Real-time Eye and Head Pose Detection Using Active Shape Models with Stasm

Zeming Li
Supervised by: Prof. K. N. Plataniotis
M.Eng Project

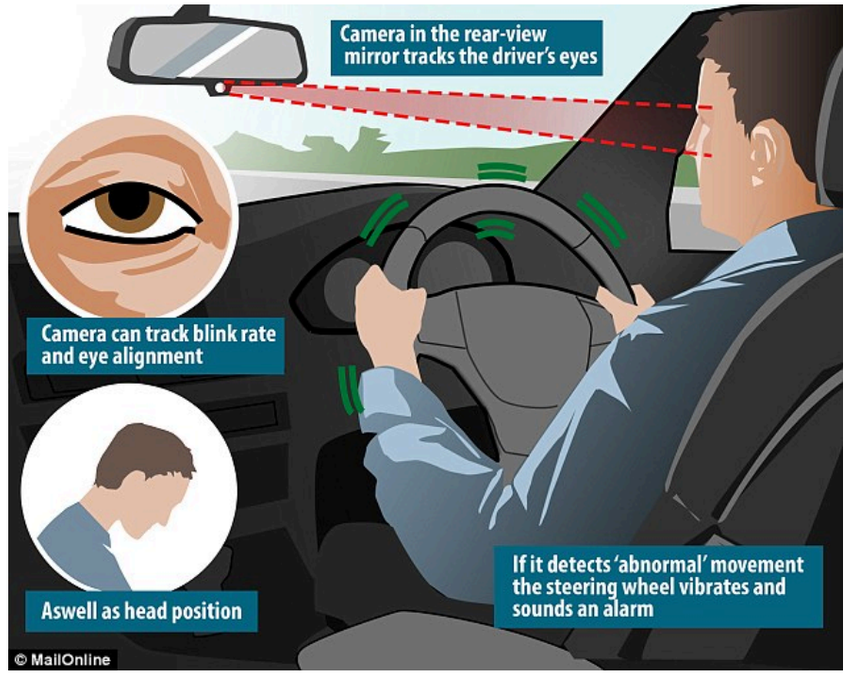
University of Toronto
Department of Electrical and Computer Engineering

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

Applications & Motivations

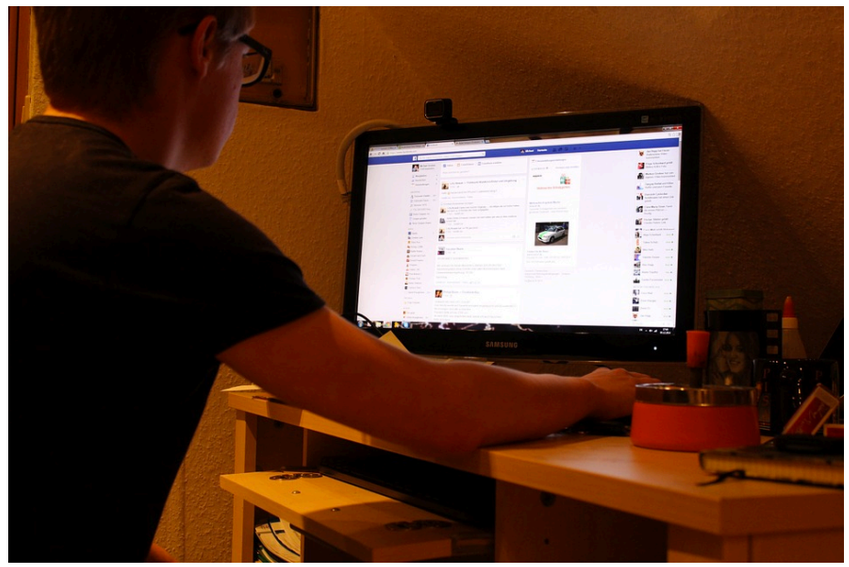
1. Eyes Off the Road Detection System

- Locates facial features under various conditions
- Estimates head movement
- Gives a reminder to the diver



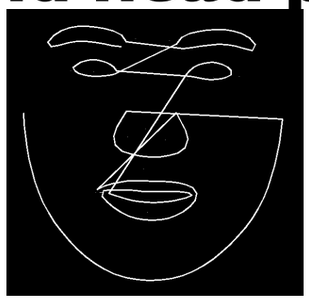
2. Reading Behavior Monitor

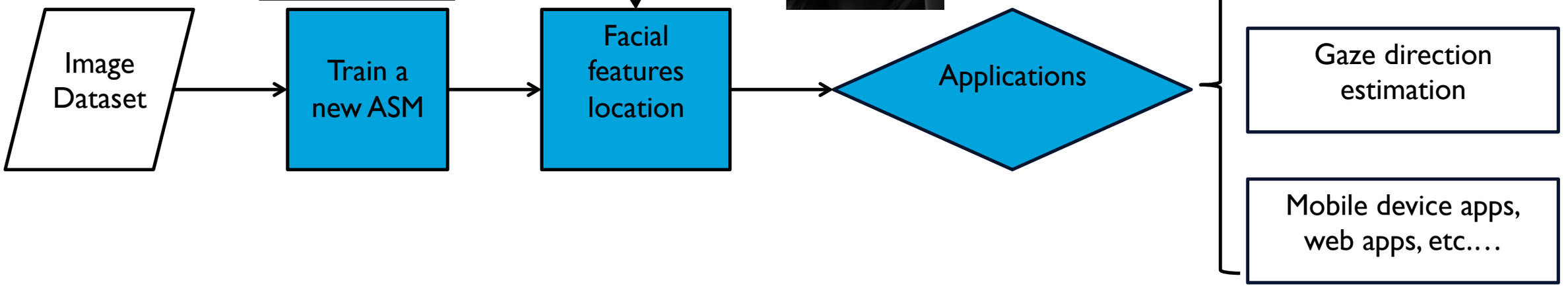
- Estimates gaze direction
- Calculates residence time on the screen
- Generates a gaze density map for analysis



Project Overview

Objective: To train a new active shape model & implement it for eye and head pose detection





- Train a new ASM
  - Use HELEN dataset and optimized methods to obtain a 194-landmarks shape model
- Facial features location
  - Locate 194 facial features on each input image

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

Problem Statement

The new ASM training

- Problem: Training a 194-landmarks model of the shape of faces on HELEN dataset
- Input: Images, shapefiles, landmark table, optimized methods
- Output: A statistical model of the 194 facial features

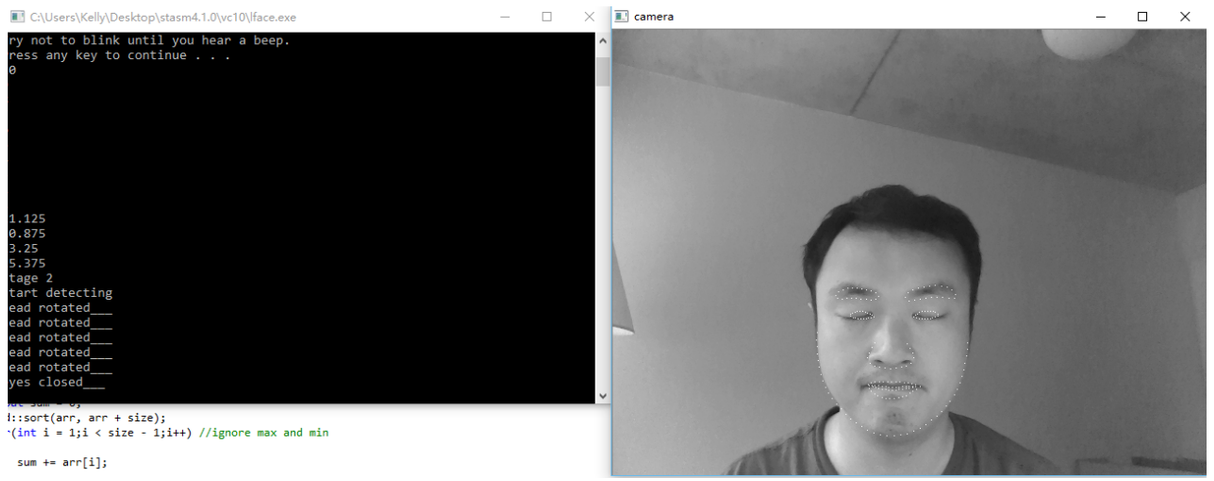
Facial features locating

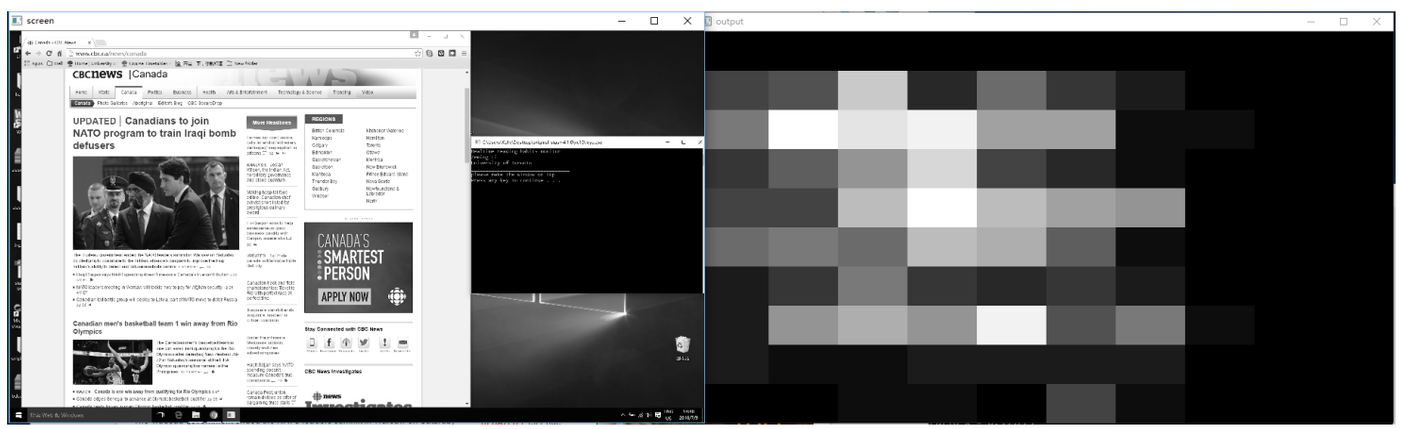
- Problem: Automatically locate each landmark on the facial features in a new image
- Input: Digital color face image
- Output: A matrix(194x2) which contains landmarks location information

Contributions

- Proposed a newly trained robust active shape model on HELEN dataset which locates 194 landmarks under various illumination conditions

- Created a real-time head pose and closing eyes detection system a reading habits monitor





UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

Organization

Contribution 1

A newly trained robust active shape model on HELEN dataset which locates 194 landmarks under various illumination conditions

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

Background: Facial Features Locating Framework

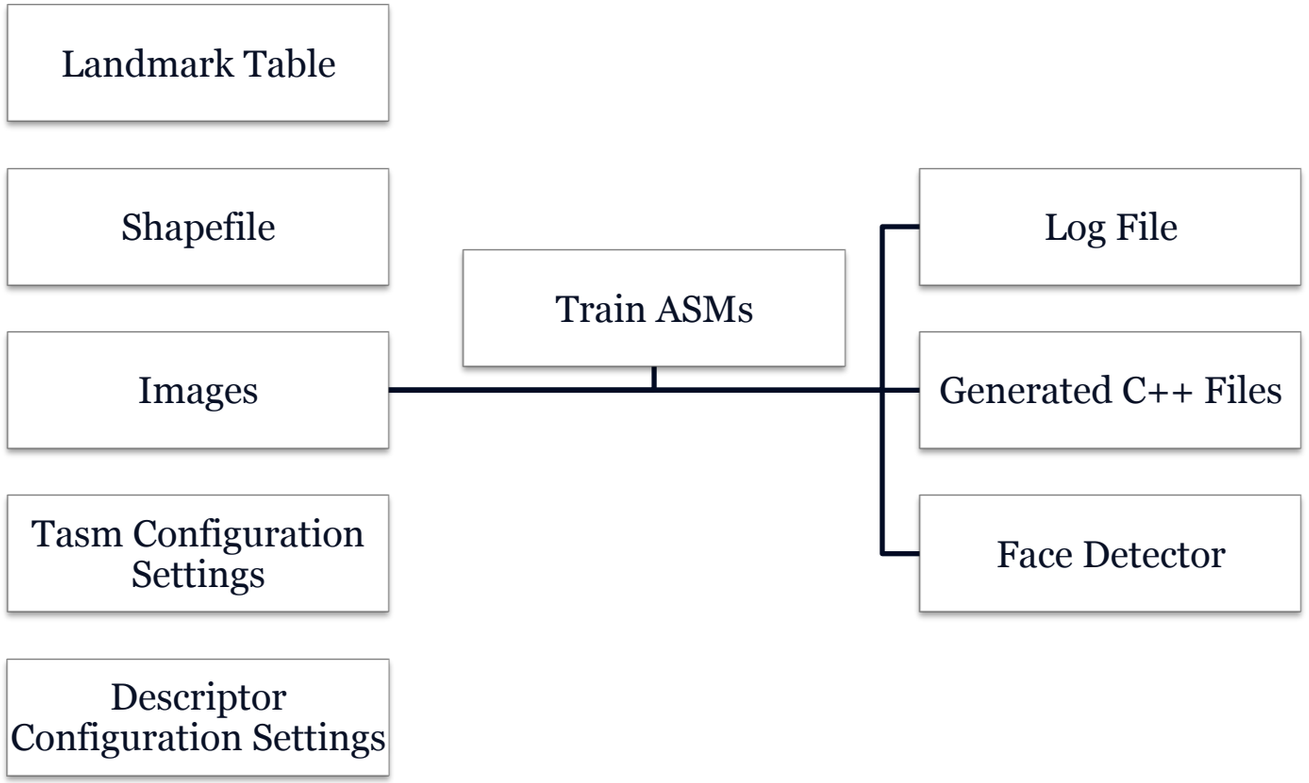
Active Shape Models(ASMs)[T. Cootes]

- Statistical Models represent the shape of objects by a set of landmarks
- Iteratively deform to match the model to a new image
- Can be trained on a set of labeled examples in images

Location framework

- Mean shape is obtained in training stage
- Different ASMs have different shapes(landmarks amount N)
- After iteration, each landmark is located to match the new object in the image

Training ASMs



- Edit the annotation file of HELEN dataset by Shell script
- Identify face conditions of face images and label them
- Train ASMs with different settings

ASMs Experiments

Objectives: to evaluate the different ASMs performance under the following scenarios

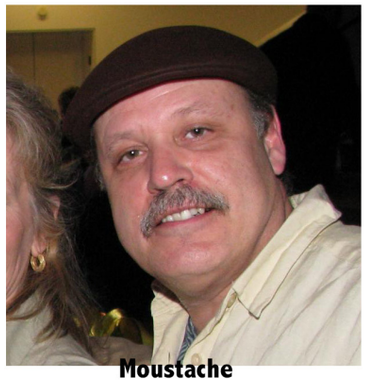
- Facial landmarks location accuracy under different environment
- Facial landmarks location speed
- The advantage of the proposed model comparing to ASM proposed by S. Milborrow(Stasm)

Glasses

Bad Image

Bad Eyes

Beard



Moustache

Experimental Setup

:

- Has 1,512 gray level images in 384x286
- Presents the frontal view of 23 different subjects' faces
- The images recorded in uncontrolled conditions by a webcam
- "Real-world" indoor environment

Evaluation Methodology:

- The Euclidean Distance *d(p1,p2)* between the annotated point 1 and estimated point 2
- Normalized by the Inter-Ocular Distance (IOD)

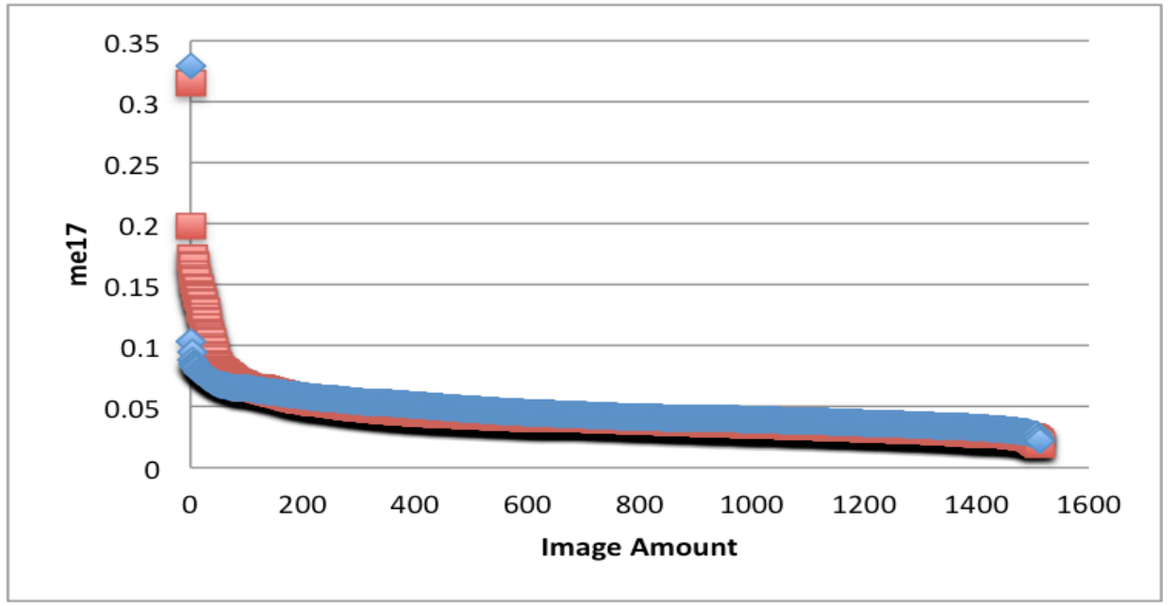
$$\delta_i^k = \frac{d\left((x_i^k, y_i^k), (\hat{x}_i^k, \hat{y}_i^k)\right)}{IOD} \qquad \mathrm{P} = \frac{\sum_{i=1}^{\iota}\sum_{k=1}^{\kappa}\delta_i^k}{\mathrm{K} \times \mathrm{I}}$$

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

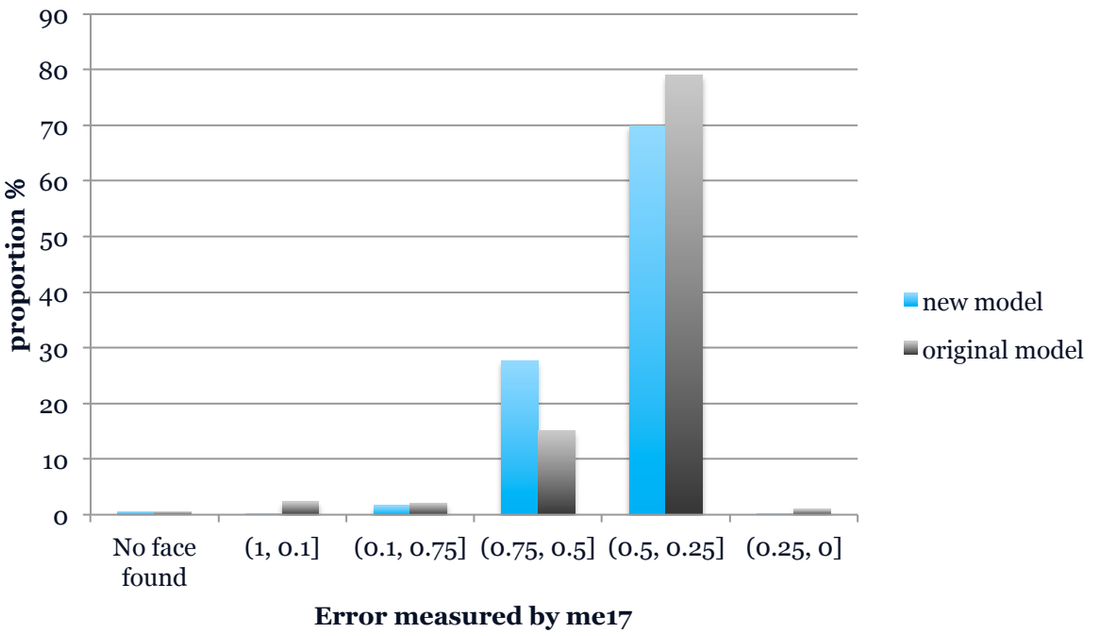
Results

Accuracy analysis

- Performance of the new model (blue) and the original Stasm model (red)
- “me17” is the average Euclidean distance over 17 points



- The proportion of landmarking results in different error ranges
- The new model has fewer results with huge errors



Results

Improvement on accuracy

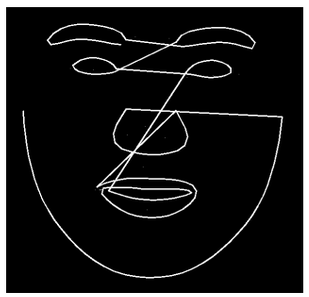
An example of how newly trained ASM improves facial landmarks locating performance

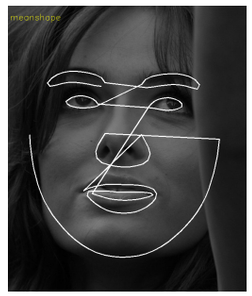
- Newly trained model (left) is better on the rotated face
- The me17 of Stasm is 0.078, about 29% higher than that of the newly trained model

Improvement on speed

Model name	me17	Mean time/number of landmarks (ms)
Stasm	0.043590	1.3377
The new model	0.045635	0.6546

Organization





Webcam

Image Dataset

Train a new ASM

Facial features location

Applications

Closed eyes and head pose detection

Gaze direction estimation

Mobile device apps, web apps, etc.…

Contribution 2

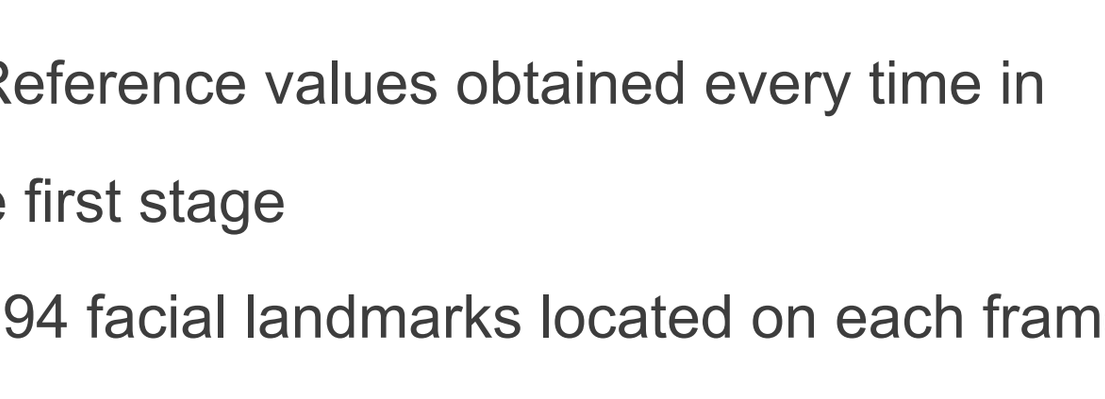
Created a real-time head pose and closing eyes detection system a reading habits monitor

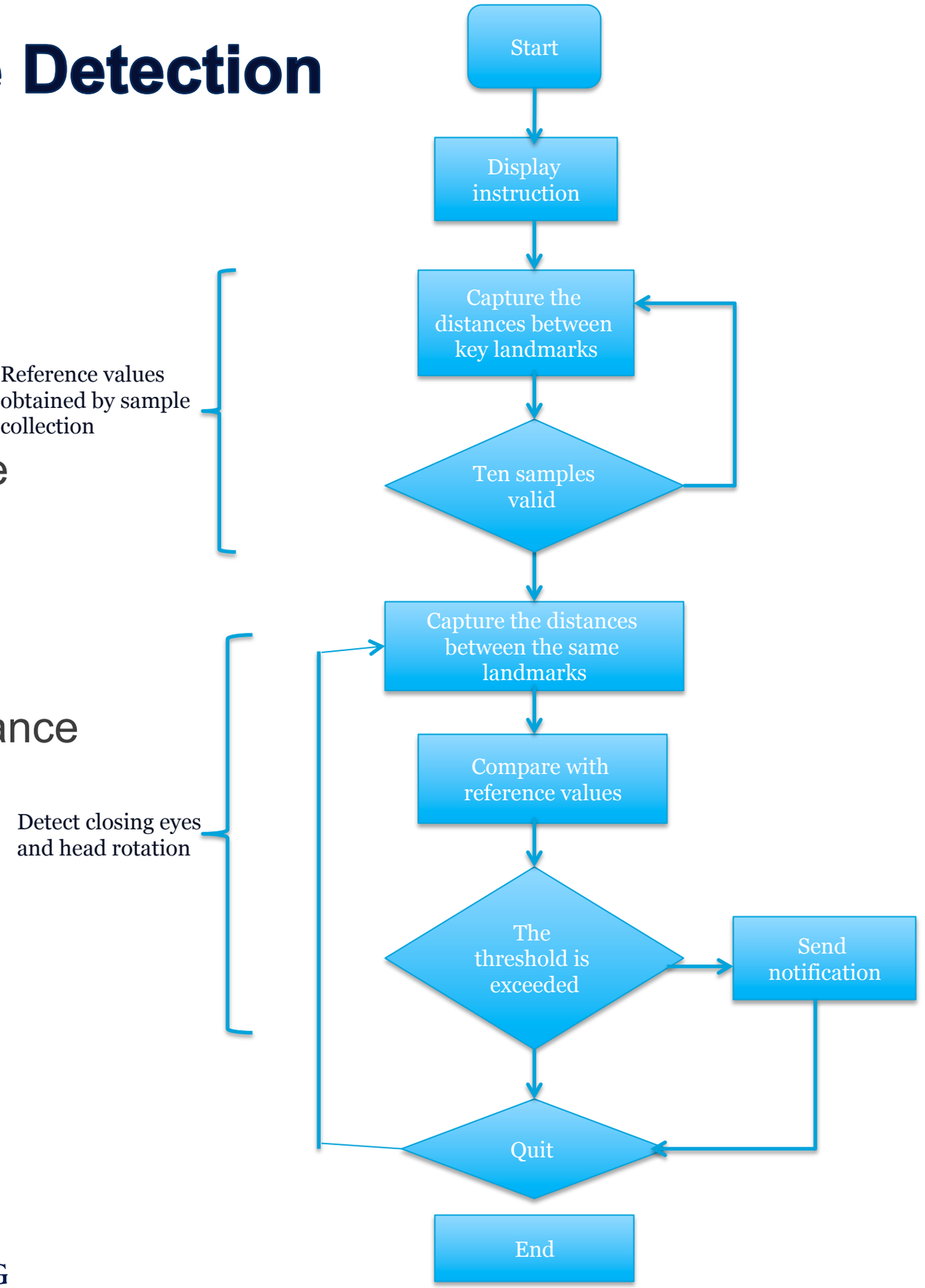
UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

Closed Eye and Head Pose Detection

Overview

- Reference values obtained every time in the first stage
- 194 facial landmarks located on each frame of video captured by webcam
- Closing eyes and head rotation detection is judged based on the normalized vertical distance between some key landmarks



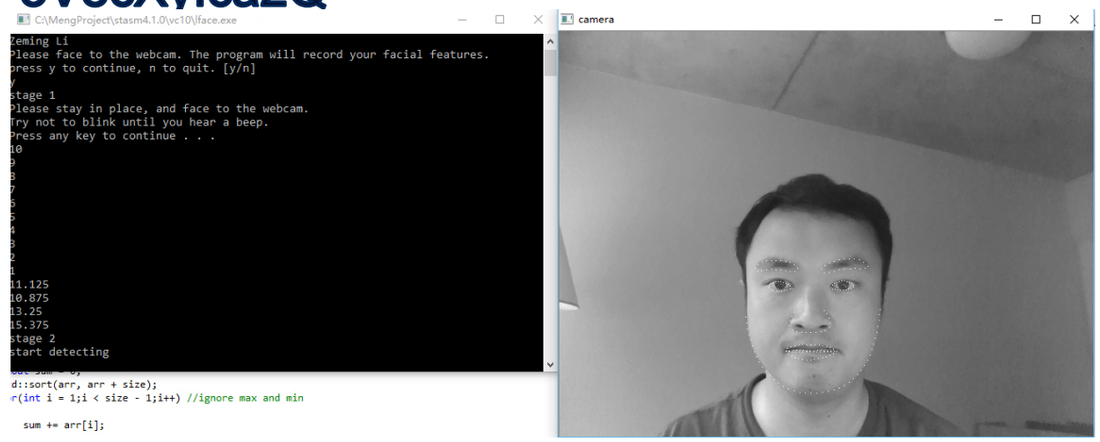


UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

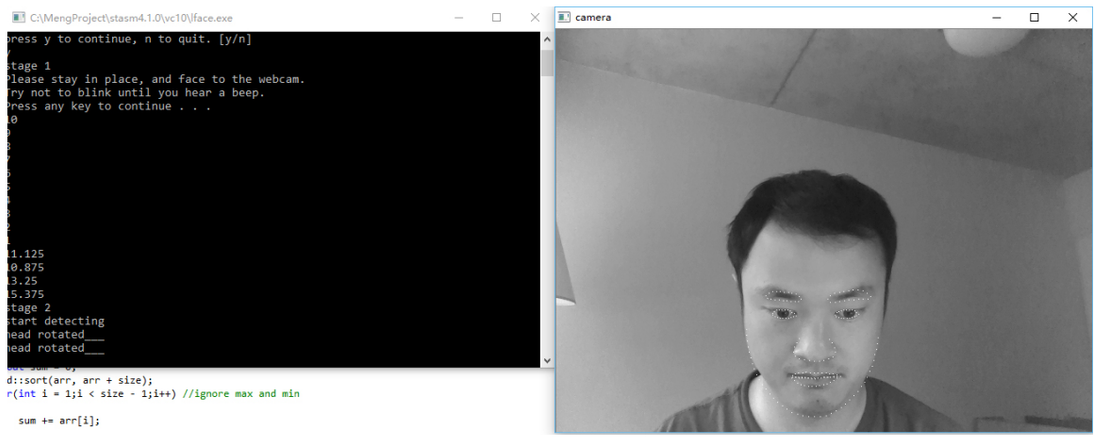
Working Process

demo video: https://www.youtube.com/watch?v=oV8cXvIca2Q

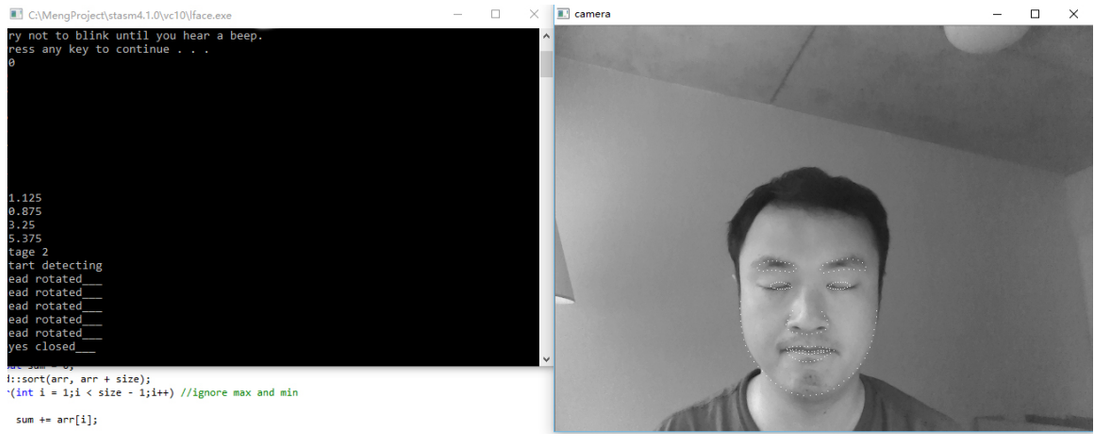
- The application locates 194 landmarks and compares the distance between key landmarks with the standard values



- When the rotated head is detected the reminders are two short beeps and a literal reminder displayed on the console



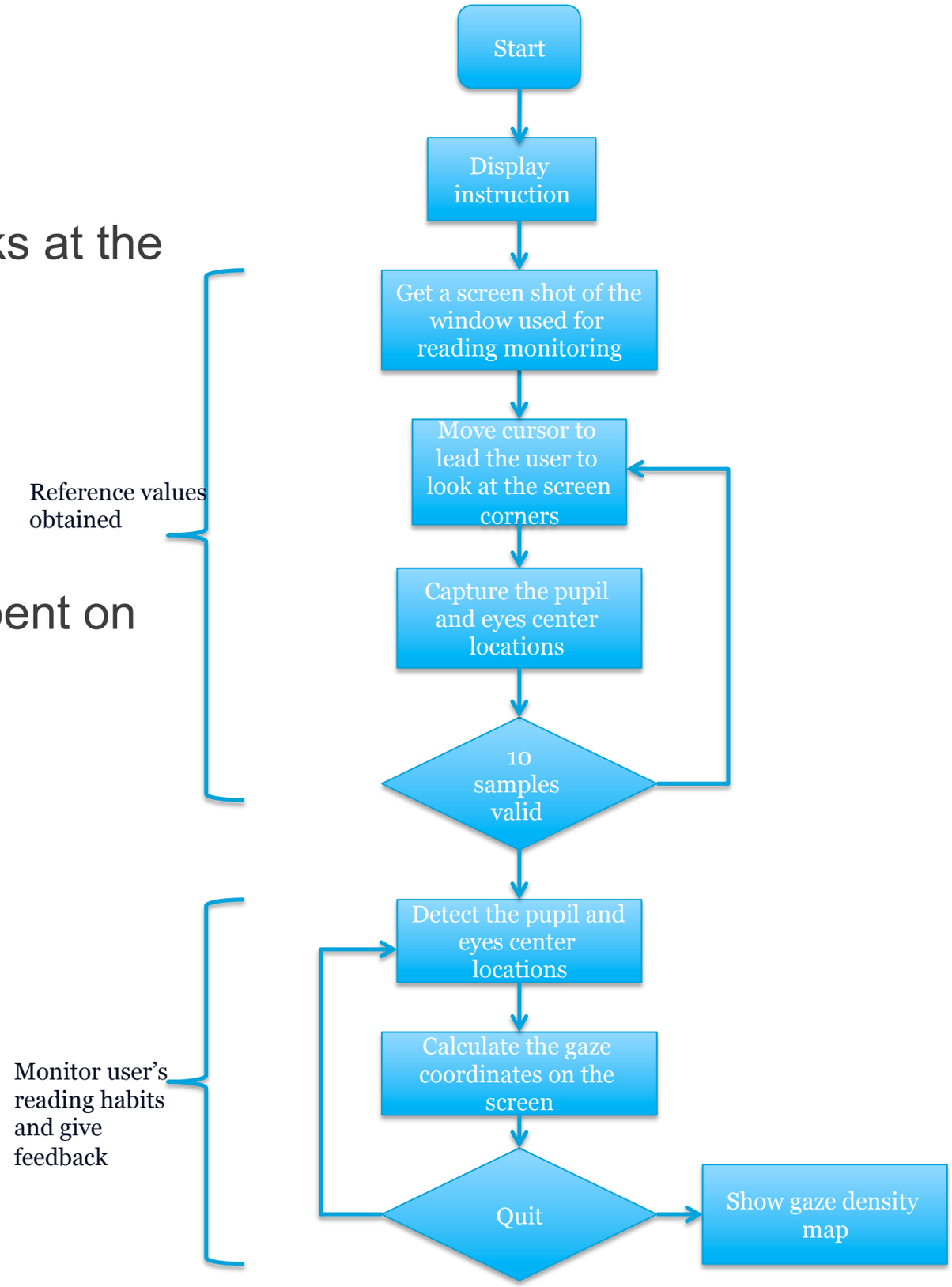
- When the closed eyes are detected, the reminders are a long beep and a corresponding literal reminder on the console



Reading Habits Monitor

Overview

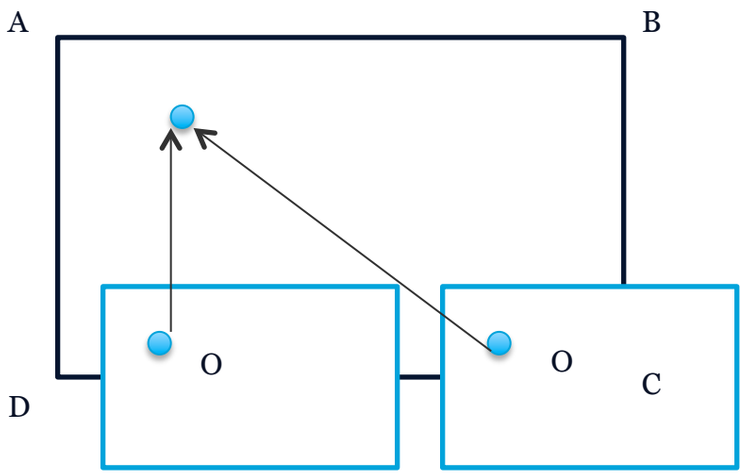
- Captures pupils location when the user looks at the 4 corners
- Locates key landmarks on each frame
- Estimates gaze direction on the screen
- Calculates gaze direction residence time spent on 1 of 100 blocks of the screen.
- Outputs a gaze density map

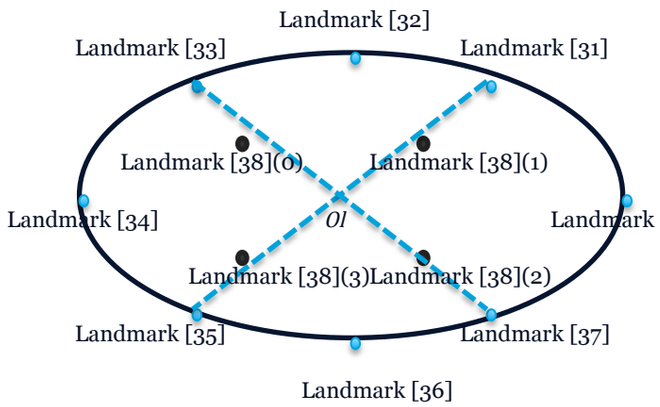


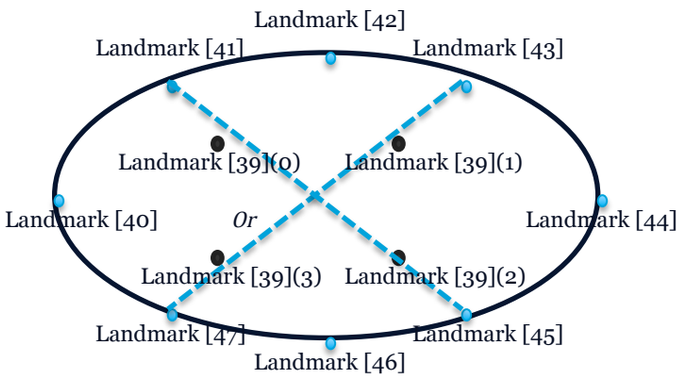
Gaze Direction Estimation

- Five landmarks used for gaze estimation around each eye. The four landmarks on the eye border are used to calculate the geometric center of the eye.

- Representation of the gaze mapping principle. The pupils' movements in a particular area in the eyes could be mapped to the gaze location on the screen.



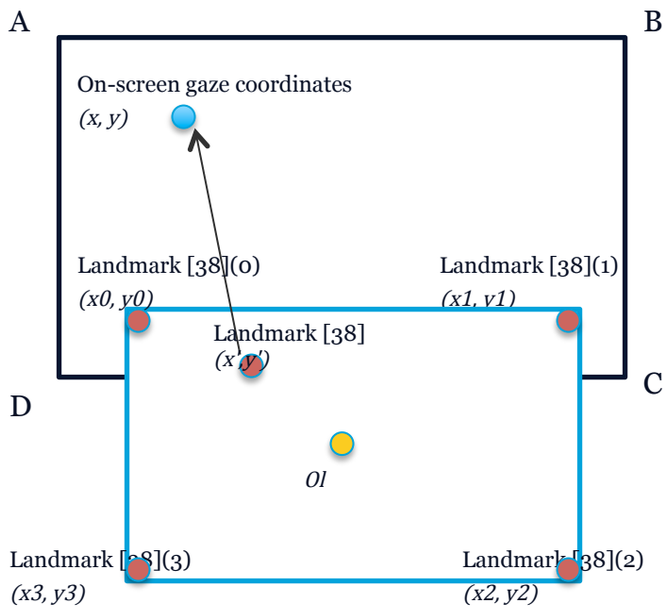




Gaze Direction Estimation

- The location of the left pupil is mapped to the estimated gaze coordinates on the screen.

- The same thing happens to the right pupil.



The on-screen gaze coordinates *(x, y)*

$$\frac{x}{l_{AB}} = \frac{x' - x_{ol} - x_{0'}}{(x_1 - x_0) \div 2 + (x_2 - x_3) \div 2}$$

$$Where\ x_{0'} = x_0 - x_{ol}, x_{ol}\ is\ a\ constant\ value\ obtained\ with\ x_0$$

$$l_{AB}\ is\ the\ width\ of\ the\ user's\ screen$$

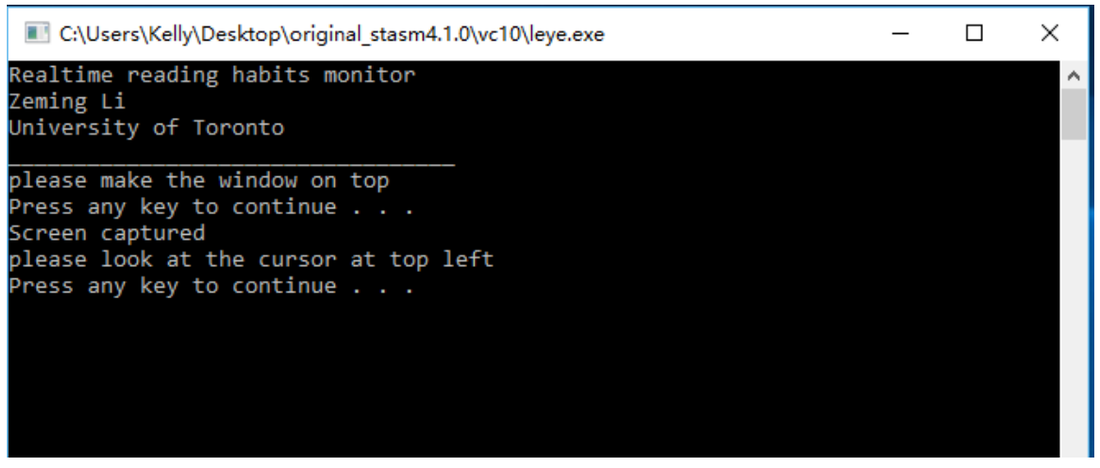
$$\frac{y}{l_{AD}} = \frac{y' - y_{ol} - y_{0'}}{(y_1 - y_0) \div 2 + (y_2 - y_3) \div 2}$$

$$Where\ y_{0'} = y_0 - y_{ol}, y_{ol}\ is\ a\ constant\ value\ obtained\ with\ y_0$$

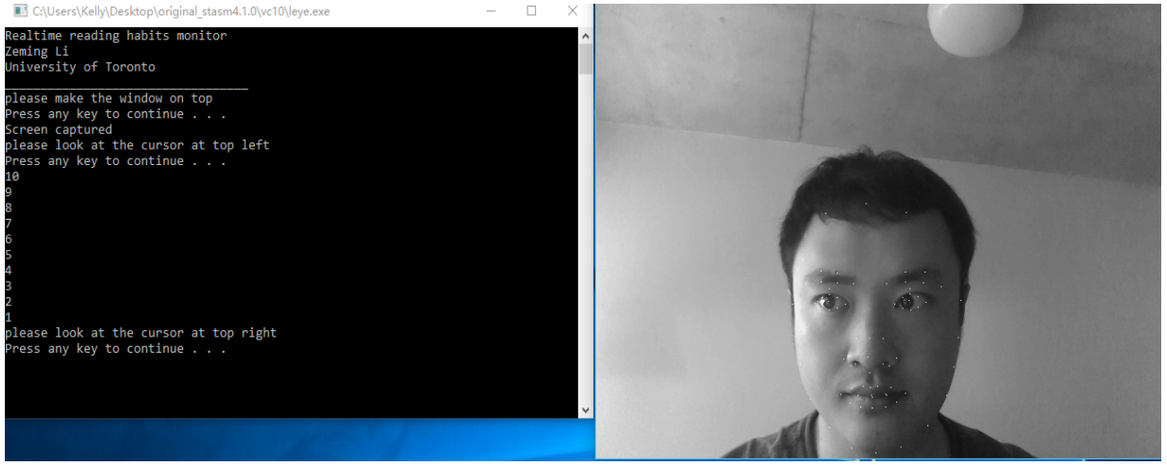
$$l_{AD}\ is\ the\ height\ of\ the\ user's\ screen$$

Working Process

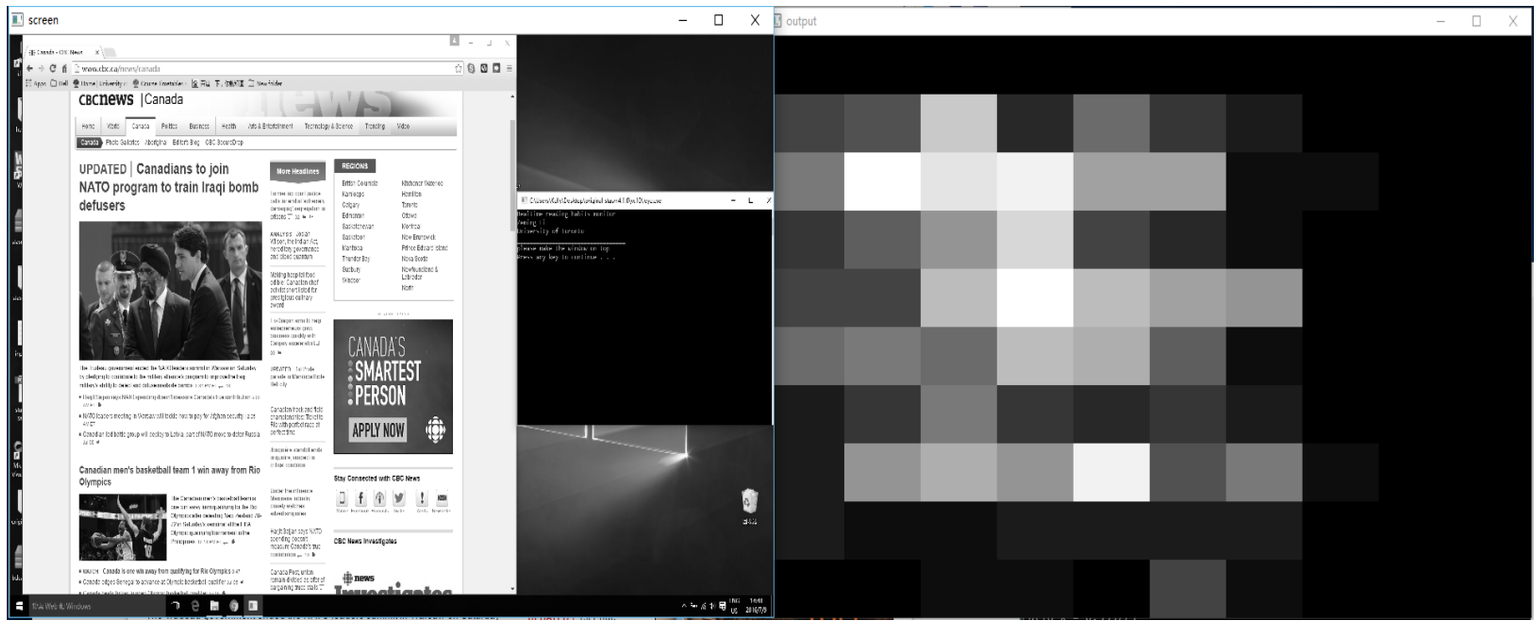
demo video: https://www.youtube.com/watch?v=p7XOMBNL2Jk



- Guides displayed on the console. The window with reading materials on the top is screen-shotted for analysis



- The cursor is set to the top left corner firstly and guides the user to look at the cursor



- Screenshot of the window used for monitoring and gaze density map with the same size
- The color of the blocks in gaze density map reflects the user's reading habits

Extensive Applications

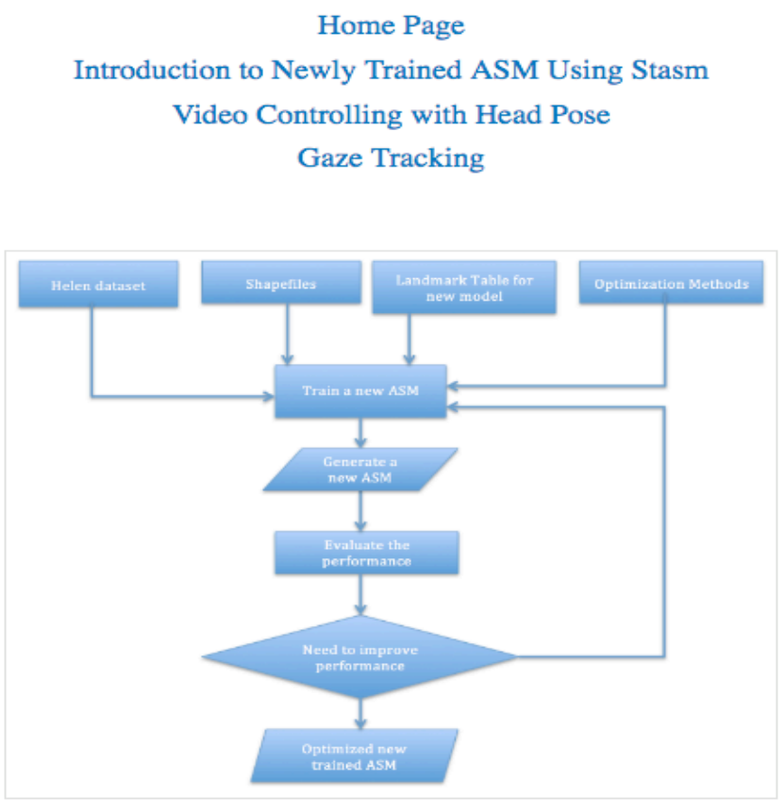
ASM WebApp (xlabs chrome extension api needed)

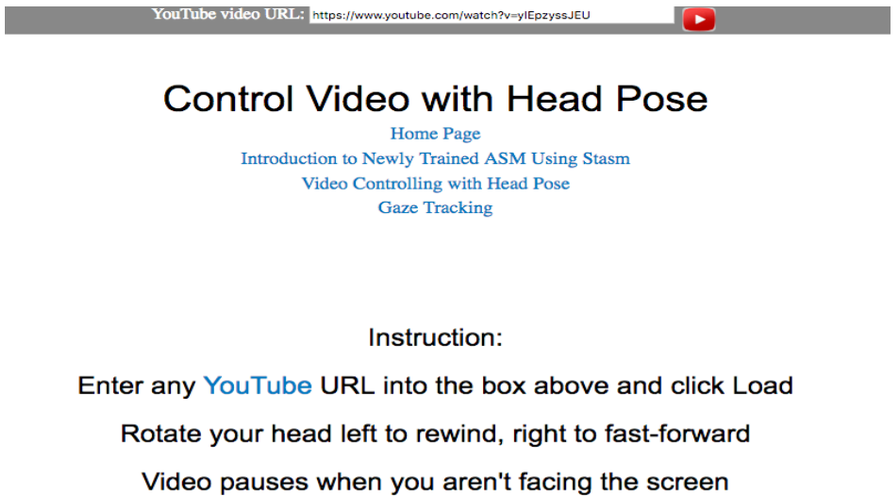
https://asm-webapp.appspot.com/

GitHub: https://github.com/zemli/ASM-WebApp

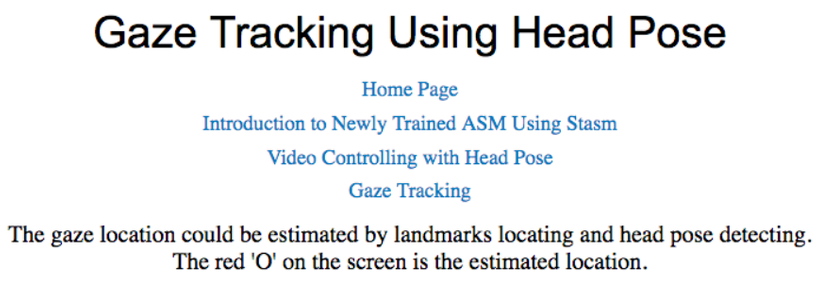
- A web app controls video with the head pose

- The homepage of the website





- A web app estimates the gaze location on the screen



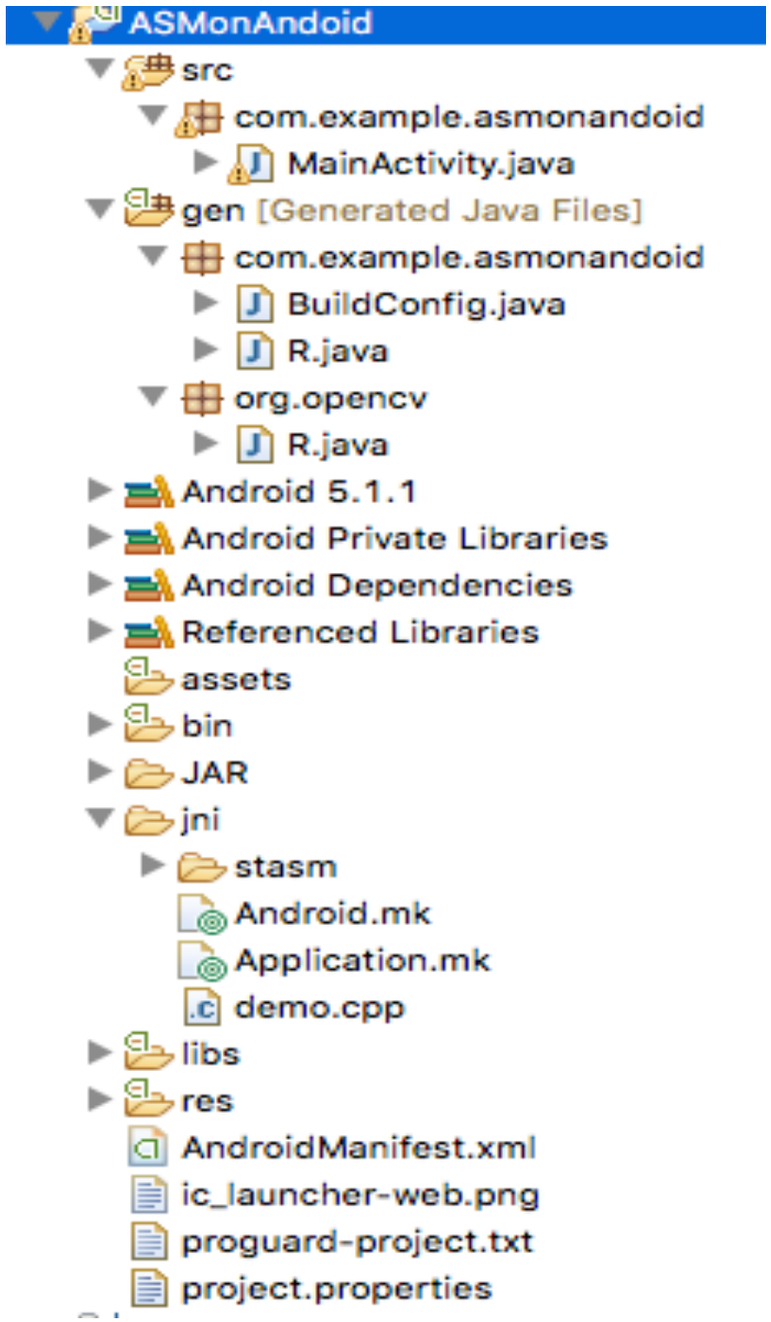
ENGINEERING

Extensive Applications

ASM on Android

- Facial features locating can be realized by using ASM with OpenCV4Android library

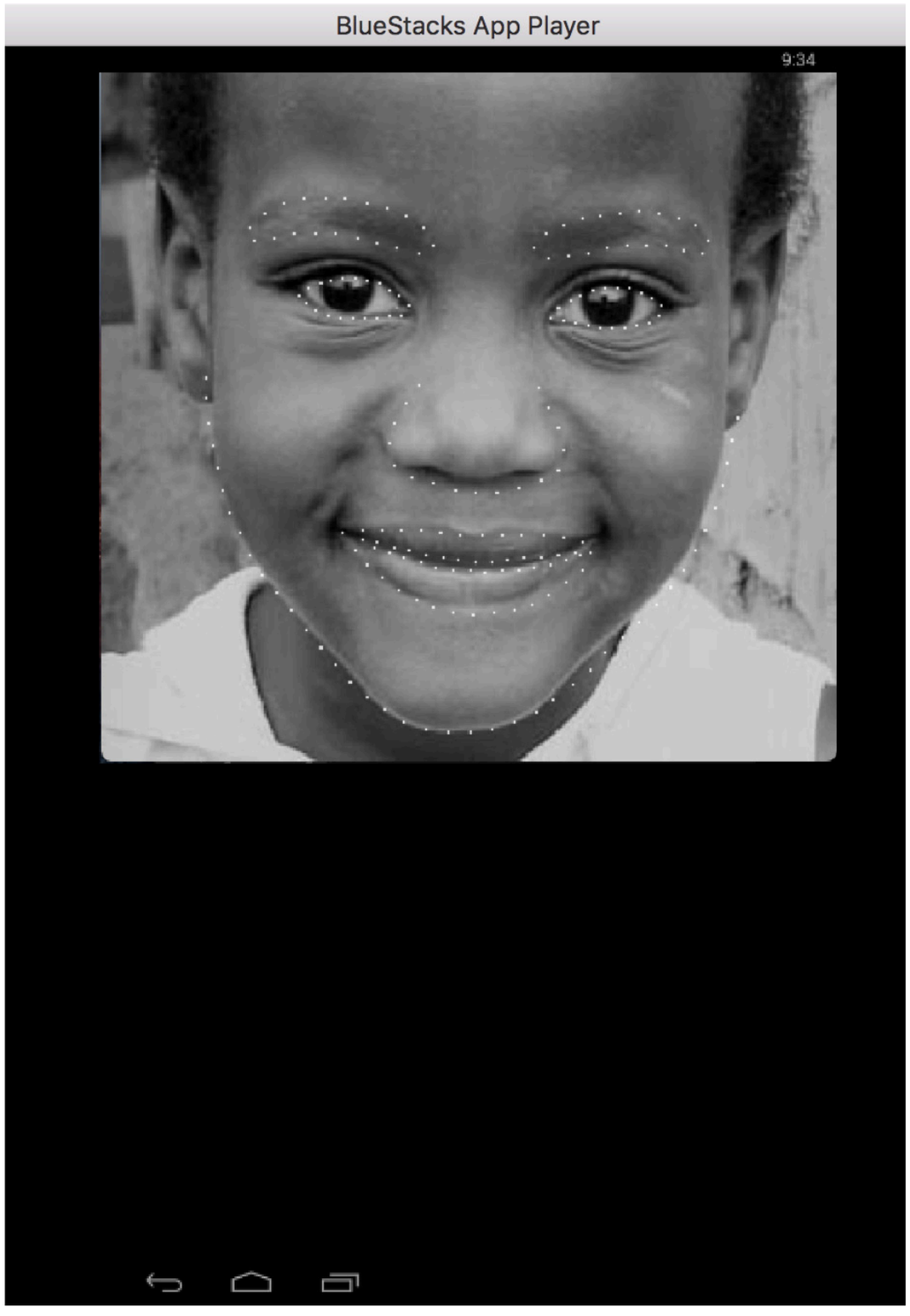
- JNI (Java Native Interface) could use an interface works as a pointer links android application in local reference with ASM source code in global reference.



Extensive Applications

ASM on Android

- Android app running on a virtual machine (BlueStacks) for MacOS. 194 landmarks are accurately located on the face in test image.



Conclusion & Future Work

The new ASM trained on HELEN dataset

- High accuracy under various complications
- Fast processing speed on each landmark
- Detailed shape model (194-landmarks model) for further processing

Real-time Eye and Head Pose Detection System

- Efficient for building real-time applications
- Robustness to different illumination conditions
- Apply to EOR system / monitor users' reading habits

Future Work

- Comprehensive algorithm to estimate gaze direction
- Apply the new model to other biometric applications need detailed shape information (i.e., face verification, facial micro-expression).

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

Thanks!

Questions?

Email: zeming.li@mail.utoronto.ca

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

References

[1] online: www.dailymail.co.uk

[2] Milborrow, S. and Nicolls, F., 2014, January. Active Shape Models with SIFT Descriptors and MARS. In VISAPP (2) (pp. 380-387).

[3] Milborrow, S., Morkel, J. and Nicolls, F., 2010. The MUCT landmarked face database. Pattern Recognition Association of South Africa, 201(0).

[4] Benlamoudi, A., Samai, D., Ouafi, A., Taleb-Ahmed, A., Bekhouche, S.E. and Hadid, A., 2015. Face spoofing detection from single images using active shape models with stasm and lbp. In Proceeding of the Troisime CONFERENCE INTERNATIONALE SUR LA VISION ARTIFICIELLE CVA 2015.

[5] Liu, L., Hu, J., Zhang, S. and Deng, W., 2014, November. Extended supervised descent method for robust face alignment. In Asian Conference on Computer Vision (pp. 71-84). Springer International Publishing.

[6] Le, V., Brandt, J., Lin, Z., Bourdev, L. and Huang, T.S., 2012, October. Interactive facial feature localization. In European Conference on Computer Vision (pp. 679-692). Springer Berlin Heidelberg.

[7] Aggarwal, G., Biswas, S., Flynn, P.J. and Bowyer, K.W., 2012, January. A sparse representation approach to face matching across plastic surgery. InApplications of Computer Vision (WACV), 2012 IEEE Workshop on (pp. 113-119). IEEE.

[8] Asthana, Akshay, Asim Khwaja, and Roland Goecke. "Automatic frontal face annotation and AAM building for arbitrary expressions from a single frontal image only." 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE, 2009.

[9] Jesorsky, O., Kirchberg, K.J. and Frischholz, R.W., 2001, June. Robust face detection using the hausdorff distance. In International Conference on Audio-and Video-Based Biometric Person Authentication (pp. 90-95). Springer Berlin Heidelberg.