Report

Wearable Gestural Interface

Master thesis

July 2009 - December 2009

Matthias Schwaller

Professors:
Elena Mugellini (EIA - FR)
Omar Abou Khaled (EIA - FR)
Rolf Ingold (UNIFR)
Abstract: This report describes my MasterThesis of the University of Fribourg. The aim of the project was to develop a Wearable Gestural Interface. It is a wearable tool on which the user navigates with hand gestures. The project was done under the collaboration of the University of Fribourg UNIFR and the University of Applied Sciences of Western Switzerland, Fribourg EIA-FR.

The Wearable Gestural Interface uses color segmentation and a Morphology operators for the preprocessing. For the tracking the Condensation algorithm was used. The recognition step was done with The Gesture and Activity Recognition Toolkit (GART) which uses The Hidden Markov Model Toolkit (HTK).
Contents

List of figures 5

I Introduction 6

1 Context 6

2 Problem description 6
   2.1 The existing situation 6
   2.2 Goals 6

3 Report structure 7

II State of the art 8

4 Augmented Reality 8

5 Wearable Computing 9
   5.1 Challenges 9
   5.2 Operational modes of wearable computing 11
   5.3 Attributes 12
   5.4 Discussion 12

6 Gesture recognition 13
   6.1 Image acquisition 13
   6.2 Preprocessing 14
   6.3 Tracking 15
   6.4 Gesture recognition 16
   6.5 Projects 18

7 Libraries 22
   7.1 OpenCV 22
   7.2 Emgu CV 22
   7.3 Hidden Markov Model Toolkit 22
   7.4 Gesture and Activity Recognition Toolkit 23

8 Discussion 23
   8.1 Image acquisition 24
   8.2 Preprocessing 24
   8.3 Tracking 24
   8.4 Gesture recognition 24
   8.5 Projects 25
9 Conclusion 25

III Design 26

10 Emplacement of the equipment 26
  10.1 Camera 26
  10.2 Beamer 26

11 The features of the application 27
  11.1 Application 27
  11.2 File browser 27
  11.3 Once the file is open 27
  11.4 PowerPoint in full screen 28
  11.5 Hand gestures 29

12 Hardware architecture 31
  12.1 Beamer 32
  12.2 Camera 32
  12.3 Ultra mobile PC [UMPC] 33

13 Software architecture 34
  13.1 WGI architecture 34
  13.2 InterFace architecture 37

14 Conclusion 39

IV Implementation and prototype 40

15 Hardware 40
  15.1 Camera 40
  15.2 Beamer 40
  15.3 UMPC 41
  15.4 Finger markers 41
  15.5 The complete wearable tool 43

16 Wearable Gestural Interface 44
  16.1 The principle class Gesture Recognition 44
  16.2 Touch mode and gesture mode 44
  16.3 Preprocessing 44
  16.4 Tracking 46
  16.5 Gesture segmentation 47
  16.6 Gesture identification 47
  16.7 Gesture execution 48
16.8 Gesture training ........................................ 48

17 Including WGI in InterFace .......................... 49
17.1 HandGestureManager ................................. 49
17.2 Modifications in InputManager ................. 49
17.3 GestureManager ..................................... 49
17.4 Presenter ............................................ 50

18 Results and prototype ............................... 51

19 Tests ..................................................... 51
19.1 Test for the number of gestures ................. 52
19.2 Test for the number of skip states ............. 52

20 Conclusion ............................................. 53

V Conclusion and future work ........................ 54

21 Occurred problems .................................. 54

22 Improvements and Future work .................. 55

23 Personal impressions ............................... 56

24 Gratitude .............................................. 58

VI List of abbreviations ................................ 59

VII Bibliography ......................................... 60

VIII Appendix ............................................. 63

25 Installation manual .................................. 63

26 Specifications ......................................... 63

27 Planning ............................................... 63

28 Comparison .......................................... 63

29 Tests ................................................... 63
List of Figures

2. HSV color space [19] .................................................. 14
3. The sixthsense [28] ..................................................... 19
4. Framework InterFace [29] ............................................. 20
5. The project VICI [31] .................................................. 21
6. Natal project [33] ........................................................ 22
7. Data collection and gesture recognition of GART [40] ........ 23
8. File browser ............................................................... 28
9. MS PowerPoint .......................................................... 28
10. MS PowerPoint in presentation mode ................................ 29
11. Gesture for clicking .................................................... 30
12. Gesture for opening a presentation in full screen (F5) ........ 30
13. Gesture for closing a presentation (ESC) .......................... 31
14. Gesture to go to the next/previous slide ......................... 31
15. Gesture to send a file to another person ......................... 32
16. Beamer: 3M MPro 110 ................................................. 32
17. QuickCam Pro for Notebooks ......................................... 33
18. Sony Vaio UX micro PC ............................................... 33
19. WGI architecture ....................................................... 35
20. InterFace architecture ................................................ 37
21. InterFace architecture with WGI ................................... 38
22. Camera with fixation .................................................. 41
23. Joby Gorilla Pod ......................................................... 41
24. Beamer with fixation .................................................. 42
25. Finger markers ......................................................... 42
26. The complete wearable tool from the front .................... 43
27. The complete wearable tool from the side ..................... 43
28. Histograms ............................................................... 45
29. Color segmentation ................................................... 46
30. Test application for gesture recognition ......................... 47
31. Graphic component PowerPointFile .............................. 50
32. Graphic component file browser (ExplorerFrame) ............ 50
33. The Wearable Gestural Interface in use ......................... 51
34. Test: number of gestures ........................................... 52
35. Test: number of skip states ....................................... 53
36. SPY sunglasses with integrated camera .......................... 57
37. LG eXpo, cell phone with integrated projector ................ 57
Part I
Introduction

1 Context
The students at the University of Fribourg in Informatics conclude their Masters degree with a Master Thesis. The project was done under the collaboration of the University of Fribourg UNIFR and the University of Applied Sciences of Western Switzerland, Fribourg EIA-FR. The duration of the project was twenty weeks.

2 Problem description
These days, laptops and smart phones are indispensable. These gadgets are not only used for work in the office but also to get information on the way. When a user would like to visualize information, he has to take the gadget out of the pocket and navigate in the menu.
The idea of this project is to simplify this kind of tasks. Therefore a smart tool is developed. It permits to project the information in front of the user either on a wall or on another object. The user can navigate with the tool via hand gestures. This makes the usability of the tool more natural and simpler to use, since hand gestures are often used by humans. However the tool is wearable so the user does not have to take it out of the pocket.
In order to bring the web closer to the user a wearable tool which permits to show additional information on objects goes in the right direction.

2.1 The existing situation
Nowadays, with the IPhone, Android and other touchable devices the touch mode becomes more and more used. But a big disadvantage of those tools is that the interface requests to take the tool out of the pocket. In order to be more comfortable for the user, this disadvantage needs to be eliminated.

2.2 Goals
In order to develop the Wearable Gestural Interface some goals have to be achieved. First the state of the art has to be analyzed in order to know what exists and what will be possible. Further the possible hardware has to be searched. As a next step an application has to be found which is able to demonstrate the wearable gestural interface. Once the above goals are achieved this Wearable Gestural Interface has of course to be developed. In order to show the functioning of the Wearable
Gestural Interface a demonstration application has to be implemented. This application has to include the Wearable Gestural Interface.

3 Report structure

The report starts with the state of the art which contains Augmented Reality, Wearable Computing and gesture recognition. Further, the report continues with the design. The design includes the emplacement of the equipment, the features of the application, the hardware architecture and software architecture.

In the part implementation and prototype the hardware is presented. It is also shown how the Wearable Gestural Interface is implemented. Further it is illustrated how the Wearable Gestural Interface can be integrated in InterFace. As next the results are presented. The report finishes with a conclusion and the possible future work.
Part II
State of the art

This chapter presents the research part of this project. Since the final product will be worn by the user, one of the main topics in the research is Wearable Computing. The user will further use hand gestures to use the product; therefore the second big research topic is Gesture Recognition.

First the topic Augmented Reality will be presented before the above main topics are presented. At the end of the Gesture Recognition section, some projects will be exposed. Before the conclusion of this chapter is presented a discussion about the treated subjects will be illustrated.

4 Augmented Reality

This project will augment the environment of the user by additional information. If the reality is augmented by additional - mostly visual - information, one speaks about augmented reality.

In the augmented reality systems, virtual images are overlaid upon a normal view of the real world. Unlike the virtual reality, in which the user is completely in a virtual world, in augmented reality the display of additional information is in the foreground.

Augmented reality in [1] is defined with the three following characteristics:

1. Combines real and virtual
2. Interactive in real time
3. Registered in 3D

Since the idea of this project is to display 2D information on flat objects, this project will not use 3D. As a future work on this project 3D is imaginable. More about this topic you find in the section [22] Improvements and Future work.

The applications in this domain are very various. It may be a walker which goes to the mountains and would like to know the names of the mountains like in figure [1]. Another application may be a relief worker which can see information about dangerous zones. [2] [3]

Another very impressing Augmented Reality tool is the software from http://www.wikitude.org/ which permits already the above example for the Android software on mobile phones.
5 Wearable Computing

Wearable Computing is the area of research which deals with wearable computers. A wearable computer is a computer which is worn on the body. So it is a computer that is subsumed into the personal space of the user. The device is controlled by the user and always with the users. This means that the device is always on and always accessible.

5.1 Challenges

Designers have to rise to significant challenges. The biggest challenges are now presented in this chapter. Each of these challenges is closely related to the others, and a design change to minimize a challenge does often affect the other challenges.

- Power use
  This is maybe the most limiting challenge in wearable computing. Often the weight of the battery is bigger than the weight of the electronics of a wearable device.
  If the battery life of the wearable device is too short, the user will soon get frustrated, because the effort to keep the system powered is too big.
  If the worn devices are spread over the body the power supply of a wearable computer becomes more complicated. Often each device has its own battery.

- Heat dissipation
  The heat dissipation is the companion problem of power use. Devices which are in contact with human skin should not exceed 40 °C. Normal desktop computers get hotter. So wearable computers have to use less watts, but according to a 1998 study, processors exceeding the 40W range cost an additional US dollar per watt per chip.

- Networking
  Most of the services are much more powerful if they have network.
In wearable computing, "bits per second per watt" is often the more meaningful measure than the maximum throughput. Since the networks for wearable computing are mostly wireless networks, there will always be some zones in which the signal is not available or just too less powerful to use the service.

- **Off-body communications**
  Off-body communications are communications from the mobile device to a fixed infrastructure. In wearable devices the antenna is often small; therefore the range of the signal is not so strong. This communication is the most researched.

- **On-body communications**
  On-body communications uses less energy than off-body communications. Energy use becomes critical because each device must have its own battery.

- **Interoperability**
  If the user need to access more than one wireless services, he has to wear extra equipment. Downloading the appropriate software can change its communication standards and protocols.

- **Communication with near-body object**
  Most of the standards which are used for near-body communication assume that the device has access to a significant energy supply.

- **Privacy**
  It is important to mention that user privacy concerns are not equivalent to security concerns. It is the right of each individual to control their personal information. Security involves the protection from unauthorized users to information.

- **Interface design**
  The interface is a human and computer interaction. When designing it, the factors of human-computer interfaces, psychophysics, human factors, ergonomics, industrial design and fashion has to be taken into account.
  The peripheral interfaces have to be designed in a way which makes simple things simple and complex things possible.

- **Intellectual tools**
  A big challenge in Wearable Computing is to create systems that augment a user’s natural abilities through computational components. An intellectual tool therefore should provide information support while the user is concentrated on his primary task.
  A wearable should be able to retrieve the context in which notes or
a task was done. This may help for indexing and also afterwards the user can easier find his information.
The wearable should provide just-in-time information. When you look around on a mountain for instance and you would like to know the name of this mountain, the information has to be just-in-time. Otherwise the names of the mountains will be always a bit too late and you are already looking to the next mountain.

5.2 Operational modes of wearable computing

The three operational modes of wearable computing are:

1. Constancy
   The System is never dead, it may sleep but it is always ready. This permits to have a constant user - interface which runs continuously. Constancy systems do not have to boot in order to use. Unlike a computer.

2. Augmentation
   The mode augmentation extends the real world of the user. The user will now do 2 things at the same time. The main task of the user is to do something else than computing and at the same time he does the computing.

3. Mediation
   The concept of the Mediated Reality is to encapsulate the user fully or at least partially. It exists two aspects of this encapsulation:

   (a) Solitude
       It can function like an information filter and therefore block out material we may not wish to experience. It may allow us to vary our perception of reality in a very mild sort of way.

   (b) Privacy
       It can allow us to block or modify information leaving our encapsulated space. Therefore the wearable computing may used to create a new level of personal privacy. Since the personal computer is always worn, it is more difficult for an attacker to access the hardware.
5.3 Attributes

In wearable computing one speaks about six information flow paths or attributes. These attributes are the following:

1. Unmonopolizing
   The user may attend to other matter while using it. So the user is not cut off from the outside world. The computing is therefore seen as a secondary activity.

2. Unrestrictive
   This attribute means that you are totally mobile. So "you can do other things while using it".

3. Observable
   The device can get the attention continuously of the user if he wants it. The output medium is constantly perceptible by the wearer.

4. Controllable
   The user can grab the control of the wearable devices at any time he wishes.

5. Attentive
   A wearable computing device has to be attentive to the environment. It has to be multimodal and multisensory.

6. Communicative
   The device can be used as communication medium. A user is expressive through the medium.

In the article of Mann it is explained that additional to the above 6 attributes a wearable computer has to be:

- Constant
  This attribute describes that the wearable devices is always on. It does not have to be opened and switched on like a laptop.

- Personal
  "Human and computer are inextricably intertwined."

5.4 Discussion

This subsection concludes the part Wearable Computing of the state of the art with a discussion. Since the prototype will be a wearable tool the challenges of Wearable Computing does also count.
The first challenge, power use, will be solved the following way. First we need a small beamer. Since all these small beamers have already a battery, therefore a beamer with a good battery lifetime has to be chosen. For the camera a webcam is sufficient. A webcam uses only a USB port which will be connected to the ultra mobile PC (UMPC). However also the UMPC has to have a good battery in order to permit a long lifetime. Since the beamer and as well the UMPC can become warm, they have to be worn on the body and cannot be putted in a pocket.

The communication which will be used is an off body communication. The camera and the beamer will be connected by cable with the UMPC. The UMPC will communicate with WLAN or with other WGI tools. In order to permit to make simple things simple, the interface has to be the simplest possible. However the gestures have to be simple as well. Like this the user has not to do some special effort in order to be able to communicate with the WGI tool.

6 Gesture recognition

To recognize a hand gesture of a human being, different steps has to be done. The video has to be cut in different images. In these images the hand has to be detected and tracked. Only after these steps the effective gesture recognition can start. This part of the report shows the necessary steps with different techniques which are already used in research projects or in industry. [12] [13]

6.1 Image acquisition

The first phase of the gesture recognition is the image acquisition. In this first phase the live stream of the webcam is splitted into pictures. For human cognition 16 to 18 images are enough for an illusion of flowing movements, as soon as the pictures are similar [14]. Since the project aims to detect hand gestures of human, there is no need to use more pictures than that. Another important point of the image acquisition is the resolution of the images. The higher the images are, the longer the computation will take. Since the hand gesture recognition has to be done in real-time, the computation time is strictly limited. The project EMOVI [15] used a resolution of about 320 x 240 pixels. The same resolution was cited in the article [13]. In the article [16], where they did hand detection, they used an image size of about 640 x 480 pixels. For a robust finger tracking method in article [17] they used a resolution of about 260 x 180 pixels. For hand detection in article [18] they use a resolution of 80 x 60 pixels.

The third point of the image acquisition is the color space. A color space is the manner how colors are represented. Most often the images are represented in the RGB (Red Green Blue) color space. This is an additive color
space which means that the three colors red, green and blue are added together. Another often used color space is the HSV (Hue Saturation Value). A big advantage of this color space is that it is often more natural to think in hue and saturation. This color space is less sensitive to shadow and uneven lighting. The articles and did also used this color space. A graphical representation is given in figure 2.

6.2 Preprocessing
The preprocessing is the phase where the hand is detected in the image. This means that with the help of image processing we will isolate the interesting for our application. Segmentation, like the preprocessing is also called, is a critical part. If the hand is not properly isolated, missing parts for instance, from the rest of the image, the next steps will not succeed too. Often this step is also used to reduce the noise in the image. In practical use, there is often a combination of different techniques which gives a better success rate.

6.2.1 Pixel level segmentation
In pixel level segmentation the hand is extracted either by color segmentation or by background subtraction. Background subtraction uses a threshold to subtract the background form the image. This technique works well if you either know the background or if it is a static background. Color segmentation can be used in two ways. First you can detect the color of the hand. This technique is not that easy because there may be shadows on the hand and not each person has the same skin color. There maybe also the problem that in the background is an object with the same color as the hand, which will therefore also be detected. The second method is to put color markers on the fingers. But also here
can be some regions which will be detected because of an object in the background with the same color.

6.2.2 Motion segmentation
"Moving objects in a video stream can detect by calculation of inter frame differences and optical flow. ... However, such a system can not detect a stationary hand or determine which of several moving objects is the hand."

6.2.3 Contour detection
An interesting technique is the contour detection or edge detection. This technique permits to detect the contours of objects in an image. The advantage of this technique is that it is not directly depending on skin color and lightning conditions. In an image there can be a large number of edges from the tracked hand and the background. Therefore this technique is often used together with another technique.

6.2.4 Correlation
A correlation is the same as a convolution by turning the image with 180°. "The convolution can be understood as a sum of shift and multiply operations." A correlation is a local operator. This means that the pixel in the resulting image is not only depending on the pixel of the original image but also from its neighborhood. For this method a sort of kernel is used which defines how big the used neighborhood is and how much each pixel is taking into account. The problem of a correlation is that it has some problem when the object is rotated or scaled. The problem can be avoided by continuously updating the template, but afterwards there is some risk to tracking afterwards something other than the hand.

6.3 Tracking
Once the object was been found, it has to be followed in order to detect which movement the user did. The hand is therefore followed form frame to frame.

6.3.1 Kalman filter
The Kalman filter is able to follow the position of a movable objects but also the uncertainties of both the dynamic model and the low level measurements. Kalman filters are easily computable in real time and are able to eliminate
noise. In the basic form a Kalman filter cannot track objects on an unknown background. A version on a controlled background with good results can be obtained. [12] [21]

6.3.2 Condensation

The condensation filter is an alternative to the Kalman filter and is one of the most used techniques for tracking. It is based on a broader class of estimators called particle filters. [22] Condensation algorithm can be used to detect and track the contour of objects moving in a cluttered environment. [13] The paper [12] references articles where condensation together with eigen tracking are used for gesture recognition. The website [23] illustrates videos, where the utilization of condensation is demonstrated.

6.4 Gesture recognition

Hand gesture recognition can be divided into two parts; Hand Posture Recognition (HPR) and Hand Gesture Recognition (HGR). HPR is static which means that it is used for a single image. HGR is dynamic. The hand will be followed over a sequence of images. [13] In this project we will use the HGR. Like this, the user will be able to point on something for instance.

6.4.1 Gesture segmentation

To be able to detect all the different gestures over the time, they have to be segmented. Of course there exist different methodologies to split these gestures. The simplest method is to have a kind of initialization posture. [13] In this kind the starting and the ending posture would be the same. The problem of this strategy is, that it will not be very natural for the user. Another possibility is to have a posture which delimits the different gestures. [13] A third method is called velocity. This means that when the hands are not moving, it is considered as the end of the gesture. [13] Last but not least there is the method called acceleration. Normally when someone stops a gesture, they often accelerate afterwards to start a new one. [13]

6.4.2 Neural Networks

A generic Neural Network (NN) is composed of the 3 parts:

1. Input

2. Hidden Layer(s). At least one, but can be more than one. Each node has a weight which represents the dependencies of the previous and the next layer. However the neurons of the same layer are independent.
3. Output

NN can be extremely robust with the possibility to online learning and parallelization depending on the model and the learning algorithm. There are some Neural Networks models allow to capture the temporal relations by using times in their connection. [13]
The Neural Networks often have the problem of modeling non-gestural patterns. [12]

Advantages of Neural Networks:
- Good in pattern recognition
- The system is developed through learning rather than programming
- NN are flexible in a changing environment

Disadvantages of Neural Networks
- Difficult to extract rules from NN
- Extremely depending on the training data
- Tends to be computer intensive
- Multi-Layer NN (multiple hidden layers) lacks memory mechanism to save past information

6.4.3 Hidden Markov Model

One of the most used technique is the Hidden Markov Model (HMM). It is also used for speech recognition and handwritten character recognition. The HMM is a statistical model which is modeled with a Markov process with unknown parameters. The extracted model parameters can then be used to perform hand gesture recognition. An HMM can be considered as the simplest dynamic Bayesian network. [13]
The variants and extensions of HMM are big. To model a class, it is common to use one HMM for each class (gesture).

Advantages of HMM:
- Adapted for continuous and online applications
- Incorporation of Prior Knowledge
- Uses established training algorithms which are computationally efficient to develop and evaluate
• Modularity (can be integrated in bigger HMM)

Disadvantages of HMM:
• Generally slow
• Require a priori notion of model topology
• Standard machine learning problems (model may not converge, avoid over-fitting)
• Need a moderate amounts of training data to use

6.4.4 Dynamic Bayesian Network

Bayesian Networks serves as representation of observed events and the out coming conclusions [24]. The Dynamic Bayesian Network (DBN) is a general-ized version of the Bayesian Network with an extension to temporal dimen-sion [13]. The DBN permits the fast addition of new gestures with a minimal training. DBN includes HMM. The referenced articles in paper [13] mention that DBN was very successful for offline Human Gesture Recognition. It references as well papers in which the real-time Human Gesture Recognition with DBN may work.

Advantages of DBN:
• Suitable for small incomplete data sets
• Structural learning possible
• Combining different sources of knowledge (with new data)
• Fast response for discrete observations
• Avoid over-fitting of data

Disadvantages of DBN:
• Not suitable for continuous observations
• No support of feedback loops

6.5 Projects

In order to permit a better understanding, this subsection presents some projects of this research area. This permits to have a better overview to what is possible and also to give inspiration and ideas for this project.
6.5.1 The sixthsense

The sixthsense was presented at the TED Conference in February 2009. This project is also known under the name "WUW - Wear Ur World - A Wearable Gestural Interface". The project was initiated under the MIT. The gadget of the project permits to get information controlled by hand gestures. It permits to access the information without having a PC in front of you.

The gadget consists of a webcam, a small beamer and a mobile device. To communicate with the system, the user uses hand gestures. To better recognize the hand gestures, the user has to put color markers on the fingers (index finger and thumb of both hands). The camera permits to see what the user sees and the beamer afterwards permits to augment the real world.

Aside from the detection of the hand gestures, the gadget is also able to detect and interpret text on objects, like the title of a book and so on. The beamer permits to display information on every surface, like the book for instance.

The current WUW prototype implements several applications. One of these applications permits to display the delay time of a flight while presenting the flight ticket in the front of the camera or to extend a topic of a newspaper with a video displayed on the newspaper.

The software of the presented WUW was developed with C#, WPF and OpenCV. On the figure the project can be seen.

![Figure 3: The sixthsense](image)
6.5.2 InterFace

The aim of the project is a toolkit for developing which allows interacting with the three technologies: acoustic, electromagnetic and optic. It is written in C#. The project permits to interact with any surface (wall, table, etc.). InterFace was developed with C# the .Net framework 3.5 and WPF. A demonstration video can be seen on the project’s website (inter-face). A photo of the framework in use is shown in figure 20.

The project is a cooperation between EIA-FR, heig-vd and eig.

Figure 4: Framework InterFace

Since the project is a framework it is possible to introduce new modalities. However it is possible to integrate the hand gesture recognition as a new modality.

6.5.3 The VICI project

The project Visualization of Immersive and Contextual Information (VICI) permits to augment the reality with additional information by portable equipment (see figure 5).

The VICI Demonstrator will use visual markers for optical tracking. These visual markers are used to identify static objects. A marker next to a drawing allows afterwards getting information about used techniques, biography of the artist and so on.

The project is a cooperation between arc, EIA-FR, heig-vd.
6.5.4 The 6sense Project

This project is in the Augmented Reality domain. The user can see information which is overlaying the real world. The goal of this project is to provide a contextual, intuitive and multimodal user interface applicable in many different domains, such as chemical plant supervision, cultural site visit or home automation.

In order to be able to overlay physical objects with contextual AR information, a framework is constituted.

The prototype is composed of headphone, microphone, see trough glasses and a video camera.

The project is a cooperation between EIA-FR and EPFL. [32]

6.5.5 Natal

Natal is a project developed by Microsoft for the controller-free gaming with the Xbox 360. A sensor device is in front of the TV, so that the user does not need to touch controller anymore. The user can navigate with gestures or with spoken commands. The device is able to make voice and body gesture recognition. It is even possible to use it with multiple users. The user can be represented by a 3D avatar.

The sensor is constituted of a camera, a depth sensor, a multi-array microphone and a processor running proprietary software. [33] See figure 6.
7 Libraries

The following libraries are used in the project a shortly explained.

7.1 OpenCV

"OpenCV is a computer vision library originally developed by Intel. It is free for use under the open source BSD license. The library is cross-platform. It focuses mainly on real-time image processing, as such, if it finds Intel’s Integrated Performance Primitives on the system, it will use these commercial optimized routines to accelerate itself." [34]

The version 1.2 was used for this project.

7.2 Emgu CV

"Emgu CV is a cross platform .Net wrapper to the Intel OpenCV image processing library. Allowing OpenCV functions to be called from .NET compatible languages such as C#, VB, VC++, IronPython etc. The wrapper can be compiled in Mono and run on Linux / Mac OS X." [36]

The version 1.1 was used for this project.

7.3 Hidden Markov Model Toolkit

"The Hidden Markov Model Toolkit (HTK) is a portable toolkit for building and manipulating hidden Markov models. HTK is primarily used for speech recognition research although it has been used for numerous other
applications including research into speech synthesis, character recognition and DNA sequencing." [37]
The version 3.4.1 was used for this project.

7.4 Gesture and Activity Recognition Toolkit

"The Gesture and Activity Recognition Toolkit (GART) (formerly Georgia Tech Gesture Toolkit) is a toolkit to allow for rapid prototyping of gesture-based applications." [38]
The Gesture and Activity Recognition Toolkit was developed with Java. GART requires HTK. The data collection process and the gesture recognition process of GART will not be analyzed in this thesis. These two processes are shown in figure 7. Earlier, the toolkit was called Georgia Tech Gesture Toolkit (GT2k).

GART uses Viterbi and Forward-Backward algorithms to train and evaluate the HMMs. [39] [40]
The version beta 0.3.0 was used for this project.

![Figure 7: Data collection and gesture recognition of GART](image)

8 Discussion

This subsection concludes the part Gesture Recognition of the state of the art with a discussion. The recognition of the gestures with this wearable tool will be the biggest challenge of this project. The projection on which information will displayed can change. The user wears a camera and therefore the background on which the gestures are made is always changing. Since the project is part of the SLR and almost all projects in the SLR are made in C#, this project will also be made in C#. Hence the language of this project will be C# as well. This gives the possibility to easy integrate this project (or parts from it) into others and vice versa. The library which
will be used for the image processing and analysis is OpenCV (see 7.1). Since this library is for C and C++, a wrapper is needed. Hence Emgu CV (see 7.2) will be used.

8.1 Image acquisition

The higher the resolution of an image, the longer the computational time will take. However the right resolution has to be found. In the section 6.1, there is mentioned an article which takes from 640 x 480 down to 80 x 60 as resolution. Since the webcam is always moving it will be better to take a resolution which is not too small. So the chosen resolution is 320 x 240 pixels. In order to be able to get good gesture recognition it will take 15 frames per second.

To be able to work with the colors, it is easier to change from RGB to HSV color space. This color space will be more robust towards illuminations changes.

8.2 Preprocessing

Since the tool is wearable the background is always changing. Another factor is that there will be some gestures which will be done without much motion. In order to arrive at results which are good, the user will wear finger marker caps. For the points mentioned before, the most adapted is a pixel level segmentation or more precise the color segmentation.

8.3 Tracking

How it is mentioned above (6.3) the most used techniques for tracking is the condensation. Since the condensation algorithm can detect and track contour of moving objects in a cluttered environment.

8.4 Gesture recognition

The last part is the task of the gesture recognition. The goal is to transform the tracked observation into a defined gesture. Also in this part, several possibilities will be exposed. To each of the possibilities some advantages and disadvantages are mentioned. If only looking at these advantages and disadvantages, the chosen technique would be Dynamic Bayesian Network for the recognition phase.

Real-time experiments with DBN are in progress but at the moment there are not many results. Therefore the chosen technology is Hidden Markov Model. HMM presents good performance and robustness for real-time gesture recognition. The DBN will be used for future development.
8.5 Projects

The main idea of the project came with the demonstration of the project sixthsense. So this project will be much lent on it. The main difference of this project will be that a device will be able to communicate with the other devices.
In order to produce a good and impressive prototype, the gesture recognition will be introduced in the framework Inter-Face. The integration of hand gestures adds a new modality in the Inter-Face framework.

9 Conclusion

The useful techniques and projects were analyzed in the discussion of the state of the art. So now we know which technology to use for which component. With these technologies the conception of the Wearable Gestural Interface can now be made and is exposed in the next part. But first the emplacement of the hardware will be illustrated.
Part III
Design

In the part II we saw all the important technologies which are necessary to produce a Wearable Gestural Interface. This techniques can now be used in the design part of the project. First we will see where the different hardware tools can be placed. Further we will have a look at the features of the prototype application and which hardware will be used.

10 Emplacement of the equipment

The idea of the project is to wear a portable beamer and camera. That is why it is necessary to think about the emplacement of the body worn beamer and the body worn camera. This section analyzes several emplacements and explains which one will be taken.

10.1 Camera

The idea of the project is, that the camera should be able to see the same things as the user sees. Therefore the camera should be placed more or less near the eyes.

The first possibility which may become to mind is to place the camera on the head. The disadvantage of this emplacement is that the head moves a lot. This may complicate the gesture recognition.

A second possibility is to wear the camera on the chest as near as possible to the throat. Since the gestures are always in front of the user, they are also always in front of the camera. However the chest is less moving than the head. Therefore the chosen emplacement must be the chest.

10.2 Beamer

The information which is presented in the users view has to be projected by a beamer.

In the paper "Mobile Interfaces Using Body Worn Projector and Camera" (41), they analyze several possibilities where to put the projector. The finally chosen location of the beamer in this paper is the lumbar mounted version.

Another possibility would be to wear the beamer on the shoulder. This emplacement would be very instable because of the hand gestures which moves also the shoulder.

The third possibility is to wear the beamer on the chest. One possibility is to put the beamer flat on the chest and use a mirror to turn the image. The
other possibility is to have it straight in front of the user. The possibilities on which the beamer lies on the chest can give shadow to the images when the user makes his hand gestures in front of him, which is a big disadvantage. The chosen emplacement of the beamer for this project is the lumber mounted version. It is a stable version which will not produce shadow on the presented information while the user makes his gestures.

11 The features of the application

This section describes the principal functionalities of the application. In order to show what kind of tools this wearable gestural interface is able to do, a smart demonstration application will be developed. The applications name is Presenter. Like the name says, the application serves to present documents. More exactly it serves to present Microsoft PowerPoint documents. This is an application which allows to navigate with hand gestures and to use this WGI-tool as well to send files to a person next to him or even to someone which has the same application. First the application is described and afterwards the different hand gestures are presented.

11.1 Application

The presented application allows the user to see presentation files and to share the files with another person. With this application the user can navigate into his slides with hand gestures. When a user wishes to share the application with the person next to him, he can only use a gesture to send the file to the other one.

11.2 File browser

When the user starts the application, a file browser will be in fornt of him. At the beginning, this file browser will only to one specific folder. This folder can not be changed directly by the user. In this window the user can click on one icon for opening the presentation. To do this he will use the Gesture for clicking described below. He can scroll down (vertical scroll) to see the files which are not visible on the screen. If the user will send the file to his friend next to him, he has to use the Gesture to send a file to another person.

A file browser is shown in figure 8.

11.3 Once the file is open

Once the user has opened a file, he will see MS PowerPoint (see figure 9). In this PowerPoint screen, he can either open the file in the presentation
mode with Gesture for opening a presentation in full screen (F5) or he can close this screen with the Gesture for closing the presentation (ESC) to close PowerPoint.

11.4 PowerPoint in full screen

In the presentation mode of MS PowerPoint (see figure 10), the user can switch the slides with the Gesture to go to the next/previous slide or simply quit this mode by using the Gesture for closing the presentation (ESC).
11.5 Hand gestures

This section presents the five gestures which will be implemented for the prototype. It is clear that for the future other gestures should be included.

11.5.1 Gesture for clicking

This gesture permits simply to click on a file to open it. The gesture can be either executed by the right hand or by the left hand. So when the user wishes to take an object he can use both hands and execute two click gestures at the same time. Note that a click contains the click down and release. To deplace an object the user clicks down, moves the object and releases. A right hand side click is shown at figure 11.

11.5.2 Gesture for opening a presentation in full screen (F5)

The gestures presented in figure 12 permits to open a presentation in the presentation mode when the user has opened MS PowerPoint.

11.5.3 Gesture for closing the presentation (ESC)

To close the presentation mode or the Powerpoint, the gesture ESC is used. It is the same gesture as the Gesture for opening a presentation in full screen (F5) but in inverse sense. It is presented in figure 13.

11.5.4 Gesture to go to the next/previous slide

The two gestures which are presented in figure 14 permit to navigate in the slides in order to go to the next or the previous slide.
11.5.5 Gesture to send a file to another person

To send a file to another person, the gesture presented in figure 11 will be used. The user has first to "take" the icon of the desired file and further to slide it out of the screen on the right side. First, the file has to be selected, in order to know which file has to be send.
12 Hardware architecture

A comparison was necessary to choose the hardware for this project. The chosen material will be presented in this chapter. For more details the comparison can be consulted in the appendix.
12.1 Beamer

The chosen beamer is the 3M MPro 110 (see figure [16]). There would be a beamer with a better resolution and autofocus. The problem is that this beamer is not yet sorted out. An evaluation kit is available, but costs about 5'000$.

12.2 Camera

QuickCam Pro for Notebooks is the smallest web cam with autofocus in the comparison. This camera is well adapted for the project since the user has

\[1\text{http://www.gizmag.com/pictures/gallery/10358_13110812813.jpg}\]
to wear it on the body. The camera is shown on figure 17.

Figure 17: QuickCam Pro for Notebooks

12.3 Ultra mobile PC (UMPC)

Sony Vaio UX micro PC is the chosen model (see figure 18) for this project. The reasons why the Sony Vaio UMPC has been chosen is on one hand the best processor and on the other hand there was already such a UMPC at the EIA-ER available.

Figure 18: Sony Vaio UX micro PC

http://www.logitech.com/repository/506/jpg/4229.1.0.jpg

2http://www.logitech.com/repository/506/jpg/4229.1.0.jpg
13 Software architecture

The chapter software architecture (check if the name of the chapter has not already changed) permits to see how the architecture of the WGI has been built. It serves also to see how it can be included in applications in order to use hand gestures to interact with the application.

The WGI should be built, in a manner that it can be used in several applications. So the WGI has not only to be running but also to inform the interested applications and components. All the gestures which are detected by the WGI should therefore be sent to all interested listeners. Therefore the WGI includes an interface which permits to listen to these gestures. The details of how the components of WGI are built can be seen in the next chapter.

In order to make a prototype, the Wearable Gestural Interface will be integrated in the InterFace (6.5.2) framework of the EIA-FR. The project InterFace was already built, to be able to combine several modalities or technologies. However the WGI will be introduced as a new modality. So first we will see how the Wearable Gestural Interface was designed. Further the components which are necessary to create in order to integrate WGI in InterFace are presented.

13.1 WGI architecture

The Wearable Gestural Interface permits to recognize hand gestures. As already written in the chapter 6, gesture recognition needs 3 different steps. Namely it needs a preprocessing, a tracking and a step for gesture identification. For this project a fourth step is needed. This fourth part permits to inform components that a gesture has occurred. These different steps can be seen in the figure 19 with different colors. To manage between those different steps the class GestureRecognition is used.

Each of these parts uses different libraries and will be described in the following subsections.

13.1.1 The principle class Gesture Recognition

The principle class Gesture Recognition is used as a controller for the entire Wearable Gestural Interface. This class permits to share and store different objects which are used by different parts of the application. It also permits to get the different images from the webcam and contains the timer which starts several times a second the preprocessing for a new image of the webcam.

If the WGI will be integrated in a project, it is only this class which has to be created by the client aside from the gesture listener which will be explained later.
13.1.2 Preprocessing

This is the first part of the gesture recognition. In these parts the image is segmented in different colors. The color segmentation permits to segment the image that it has only the colors which are used for the gestures. The second step of the preprocessing is the cleaning of the image. The cleaning step permits to eliminate some irregularities in the edges of the color rectangles. Therefore the opening and closing Morphology cleaning operators are used.

As a third step, the contours of the color rectangles are created. Since the color markers, which the user will wear to make the gestures, are rectangular, the simplest and most efficient way is to detect rectangles. Of these rectangles it is afterwards possible to take the center. The coordinates for the centers can later be used as input for the tracking and gesture identification.

It is important to mention, that when this first step is not well done, it will have consequences later on. It would be more difficult to identify the different gestures.

For the entire solution, the Emgu CV 7.2 library will be used. This library permits to wrap the C++ library OpenCV 7.1 into C#.

13.1.3 Tracking

The tracking part permits to track the different color rectangles and to predict their locations. The algorithm which is used for this part is the Condensation algorithm 6.3.2. This algorithm is already implemented in OpenCV 7.1 but not in Emgu CV 7.2. Therefore this part of the project has to be implemented in C++ (EmguCVExtension) and afterwards be imported as a DLL.

The C# Condensation class permits the wrapping to the C++ functions.

13.1.4 Gesture identification

The gesture identification part of the gesture recognition permits to train the different gestures as well as to detect them.

The library which is used for this part is HTK 7.3. HTK is made for voice recognition and can be adapted for the gesture recognition. GART 7.4 is an extension of HTK. GART permits to detect mouse gestures. So for this project GART was extended to 8 coordinates (2 per color).

13.1.5 Gesture execution

The fourth part of the Wearable Gestural Interface permits to inform the interesting components. Therefore this package contains a class which stores all the interested components which are listening to the WGI events.

This package contains also an interface called WGIListener. This interface
has to be implemented by the components which would like to be informed if a gesture has occurred.

## 13.2 InterFace architecture

For the demonstration application the framework InterFace will be used. This framework permits to transform a table in a touchable interface for instance. A big advantage of this framework is that it includes several modalities or technologies. In the following image 20, the architecture of the framework is presented.

![InterFace architecture](image)

**Figure 20: InterFace architecture**

### 13.2.1 The existing

The blue rectangles at the bottom of the schema are the different modalities or technologies. They represent the mouse, keyboard etc. So these parts are not software or at least hardware drivers. The next layer from the bottom contains the specific managers for the specific modalities. In other words they transform the specific inputs from the modalities in a common format for the InterFace framework.

From this inputs, it goes up to the plugins which represents the applications.
13.2.2 Add of new components

To include the WGI as a new modality or technology it is necessary to include the WGI as a DLL. A manager needs to be implemented as well. Since in the InterFace framework at the moment there are only events for "touch" like "touch down", "touch up" and so on. There has also to be included some other events specific for gestures. The reason why new events have to be introduced is that with hand gestures there are not only "touch" which can be occurring but also a gesture to close a component for instance.

The architecture of the InterFace framework with the integrated WGI is illustrated in figure 21.

The blue box WGI at the bottom of the image represents the Wearable Gestural Interface. This component is in the figure 19. It is this component which can be included in other applications and he is not specific to the framework InterFace.

As next a Manager for the WGI has to be implemented. This is the component which is listening to the occurred gestures of WGI and informs the InputManager that a gesture has been occurred. Of course this component has, like the other managers, to include the IInput interface.

In the next layer the InputManager has to be extended. This means that it has to be modified in order to be able to execute the forward gestures to a new component named Gesture Manager.

The GestureManager will not be implemented the same way as the Modal-
ityManager. The ModalityManager on one hand sends the event to the component on which was clicked, since it has the coordinates on which was executed the event. The GestureManager on the other hand has no coordinates which are available. This is because the gestures are not executed on one selected point, but on more or less the whole camera image size. Another problem is that the application will open the PowerPoint, but it will not be possible to send specific gestures events to this application. The GestureManager will execute at the moment only key strokes to navigate in the PowerPoint and one event which is to send a file.

13.2.3 Demonstration application

The demonstration application was described in section 11. The application is of course using the InterFace framework, therefore the integration of WGI into the framework was done. In order to be able to show such a file browser, which is shown on figure 8, two new components have to be implemented. First a kind of file browser has to be implemented in order to let the user choose a PowerPoint file to display. To display a specific PowerPoint file a new component is created which contains a picture and the name of the file.

14 Conclusion

In the chapter design we saw where the hardware will be fixed on the user and which hardware will be used. It was also exposed which features the application, which will be developed, will contain. Like the name of the chapter says, the design of the Wearable Gestural Interfaced was presented as well. It was also shown how the WGI can be integrated in the InterFace framework.
Part IV
Implementation and prototype

The chapter design showed where the hardware can be placed and now we will see how it can be fixed at the emplacements. Further we will see how the Wearable Gestural Interface is implemented and how it was included in the InterFace framework. At the end, the results of the prototype are presented.

15 Hardware

In order to have a wearable interface the hardware has to be fixed on the body of the user. This section serves therefore to show how the used hardware of the prototype can be fixed. Since this is only a prototype, the hardware components are fixed each one at the body of the user. The different hardware fixations had to be invented and further constructed. The invention was made by the author of the report but some construction was made by a collaborator of industrial technology department.

15.1 Camera

The camera is fixed on a plastic panel. A band around the neck helps to wear the plastic panel. To give more stability to the camera, the panel is that big. In order that the user does not have so much cable in front of him, the cable goes up on the band in front of the user and goes down on the back of the user to the UMPC. For better understanding, the camera fixation is shown on figure 22.

15.2 Beamer

The fixation of the beamer is a bit more complicated than the one of the camera. It is because of the fact, that the beamer should be a bit in the air to guarantee a good cooling. Furthermore it should be possible to turn the beamer in the X-axis and the Y-axis. Like this the user is able to move the projected image in front of him to a location which is the most convenient for him.

A Joby Gorilla Pod was bought by the author (see figure 23). Further this pod was modified in order to be wearing on a belt. Since the Joby Gorilla Pod is flexible the beamer can be turned in almost every direction. The figure 24 shows, how it looks like.
15.3 **UMPC**

In the shipment of the ultra mobile PC a case was already thereby. The case is made to wear at the belt. Therefore no modification need to be done for the case.

15.4 **Finger markers**

The prototype uses some finger markers in order to detect the four significant fingers for the applications. The fingers are the two index fingers and the two thumbs.

For the first attempt at walking some color pen was used in order to start
with the segmentation in different colors.
Further the first version of finger markers was created. This first version
was just cut from color balloons. After five or six times of using these color
markers, they became very gluey. Therefore they were either not reusable
or it has taken to much time to bring them back in a state which they were
reusable. The reason why is, that at the beginning this balloons have some
magnesium which protects the balloons from being gluey.
So a next generation of finger markers had to be found. Since the balloons
have already good colors which are good recognized, the idea was to keep
this colors. The solution was to take a foxglove and afterwards to put the bit
of the balloon on the foxglove. Now they can easily be putted on the fingers
and the colors are not too much reflecting and are good recognizable. The
figure 25 shows the foxgloves.
15.5 The complete wearable tool

The complete wearable tool with the camera, the beamer and the UMPC can be seen on figure 26. The UMPC and the beamer are as said above, worn on the belt, which can be seen on figure 27.

Figure 26: The complete wearable tool from the front

Figure 27: The complete wearable tool from the side
16 Wearable Gestural Interface

The gesture recognition is coded in C#, in exception of the tracking, with Microsoft Visual Studio 2008. The wearable Gestural Interface is compiled as a DLL which made it easy to integrate it in other applications. This section permits to illustrate the implementation of the different parts of the WGI.

16.1 The principle class Gesture Recognition

To be able to use the Wearable Gestural Interface it is this class from which an object has to be created. However the class contains the objects which are used for all the parts of the application. So it contains the different images which are used for the different steps. It contains also the object of the other parts of the Wearable Gestural Interface. It can be seen as a sort of controller of the WGI. This class contains the timer which executes fifteen times a second the pre-processing of a new iteration. This means that fifteen times a second an image of the camera is taken. Of course this value is stored in a constant. It is also this class which says for the gesture recognition which videos has to be taken to train the HMM. Further it is also this class that starts the training.

16.2 Touch mode and gesture mode

The Wearable Gestural Interface offers two different modes: the touch mode and the gesture mode. In the touch mode the two hands can be used like two mice. The two index fingers are used as pointers and the thumbs as clicks. The touch mode permits to execute the programmed gestures. So the touch modes don’t need training.

The mode can be switched by speech recognition. The used speech recognition is the one from Microsoft (System.Speech.Recognition.SpeechRecognizer). This speech recognition detects only one command named "switch mode". Like it already says this permits to change the mode between touch mode and gesture mode.

16.3 Preprocessing

The preprocessing is the first step executed in each iteration. First the image has to be converted to the HSV (see 6.1) format.

From this converted image four new images are created in the four colors of the finger markers (red, blue, yellow and green). To segment the image in different color images there are three different thresholds used, one in each component of HSV (hue, saturation and value).

In order to find the specific thresholds for the different color markers, a first
application had to be developed. This application permits to show the histograms of each component of the [HSV] image. The user can therefore put the camera in the front of a color marker and can see which values changes. The screenshot of this application is shown in figure 28.

The application was extended in order to be able to see which parts of the

![Figure 28: Histograms](image)

image are taken into account for the specific colors. The figure 29 permits to see the different images of the application. On the first image the color markers are on a white surface. The second image (top right) of the same figure the red part of the image can be seen. The same thing can be seen for each color. At the bottom of the application on the left, the 4 colors are detected. This picture was made by and of the 4 segmented images. Finally, the last image permits to see only the detected colors of the original image. It has to be remarked that the application for the gesture recognition only uses the image at the bottom left.

To recognize the color markers, it is not enough to just segment the images in the four color marker images. As next these four images are again preprocessed. An opening and closing morphology operator is executed on these images. This permits to plain the edges in the images. In each of these color images we will detect the edges. From these edges afterwards it is possible to create rectangles. The finger markers look very similar to rectangle. The rectangles permit to find the centers. These centers can further be used as locations which are used as input for the tracking.
16.4 Tracking

As explained above, the input for the tracking is the location of each the color marker. The used tracking algorithm is the Condensation (see section 6.3.2).

The centers are calculated from the detected color rectangles of the color markers. It is possible that those color rectangles are not detected in an image. This can have several reasons, like the adjustment of the camera for instance. If one of these color rectangles is lost, the tracking will give his most probable location.

The WGI contains a variable which determines, in how much images a color can be missed before the color is considered as not present. If there are more images than the variable, the color is considered as not detected in the image. If there are fewer images than the variable, the prediction of the condensation is taken as the color location.

This tracking mechanism permits to follow a gesture without interruption. Afterwards the gesture can be identified.

The code which was implemented in C++ was much inspired by the publication of the Speech Vision and Robotics group Cambridge University Engineering Department [42].
16.5 Gesture segmentation

The camera takes nonstop images. So a way has to be found in order to be able to say when a gesture starts and when a gesture is finished. In other words the gestures have to be segmented. As already explained in section 16.2 there are two different modes (touch mode and gesture mode). In the touch mode there is no segmentation because there are no gestures. In the gesture mode the starting of a gesture is considered as soon as colors are detected. For this prototype the starting is considered only when a color combination is detected.

This permits that no gesture is found if there is none. Further for a general usage it is possible to just start the gesture if one color is detected. As soon as the color combination (or for the general version as soon as no color) is detected the gesture is considered to be finished.

It is possible, that in one image a color has disappeared. Therefore a color is considered to be lost after a value of a constant to be lost. If a color is not detected the prediction of the tracking is considered as the location before the color is definitely lost.

16.6 Gesture identification

The gesture identification is the part where a gesture is detected by the system. For this gesture recognition, as said in the section 13.1.4, the library GART is used.

In order to start the gesture recognition with GART a sensor has to be started as well as the collecting. For each step, so more times a second, the locations of the colors have to be sending to the sensor. As soon as the gesture is finished the founded gesture can be asked on the model which is a specific object of the GART library. However in order to look if this part of the application is working an application with the figure 30. This screenshot was made after the execution of the gesture "esc" which

![Figure 30: Test application for gesture recognition](image)
In recognition processes often binary HMM are used. This means that for each gesture an HMM is trained. For the gesture recognition you will test the gesture on each HMM with a predefined barrier. The system is able to say if a gesture occurred and if one of these occurred which one it was. Such a binary system is not possible with GART. GART manages himself how much HMM are used. It demands more than 2 gestures and more than 5 videos per gesture. Another possibility is to define how much skip states you would like to have. A test of such skip states is made in section 19.

16.7 Gesture execution

With the gesture identification the name of the gesture can be found. The next step is afterwards to inform the interested components in order to be able to execute the desired action after a gesture. For this step a class called WGIEventDistribution was developed. As described already in the design chapter (see 13.1.5), a client class has to implement the WGIListener. The class, which includes the listener, has additionally to subscribe on the WGIEventDistribution. This permits afterwards to inform each subscriber which gesture occurred. Afterwards it is the task of the subscriber to execute the specific action. The advantage of this technique is that each client application can do individual actions for the gestures.

A sample of such a gesture execution is described in the section 17.1.

16.8 Gesture training

To realize good gesture recognition, the gestures need to be trained. For In order to train the system the gesture name has to be given followed by sample data. This permits the system to make relations between the name and the samples.

However for this training phase several persons have done the gestures on several backgrounds. This permits to make a more stable application. If one person does a gesture several times, they become very identical. This could as a consequence that the application later only detects the gesture of this person.

Sample videos of the gestures have been done by the author, so that the gestures could have been shown to other persons. This permitted the persons learn the gesture. If the videos are not done by the author it would have taken much more time.

There were 9 persons which made the different gestures. The number is this small, because it took much more time than preview to make the videos of the different gestures with all those persons. In the future this training should be made with much more test persons.
17 Including WGI in InterFace

The prototype uses the InterFace framework. The framework was therefore extended. The Wearable Gestural Interface is included in the InterFace as a new technology or modality. The InterFace framework was built very modular which permits to extend it easily. This is also the reason why there are not so much changelings which needed to be done. The schema of the extended InterFace framework can be seen on figure 21. The Wearable Gestural Interface is included as a DLL like already explained in the chapter 13.2.2.

17.1 HandGestureManager

The HandGestureManager is the equivalent as the MouseManager for the mouse. These managers are listening to the different events of their devices or modalities and transform it later to the events which are specific to the InterFace framework. So it raises the events on the InputManager. The HandGestureManager is implementing the WGIListener. Therefore the HandGestureManager is listening to the gestures and mice actions of the WGI. The mice part is very identical to the cursor parts of the input devices. It permits to move the cursor and to click on several elements on the screen. The gestures execute new events which had to be extended in the InputManager.

17.2 Modifications in InputManager

In order to be able to inform the ModalityManager that a gesture has raised the InputManager had to be modified. Before the integration of WGI in the InterFace framework the InputManager sent the events only forwards to the ModalityManager. With the gestures this will now be changed. On one side the cursor events are still sending to the ModalityManager. These are the events of changing the cursor coordinates and the touch down and touch up. On the other side the gesture events are sent to a new class called GestureManager which is explained just below.

17.3 GestureManager

The class GestureManager is much different from the ModalityManager. The InterFace framework permits multiplugin. This means that you can have several applications or plugins which are running at the same time. Therefore the cursors permit to give also the coordinates when a touch down and a touch up were raised. With the coordinates it is possible to find the corresponding object. For the gestures this is not possible since the gestures are not executed on
one point. Four of the five gestures are executed as key simulations. The fifth one is sent to the application. It is therefore not possible to execute several applications at the same time. For the future development a method to find the corresponding application needs to be developed in order to permit multi plugin.

17.4 Presenter

For the prototype application which was described in section 11 two new graphic components in the InterFace framework had to be created. The first new graphic component which was created is the PowerPointFile. Like the name already says, it is the component which represents the PowerPoint file. A preview of the file is given on the left of the component and on the right side the name of the file is displayed (see figure 31). The second new graphic component is kind of folder or file browser in which the above PowerPointFiles are displayed. The folder contains a reference of each PowerPointFile, which permits him afterwards to send the selected file as soon as a send gesture was executed. This can be seen on figure 32.

![Figure 31: Graphic component PowerPointFile](image1)

![Figure 32: Graphic component file browser (ExplorerFrame)](image2)
18 Results and prototype

At the beginning of this project, there was the idea of the presenter demonstration application. This demonstration application has been realized and was designed during this project. Now the first prototype of the Wearable Gestural Interface exists. The project can be worn and no power cable is needed. In figure 33 the WGI can be seen in use.

The project is functional and can be used. Of course there are some points which have to be improved. This points are listed in section 22.

![Image](image.png)

Figure 33: The Wearable Gestural Interface in use

19 Tests

The Wearable Gestural Interface is very difficult to test. The gesture recognition has been tested and the values were written down. The entire system cannot be written down with values. It would be possible to test by hand and write down the statistics. But this process would take lots of time and has therefore not been done in this Master Thesis.

The system is not so easy to use, because the user has no feedback in the gesture mode (see section 16.2) where his hands are. If the user does a gesture not really in front of the camera, not the entire gesture can be seen by the camera. Therefore the user has to be trained.

Two tests for the gesture recognition were made. The first one uses only gestures of the author and the second one uses the gestures of 9 test persons.
It can be notated that the first test gives the better results. This is why the author has done very often the gestures and does it very identical in every video. The other persons do these gestures slightly different. So in the best recognized mode, the application is only well adapted for the author.

### 19.1 Test for the number of gestures

In this test, only gestures of the author have been used. For each gesture 14 videos were trained. GART uses 70% of the gesture for the training and 30% for the tests. Since there cannot be less than 3 gestures, the number starts at 3.

It can be remarked (see figure 34) that the best recognition rate of 85.94% is achieved with 5 gestures.

![Figure 34: Test: number of gestures](image)

### 19.2 Test for the number of skip states

In this test, the gestures of the 9 test persons have been taken which made 31 gestures per video. The skip states starts at one and goes up to 10.

It can be remarked (see figure 35) that the best recognition rate of about 73.91% was achieved with 4 and 5 skip states.
20 Conclusion

The part implementation and prototype showed us how the hardware is fixed on the user. It presented how the Wearable Gestural Interface was implemented. This chapter permitted also to see which different steps are necessary to include it in other software like the InterFace framework. At the end we saw the results of the prototype and that the Wearable Gestural Interface is a functional hand gesture input for applications.
Part V
Conclusion and future work

In this last chapter we will do a synthesis on the project. It contains the occurred problems, which appeared during this project. The next section describes the improvements and the possible future work. At the end the personal impressions are written down.

21 Occurred problems

During the entire project there were several problems which occurred. The most important ones and those which have taken the most time will be explained.

- The Condensation algorithm (see \[6.3.2\]) was only implemented in OpenCV 7.1 and not in Emgu CV 7.2. This is the reason why the tracking part of the project is written in C++ and not like the rest in C#.

- In the C++ code the state (a matrix) of the condensation was always on null. In order to resolve the problem, the Dyn matrix which is used for the state computation has to be initialized with the identity matrix.

- The first idea for the gesture identification was realized with the HMM of OpenCV 7.1. The HMM of OpenCV is in the auxiliary package. It is very difficult to find samples and help. All the posts in the OpenCV group (\[35\]) were not answered and also mine did not receive any answer.
In the execution of one method was an error. The message was only "external component has thrown an exception". I had to change between different OpenCV versions, just to find out that he needs a black and white image and not a color one.
Therefore this approach was given up since there were not any productive results after several days of working.

- In the project the gestures are identified by GART \([7.4]\). GART can be downloaded as a jar file. In order to be able to use it in C# it has to be transformed in a DLL. The program which was used for the transformation was IKVMC.

- By the execution of the test applications there was an error on the UMPC (Emgu.CVInvoke exception). The reason why it did not work was that the Emgu CV 7.2 has not been installed on the UMPC. The libraries were not directly given by the installer of the test application.
If you do not give a specific folder for the GART component, it will take the default folder which is C:\Documents and Settings\User. With this folder the files could not be saved. By changing the folder to C:\wgi the problem has been solved.

In the PowerPointFile component a PowerPoint file can be opened. At the beginning, this raised always an exception. The following line solved the problem:

```csharp
oApp.Visible = Microsoft.Office.Core.MsoTriState.msoTrue;
```

In the InterFace framework, in some regions on the screen can be seen a flickering of the cursors. This problem occurred only on the UMPC. This problem was not further analyzed, aside from checking if the right coordinates where transmitted which was the case.

Since it is not so easy for the user to find the center of the camera vision, it is difficult to use the defined gestures and the user has to be trained. Therefore some other gestures have been tested. A gesture which was tested was with the four color markers from below the camera vision to above of the camera vision. It seems that when the four markers does only a vertical movement, it is not supported by GART. Even if the gesture was not trained it raised an error. So this gesture was not compatible with GART.

The used version of GART is the newest beta at the moment. Sometimes the training of the gestures was not possible. Once an error occurred, it has taken a moment to bring the system running. The only way was to change the training files and some parameters. This had to be repeated several times. Sometimes it has taken a moment and once it worked again it worked for many repetitions.

## 22 Improvements and Future work

The Wearable Gestural Interface has been presented in this report. Like described, the application for the WGI is just a prototype. So the software and also the hardware are just in a state that it is possible to say that it works but it is very depending on the light conditions. The user also has to train a little bit in order to be really able to use the application. Possible ameliorations:

- In order to be more general, an additional application should be made for the training. At the moment, it is the same application on which only a boolean is changing in order to make a new training or to use the existing files.
• To make a better recognition of the hand gestures a more extensive training should be done. It has taken lots of time since I had to show the gestures on a video to each person which was filmed. Since the time of the project was limited, many things needed to be done; only a few persons were filmed by making the gestures. In order to make the recognition better more persons have to be taken on video.

• The selected portable beamer is not very bright. On one hand when you do not have enough light, the color markers are not well detected. On the other hand when it is to bright in the room the projected images of the beamer cannot be seen. At the end of the project 3M (the manufactory of the beamer) and as well Benq sorted out a new beamer with more lumens. The used beamer has 10 lumens. In order to make the project more stable and more agreeable, it would be better to use a brighter beamer.

• At the moment there are only two dimensional gestures which are used. When you add a second webcam it would be possible to detect three dimensional gestures. There dimensional gestures are more natural for human beings. Therefore the application would be more comfortable and simpler to user for the user.

• For a user it is difficult to know in which area of the camera vision he has his hands. It would be therefore very convenient for the user to have a feedback on the screen or the projected image respectively. At the moment, this was not possible since most of the gestures are used in the PowerPoint application. In a further version it is imaginable to develop 4 color mouse pointers which helps the untrained users to use the application.

• Another approach to make the use of the application easier for an untrained user is to use a camera which is already integrated in the sunglasses (see figure 36). If the user is looking straight forward, he would see the same thing as the camera. Therefore it would be easier for him to do the gestures in the camera vision.

• At the moment the user requires a UMPC. For further prototypes it is imaginable to use a cell phone like the IPhone or an Adroid phone. Since almost every person has today a cell phone, the user will not need an additional device.

23 Personal impressions

This project was inspired by the sixthsense project[6.5.1]. This project gives some very nice ideas of what would be possible. They say that their project
is more a concept than a really prototype which is usable. I think another nice approach would be to start with an even tinier smart tool which proposes only very small applications. About two weeks before the project ended I saw some sunglasses which permit to take photos (see figure 36). Last week I discovered that LG will sorted out the first cell phone with an integrated beamer (see figure 37). With this two hardware tools, the user will not look very strange by wearing the Wearable Gestural Interface. This project gave me the opportunity to acquire knowledge in a variety of domains. I had the possibility to immerse myself in image processing, to see the tracking and of course to execute a recognition step. Therefore the project was very complex and a big challenge for me, but it made this project very interesting. Because of this complexity, I am happy with the results even it is not that revolutionary. So the project was very enriching for me.

Figure 36: SPY sunglasses with integrated camera

Figure 37: LG eXpo, cell phone with integrated projector

4http://www.geekalerts.com/u/spy-sunglasses.jpg
24 Gratitude

I could not have accomplished my studies without the support of my family. That is why I want to thank my mother, my father, my grandmother, my sister and my brother for their abiding support. As well, I would like to thank my girlfriend Aline and her family for their backup. For the supervision of the project and the hints I want especially to thank Elena Mugellini, Omar Abou Khaled and Rolf Ingold. I want also express my gratitude to my friends and the team of the office which I had at the EIA-FR during my Master Thesis.
Part VI

List of abbreviations

AR Augmented Reality
arc Haute Ecole Neuchâtel Berne Jura
DBN Dynamic Bayesian Network
DLL Dynamic Link Library
EIA-FR École d’ingénieurs et d’architects de Fribourg
eig Ecole d’ingénieurs de Genève
EPFL Ecole polytechnique fédérale de Lausanne
GART Gesture and Activity Recognition Toolkit
heig-vd Haute Ecole d’ingénierie et de Gestion du Canton de Vaud
HMM Hidden Markov Model
HSV Hue Saturation Value (color space)
HTK Hidden Markov Model Toolkit
MIT Massachusetts Institute of Technology
NN Neural Network
PC Personal Computer
RGB Red Green Blue (color space)
SLR Smart Living Room
TED Technology, Entertainment, Design
UMPC Ultra Mobile PC
UNIFR University of Fribourg
WGI Wearable Gestural Interface
WLAN Wireless Local Area Network
WPF Windows Presentation Foundation
WUW Wear Ur World - A Wearable Gestural Interface (The sixthsense)
Part VII
Bibliography

References


[2] *German Wikipedia website for Augmented Reality*
http://de.wikipedia.org/wiki/Erweiterte_Realt%C3%A4t


[4] *German Wikipedia website for Wearable Computing*

http://en.wikipedia.org/wiki/Wearable_computer


[14] *German Wikipedia website for moving images*
http://de.wikipedia.org/wiki/Bewegte_Bilder


Website about Condensation algorithm. http://www.robots.ox.ac.uk/~misard/condensation.html


Continuous Gesture Recognition using a Sparse Bayesian Classifier. Shu-Fai Wong, Roberto Cipolla. 2006.


InterFace website. http://www.interactive-surface.ch


VICI website. http://www.vici-project.ch

6th sense website. http://www.6sense.ch
[33] *Natal project website.*  

[34] *English Wikipedia website for OpenCV*  
http://en.wikipedia.org/wiki/OpenCV

[35] *Website of OpenCV*  
http://opencv.willowgarage.com/wiki/Welcome

[36] *Website of Emgu CV*  

[37] *Website of HTK*  
http://htk.eng.cam.ac.uk/

[38] *Website of GART*  
https://wiki.cc.gatech.edu/ccg/projects/gt2k/gt2k


Part VIII
Appendix

25 Installation manual
26 Specifications
27 Planning
28 Comparison
29 Tests
Installation of the Wearable Gestural Interface on a Windows XP

1. EmguCV has to be installed and the path of the installation has to be added in the system paths

2. The C++ redistributable package has to be installed on the client (is published by Visual Studio at the same time as the WGI installer)

3. Install HTK

4. DirectX (needed by the InterFace framework)

5. Install the WGI installer
Master work
Wearable Gestural interface for SLR/SMR

Student: Matthias SCHWALLER

University of Fribourg
Departement of Informatics
Document, Image and Voice Analysis Group
Boulevard de Pérolles 90
CH-1700 Fribourg

Rolf Ingold

University of Applied Sciences of Western Switzerland - Fribourg
Information & Communication Technologies Institute
Multimedia and Information System Group
Bd de Pérolles 80
CH-1705 Fribourg

Omar Abou Khaled
Elena Mugellini
Contents
Project requirements & specifications................................................................. 3
  Project logistics.................................................................................................. 3
  Project topic .................................................................................................... 3
  Deliverables..................................................................................................... 3
State of the Art .................................................................................................. 4
Environment & Context....................................................................................... 4
Goals.................................................................................................................... 4
References.......................................................................................................... 5
Project requirements & specifications

Project logistics
- Project start date: Monday 20.07.2009
- Duration: 30 ECTS (5 months)
- Presentation date: to be defined

Project topic
- Topic: Gestural Interaction, Augmented Reality, Wearable Interface, Tangible Computing, Object Augmentation
- Aims:
  - Develop a wearable device allowing gesture recognition to augment physical world around the person in smart environment.
- Technologies used:
  - To be defined
- Possible Tools: to be defined

Deliverables
- Report
- Demonstration
- Presentation slides
- CD/DVD with all resources
State of the Art
- The sixthsense project
- The Inter-Face Project
- The VICI / 6sense Projects

Environment & Context
Smart living room & Smart meeting Room

Goals
This project aims to develop a wearable gestural interface for collaborative meeting with the smart living room & smart meeting room. The design of the system includes two parts: hardware & application.

1) For the hardware part, the following goals have to be achieved:
   a. State of the art of existing devises such as camera, beamer, gloves, ultra mobile PC, etc.
      For this project we need small devices like the following ones:
      Beamers (uses battery supply and have a length of about 12 cm):
      Camera with Autofocus
      Netbook as ultra mobile PC
      Finger marker caps or gloves to detect easier the significant fingers at the beginning may be helpful.

   b. Design and realization of the wearable demonstrator

2) For the application part, the following tasks have to be carried out:
   a. State of the art on gesture recognition algorithms & approaches
      Gesture recognition can be divided in the following 3 parts:
      - Image acquisition and preprocessing (suppressing noise, extracting important clues)
      - Tracking (detecting the moving)
      - Gesture recognition

1 http://www.gizmag.com/pictures/gallery/10358_13110812813.jpg
2 http://media.conrad.com/m/3000_3999/3400/3440/3441/344183_LB_00_FB.EPS.jpg
3 http://www.optoma.co.uk/optomapico/uploads/Gallery/100dpi/PK101-100-1-1.jpg
4 http://www.logitech.com/repository/506/jpg/4229.1.0.jpg
5 http://jrgcomputers.com/laptops/umpc-ultra-mobile-pc.jpg
These techniques can be used (ref. Course Image Processing and Analysis, Prof Ingold)
- Contour detection
- Texture segmentation
- Edge detection
- Pattern recognition
- Dynamic Bayesian Network
- Hidden Markov Model
- Neural Networks
- …

b. Specification of the application scenario and identification of main functionalities within SLR & SMR and the following some proposed ideas

**SLR**
- Map sharing; one person shows the way to another person
- Gesture for exchange files between persons (while sorting the pictures at a wall, exchange of contacts, etc.)
- 2 or more persons can play games together like Tic Tac Toe (cross and circle)
- 2 or more persons can draw an picture together
- Documentation annotation
- …

**SMR**
- Virtual keyboard
- Presentation can be projected on every persons beamer
- 1 person writes the PV of the meeting while the other persons can read it on their own beamer
- People can see simultaneously the slides and add their annotations, drawing while the presentation (on each person’s beamer)
- …

**Other application**
- Virtual analog watch on the wrist
- Video on newspaper
- Additional information on a flight or train ticket (delay, gate)
- Additional information about a book, while presenting it in front of the camera
- …

c. Design of the application
d. Implementation & test

References
[3] [http://www.vici-project.ch/](http://www.vici-project.ch/)
[4] [http://www.6sense.ch/](http://www.6sense.ch/)
[5] [http://project.eia-fr.ch/aramis/](http://project.eia-fr.ch/aramis/)
[6] [http://portal.acm.org/citation.cfm?id=1520626](http://portal.acm.org/citation.cfm?id=1520626)
<table>
<thead>
<tr>
<th>ID</th>
<th>Task Name</th>
<th>Duration</th>
<th>Start</th>
<th>Finish</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Master thesis WGI - Wearable Gestural Interface</td>
<td>100 days</td>
<td>Mon 20.07.09</td>
<td>Fri 04.12.09</td>
</tr>
<tr>
<td>3</td>
<td>Analyse</td>
<td>32 days</td>
<td>Mon 20.07.09</td>
<td>Tue 01.09.09</td>
</tr>
<tr>
<td>4</td>
<td>Project starting</td>
<td>5 days</td>
<td>Mon 20.07.09</td>
<td>Fri 24.07.09</td>
</tr>
<tr>
<td>5</td>
<td>Administratif tasks (login, PC etc.)</td>
<td>1 day</td>
<td>Mon 20.07.09</td>
<td>Mon 20.07.09</td>
</tr>
<tr>
<td>6</td>
<td>Specifications</td>
<td>3 days</td>
<td>Tue 21.07.09</td>
<td>Thu 23.07.09</td>
</tr>
<tr>
<td>7</td>
<td>Planning</td>
<td>1 day</td>
<td>Fri 24.07.09</td>
<td>Fri 24.07.09</td>
</tr>
<tr>
<td>9</td>
<td>State of the art</td>
<td>27 days</td>
<td>Mon 27.07.09</td>
<td>Tue 01.09.09</td>
</tr>
<tr>
<td>10</td>
<td>Wearable computing</td>
<td>10 days</td>
<td>Mon 27.07.09</td>
<td>Fri 07.08.09</td>
</tr>
<tr>
<td>11</td>
<td>Gesture recognition</td>
<td>10 days</td>
<td>Mon 10.08.09</td>
<td>Fri 21.08.09</td>
</tr>
<tr>
<td>12</td>
<td>Existing projects</td>
<td>5 days</td>
<td>Mon 24.08.09</td>
<td>Fri 28.08.09</td>
</tr>
<tr>
<td>13</td>
<td>Comparaison of the material</td>
<td>2 days</td>
<td>Mon 31.08.09</td>
<td>Tue 01.09.09</td>
</tr>
<tr>
<td>15</td>
<td>Conception 1</td>
<td>5 days</td>
<td>Wed 02.09.09</td>
<td>Tue 08.09.09</td>
</tr>
<tr>
<td>16</td>
<td>Architecture design</td>
<td>3 days</td>
<td>Wed 02.09.09</td>
<td>Tue 04.09.09</td>
</tr>
<tr>
<td>17</td>
<td>First application</td>
<td>2 days</td>
<td>Mon 07.09.09</td>
<td>Tue 08.09.09</td>
</tr>
<tr>
<td>19</td>
<td>Implementation 1</td>
<td>6 days</td>
<td>Wed 09.09.09</td>
<td>Wed 16.09.09</td>
</tr>
<tr>
<td>20</td>
<td>Preparing menus…</td>
<td>2 days</td>
<td>Wed 09.09.09</td>
<td>Thu 10.09.09</td>
</tr>
<tr>
<td>21</td>
<td>Communication module</td>
<td>2 days</td>
<td>Fri 11.09.09</td>
<td>Mon 14.09.09</td>
</tr>
<tr>
<td>22</td>
<td>First application</td>
<td>2 days</td>
<td>Tue 15.09.09</td>
<td>Wed 16.09.09</td>
</tr>
<tr>
<td>24</td>
<td>Gesture recognition</td>
<td>34 days</td>
<td>Thu 17.09.09</td>
<td>Tue 03.11.09</td>
</tr>
<tr>
<td>25</td>
<td>Image acquisition</td>
<td>2 days</td>
<td>Thu 17.09.09</td>
<td>Fri 18.09.09</td>
</tr>
<tr>
<td>26</td>
<td>Conception</td>
<td>1 day</td>
<td>Thu 17.09.09</td>
<td>Thu 17.09.09</td>
</tr>
<tr>
<td>27</td>
<td>Implementation</td>
<td>1 day</td>
<td>Fri 18.09.09</td>
<td>Fri 18.09.26</td>
</tr>
<tr>
<td>28</td>
<td>Segmentation</td>
<td>9 days</td>
<td>Mon 21.09.09</td>
<td>Thu 01.10.09</td>
</tr>
<tr>
<td>29</td>
<td>conception</td>
<td>4 days</td>
<td>Mon 21.09.09</td>
<td>Thu 24.09.09</td>
</tr>
<tr>
<td>30</td>
<td>Implementation</td>
<td>3 days</td>
<td>Fri 25.09.09</td>
<td>Tue 29.09.29</td>
</tr>
<tr>
<td>31</td>
<td>Test and corrections</td>
<td>2 days</td>
<td>Wed 30.09.09</td>
<td>Thu 01.10.30</td>
</tr>
<tr>
<td>32</td>
<td>Tracking</td>
<td>9 days</td>
<td>Fri 02.10.09</td>
<td>Wed 14.10.09</td>
</tr>
<tr>
<td>33</td>
<td>conception</td>
<td>4 days</td>
<td>Fri 02.10.09</td>
<td>Wed 07.10.31</td>
</tr>
<tr>
<td>34</td>
<td>Implementation</td>
<td>3 days</td>
<td>Thu 08.10.09</td>
<td>Mon 12.10.33</td>
</tr>
<tr>
<td>35</td>
<td>Test and corrections</td>
<td>2 days</td>
<td>Tue 13.10.09</td>
<td>Wed 14.10.34</td>
</tr>
<tr>
<td>36</td>
<td>Gesture recognition</td>
<td>14 days</td>
<td>Thu 15.10.09</td>
<td>Tue 03.11.09</td>
</tr>
<tr>
<td>37</td>
<td>conception</td>
<td>9 days</td>
<td>Thu 15.10.09</td>
<td>Tue 27.10.35</td>
</tr>
<tr>
<td>38</td>
<td>Implementation</td>
<td>3 days</td>
<td>Wed 28.10.09</td>
<td>Fri 30.10.37</td>
</tr>
<tr>
<td>39</td>
<td>Test and corrections</td>
<td>2 days</td>
<td>Mon 02.11.09</td>
<td>Tue 03.11.38</td>
</tr>
<tr>
<td>41</td>
<td>Hardware</td>
<td>2 days</td>
<td>Wed 04.11.09</td>
<td>Thu 05.11.09</td>
</tr>
<tr>
<td>42</td>
<td>Body fixation</td>
<td>2 days</td>
<td>Wed 04.11.09</td>
<td>Thu 05.11.39</td>
</tr>
<tr>
<td>ID</td>
<td>Task Name</td>
<td>Duration</td>
<td>Start</td>
<td>Finish</td>
</tr>
<tr>
<td>----</td>
<td>-----------------------</td>
<td>----------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>43</td>
<td>Conception 2</td>
<td>4 days</td>
<td>Fri 06.11.09</td>
<td>Wed 11.11.09</td>
</tr>
<tr>
<td>44</td>
<td>Applications</td>
<td>4 days</td>
<td>Fri 06.11.09</td>
<td>Wed 11.11.09</td>
</tr>
<tr>
<td>45</td>
<td>Conception 2</td>
<td>4 days</td>
<td>Fri 06.11.09</td>
<td>Wed 11.11.09</td>
</tr>
<tr>
<td>46</td>
<td>Applications</td>
<td>4 days</td>
<td>Fri 06.11.09</td>
<td>Wed 11.11.09</td>
</tr>
<tr>
<td>47</td>
<td>Implementation</td>
<td>2 days</td>
<td>Thu 12.11.09</td>
<td>Fri 13.11.09</td>
</tr>
<tr>
<td>48</td>
<td>Applications</td>
<td>2 days</td>
<td>Thu 12.11.09</td>
<td>Fri 13.11.09</td>
</tr>
<tr>
<td>49</td>
<td>Test and corrections</td>
<td>3 days</td>
<td>Mon 16.11.09</td>
<td>Wed 18.11.09</td>
</tr>
<tr>
<td>50</td>
<td>Finishing report</td>
<td>8 days</td>
<td>Thu 19.11.09</td>
<td>Mon 30.11.09</td>
</tr>
<tr>
<td>51</td>
<td>Final presentation</td>
<td>2 days</td>
<td>Tue 01.12.09</td>
<td>Wed 02.12.09</td>
</tr>
<tr>
<td>52</td>
<td>Demo video</td>
<td>2 days</td>
<td>Thu 03.12.09</td>
<td>Fri 04.12.09</td>
</tr>
<tr>
<td>53</td>
<td>Test and corrections</td>
<td>3 days</td>
<td>Mon 16.11.09</td>
<td>Wed 18.11.09</td>
</tr>
<tr>
<td>54</td>
<td>Finishing report</td>
<td>8 days</td>
<td>Thu 19.11.09</td>
<td>Mon 30.11.09</td>
</tr>
<tr>
<td>55</td>
<td>Final presentation</td>
<td>2 days</td>
<td>Tue 01.12.09</td>
<td>Wed 02.12.09</td>
</tr>
<tr>
<td>56</td>
<td>Presentations</td>
<td>107 days</td>
<td>Mon 20.07.09</td>
<td>Wed 16.12.09</td>
</tr>
<tr>
<td>57</td>
<td>1st presentation (10.00)</td>
<td>0 hrs</td>
<td>Fri 21.08.09</td>
<td>Fri 21.08.09</td>
</tr>
<tr>
<td>58</td>
<td>2nd presentation</td>
<td>0 days</td>
<td>Fri 02.10.09</td>
<td>Fri 02.10.09</td>
</tr>
<tr>
<td>59</td>
<td>Final presentation</td>
<td>0 days</td>
<td>Wed 16.12.09</td>
<td>Wed 16.12.09</td>
</tr>
<tr>
<td>60</td>
<td>Report &amp; administratif tasks</td>
<td>100 days</td>
<td>Mon 20.07.09</td>
<td>Fri 04.12.09</td>
</tr>
<tr>
<td>61</td>
<td>Presentations</td>
<td>107 days</td>
<td>Mon 20.07.09</td>
<td>Wed 16.12.09</td>
</tr>
<tr>
<td>62</td>
<td>1st presentation (10.00)</td>
<td>0 hrs</td>
<td>Fri 21.08.09</td>
<td>Fri 21.08.09</td>
</tr>
<tr>
<td>63</td>
<td>2nd presentation</td>
<td>0 days</td>
<td>Fri 02.10.09</td>
<td>Fri 02.10.09</td>
</tr>
<tr>
<td>64</td>
<td>Final presentation</td>
<td>0 days</td>
<td>Wed 16.12.09</td>
<td>Wed 16.12.09</td>
</tr>
<tr>
<td>65</td>
<td>Deadline of the project</td>
<td>0 days</td>
<td>Fri 04.12.09</td>
<td>Fri 04.12.09</td>
</tr>
</tbody>
</table>
Comparison beamers

<table>
<thead>
<tr>
<th>Brand &amp; Type</th>
<th>Wight [g]</th>
<th>Beamer size [cm]</th>
<th>Image size [cm]</th>
<th>Battery lifetime [min]</th>
<th>Resolution [pixel]</th>
<th>Light source</th>
<th>Brightness [Lumens]</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>3M MPro110</td>
<td>152</td>
<td>11,5 x 5 x 2,2</td>
<td>15 – 127</td>
<td>60</td>
<td>640 x 480</td>
<td>LED</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>AIPTEK V10+</td>
<td>155</td>
<td>12,5 x 5,5 x 2,3</td>
<td>15 – 127</td>
<td>90</td>
<td>640 x 480</td>
<td>LED</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Optoma PK101</td>
<td>115</td>
<td>10,3 x 5 x 1,5</td>
<td>15 – 150</td>
<td>120</td>
<td>480 x 320</td>
<td>LED</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Microvision SHOWWX</td>
<td>122</td>
<td>11.8 x 6 x 1.4</td>
<td>15 – 250</td>
<td>Movie capable ?</td>
<td>848 x 480</td>
<td>Laser (Class 2)</td>
<td>10</td>
<td>Autofocus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not yet out</td>
</tr>
</tbody>
</table>

Product to order:

3M MPro110

Microvision would be the best beamer for this project. They offer at the moment an evaluation kit, but it costs about 5’000 $. Therefore we will take the 3M MPro 110.
## Comparison cameras

<table>
<thead>
<tr>
<th>Brand &amp; Type</th>
<th>Resolution [pixel]</th>
<th>Color depth [bit]</th>
<th>Frames per second</th>
<th>Connection</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logitech QuickCam® Pro for Notebooks</td>
<td>1600 x 1200</td>
<td>24</td>
<td>30</td>
<td>USB</td>
<td>Autofocus, Carl Zeiss lens</td>
</tr>
<tr>
<td>HP Premium Autofocus Webcam</td>
<td>1600 x 1200</td>
<td></td>
<td>30</td>
<td>USB</td>
<td>Autofocus</td>
</tr>
<tr>
<td>Logitech Pro 9000</td>
<td>1600 x 1200</td>
<td>24</td>
<td>30</td>
<td>USB</td>
<td>Autofocus, Carl Zeiss lens</td>
</tr>
</tbody>
</table>

**Product to order:**

Logitech QuickCam Pro for Notebooks.

This webcam has the autofocus technology and a very good lens. Aside it is the smallest webcam of the above three.
## Comparison ultra mobile PC

<table>
<thead>
<tr>
<th>Brand &amp; Type</th>
<th>Processor</th>
<th>Memory [MB]</th>
<th>Disk [GB]</th>
<th>Battery lifetime [h]</th>
<th>Weight [g]</th>
<th>Size [cm]</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sony Vaio UX1</td>
<td>Intel Core Solo Prozessor U1500 ULV 1.33 GHz</td>
<td>1024</td>
<td>32</td>
<td>2.6</td>
<td>486</td>
<td>15 x 10 x 38</td>
<td>With keyboard</td>
</tr>
<tr>
<td>Q1-900 Ceegoo</td>
<td>Intel Celeron M Prozessor 353, 900 MHz</td>
<td>512</td>
<td>40</td>
<td>3</td>
<td>779</td>
<td>23 x 14 x 26</td>
<td></td>
</tr>
<tr>
<td>Asus R2H</td>
<td>Intel ULV Celeron M Processor 900MHz</td>
<td>256</td>
<td>20 - 60</td>
<td>2</td>
<td>830</td>
<td>23 x 13 x 3</td>
<td></td>
</tr>
</tbody>
</table>

http://www.umpc.ch/umpc/umpc-vergleich.html

**Product to order:**

**Sony Vaio UX1**

The UMPC of Sony Vaio has the best processor, the most ram and aside from this it is much lighter than the others.
Tests

Test for the number of gestures

3 Gestures

<table>
<thead>
<tr>
<th>n</th>
<th>e</th>
<th>f</th>
<th>p</th>
<th>e</th>
<th>s</th>
<th>r</th>
<th>x</th>
<th>c</th>
<th>e</th>
<th>t</th>
<th>v</th>
<th>Del</th>
<th>%c / %e</th>
</tr>
</thead>
<tbody>
<tr>
<td>next</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38.5/17.0</td>
</tr>
<tr>
<td>esc</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f5</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prev</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>91.7/2.1</td>
</tr>
<tr>
<td>Ins</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 Gestures

<table>
<thead>
<tr>
<th>n</th>
<th>e</th>
<th>f</th>
<th>p</th>
<th>e</th>
<th>s</th>
<th>r</th>
<th>x</th>
<th>c</th>
<th>e</th>
<th>t</th>
<th>v</th>
<th>Del</th>
<th>%c / %e</th>
</tr>
</thead>
<tbody>
<tr>
<td>next</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80.0/5.3</td>
</tr>
<tr>
<td>esc</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prev</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50.0/10.5</td>
</tr>
<tr>
<td>Ins</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5 Gestures

Test for the number of skip states

1 Skip state
### 2 Skip states

---

**HTK Results Analysis**

**Date:** Fri Dec 04 08:58:46 2009  
**Ref:** c:\gart\gart_workspace\labels.mlf  
**Rec:** c:\gart\gart_workspace\output.mlf

---

**Overall Results**

**SENT:** %Correct=72.00 [H=18, S=7, N=25]  
**WORD:** %Corr=72.00, Acc=72.00 [H=18, D=0, S=7, I=0, N=25]

---

**Confusion Matrix**

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>e</th>
<th>f</th>
<th>p</th>
<th></th>
<th>Del</th>
<th>[%c / %e]</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>[66.7/8.0]</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>[80.0/4.0]</td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>[0.0/16.0]</td>
</tr>
<tr>
<td>v</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

---

### 3 Skip states

---

**HTK Results Analysis**

**Date:** Fri Dec 04 08:52:40 2009  
**Ref:** c:\gart\gart_workspace\labels.mlf  
**Rec:** c:\gart\gart_workspace\output.mlf

---

**Overall Results**

**SENT:** %Correct=72.00 [H=18, S=7, N=25]  
**WORD:** %Corr=72.00, Acc=72.00 [H=18, D=0, S=7, I=0, N=25]

---

**Confusion Matrix**

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>e</th>
<th>f</th>
<th>p</th>
<th></th>
<th>Del</th>
<th>[%c / %e]</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>[66.7/8.0]</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>[40.0/12.0]</td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>[50.0/8.0]</td>
</tr>
<tr>
<td>v</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

---

### 4 Skip states

---

**HTK Results Analysis**

**Date:** Fri Dec 04 08:43:52 2009  
**Ref:** c:\gart\gart_workspace\labels.mlf  
**Rec:** c:\gart\gart_workspace\output.mlf

---

**Overall Results**

**SENT:** %Correct=73.91 [H=34, S=12, N=46]  
**WORD:** %Corr=73.91, Acc=73.91 [H=34, D=0, S=12, I=0, N=46]

---

**Confusion Matrix**

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>e</th>
<th>f</th>
<th>p</th>
<th></th>
<th>Del</th>
<th>[%c / %e]</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>[90.9/2.2]</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>[33.3/13.0]</td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>[88.9/2.2]</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>[55.6/8.7]</td>
</tr>
<tr>
<td>t</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
### 5 Skip states

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>n</th>
<th>e</th>
<th>f</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>e</td>
<td>s</td>
<td>5</td>
<td>r</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>x</td>
<td>c</td>
<td></td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>t</td>
<td>v</td>
<td>Del</td>
<td>%c / %e</td>
<td></td>
</tr>
<tr>
<td>send</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>next</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>esc</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>f5</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>prev</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Ins</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

---

Project WGI – Wearable Gestural Interface

Matthias Schwaller