# Eyebrow Detection for Expression Analysis Using the Kinect

Baltac Mihai

*m.baltac@jacobs-university.de*

JACOBS UNIVERSITY

Bachelor Thesis in Computer Science

May 16, 2014

# Overview

# Origins

Psychologist Ekman and Friesen pursued research in facial expression and emotion analysis. They discovered that regardless age, sex, racial and cultural backgrounds, there are seven different facial expressions that are displayed equally by people: anger, disgust, fear, sadness, happiness, surprise and contempt[Ekman, 1978], [Ekman, 1999]:

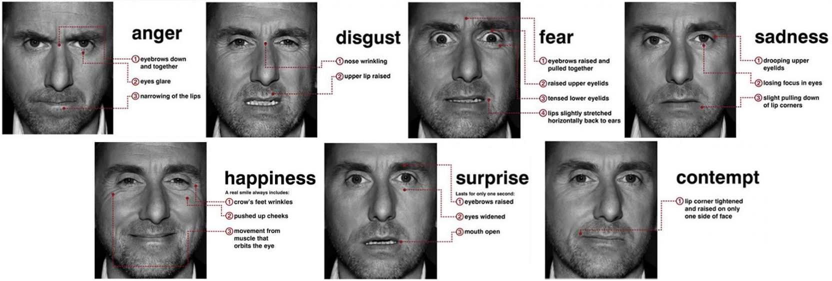



Figure: The 7 emotions. *Actor Tim Roth in 'Lie to me' series* [Expressions]

# Why eyebrows?

Eyebrows are very important facial elements. They provide a relevant interpretation of one's emotions.

The experiment pursued by John T. Cacioppo [Cacioppo, 2003] provides a clear idea of the importance of eyebrow movement by comparing it with another expression: the smile.

# Cacioppo's results

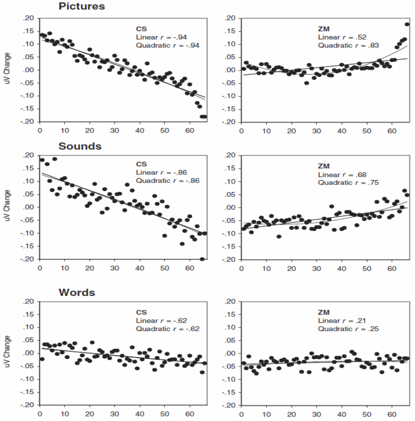



Figure: Experiment results on each type of stimuli: the linear evolution of the *corrugator supercilii*(left) and the quadratic evolution of the *zygomaticus major*(right) [Cacioppo, 2003]

# Detecting eyebrows

In recent years, the idea autonomous facial expression analysis became of interest in different fields of Computer Science such as robotics, computer vision or artificial intelligence, with interesting applications:

- improved machine-human interaction
- virtual avatar animation

Detecting eyebrows is just a small, yet important, part of this technology under development and it represents the main focus of our research.

# The method

Previous work has been done in the area of eyebrow detection, however only with focus on RGB image input data. Our solution uses the depth information from the Kinect to cover a wider rage of cases: detecting eyebrows when face is not frontally oriented towards the camera. There are two main parts that compound the method:

- Eyebrow detection
- Frontal pose image extraction

# ROI estimation

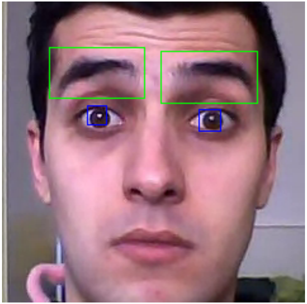


Figure: Face region with detected eyes and estimated eyebrow ROI

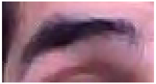
# Pseudo-hue extraction

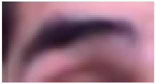
## **Algorithm 1**: PseudoHue

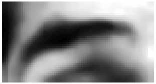
**Data**: eyebrow rgb image - *browROI*
**Result**: pseudo-hue matrix - *pseudoHue*

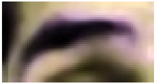
For less noise and better results blur the eyebrow ROI by a $5 \times 5$ kernel .
Convert RGB image into HSI. Apply histogram equalization on the intensity plane for better colouring and convert HSI back to RGB. Keep the equalized intensity plane.
**forall** *pixels in browROI* **do**
    *compute the 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**
    $val = val^k$ ▷ since all values are in [0,1]
Re-normalize

# Pseudohue


(a) original ROI


(b) blured ROI


(c) contrasted intensity plane


(d) HSI to RGB

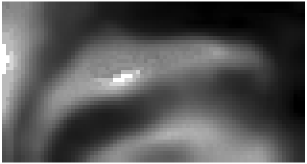


Figure: Pseudohue.

# Thresholding

We apply *adaptive thresholding algorithm* [Majumder, 2012] on the pseudohue matrix. For each value in our pseudo-hue we apply *localTreshold* as shown in equations 1, 2 and 3 and in the end, we threshold the result once more over the global mean.

$$localTreshold(x, y) = \mu_l(x, y) + k \cdot \sigma^2(x, y) \tag{1}$$

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 $k$.

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

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

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).

# Thresholding

Figure: The binary tresholded image (after erode-dilate)

2-4 iterations of *erode + dilate* operation are applied for better separation of blobs and noise removal. The largest blob in the image represents the eyebrow; it is segmented and then the contour is extracted.

The key eyebrow point of interest is the inner point. It is actioned by the *corrugator supercilii* muscle [Cacioppo, 2003] and is the most moving part of the eyebrow.

# Contour and key-points

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

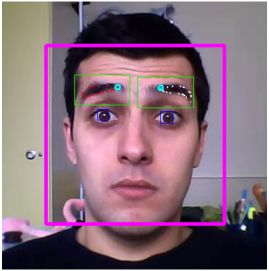


Figure: Face with detected eyebrows.

# Frontal pose image extraction

This is necessary because eyebrows would be segmented inaccurately in normal RGB image if the head has horizontal or vertical tilt.

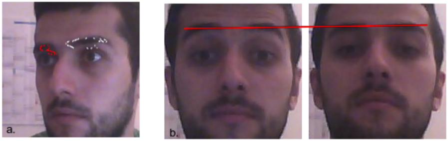



Figure: Limitations of eyebrow detection in RGB: a. bad segmentation, hard to interpret due to horizontal tilt; b. eyebrows seem to be at the same level although they are in raised position in the left picture and in normal position in the right picture where the head has vertical tilt.

Another reason is that the psychological perception of facial expressions from different angles of regard can be altered [Kappas, 1994].

# Head orientation

The *Random Forests for Real Time Head Pose Estimation* method described in [Faneli] for obtaining the head orientation.

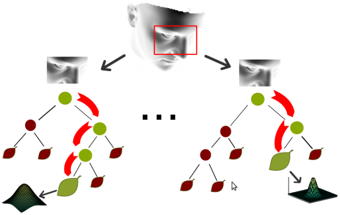


Figure: Random forests

# Point cloud rotation

Given its position, everything but the face is filtered from the point cloud ($18cm$ X $20cm$). The face is then translated to origin and rotated towards $X - axis$.

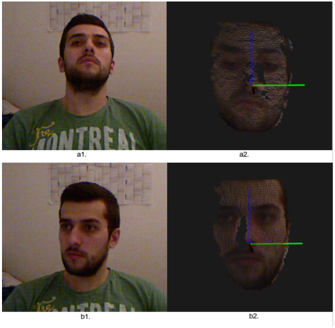



Figure: a. Vertical head tilt and rotated face point cloud; b. Horizontal head tilt and rotated face point cloud. It can be clearly seen that a non-frontal face pose was rotated so that it becomes frontal.

# Image extraction

The image is extracted in a parallel projection of 90X100 pixels. Due to rotation, the image is fragmented. The color of each black pixel is interpolated over the non-black surrounding pixels and the image is scaled by a factor of 2. Eyebrow detection is applied.

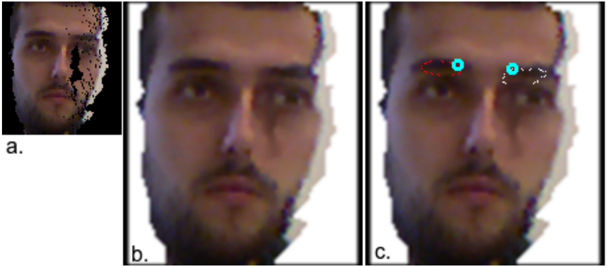



Figure: a. Fragmented image extracted from point cloud; b. Repaired image; c. Segmented eyebrows on the respective frame.

# Evaluation

The solution was evaluated by pursuing two experiments:

1. has the purpose to test the eyebrow key-point estimation in RGB images
2. has the purpose to evaluate the two different parts of the approach as a whole.

# Setup

20 picture of eyebrows were selected. The target position of the key points were pre-recorded by hand and compared with the results obtained after applying our method.


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

# Result

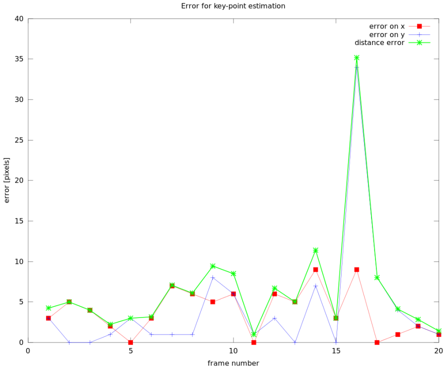
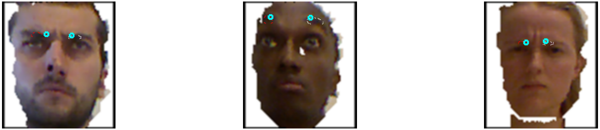



Figure: This plot displays the error in pixels on x coordinate, y coordinate and direct distance between the pre-recorded point and the extracted point. The fail displayed in Figure 12 is also visible here. The average error was also computed: *x: 3.85 pixels, y: 4.2 pixels and distance: 6.568 pixels*.

# Setup

Three subjects with different skin and hair color participated to the experiment. They were recorded with the Kinect under 5 different head orientations: frontal, $22.5^o$ upward, $10^o$ downward, $20^o$ leftward and $20^o$ rightward.

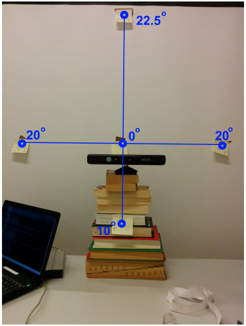


(a) (b) (c)

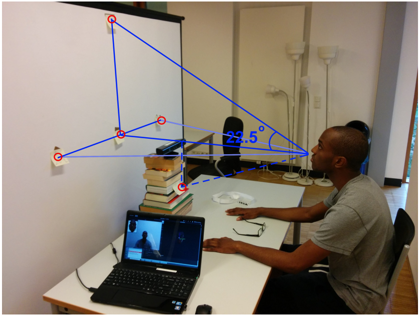
Figure: Participants of the second experiment.

# Setup

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.



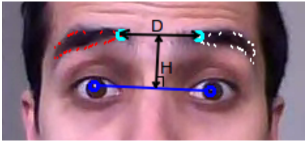

(a) From the participant's perspective.



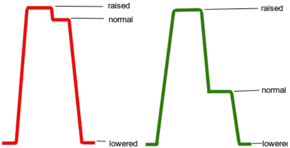

(b) Recording participant - upward head tilt.

# Values to be recorded

The two values of interest that are going to be recorded are D and H:




(a) The two values of interest for the second experiment.




(b) Expected plot shapes.

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 b were recognized in the plots constructed over the recording.

# Good results sample

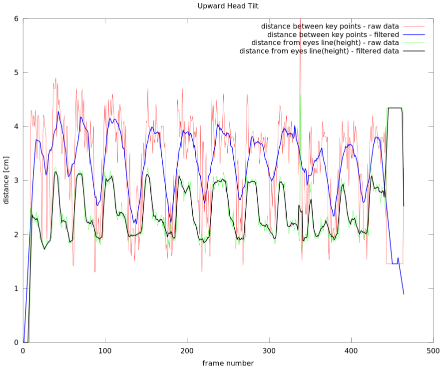



Figure: Plots for upward head tilt. The window size used for filtering was 15 for D(blue) and 5 for H(black).Positive results.

# Bad results sample

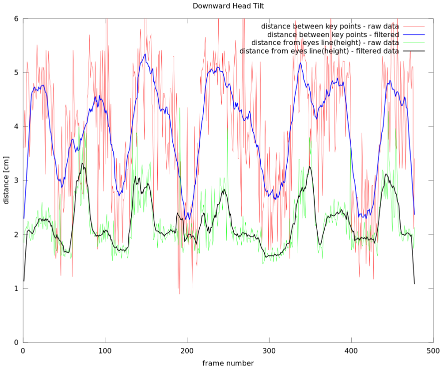


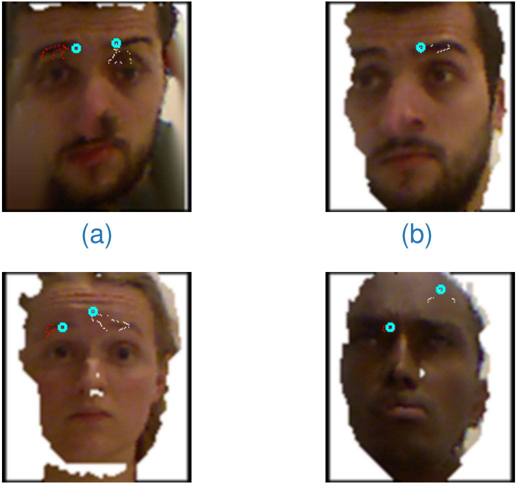

Figure: Plots for downward head tilt. The window size used for filtering was 19 for D(blue) and 9 for H(black). The quality of the results is significantly lower: high level of noise and inaccuracies in following the expected pattern. The reason is that at downward tilt, the extracted frontal pose is 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.

# Final results

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

Table: 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.

# Bad Segmentation examples



(a) (b) (c) (d)

# Future Work

- Increase the robustness of the solution for extreme cases
- Consider introducing other approaches such as skin color segmentation
- Use a sensor with a better resolution for further research
- Expand our research to a full face feature detection and tracking

# Conclusion

Our solution for eyebrow detection is innovative in the way that it uses the 3D perception of the Kinect to cover more real-world cases when the subjectś head is not frontally oriented towards the camera.

Further research in the area could bring important benefits in the expression analysis technology and therefore in different fields of Computer Science such as robotics and artificial intelligence. And also psychological studies of facial expressions could benefit of such an autonomous approach.

# References I


- P. Ekman and W. Friesen "Facial Action Coding System: A technique for the measurement of facial movement", Palo Alto, CA, USA, 1978
- P. Ekman "Basic emotions", University of California, San Francisco, CA, USA in *Handbook of Cognition and Emotion* by T. Dalgleish and M. PowerS, 1999
- J. T. Larsen[a], C. J. Norris[b] and J. T. Cacioppo[b] "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
- P. Viola and Michael J. Jones "Robust real-time face detection", International Journal of computer Vision, vol. 54(2), p. 137-154, 2004
- 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
- W. Niblack "An introduction to digital image processing", Strandberg Publishing Company Birkeroed, Denmark, Denmark, 1985.
- 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.
- 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

# References II


- G. Fanelli, J. Gall and L. Van Gool Leuven "Real Time Head Pose Estimation with Random Regression Forests"
- 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
- 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
- The source for images in Figure 1 can be found here
- The ROS(Robot Operating System) documentation on Kinect tools can be found here
- The source code and training data sets for the head pose estimation can be found here
- The ROS adapter node for the head pose estimation can be found here
- The PCL(Point CLoud Library) documentation can be found here
- The OpenCv documentation on face and eye detection can be found here
- The eyebrow movement interpretations can be found at links one and two

# The End