen, Jakob E. Bardram, Mads Søgaard

Pervasive Healthcare
Department of Computer Science, University of Aarhus
8200 Aarhus N, Denmark
{thomasr, madss, bardram}@daimi.au.dk

ABSTRACT

In this paper we will focus on pervasive interaction that allows the users to interact with computers when needed and not just when they are sitting in front of a computer. Inspired by the work done in hospitals we describe several projects and frameworks that address pervasive interaction. We focus on two types of interaction: Movement based interaction and speech based interaction. Finally, we suggest that the role of *context* and *uncertainty* would be relevant to elaborate further on at the workshop.

Keywords

Pervasive interaction, movement based interaction, speech based interaction, mixed interaction spaces.

INTRODUCTION

Pervasive interaction is interaction with computers that takes place at the time and place decided by the use context and the interaction is not limited and dictated by the presence of a personal computer.

The mouse and keyboard have for a long time been the only mean of communicating with a computer, but both interaction devices require the presence of a flat surface, a desk. Moving from desktop interaction to more pervasive interaction, e.g. interaction in busses, trains, cars, hospitals, and when walking or running, the assumption that the user is sitting or standing at a flat surface breaks. A number of new interaction techniques have appeared to support this non-desktop situation. Mobile phone interaction, pen based interaction, large wall display interaction, augmented interfaces, and tangible interfaces are just some examples of interfaces that promise to carry interaction away from the desktop.

In this paper we focus on movement based interaction and speech based interaction, two non-desktop interaction technologies. We present a set of projects and frameworks we have developed and the main contribution of the paper is to highlight the possibilities for pervasive interaction with these two interaction technologies.

Our main focus area is interaction with computers in hospitals and within the hospital we have been working with two scenarios. How two interact with computers while been mobile and how two interact with computers during surgery.

MOVEMENT BASED INTERACTION

Movement based interaction is interaction where the movement of the human body is used as input to the computer. In this section two projects are presented that uses movement.

In the first project we focus on interaction with large wall displays. We track the location of both of the user's hands and overlay the image of the tracked hands as a dual cursors.

In the project mixed interaction spaces we use the camera in a mobile phone to track the position of the device in relation to a feature. The movement of the device is thus registered and used as input to a large variety of different applications.

Finally we have been working on a framework for movement based interaction that uses cameras as tracking devices.

Hands2Cursor

In the first project we worked with how a surgeon can interact with computers while operating.

Figure 1: A typical operation room setup

Figure 1 is from a typical surgery. The surgeons are sterile and situated next to the patient. How can you interact with a computer in this situation? In a first attempt to address this situation we look at gesture interaction. Would it be possible for a computer to recognize the surgeon's hands and if so could this movement information be used to control a user interface?

Based on this initial research question we developed a system based on a PC and a webcam. We have worked with a number of different tracking algorithms e.g. different variants of Hough algorithms, different edge and feature detection algorithm, flood fill algorithms and the CamShift algorithm. The outcome was a system that found the position and rough rotation of both of the user's hands under most conditions.

Figure 2: Hands interaction with a large wall display. The tracked hands are projected on the interface.

However, a major finding (not surprisingly) from this work was that it was really difficult to find a robust algorithm that performed well in all conditions. Therefore it was really important to visualize to the user the performance of the system. Did the system work as planned or had it lost track of the hands? We worked with an approach where we projected the images of the tracked hands as the computer saw them on top of the interface at the position where the computer thought they were. This approach proved to be rather successful because it provided the user both with a feeling of actually manipulating the interface with the hands and at the same time visualized how well the system was able to track and locate the user's hands.

After a period of time we paused the project temporary because other interaction techniques, e.g. speech as described in the next section, seemed more suitable for this type of environment. However, some of the findings were transferred to a mobile platform and used in the next project. A video of the tracking system and further information is available at [2, 6].

Mixed Interaction Spaces

With the project Mixed Interaction Spaces we looked at how to interact with computers while been mobile. A mobile device is required to fit into the pocket of the user, which limits the possibilities for interaction. However, at the same time a mobile device is almost always present and at hand. In the mixed interaction space project we looked at how to use the fact that a device is mobile to facilitate new ways of interactions. In the project we use the embedded camera most mobile devices are equipped with to track one or more features. By tracking a feature with the mobile device we get a fixed-point or reference point we can use to determine the position of the mobile device and its movement in relation to this feature.

In the current version of the prototype we have three different algorithms running on a Nokia Series 60 platform. We have a variant of Hough circle detection algorithm for tracking printed or hand drawn circle and we have two algorithms based on the CamShift algorithm for tracking either a colored object or the users face.

Figure 3 shows what happens when the face tracking algorithm is used (this is only possible on mobile phones with a camera pointing towards the user). The face works as a fixed point. By looking at the size of the face the mobile device can determine if it is being moved closer or further away from the face. The algorithm can also determine if the phone is moved up or down, left or right and to some extend rotation. This is however, only possible if the face keeps within the visible range of the camera. The space in which the phone can track the feature is highlighted on figure 3 (not to be confused with the cameras field of view).

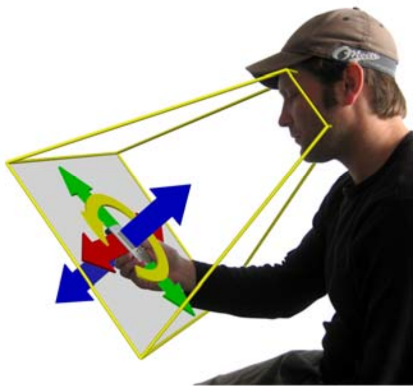

Figure 3: The Mixed Interaction Space with face tracking

Based on this technique we have developed several different applications. Some of the applications run on the phone and use the interaction technique to pan and zoom on maps or images or use it as input to different games. Other applications run on a nearby large wall display where e.g. a cursor can be controlled with the interaction technique. Bluetooth is used to transmit the input information to the large wall display and we have developed different multi-user applications that also use this technique as input.

A typical use scenario is that a user starts the program, connects to e.g. a large wall display, look for a feature (e.g. the top of a Coca-Cola bottle as in figure 4), and use this feature as fixed point.

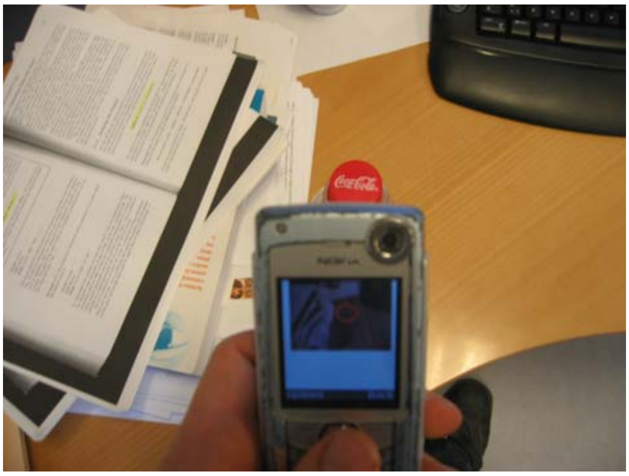

Figure 4: The top of a Coca-Cola bottle is used as reference point for the tracking application.

The main findings in the project is described in [3, 4], however the technique seems really promising. It works really robust and you can use any nearby object, draw a circle or use your own head to interact with the system. Besides most people carry a mobile phone and using it as an interaction mediator between interfaces in the world allows the interface to remain simple and robust while allowing several users to hook up to the interface with their own devices.

Movement based framework

Based on the above present projects and a range of other camera based applications we have identified three important concepts and build a framework around these three concepts.

The three main components are space, relation and feedback.

Space: We found all the projects that used cameras to track the movement of the user somehow worked with the notion of space. The scale, orientation and the number of used spaces seemed to be both characteristic and important. Some of the projects use large spaces that are able to contain multiple users, while other e.g. the space between a mobile phone and the users face is much smaller. Some uses the floors as main interaction areas while others use walls etc.

Relations: When a camera tracks a feature, a relationship can be described between the camera and the feature. Depending on the tracking algorithm multiple features can be tracked thus spanning multiple relations. Also the information an algorithm can extract from a relation can differ. Some algorithms are e.g. only able to identify colored objects while others are able to uniquely identity the different features through 2D barcodes.

Feedback: Finally, as hinted in the section Hands2Cursor, camera based interaction rely heavily on user feedback. Being able to describe how, what kind and where feedback occur is important. Is an application using auditory feedback or visual feedback? Is the information presented on a hand held device or a shared surface? How is multi-user feedback addressed?

We have used the above summarized framework to describe and analyze nine very different camera based interfaces. The full framework and the analysis can be found in [1].

SPEECH BASED INTERACTION

Speech based interaction is another area we have been working heavily with. As with your body, speech is always there and in many situation where your hands are occupied it is the only alternative. Speech interaction has been around for decades however, the performance and accuracy have been greatly improved lately and speech recognition is now also available on even handheld devices.

ActiveTheatre

We have been working extensively with speech interaction in the project named ActiveTheatre. The main problem we address in this project is the same as with the Hands2cursor. How do you interact with computers while operating? Speech seemed much more promising than gestures for interacting while operating. When operating the surgeon normally use his hands for operating and multiple users occludes the cameras field of vision.

The idea behind the active theatre project is to use large wall display and speech interaction to recall medical images and data from the electronic patient record while operating. However, we wanted to not only facilitate accessing the data, but also actively creating new data while operating. Examples would be to capture video and images of the operation or using speech recognition to access medical information.

We have build and evaluated several different prototypes that use speech as input. Figure 5 shows a surgeon from a nearby hospital trying the system out.

Figure 5: A surgeon is testing the interface. The interface runs on both a large wall shared display and an arm mounted 17" inch screen.

In the application we use two different modes of speech recognition. We use speech commands to control the application e.g. to zoom in on pictures or text, to take pictures and to navigate in the interface. The speech commands are based on a predefined syntax. The second mode is free dictation where every word is recognized. We use this e.g. for annotation the pictures or for writing notes.

In the first versions we use Microsoft Speech Recognition engine as a proof of concept, but are currently working with a Danish speech recognition engine focusing on the medical domain based on Phillips Speech Engine.

We are currently developing the third version of the system, which is going to be tested during real life surgeries later this spring. The project is further described in [5].

CONCLUSION

Pervasive interaction requires novel form for interfaces that supports new use situations. There are many forking paths that move forward. In this paper we have presented some of the work we have been doing focusing on movement and speech interaction.

Many of the findings are published elsewhere, however there are some common themes in the above presented work that could be interesting to discuss at the workshop.

Context: When moving away from the desk into the real world the number of different contexts explodes and the importance of taking the context into consideration is getting more and more important. What are the important findings with novel interaction techniques that allow them to work both in operating theatres, school, busses, while running etc. or is the question not to find cross context principles, but to find specialized interaction techniques that are tailored to a specific work setting?

Uncertainty: Many of the novel interfaces that appear use sensors with uncertainty associated with them. How do you handle uncertainty in the interaction technique? Movement based systems are not always able to recognize your gesture and speech recognition system might hear wrong. Maybe the computer heard what you said, but that was not what you actually meant. How to handle uncertainty with interaction could be another interesting topic.

REFERENCES

1. Eriksson, Eva, Hansen, Thomas Riisgaard, Lykke-Olesen, Andreas, Space, Relation and Feedback - describing, discussing and analysing movement based interaction using video tracking, submitted for publication.
2. Hansen, T.R., Interaction without Keyboard and Mouse, UbiHealth workshop, Ubiquitous Computing Conference 2004, Nottingham, England.
3. Hansen, T. R., Eriksson, E. & Lykke-Olesen, A., Mixed Interaction Space — Designing for Camera Based Interaction with Mobile Devices, in CHI'05 Extended Abstracts of the Conference on Human Factors in Computing Systems, ACM Press, pp.1933–6, 2005.
4. Hansen, Thomas Riisgaard, Eriksson, Eva., Lykke-Olesen, Andreas, Mixed Interaction Space – Expanding the Interaction Space with Mobile Devices. In Proceedings of British HCI 2005, 2005.
5. Hansen, Thomas Riisgaard, Bardram, Jakob E., ActiveTheatre – a Collaborative, Event-based Capture and Access System for the Operating Theatre, accepted as full paper at UbiComp 2005.
6. Videos at Pervasive Interaction.org: http://www.pervasive-interaction.org/videos/