Lip Reading to Text

Waleed Deaney

Supervisor: Prof. I M Venter

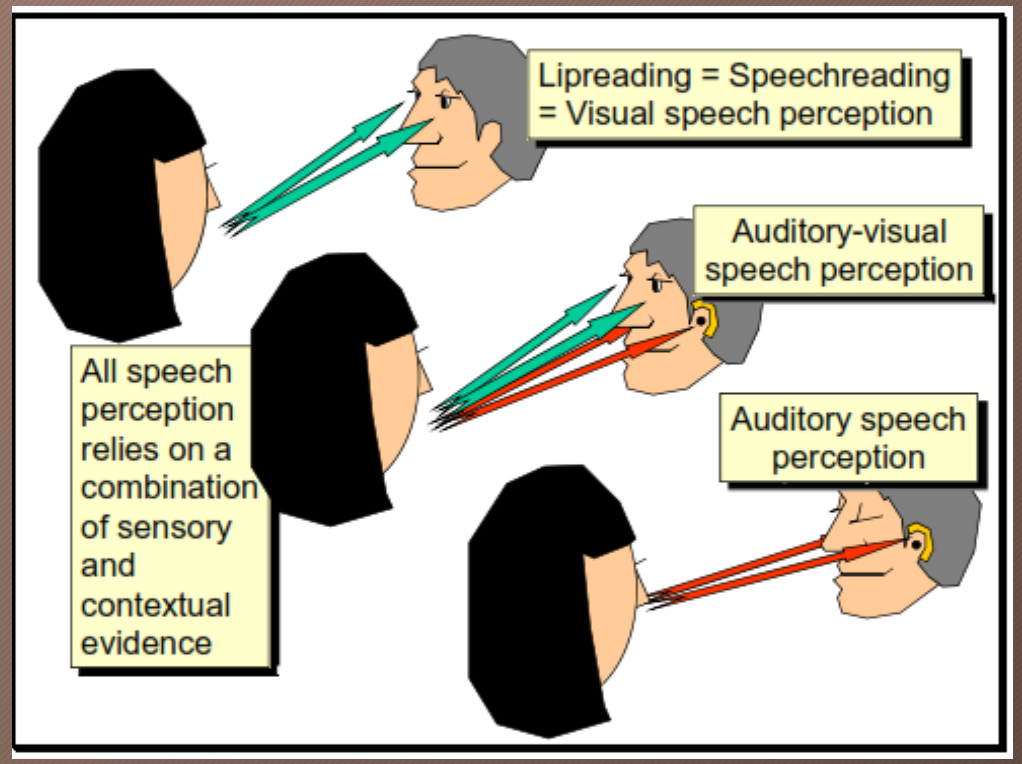
Co-Supervisor: Mr M Ghaziasgar

Mentor : Mr K Abrahams

Co-mentor: Mr N de la Cruz

A Quick Recap…

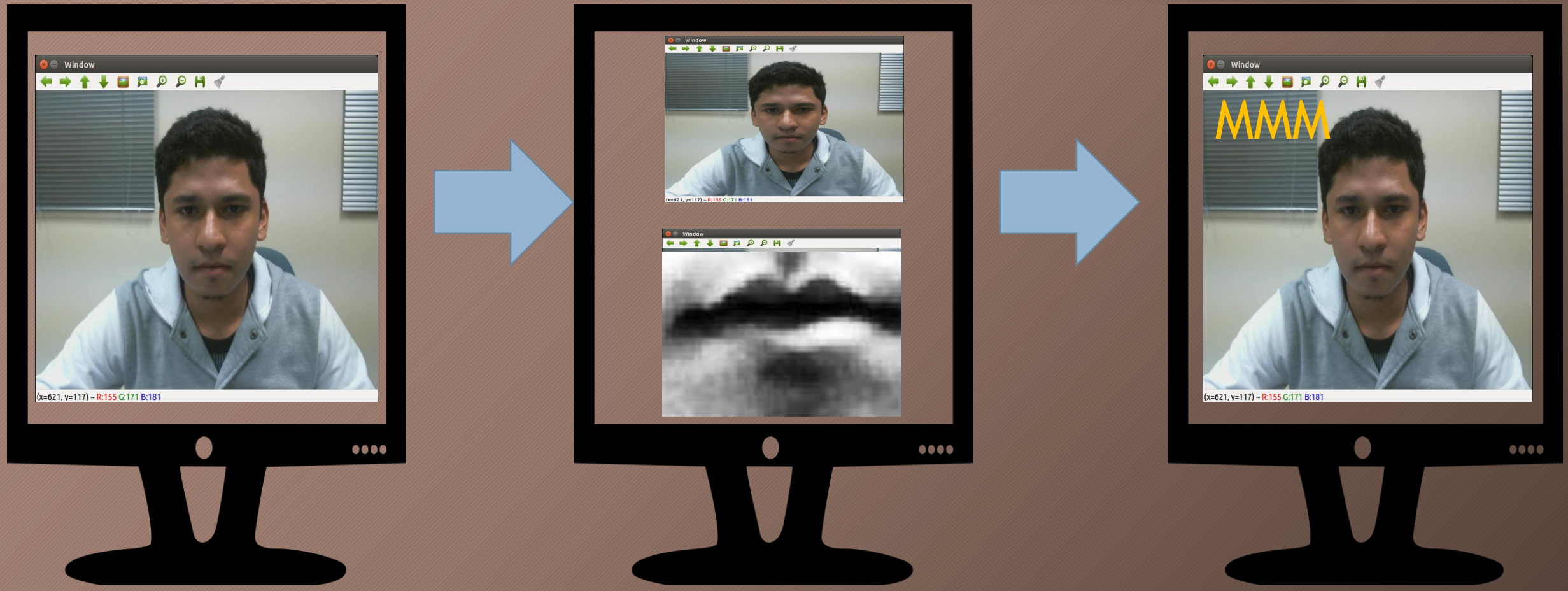
- Speech recognition software:
  - Limitations:
    - Noisy environments
    - Multiple speakers
- Visual speech recognition
  - Simple sounds/letters
  - Results should be displayed as TEXT



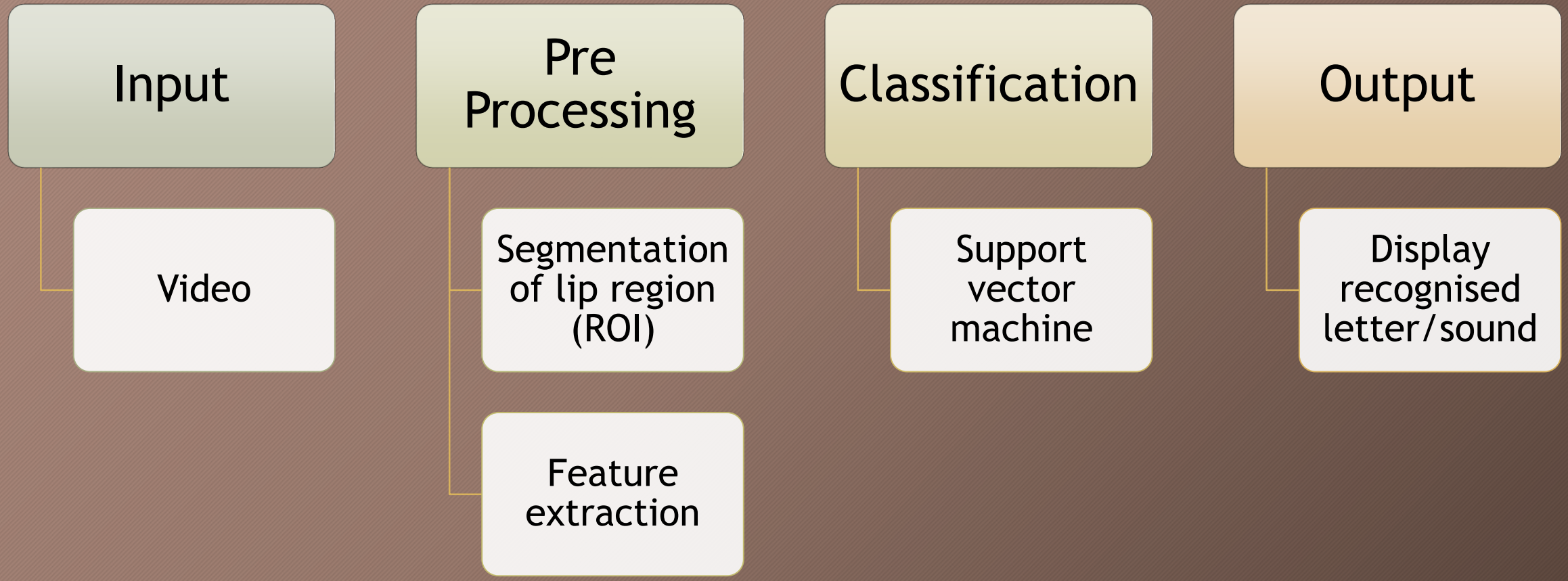
Overview

- User Interface Specification
- High Level Design
- Low Level Design
- Prototype Demo

User Interface Specification



High Level Design



Low Level Design

- Video
  - Capture frames
- Lip Segmentation
  - Face detection
  - Eye detection
  - Segement ROI
- Feature Extraction
  - Local Binary Patterns
  - Form Histograms
- Classification (SVM)
  - Training
  - Testing
- Output
  - Display recongised sound/letter

Input Frames

- Frames from webcam

Capture from camera:

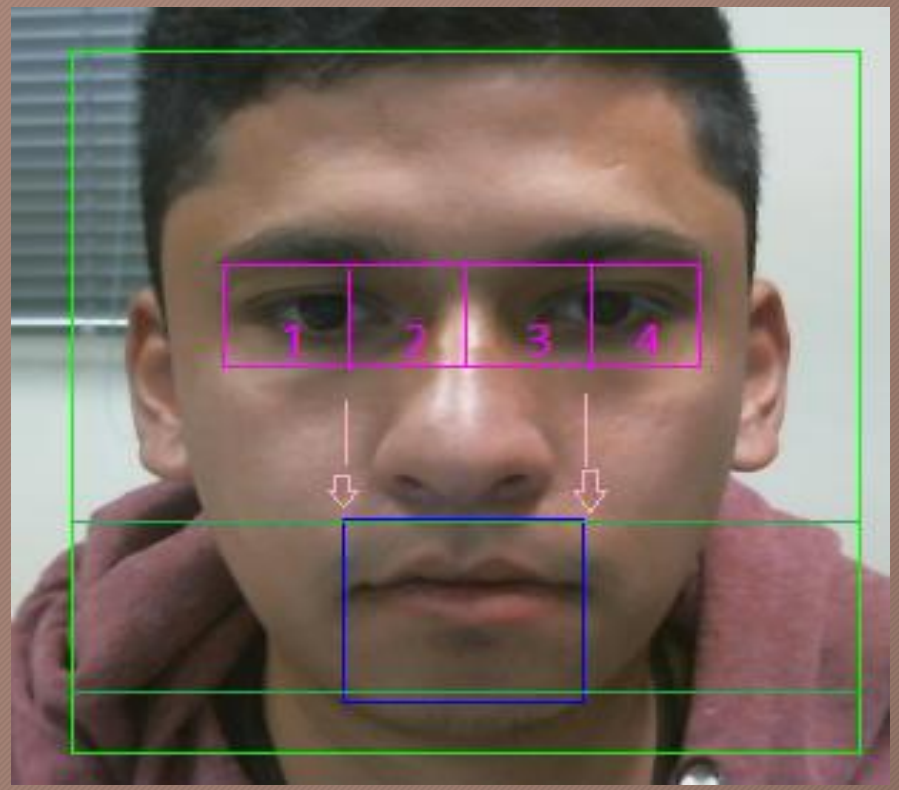
- ***cvCaptureFromCAM();***

Face and Eye Detection

- Object Detection
  - Viola & Jones Framework using Haar-Like Features

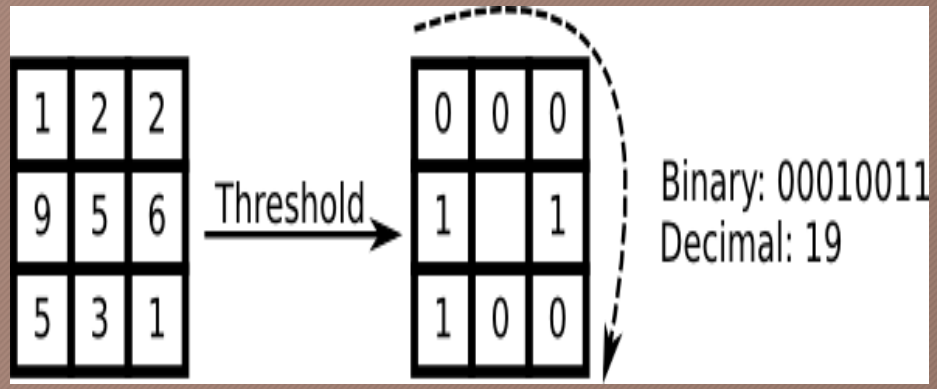
- *face_cascade.detectMultiScale();*
- *eye_cascade.detectMultiScale();*

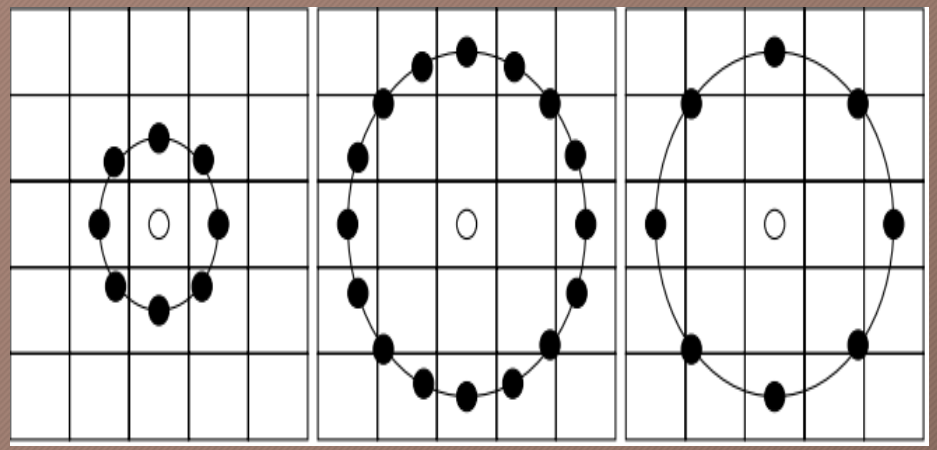
Segmentation of Region of Interest



- Split eyes into quadrants
  - X-axis
- Face height
  - Y-axis

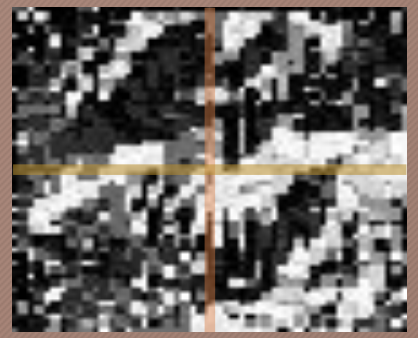
Local Binary Patterns

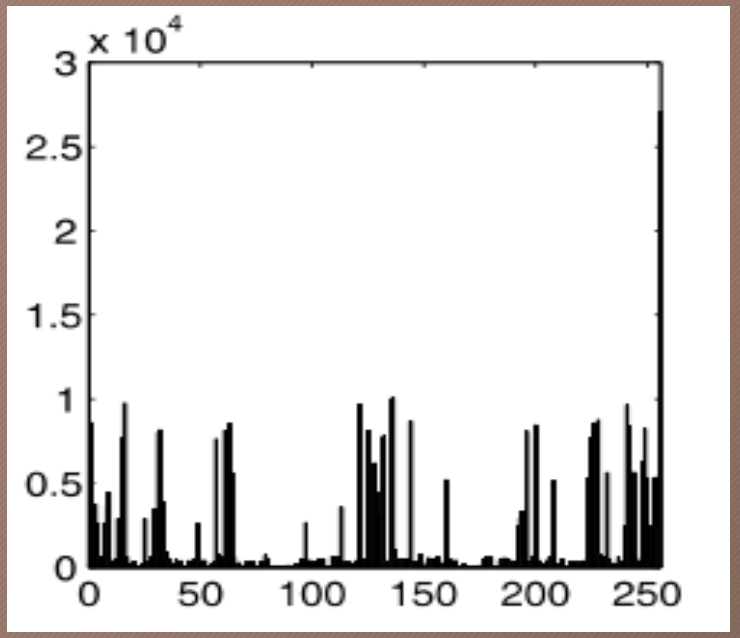




- Basic 3x3
  - Center Pixel => Neighbour = 1
  - Center Pixel < Neighbour = 0
- LBP_{PR}
  - Where P is number of neighbours
  - Where R is radius

Form Histogram





- Split LBP image (windows)
  - From Histogram
  - Concatenate
- 3x3 LBP
  - Bin Size: 2^8 = 256

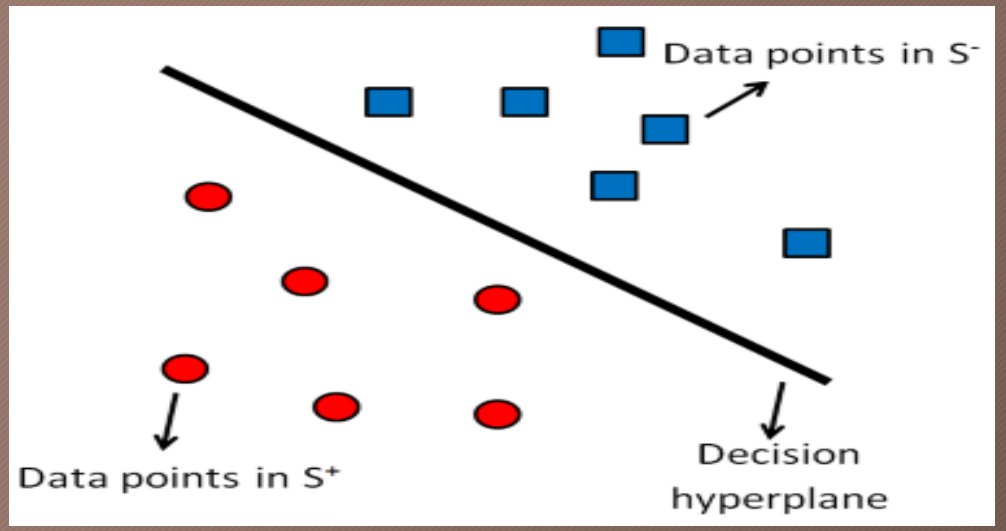
Support Vector Machine

TRAINING

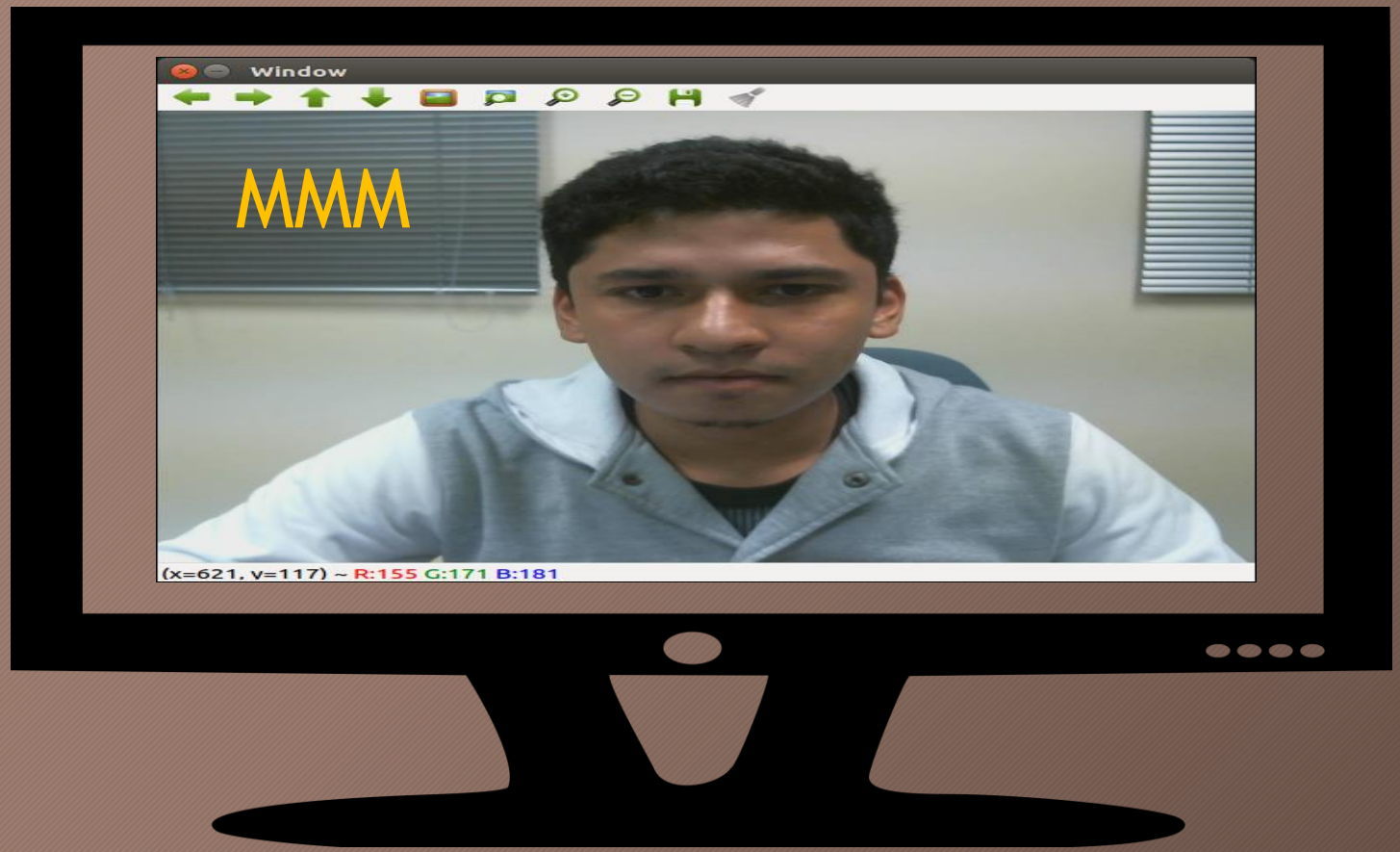
- Labeled Sounds
- Different Subjects

TESTING

- User system hasn't seen before
- SVM "Guesses answer"



Display Output



Project Plan

GOAL	Due Date
Research • Learn how to use OpenCV	End of Term 1
• Accurately locate mouth and extract features	End of Term 2
Implementation • Train the system to recognize a sounds or letters • Optimize image for better recognition	End of Term 3
Test and Evaluate • Add more training and testing data	End of Term4

References

- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. Nature,264:746-748.
- Paul Viola, M. J. (2001). Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference, volume 1, pages 511-518. IEEE.
- Matti Pietikinen, Guoying Zhao, A. H. T. A. (2011). Computer Vison Using Local Binary Patterns, volume 40. Springer.

Prototype Demo

- Detect Face
- Detect Eyes
- Segment Mouth
- LBP operation on Mouth

Questions?