



# **Eyebrow Detection for Expression Analysis Using the Kinect**

by

**Mihai Baltac**

Bachelor Thesis in Computer Science

Prof. Andreas Birk  
Supervisor

Date of Submission: May 16, 2014

---

Jacobs University — School of Engineering and Science

With my signature, I certify that this thesis has been written by me using only the indicates, resources and materials. Where I have presented data and results, the data and results are complete, genuine, and have been obtained by me unless otherwise acknowledged; where my results derive from computer programs, these computer programs have been written by me unless otherwise acknowledged. I further confirm that this thesis has not been submitted, either in part or as a whole, for any other academic degree at this or another institution.

Signature

Place, Date

## **Abstract**

Real time measurement of facial activity and emotion detection is a challenging technology, currently under development. It is considered to become a key factor in machine-human interaction, as it will move the way we now communicate with computers to a whole new level. Interpreting facial activity and emotions autonomously implies accurate detection and tracking of various elements of the face such as eyes, mouth and eyebrows. This paper is a research report that describes just one of the small steps into achieving the goal of autonomous emotion detection, which is detecting eyebrows with the Kinect. Eyebrow detection was chosen for research because eyebrows play a very important role in emotion expression, as it can be seen further in this paper.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Statement and Motivation of Research</b>	<b>1</b>
2.1	Facial Expressions and the importance of Eyebrows . . . . .	1
2.2	RGB-D Sensors. The Kinect . . . . .	4
2.3	Related work . . . . .	4
<b>3</b>	<b>Research Investigation</b>	<b>5</b>
3.1	Eyebrow detection . . . . .	5
3.2	Frontal-pose image extraction . . . . .	8
3.3	Implementation aspects . . . . .	11
<b>4</b>	<b>Evaluation and Results</b>	<b>11</b>
4.1	First experiment . . . . .	11
4.2	Second experiment . . . . .	12
<b>5</b>	<b>Future Work</b>	<b>19</b>
<b>6</b>	<b>Conclusion</b>	<b>20</b>

# 1 Introduction

The concepts of facial expression recognition and emotion detection have their basis in the studies and research pursued by psychologists Ekman and Friesen. They reached to the conclusion that regardless age, sex, racial and cultural matters, there are seven different facial expressions that are displayed equally by humans: anger, disgust, fear, sadness, happiness, surprise and contempt[1], [2], [3]. These facial expressions are displayed via well defined movements of different facial muscles, as it will be presented later in this paper.

In recent years, this idea became of increasing interest in different fields of Computer Science such as robotics, computer vision, artificial intelligence and affective computing, a more recent branch of computer science introduced by Rosalind Picard[4]. The idea of a computer unit able to accurately detect facial expression would have a great impact on technology, followed by several, very interesting applications such as a improved computer-human interaction, virtual avatar animation, autonomous emotion analysis and others.

Detecting eyebrows is small, yet very important part for autonomous emotion analysis. In our research we worked on a method for detecting eyebrows with the Kinect recording the subject from different angles, not only frontal. The final purpose of our research is to record data for emotion interpretation. Therefore, our focus is to accurately estimate the inner eyebrow key points as it will be described further in this report.

The paper is structured in six sections. The first one is the introduction that has the purpose of giving a rough idea of the concepts and the scientific area under which this document is located. Section 2, 'Statement and Motivation of research', will explain the main concepts and describe the theoretical basis and purpose of the research problem. In this section the reader will learn about facial expressions and the importance of eyebrows, find out some information about the Kinect sensor but also about previous work done regarding eyebrow detection and tracking. Further, 'Research Investigation' will describe in detail the individual steps in the method that we adopted for eyebrow detection. Then, 'Evaluation and Results' will describe the experiments under which our solution was tested and the obtained results and conclusions. Finally, 'Future work' will present a rough idea about the future plans with this project and the 'Conclusion' will summarize the main aspects of our report.

## 2 Statement and Motivation of Research

### 2.1 Facial Expressions and the importance of Eyebrows

As stated in the introduction, the facial expression concept has at its base the study made by psychologists Ekman and Friesen[1] on human faces and how they react under different emotions that the subject encounters. It was discovered that there are seven main emotions that create consistent facial expressions displayed universally by subjects

regardless sex, age, race or culture. These are: *anger, disgust, fear, sadness, happiness, surprise, and contempt* (shown in Fig. 1).

Eyebrows have a very important role in facial expressions, as they display relevant movements under certain emotions[6]. Here are some examples of interpretations:

- **Lowered** - usually indicates *deception* when accompanied by a lowered head; it is also considered a sign of *dominance*
- **Raised** - indicate *surprise*
- **Middle raised** - this can show *sadness* or *fear*, but sometimes also *relief*
- **Middle lowered** - usually represents *anger* or *frustration*
- **Middle together**(pulled together but not sloping) - can be a sign of *confusion*

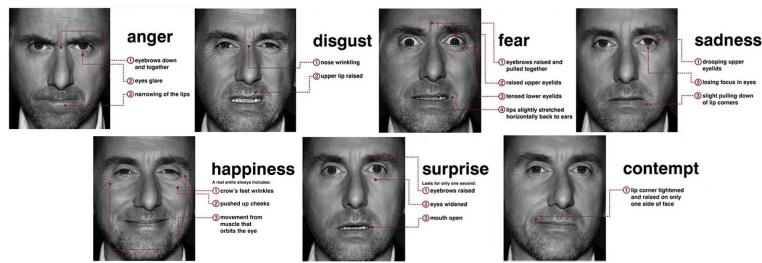


Figure 1: The 7 emotions. Actor Tim Roth in 'Lie to me' series [20]

Although not much attention was given to eyebrows regarding facial expressions(most theories had more focus on eyes and mouth), their true importance was demonstrated in some(more recent) psychological experiments. Two of these experiments will be further highlighted in this paper.

The first one, described in [5], creates a contrast between eyes and eyebrows in terms of which of these elements is a better tracker for face recognition. The subjects were shown a large number of photos with faces of famous Caucasian man an women of different ages, that were expressing different emotions. From each photo, there were two samples: in one the eyes were removed from the face, leaving only the eyebrows visible, and in the other, they eyebrows were removed, leaving the eyes (as you can see in Fig. 2 - left).

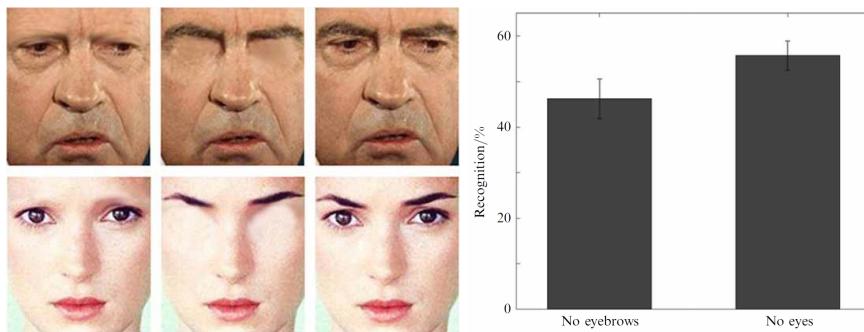


Figure 2: Samples for the experiment(left). Results(right). [5]

The result was that in most cases the participants were more accurate in recognizing the faces from the samples with missing eyes(judging only by the position of eyebrows)

than from the samples where eyebrows were missing (as described in Fig. 2 - right). This proves that eyebrows are a better recognizable pattern when it comes to human face than eyes are. It is easier to remember the shape of one's eyebrows than the shape and color of its eyes. Relating this idea to the presented topic, this should mean that the eyebrows shape can be detected and tracked accurately.

The second experiment[6] is the theoretical core of our research. This experiment focuses on how two important face muscles act under different emotional stimuli. These muscles are the *corrugator supercilii* which manipulates the inner part of the eyebrow and the *zygomaticus major* which lies from the corner of the mouth to the temple and generates the smile. The experiment consisted in testing which of these muscles describes a more relevant electromyographic activity under different emotions. The subjects were connected to an electromyograph <sup>1</sup> that was recording the activity of the two muscles and were presented different stimuli: pictures, sounds and words. The result was that in most cases the *corrugator supercilii* had a linear level of contraction, growing from 'good' to 'bad' stimuli, whereas the *zygomaticus major* had a quadratic evolution under the same stimuli(Fig 3). In other words people were tending to smile at both extremes of their feelings (good and bad). This means that interpreting smile as a good emotion can be deceiving as smiling can also express a bad emotion(such as shyness or contempt). In contrast, eyebrows having a more linear evolution, can give a more relevant approximation of bad feelings. This proves that the *corrugator supercilii* would have the best resolution for negative evaluations of the affective state of the subject(good or bad).

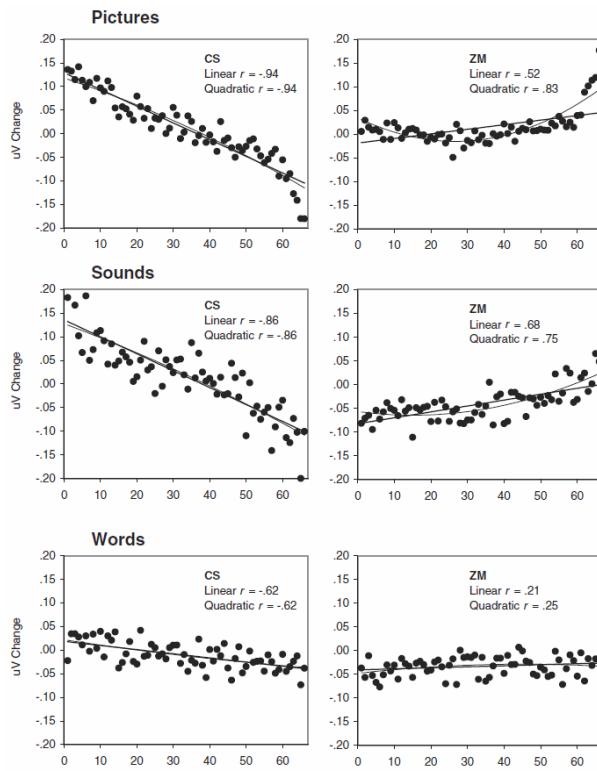


Figure 3: Experiment results on each type of stimuli: the linear evolution of the *corrugator supercilii*(left) and the quadratic evolution of the *zygomaticus major*(right) [6]

<sup>1</sup>An **electromyograph** is a device used for recording the electrical activity of muscles.

## 2.2 RGB-D Sensors. The Kinect

RGB-D sensors combine the RGB color data from a normal (mono) camera with depth data per-pixel<sup>2</sup>. This upgrade from a normal camera came with multiple use-cases such as *3D mapping and modeling, natural user interface, accurate image segmentation*, and others.

The Kinect was introduced by Microsoft in 2010 as an accessory of the game console XBOX-360. This device was a revolution in the world of RGB-D cameras, not necessarily technologically but in terms of price. It costed 150\$, which was considerably less than any device of its kind at the time. This drastically increased developer's interest in RGB-D cameras and their applications [7].



Figure 4: The Kinect and components [7]

The Kinect device consists of: a normal RGB camera, an IR laser emitter and receiver, a moving motor, an accelerometer, and a series of 4 microphones (Fig. 4). The camera can capture 640x480 images at 30 fps. It also disposes of features including color saturation and flicker avoidance. The IR emitter and receiver compute the depth data for each pixels. The emitter sends IR light that reflects on objects and is captured back by the receiver, which records 1200x960 resolution at 30 Hz and has a field of view of 45° vertically and 58° horizontally and a range between 0.8 and 3.5 meters. The motor moves the device head to better capture the objects that need to be present in the frame. The accelerometer is aware of the position of the device, taking into account the environment and user's position. The microphones process 16 bit audio at 16kHz [7].

Such a device would be the best available option for the purpose presented in this paper as the color and depth data combined would help for a more accurate detection of face and eyebrows.

## 2.3 Related work

Previous work has been done in the area of eyebrow detection that resulted in several algorithms and methods for their segmentation. The method presented in [8] uses a seed fill procedure for segmenting eyebrows in grayscale images. After detecting the eyes and approximating an eyebrow region of interest(ROI), they subtract five columns of pixels, equally distanced, from the eyebrow ROI. Further, they use a weighing procedure on

<sup>2</sup>the distance from the sensor to what is displayed on each particular pixel.

each of the rows to plant seeds in the eyebrows from which they will start the segmentation. The weighing procedure regards both the darker color of the eyebrow, but also the distance from the middle of the pixel column, assuming that the eyebrow is more probable to be in the middle of the ROI. We found this method not very robust since, although it gave good results when the eyebrow was in normal position, the results were not as good when the eyebrow was highly raised or lowered.

Another method is described in [10]. It processes the image by the level of red in the eyebrow region: a skin pixel has a higher level of red than an eyebrow pixel[13]. With this approach, they create a grayscale pseudo-hue matrix that would highlight the eyebrow shape. The matrix is further processed into a binary image with the segmented eyebrow. This is done by applying an adaptive thresholding algorithm which was first introduced in [11] for processing document images and later modified for lip segmentation[12]; version that was also adopted for eyebrow segmentation in [10]. The original algorithm is based on k means summation of local mean and constant local standard deviation. The modification consist in applying global mean thresholding on the final result. We found this a very good approach. However, in light changing conditions the colouring is also bad and therefore the pseudo-hue will not be properly constructed. Our solution for segmenting eyebrows in color image is a slightly improved version of this method as we also take into consideration the intensity plane of the image.

### 3 Research Investigation

Most of the eyebrow detection existing approaches are only giving good results on RGB images of subjects in frontal pose. Our solution is more complete, covering cases when subjects have slight face inclination both vertically and horizontally. By using the depth information from the Kinect, we extract the point cloud of the face, rotate it in frontal pose and then apply the eyebrow detection method on the image extracted from that frontal position. Therefore, the two main parts of our solution that will be discussed in this paper are eyebrow detection and frontal pose image extraction.

#### 3.1 Eyebrow detection

As stated earlier our method is based on the solution proposed in [10]. Asuming a color face image given in frontal pose, the first step in our approach is to estimate the eyebrow ROI(region of interest - the smallest region within the face that contains the eyebrow no matter its position: lowered, raised). To do that, we first detect the eyes inside the face using the OpenCv haar-cascade methods[25]. Given their position, estimate the eyebrow ROIs as it can be shown in Figure 5.

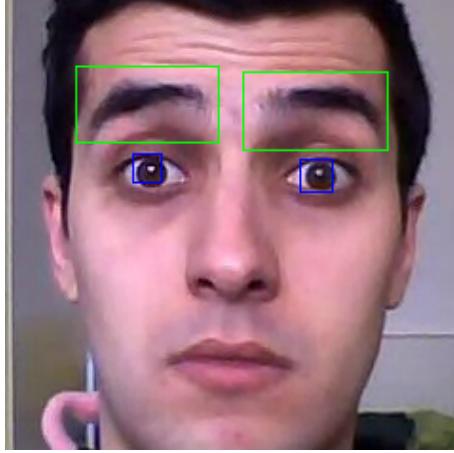


Figure 5: Face region with detected eyes and estimated eyebrow ROI

From this point on all the processing will be done on these regions. From the eyebrow region we further compute a pseudo-hue matrix to highlight the distinction between eyebrow and skin. Skin color has commonly a larger red value than hair color[13], therefore this is one of the criteria for computing the pseudo-hue. However, image color is not always ideal and thus we also take into consideration the image intensity(eyebrows being usually darker than the skin). The pseudo-hue matrix is computed as follows:

**ALGORITHM 1: PSEUDOHUE**

**Data:** eyebrow rgb image - *browROI* (Figure 6a )

**Result:** pseudo-hue matrix - *pseudoHue* (Figure 7)

- For less noise and better results blur the eyebrow ROI by a  $5 \times 5$  kernel (Figure 6b).
  - Convert RGB image into HSI. Apply histogram equalization on the intensity plane (Figure 6c) for better colouring and convert HSI back to RGB (Figure 6d). Keep the equalized intensity plane.
  - forall pixels in browROI do**
  - compute the pseudoHue value of the pixel  $h = \frac{r}{g+b} \cdot i$  ▷ where  $r, g, b$  are the RGB components and  $i$  is the intensity value of the pixel. Keep track of the maximum and minimum values  $h_{min}$  and  $h_{max}$  within the matrix*
  - forall pixels in pseudoHue do**
  - normalize the pseudoHue value  $h_{norm} = \frac{h - h_{min}}{h_{max} - h_{min}}$*
  - For a better result, contrast the normalized pseudo-hue exponentially by a chosen factor  $k$  (we set  $k = 3$ ):
  - forall pixels in normalized pseudoHue do**
  - val = val<sup>k</sup>* ▷ since all values are in  $[0,1]$
  - Re-normalize

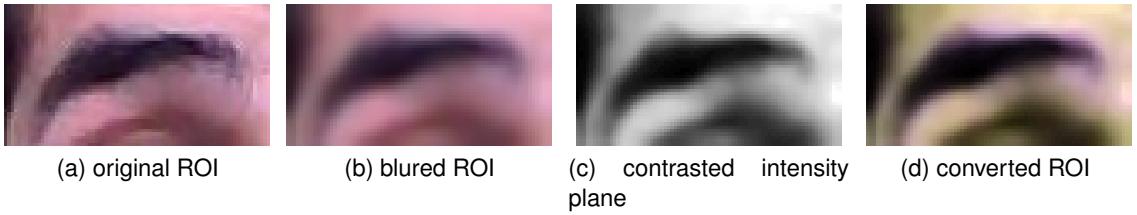


Figure 6

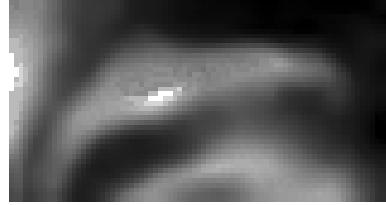


Figure 7: Pseudohue.

We have now created our pseudo-hue as in Figure 7. The next step in our process is to obtain a binary image with the segmented eyebrow. We do this by applying the *adaptive thresholding algorithm* described in [10]. For each value in our pseudo-hue we apply *localThreshold* as shown in equations 1, 2 and 3 and in the end, we threshold the result once more over the global mean.

$$localThreshold(x, y) = \mu_l(x, y) + k \cdot \sigma^2(x, y) \quad (1)$$

where  $\mu_l$  is the local mean(described in equation 2) of a  $w \times w$  local threshold window and  $\sigma^2$  is the local variance(described in equation 3) with the constant parameter  $k$ . One should calibrate this constant factor for better results. We used different values in the range [-0.3, 0.5] for different subjects and conditions.

$$\mu_l(x, y) = \frac{1}{w^2} \left[ \sum_{j=y-\frac{w}{2}}^{y+\frac{w}{2}} \sum_{i=y-\frac{w}{2}}^{x+\frac{w}{2}} (f(i, j)) \right] \quad (2)$$

$$\sigma^2(x, y) = \frac{1}{w^2} \left[ \sum_{j=y-\frac{w}{2}}^{y+\frac{w}{2}} \sum_{i=y-\frac{w}{2}}^{x+\frac{w}{2}} (\mu_l(x, y) - (f(i, j))^2) \right] \quad (3)$$

where  $w$  is the width of the local threshold window. The optimal size of the window can vary with regards to the size of the eyebrow ROI(e.g we used  $w = 10$  for an average of  $120 \times 60$  ROI size).

We further compute the global mean of the resulted matrix and threshold it into a binary image over the global mean(Figure 8). This is a meaningful approach because the number of "eyebrow" pixels within the eyebrow is larger than the number of "eyebrow" pixels outside the eyebrow and also the eyebrow area is smaller than 50% of the total ROI. In most cases there are multiple blobs in the resulted binary image(due to portions of



Figure 8: The binary thresholded image (after erode-dilate)

the hair or the darker eye-lid area), but the eyebrow blob is the largest. Therefore, after 2-4 iterations of applying *erosion* and *dialtion* operation for better separation and noise removal, we select the largest blob, aka. the eyebrow, and extract its contour(Figure 9).



Figure 9: Extracted eyebrow blob. Segmented eyebrow. Extracted key-point.

The key eyebrow point that we are interested in is the inner point. It is actioned by the *corrugator supercilii* muscle[6] and is the most moving part of the eyebrow. Therefore, the final step is it's estimation. We iterate through the contour and take the farthest point from the eyebrowROI origin(down-left for the left eyebrow, and down-right for the right eyebrow).

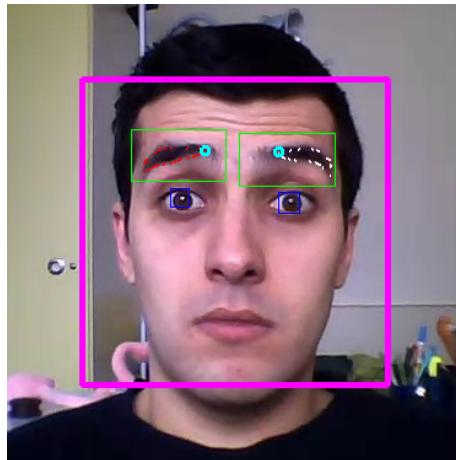


Figure 10: Face with detected eyebrows.

### 3.2 Frontal-pose image extraction

As stated earlier in this paper, extracting eyebrows in RGB images where the subjects have head inclination would give inaccurate results: only one eyebrow is properly segmented or the eyebrow movement is not perceived in the local face reference system, but globally in the image(Figure 11). To overcome this issue we use the depth information from the Kinect sensor to extract the frontal-pose face image and then apply eyebrow

detection on it. This enables us to get a better eyebrow segmentation with the head tilted in different angles (both vertically and horizontally). Besides the technical improvements in eyebrow segmentation, the capability of observing the face frontally regardless its real angle, will also provide great benefits for the psychological interpretation of eyebrows. As described in [19], the head angle of regard can significantly alter the emotional interpretation of facial expressions.

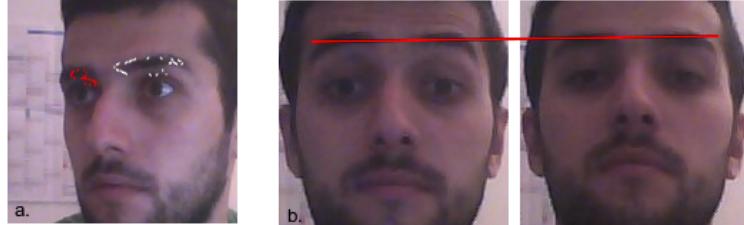
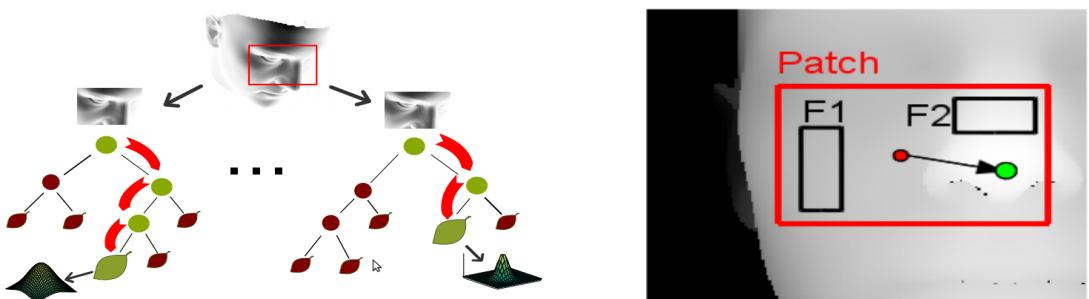


Figure 11: Limitations of eyebrow detection in RGB: a. bad segmentation, hard to interpret due to horizontal tilt; b. eyebrows seem to be at the same level although they are in raised position in the left picture and in normal position in the right picture where the head has vertical tilt.

The first step is to rotate the color point cloud extracted from the Kinect so that the face is translated with its center in the origin and rotated in frontal position with respect to the  $X$ -axis. For this to be possible, the position and orientation of the face in the original point-cloud need to be known. We use the *Random Forests for Real Time Head Pose Estimation* method described in [14] for this matter. The method uses Random Regression Forests to map large training datasets. This approach is adopted because, unlike a standard decision tree, a collection of randomly trained trees has a high generalization power, and therefore provides a more homogeneous decision making. Figure 12a exemplifies the idea. The training is done by filling the trees with patches sampled randomly from the training data(Figure 12b).



(a) " Example of regression forest. For each tree, the tests at the non-leaf nodes direct an input sample towards a leaf, where a real-valued, multivariate distribution of the output parameters is stored. The forest combines the results of all leaves to produce a probabilistic prediction in the real-valued output space." (source for image and description in [14])

(b) " Example of a training patch (larger, red rectangle) with its associated offset vector (arrow) between the 3D point falling at the patch center (red dot) and the ground truth location of the nose (marked in green). The rectangles F1 and F2 represent a possible choice for the regions over which to compute a binary test." (source for image and description in [14]).

Figure 12: Random forests

The *Head Pose estimation* method will return the pose of the head(or face) which is compounded of the head position and head orientation. With these values, we filter everything but the face from the pointcloud (we assumed size of a face to be  $18\text{cm}$  width and  $20\text{cm}$  height and filtered everything outside this area). Then we translate the filtered point cloud such as the face is centred in the origin and rotate it so the face orientation is towards the  $X - axis$ . The face point cloud will now always face the  $X - axis$  not matter its actual orientation(see Figure 13). Further, the RGB information of the point cloud is extracted in a parallel projection image of resolution  $90 \times 100$  pixels from the  $X - axis$  perspective(front of the face). As it can be seen in Figure 14a, the result image is highly fragmented, specially when the head was tilted.

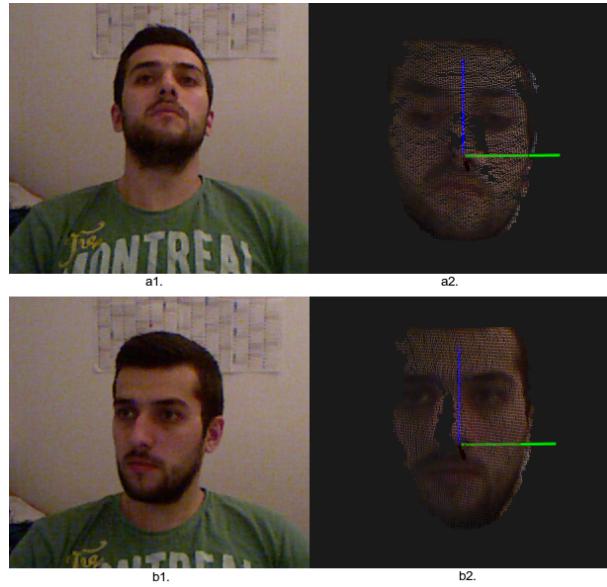


Figure 13: a. Vertical head tilt and rotated face point cloud; b. Horizontal head tilt and rotated face point cloud. It can be clearly seen that a non-frontal face pose was rotated so that it becomes frontal.

To increase the quality of the image we compute the color of every black (fragmented) pixel by mean interpolation of the non-black surrounding pixels. In the end, the eyebrow detection method is applied on a scaled version of the frontal-pose image (Figure 14b)

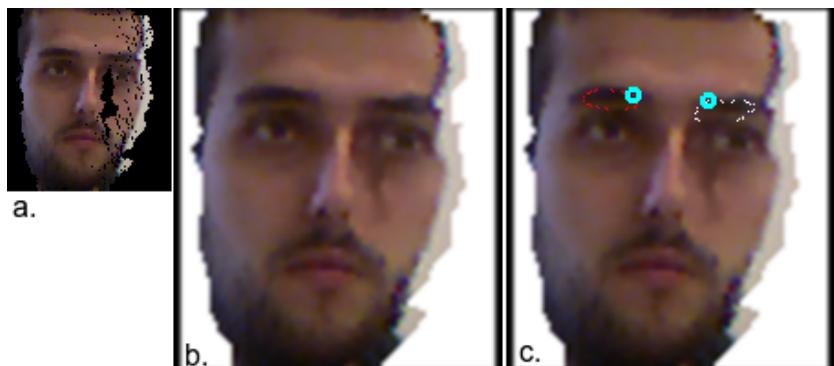


Figure 14: a. Fragmented image extracted from point cloud; b. Repaired image; c. Segmented eyebrows on the respective frame.

### 3.3 Implementation aspects

The implementation of our solution was written in C++. We used the OpenCv library for implementing the eyebrow segmentation part. The face and eye detection methods from OpenCv are based on Viola-Jones[9] solution with trained *haar cascades*. Also, OpenCv has different image processing functionality such as blurring and contrasting(histogram equalization) that were used for segmentation. The *openni* driver from ROS(Robotic Operating System - [21]) was used for capturing point cloud data from the Kinect, that was further manipulated using PCL(Point Cloud Library - [24]). For the head pose estimation we used a modified version of the ROS node *head\_pose\_estimation* [23], a ROS adaptation of the method described in [14] [22].

## 4 Evaluation and Results

We evaluated our solution by pursuing two experiments. The first one has the purpose to test the eyebrow key-point estimation in RGB images and the second one to evaluate the two different parts of our method as a whole.

### 4.1 First experiment

For the first experiment 10 different images of variate quality with faces in frontal pose were selected. For each image we cropped the eyebrow regions (20 eyebrows) and pre-recorded the coordinates of the eyebrow inner key-points as a target result. All the cropped eyebrow images are further run through our algorithm for eyebrow segmentation and key-points estimation(no frontal-pose extraction) and the results were collected by comparing the values for the extracted key-points with the pre-recorded target values (see Figures 15 and 16).

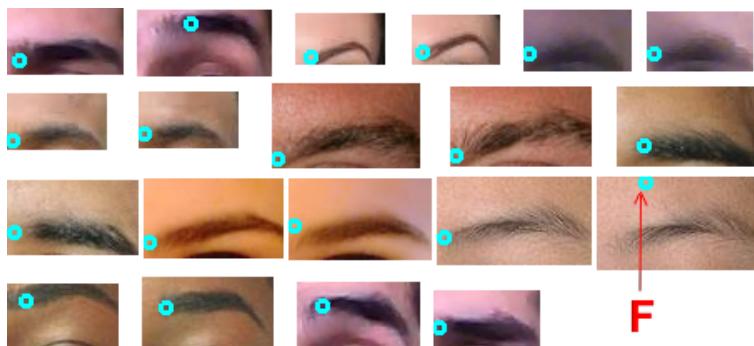


Figure 15: Test eyebrow images after *key-point extraction*. A fail can be remarked in the picture labelled with *F*.

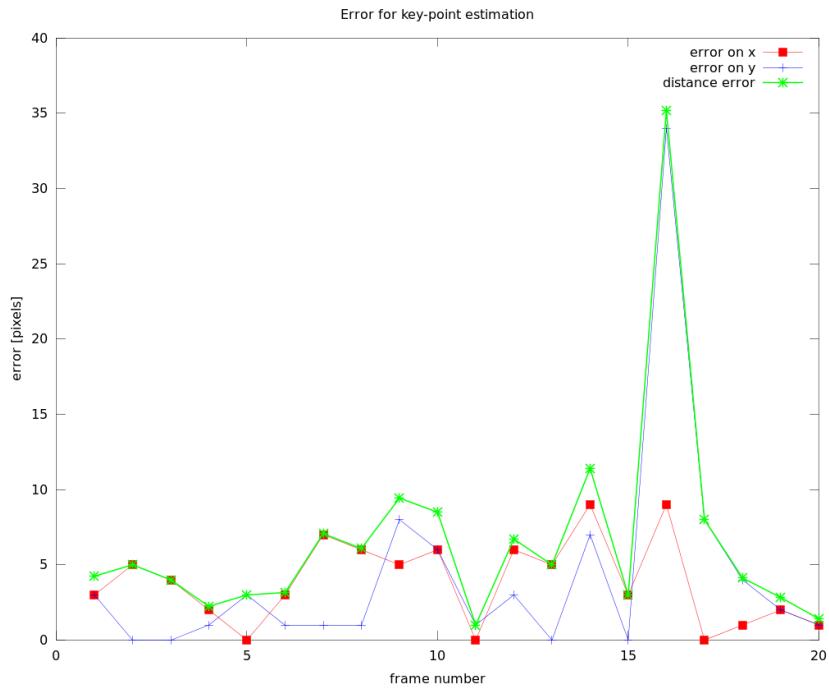


Figure 16: This plot displays the error in pixels on x coordinate, y coordinate and direct distance between the pre-recorded point and the extracted point. The fail displayed in Figure 15 is also visible here. The average error was also computed: *x*: 3.85 pixels, *y*: 4.2 pixels and *distance*: 6.568 pixels.

## 4.2 Second experiment

The second experiment was designed to evaluate our method as a whole: eyebrow segmentation and key-point estimation together with the frontal-pose extraction. Three subjects with different skin and hair color participated to the experiment(figure(17)). They were recorded with the Kinect under 5 different head orientations: frontal, 22.5° upward, 10° downward, 20° leftward and 20° rightward (we have observed that tilts larger than these would lead to bad results). For a better control of the head-tilt angles of the participants, we set up markers at measured positions vertically and horizontally around the Kinect as in Figure 18.

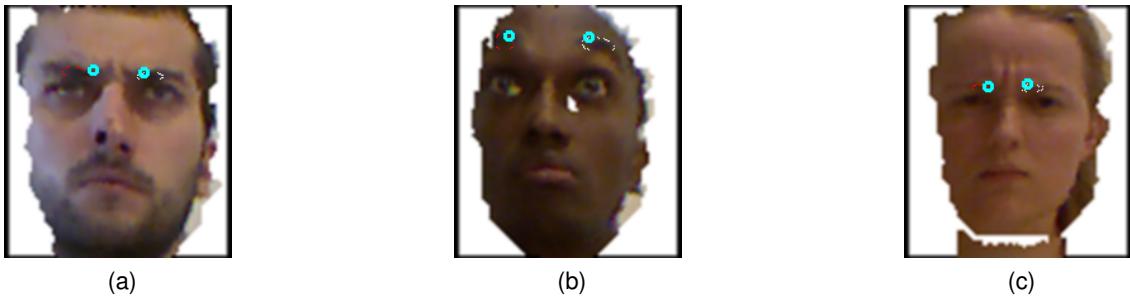
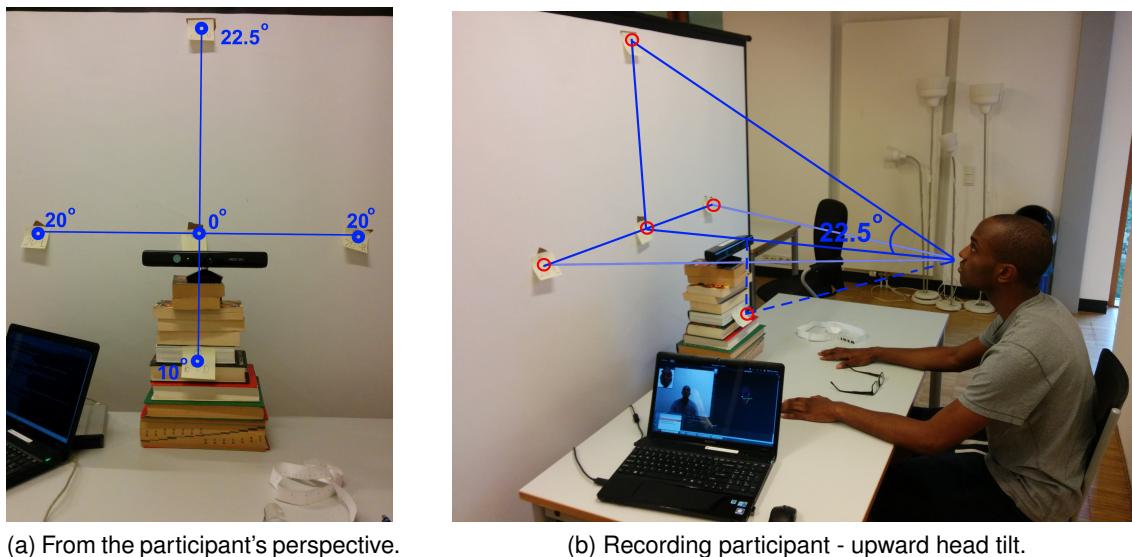


Figure 17: Participants of the second experiment.



(a) From the participant's perspective.

(b) Recording participant - upward head tilt.

Figure 18: Setup of the second experiment.

The evaluation was done over two variables of interest that we recorded: the distance between they eyebrow key-points and the average height of the key points, with correspond to the eyes line. The two values are described as  $D$  and  $H$  in Figure 19. Further in this paper, the two values will be noted as D and H.

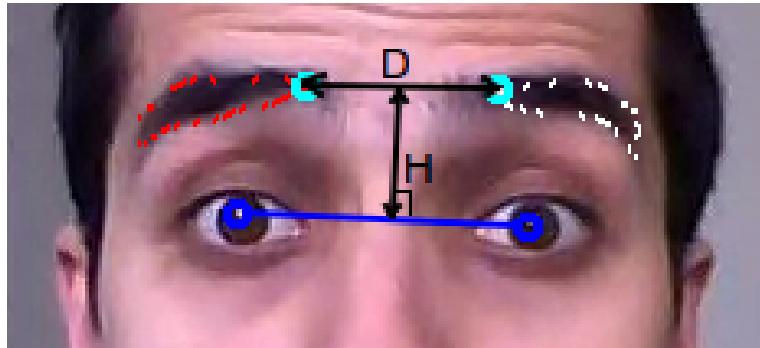
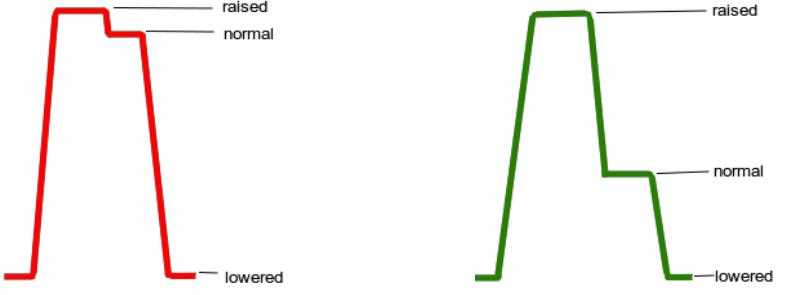


Figure 19: The two values of interest for the second experiment.

For each recording, participants were instructed to repeatedly move their eyebrows by following a pre-defined pattern: **lowered** for a couple of seconds, then **raised, normal position**(middle) and back to **lowered**. Therefore, a set of results was considered positive if the target shapes from Figure 20 were recognized in the plots constructed over the recording.



(a) Shape for D. The distance between the eyebrows usually does not vary much between *raised* and *normal* positions.

(b) Shape for H. Each different position is clearly visible in the shape.

Figure 20: The expected evolution of values of interest D and H (19) over the pre-defined movement pattern. *Note: These shapes are not drawn at scale. Their purpose is just to give the reader an idea about the expected results.*

A sample set of results is plotted in Figures 21, 22, 23, 24 and 25 for frontal pose, head tilted upward, downward, leftward and rightward respectively. In each figure, there are four plots: red, blue, green and black. The red and green plots represent the raw data recorded from the Kinect(with noise) and the blue and black plots are a filtered version of the previous two. *Sliding window average filter* with windows of different sizes was applied to obtain the filtered plots.

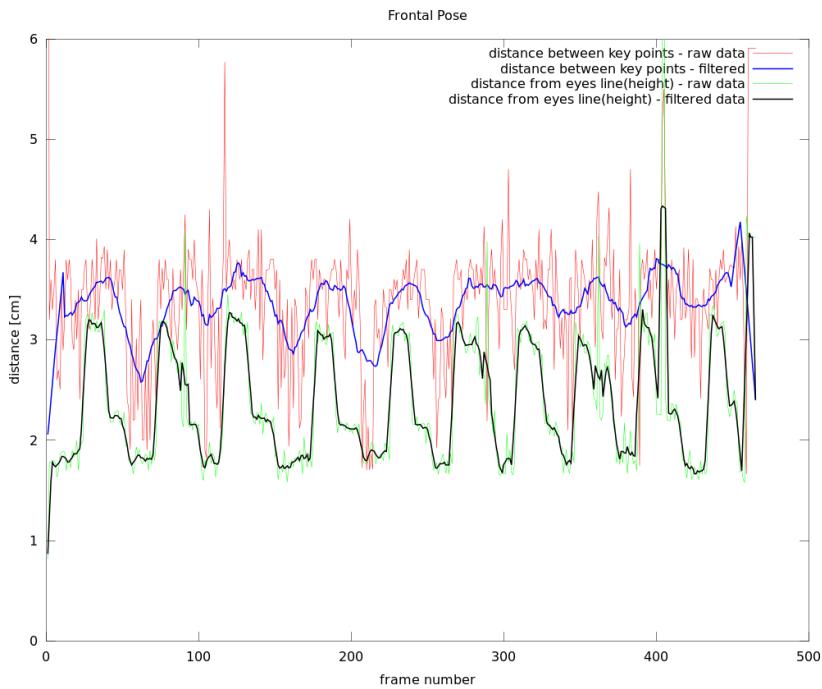


Figure 21: Plots for frontal pose. The window size used for filtering was 21 for D(blue) and 5 for H(black). The results are positive: the expected pattern for H is clearly followed. Also the plot for D is relevant, although less accurate.

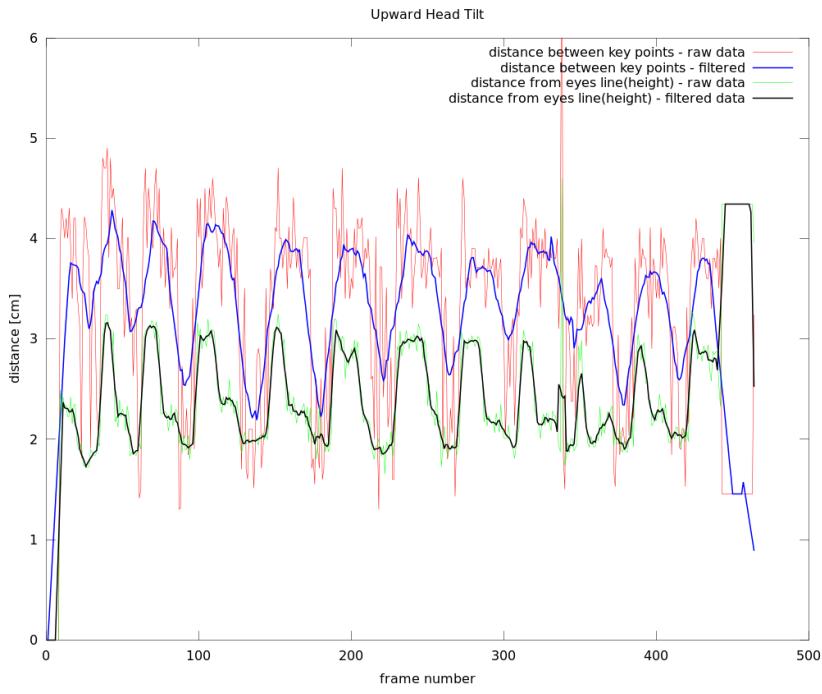


Figure 22: Plots for upward head tilt. The window size used for filtering was 15 for D(blue) and 5 for H(black).Positive results.

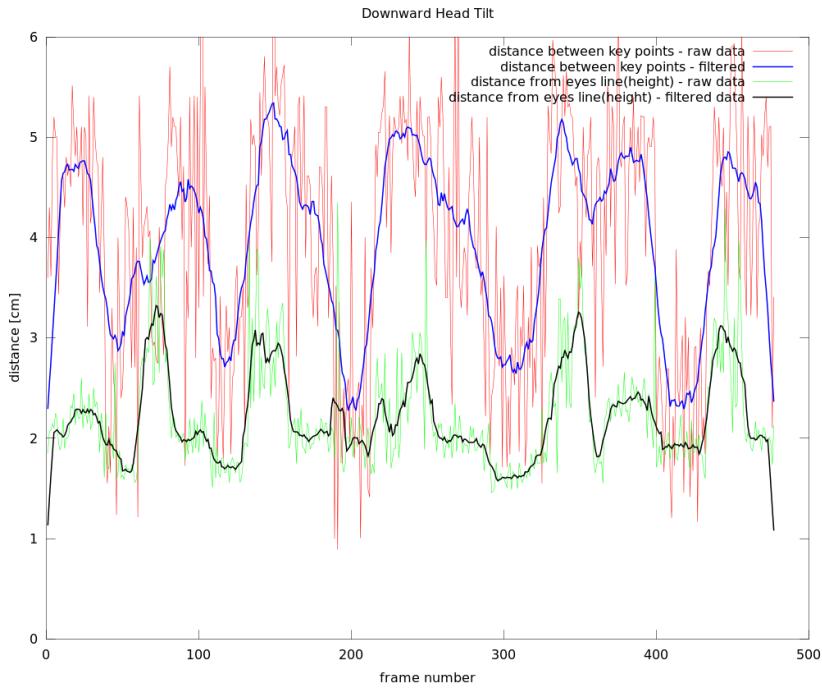
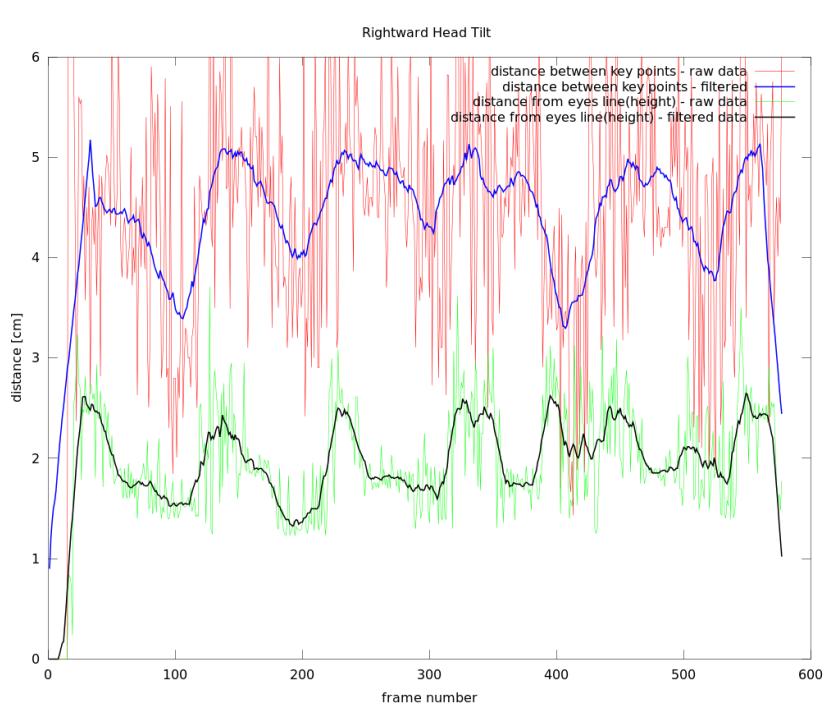
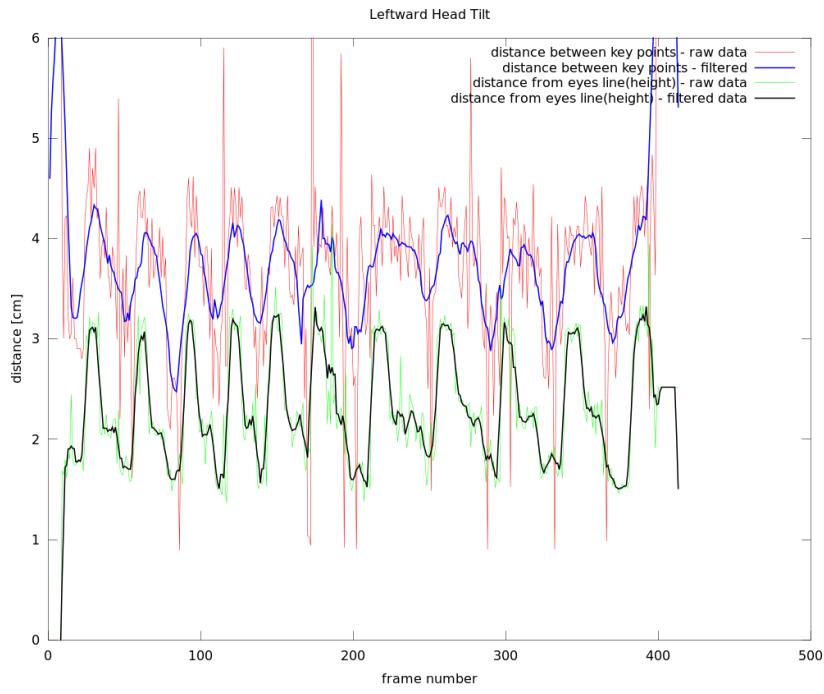


Figure 23: Plots for downward head tilt. The window size used for filtering was 19 for D(blue) and 9 for H(black). The quality of the results is significantly lower: high level of noise and inaccuracies in following the expected pattern. The reason is that at downward tilt, the extracted frontal pose is highly fragmented(lots of points in the eye and eyebrow area are lost), and therefore the image is distorted. This leads to often bad segmentation.(see Figure 26a)



After plotting the results, we pursue further evaluation on a representative interval of data from each recording. This interval should contain the evolution of one single period-pattern:lowered-raised-middle-lowered(e.g from the data plotted in Figure 21 we have chosen the interval [160,220] which clearly contains one period only). Then, standard deviation for D and H is computed over the raw data with correspond to the filtered data(equation 4). This should evaluate the level of noise and therefore the quality of the results. Moreover, we approximate a failure rate(for both D and H). This is done by estimating on what level does the plot follow the expected shapes (Figure 20) and also by eliminating the unrealistic results(the distance between eyebrows and their height should not exceed 6cm and 4cm respectively). Tables 1 and 2 display our final results.

$$\sigma = \sqrt{Variance} = \sqrt{\frac{\sum_{i=a}^b (r(i) - f(i))^2}{b - a}} \quad (4)$$

where  $\sigma$  is the standard deviation,  $[a, b]$  is the chosen interval and  $r(i)$  and  $f(i)$  are the raw value and filtered value respectively, at a certain position.

It can be clearly observed from both the presented plots and Table 1 that our method returns more accurate results for H than for D. The reason is that the color of the eyebrow usually varies over it's length, being lighter at the extremities and darker in middle and sometimes only the middle of he eyebrow is segmented(and therefore a large distance between the eyebrows).

Participant	Value	Frontal	Upward tilt	Downward tilt	Leftward tilt	Rightward tilt
1	D	0.50831 (v)	0.53682 (v)	1.0526 (v)	0.59521 (v)	1.8725 (i)
	H	0.12897 (v)	0.11860 (v)	0.27971 (v)	0.19371 (v)	0.33052 (i)
2	D	0.92573 (i)	0.76221 (v)	2.00488 (i)	1.0306 (i)	2.29773 (i)
	H	0.30174 (v)	0.22611 (v)	1.21569 (i)	0.6448 (i)	0.5239 (i)
3	D	0.88578 (v)	0.78736 (v)	1.3853 (i)	0.80216 (i)	1.95527 (i)
	H	0.25286 (v)	0.28915 (v)	0.36217(i)	0.3196(v)	0.31433(i)

Table 1: Standard Deviation Results(in cm). The *v* and *i* decide whether the result is valid or invalid. This binary evaluation has the purpose of showing the reader which data is more relevant; it follows subjective criteria decided over the data plot: there should be a clear difference between the three height levels in the hight pattern and the width should also follow its pattern(smaller at *lowered* and larger at *middle* and *raised*) for the result to be valid.

Participant	Value	Frontal	Upward tilt	Downward tilt	Leftward tilt	Rightward tilt
1	D	25%	25%	45%	30%	60%
	H	15%	10%	35%	15%	45%
2	D	50%	50%	80%	60%	70%
	H	50%	40%	70%	70%	80%
3	D	40%	40%	70%	50%	75%
	H	30%	25%	50%	30%	70%

Table 2: Failure rate.

The two tables show that for the second and third participants the results are less positive: high level of noise and high failure rate. These participants represent extreme cases for which our method needs to be improved. Unlike the first participant, for which the results are positive, the latter two do not have enough contrast between their skin color and eyebrow color and therefore our method often fails in segmenting the eyebrows(see Figures 26c, 26d)

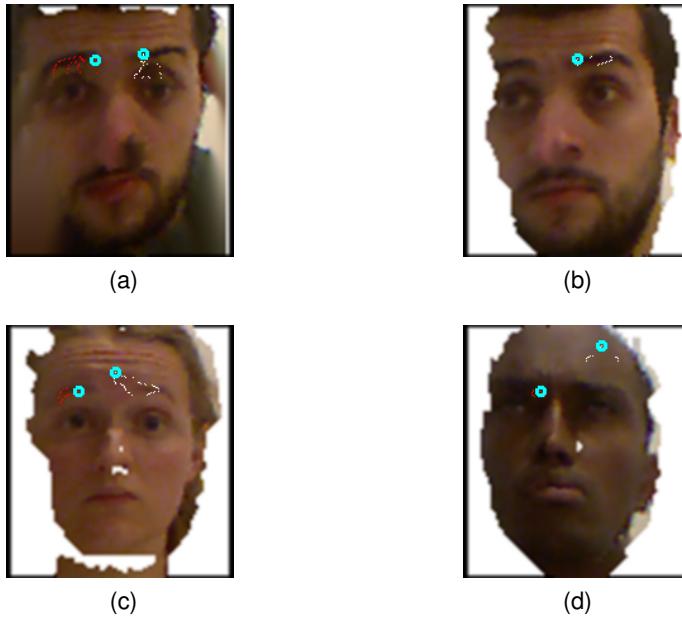


Figure 26: Limitations. Segmentation failures.

This experiment provided us with a fair view about the capabilities and limitations of the current state of our method for eyebrow detection with the Kinect. Further research is needed for better results. Moreover, an amount of 100 GB of data were recorded during this experiment as ROS bag-files.

## 5 Future Work

Different things are in plan for the future of our project. First of all, we plan on increasing the robustness of our solution by pursuing more tests and experiments to obtain a wider set of results and adapt our approach accordingly. The time efficiency of our method could be also improved(the current processing time is approximately 0.7 seconds per frame). We consider that research in skin and hair color detection could lead to considerable improvements in eyebrow segmentation. Methods such the ones described in [18] and [13] are in our interest. Also, since eyebrows are rather fine elements inside the face, using a sensor with a higher resolution such as the second version of the Kinect would drastically improve our results. Due to the limited resolution of the Kinect sensor used in our experiments, the processed and scaled image data used for eyebrow segmentation was of low quality and even distorted, and therefore resulted in negative results(see section *Evaluation and Results*). Finally, on a long term perspective, we plan to expand or research to autonomous full face feature detection and tracking for creating robust solutions for emotion detection and other purposes.

## 6 Conclusion

This paper presents a method for *eyebrow detection for emotion analysis using the Kinect* as a Guided Research topic for B.Sc. Computer Science major at Jacobs University Bremen. We presented the importance and role of eyebrows in facial expression interpretation. The two main parts of our method: eyebrow detection and frontal-pose face image extraction are rigorously described. We have shown the capabilities and limitations of our method by pursuing different experiments that provided relevant results. Our solution comes with a new touch for eyebrow detection technologies, which is to use the depth data from the Kinect to segment eyebrows when the face is not oriented towards the camera. This provides more relevant results for real-world situations when the recorded subjects might not be frontally facing the camera. Further research in this area will come with important benefits in different fields of Computer Science such as robotics and artificial intelligence, as it will improve machines' understanding of humans, but also provide the field of Psychology with a better research and development in the branch of facial expressions.

## References

- [1] P. Ekman and W. Friesen "Facial Action Coding System: A technique for the measurement of facial movement", Palo Alto, CA, USA, 1978
- [2] P. Ekman "Are there basic emotions?", American Psychology Association, University of California, San Francisco, CA, USA, 1978
- [3] P. Ekman "Basic emotions", University of California, San Francisco, CA, USA in *Handbook of Cognition and Emotion* by T. Dalgleish and M. PowerS, 1999
- [4] R. Picard "Affective Computing", MIT Press, Cambridge, Massachusetts, USA
- [5] J. Sadr, I. Jarudi and P. Sinha "The role of eyebrows in face recognition", Department of Brain and Cognitive Sciences, MIT, 7 March 2003, Massachusetts USA
- [6] J. T. Larsen<sup>a</sup>, C. J. Norris<sup>b</sup> and J. T. Cacioppo<sup>b</sup> "Effects of positive and negative affect on electroencephalographic activity over 'zygomaticus major' and 'corrugator supercilii'" , **a:** Department of Psychology, Texas University, Lubbock, Texas, USA, **b:**Department of Psychology and Institute for Mind and Biology, University of Chicago, Chicago, Illinois, USA, 2003
- [7] L. Cruz, D. Lucio and L. Velho "Kinect and RGBD Images: Challenges and Applications", IMPA-VUSGRAF Lab, Rio de Janeiro, Brazil, 2012
- [8] L. Florea and R. Boia "Eyebrows Localization for Expression Analysis", University Politehnica of Bucharest, Bucharest, Romania, 2011
- [9] P. Viola and Michael J. Jones "Robust real-time face detection", International Journal of computer Vision, vol. 54(2), p. 137-154, 2004
- [10] A. Majumder, M. Singh and L. Behera "Automatic Eyebrow Detection and Realization of Avatar for real time Eyebrow Movement", Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India, 2012
- [11] W. Niblack "An introduction to digital image processing", Strandberg Publishing Company Birkeroed, Denmark, Denmark, 1985.
- [12] A. Majumder, L. Behera, and V. Subramanian, "Novel techniques for robust lip segmentations, automatic features initialization and tracking, in Signal and Image Processing", ACTA Press, 2011.
- [13] G. Osman, M. S. Hitam and M. N. Ismail "Enhanced skin colour classifier using rgb ratio model", *International Journal on Soft Computing (IJSC)* Vol.3, No.4, November 2012
- [14] G. Fanelli, J. Gall and L. Van Gool Leuven "Real Time Head Pose Estimation with Random Regression Forests"
- [15] J. L. Moreira, A. Braun and Soraia R. Musse "Eyes and Eyebrows Detection for Performance Driven Animation", Pontifica Universidade Catolica do Rio Grande de Sul, Graduate Program in Computer Science, Virtual Human Laboratory, Porto Alegre, Brazil, 2010

- [16] Q.R. Chen, W.K. Cham and H.T. Tsui "A method for estimating and accurately extracting the eyebrow in human face image", Department of Electronic Engineering, The Chinese University of Hong Kong Shatin, New Territories, Hong Kong, 2002
- [17] K. Lam and H. Yan "An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View", *Transactions on Pattern Analysis and Machine Intelligence*, vol.20(7), p. 673-686, 1998
- [18] H. K. Al-Mohar, J. Mohamad-Saleh and S. A. Suandi "Human Skin Color Detection: A Review on Neural Network Perspective", *International Journal of Innovative Computing, Information and Control*, vol 8, no.12, 2012
- [19] A. Kappas, U. Hess, C. L. Barr and R. E. Kleck "Angle of regard: The effect of vertical viewing angle on the perception of facial expressions", *Journal of Nonverbal Behaviour*, vol. 18(4), 1994
- [20] The source for images in Figure 1 can be found [here](#)
- [21] The ROS(Robot Operating System) documentation on Kinect tools can be found [here](#)
- [22] The source code and training data sets for the head pose estimation can be found [here](#)
- [23] The ROS adapter node for the head pose estimation can be found [here](#)
- [24] The PCL(Point Cloud Library) documentation can be found [here](#)
- [25] The OpenCv documentation on face and eye detection can be found [here](#)

(NOTE: The source of some of the listed references is the "IEEE Computer Society" - [ieeexplore.ieee.org](http://ieeexplore.ieee.org))