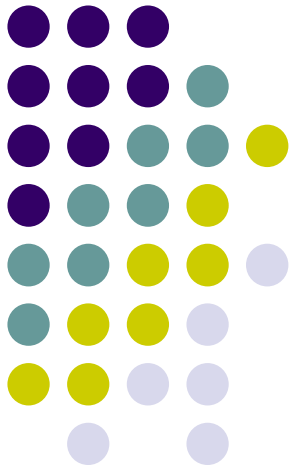
Advanced Computer Graphics
CS 525M: Visage: A Face Interpretation Engine for Smartphone Applications

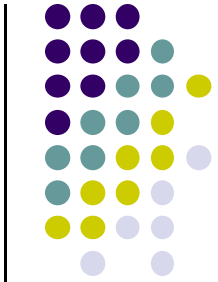
Zahid Mian

Computer Science Dept.

Worcester Polytechnic Institute (WPI)

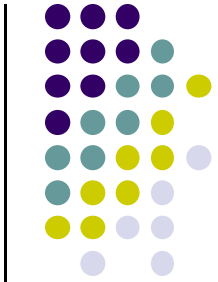


Problem/Motivation



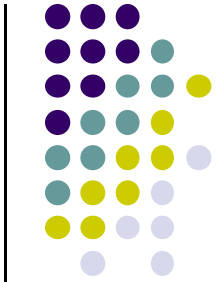
- Camera as Another Sensor
- Use Mobile Devices to …
  - Position of head
  - detect/analyze facial expressions
- Ultimately Build “smart” Apps that …
  - Use this information to provide an integrated experience
  - Provide Feedback to User
  - Others

Related Work



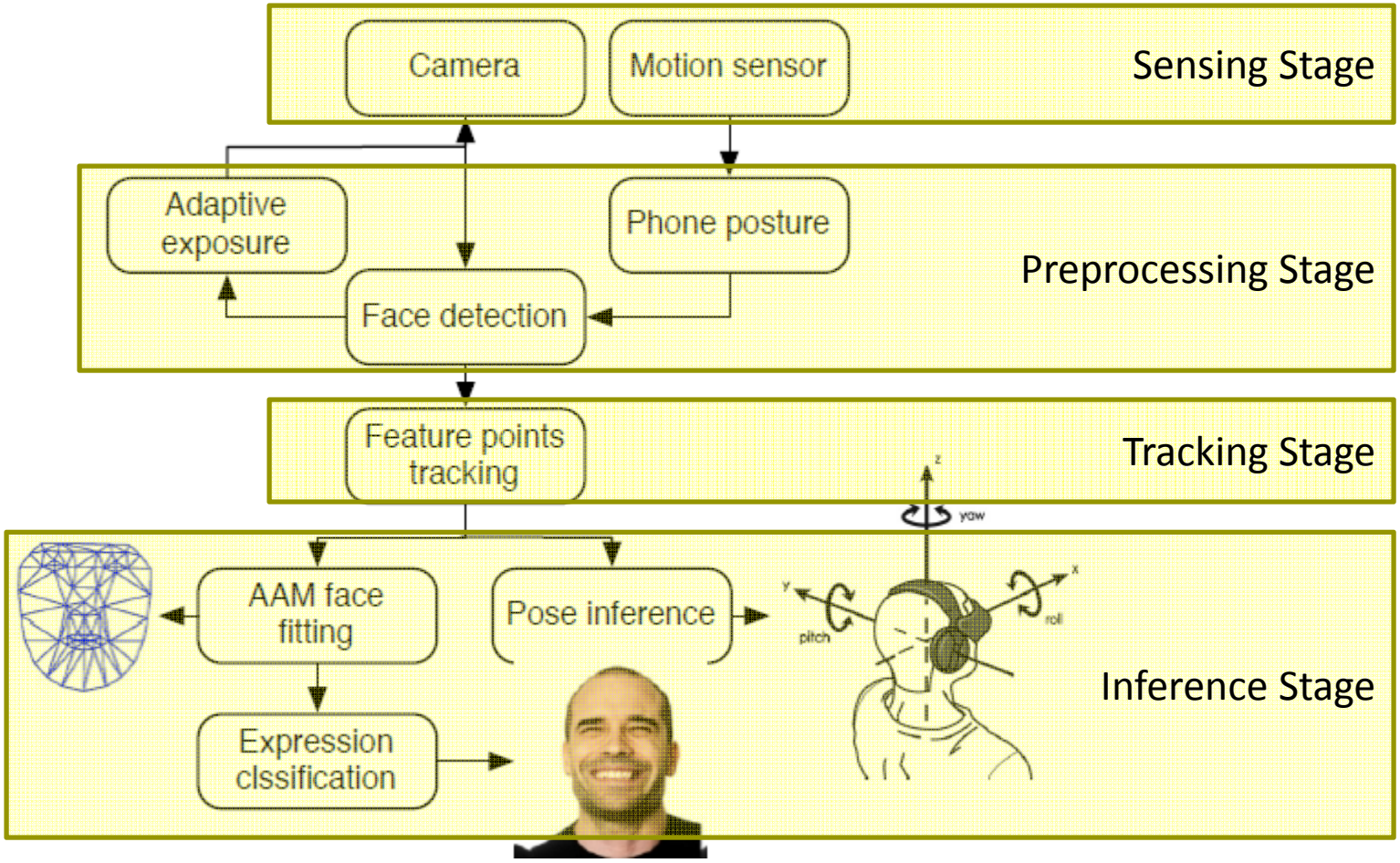
- Face Detection Mostly Limited to Desktop
  - Doesn’t take into account environment/context
- SenseCam
  - Simply takes pictures of everyday life (no processing)
- MoVi
  - Send Images to server and mine for common interests
- Google Goggles (Glass Project)
  - Mostly Server Side Processing

Limited Phone Resources

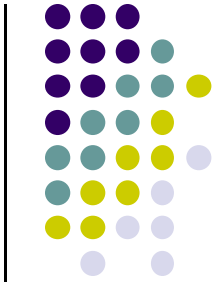


- Key Considerations:
  - Image Data Larger Compared to Other Sensors
  - Offloading Data a Transmission/Privacy Concerns
- Process Realtime, but
  - Downsampling images (192x144)
  - Larger Window Size for Sampling
  - Skip frames, if necessary
  - High CPU Usage

Visage System Architecture

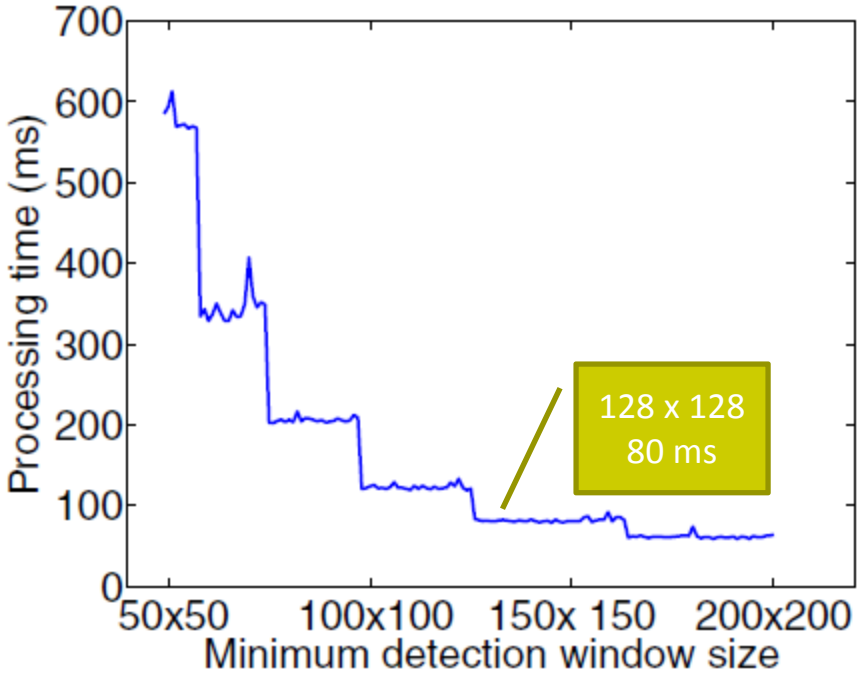


Preprocessing Stage

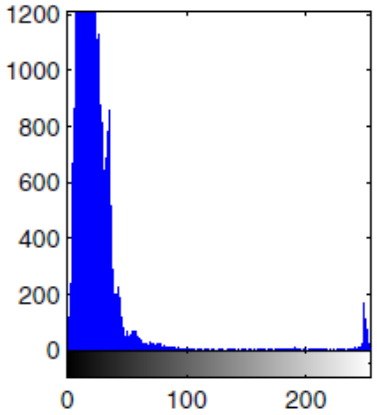


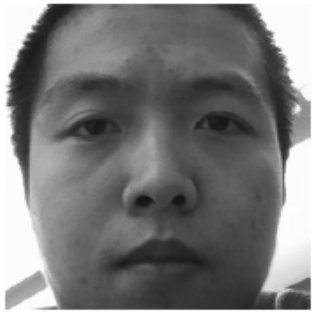
- Phone Posture Component
  - Identifies frames that contain user's face
  - Uses accelerometer/gyroscope data to determine gravity direction (phone's motion intensity)
- Face Detection with Tilt Compensation
  - AdaBoost Object detector (scan until face identified)
  - Visage compensates for phone's tilt
- Adaptive Exposure Component
  - Correct camera exposure level

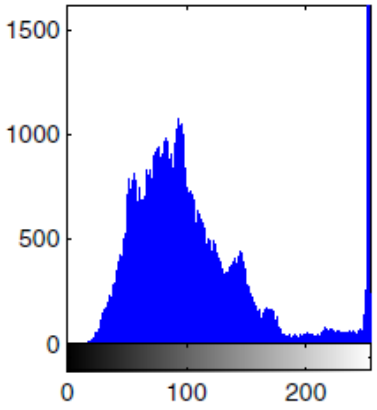
Detection Time and Window Size



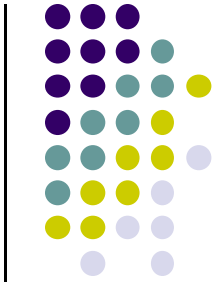
Example of Adaptive Exposure







Tracking Stage

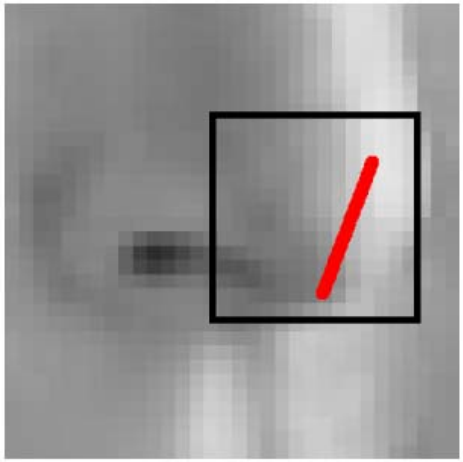


- Feature Points Tracking Component
  - Landmarks on face (eye corners, edges of mouth)
  - Lucas-Kanade method to track movement
  - CAMSHIFT allows for larger motion
- Pose Estimation Component (POSIT)
  - Pose from Orthography and Scaling with Iterations
  - Estimate 3D pose of user's head
  - Use cylinder as a baseline for head
    - x,y from 2D image; z from shape of cylinder
  - Determine rotation of cylinder
  - Use Calibration to compensate for modeling errors

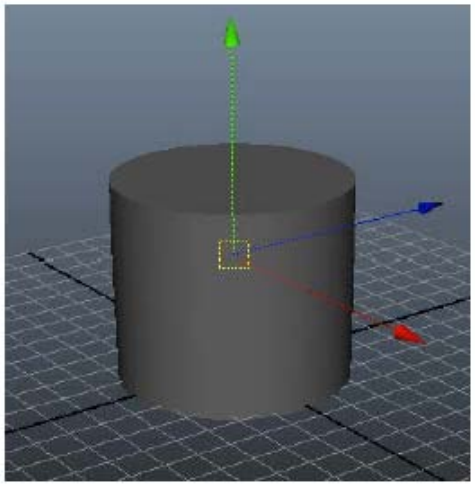
Example Lucas-Kanade method

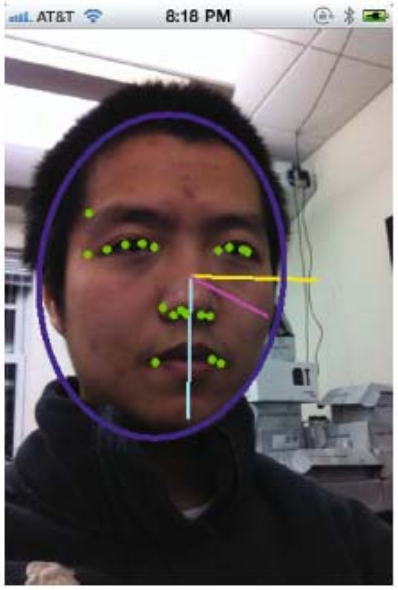
We want to track down her nose from here...

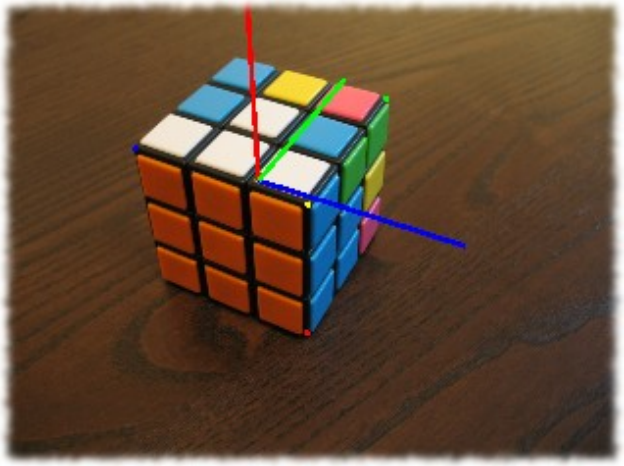
We track at first on a 2^3 scaled picture...



Examples of Pose Estimation

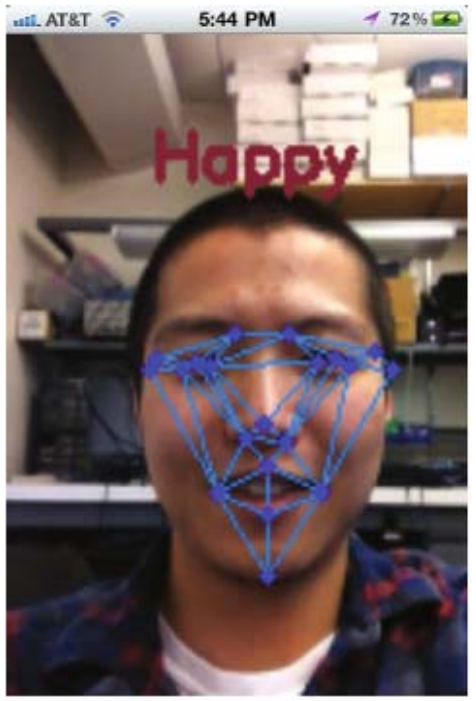




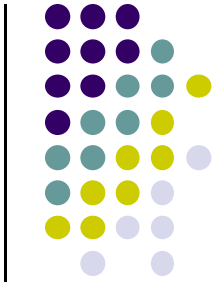


Inference Stage

- Active Appearance Models
  - Statistical method
  - Require training images (fitting process)
  - Triangular mesh, landmark points
  - Capture pixel color intensities
- Expression Classification
  - Anger, Disgust, Fear, Happy, Neutral, Sadness, Surprise
  - Fisherface technique for classification

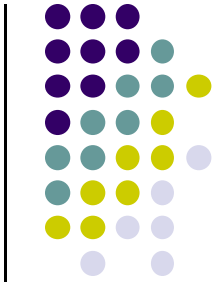


Implementation



- Apple iPhone 4
- Objective C (GUI)
- Core Processing in C
- OpenCV (Visage pipelines)

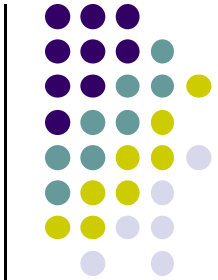
Performance Benchmarks



Tasks	Avg. CPU usage	Avg. memory usage
GUI only	$< 1\%$	3.18MB
Pose estimation	58%	6.07MB
Expression inference	29%	4.57MB
Pose estimation & expression inference	68%	6.28MB

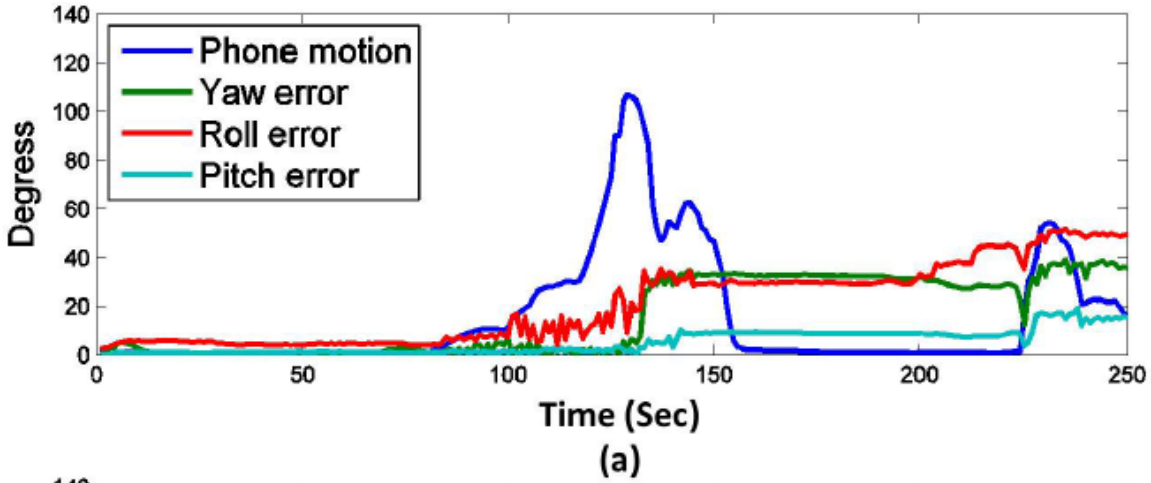
Component	Average processing time(ms)
Face detection	53
Feature points tracking	32
AAM fitting	92
Facial expression classification	3

Tilted Face Detection

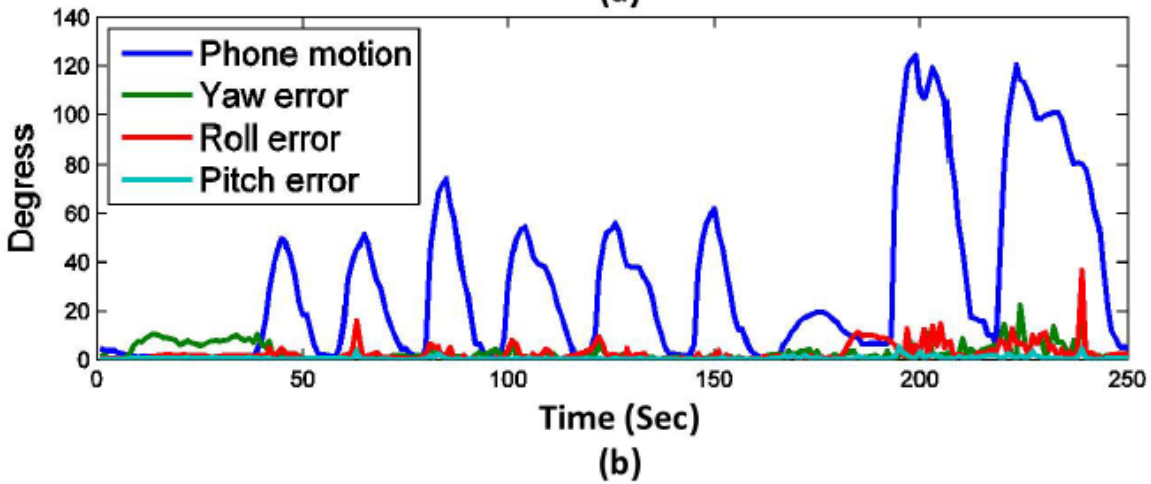


- Red-Colored Box indicates Detection
- Top Row: Default AdaBoost algorithm
- Bottom Row: Tilt Compensation (much better)
  - -90 ~ 90 degrees (range)

Phone Motion and Head Pose Estimation Errors

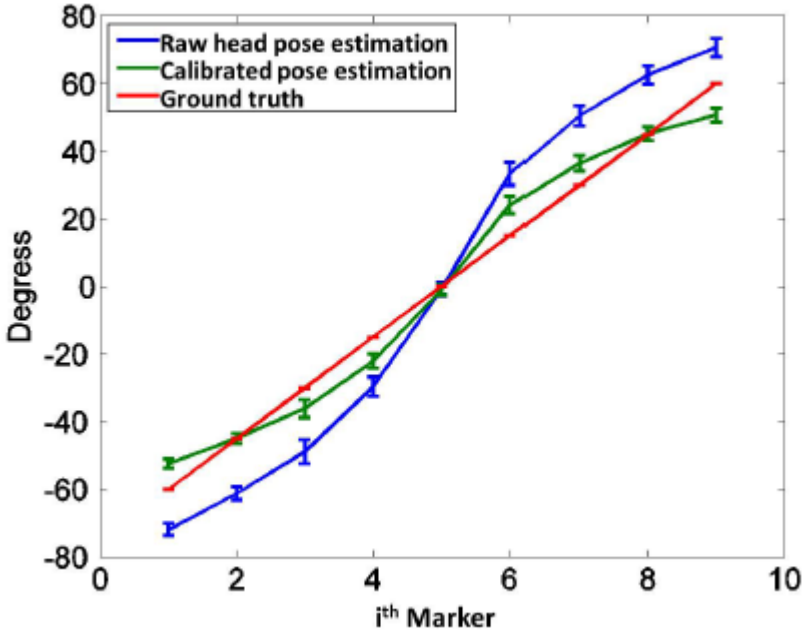


Without motion-based reinitialization



With motion-based reinitialization

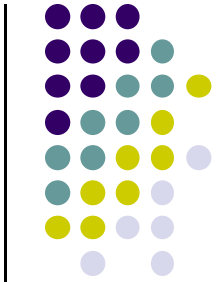
Accuracy of Head Pose Estimation



* 1-Meter Radius
* Several evenly spaced markers
* Volunteers asked to move head towards marker

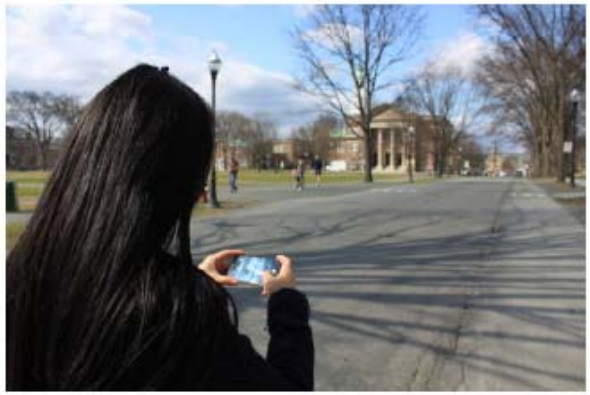
- Calibrated pose is close to ground truth

Facial Expression Confusion Matrix



Expressions	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise
Anger	93.33	6.67	0	0	0	0	0
Disgust	6.90	75.86	17.24	0	0	0	0
Fear	0	7.41	92.54	0	0	0	3.23
Happy	0	0	0	87.10	6.45	3.23	0
Neutral	0	0	0	0	90.00	10.00	0
Sadness	0	6.45	9.68	3.23	9.68	70.97	0
Surprise	0	0	3.33	3.33	0	0	93.33

Using Head Rotation – Streetview+



(a) Streetview+ on the go

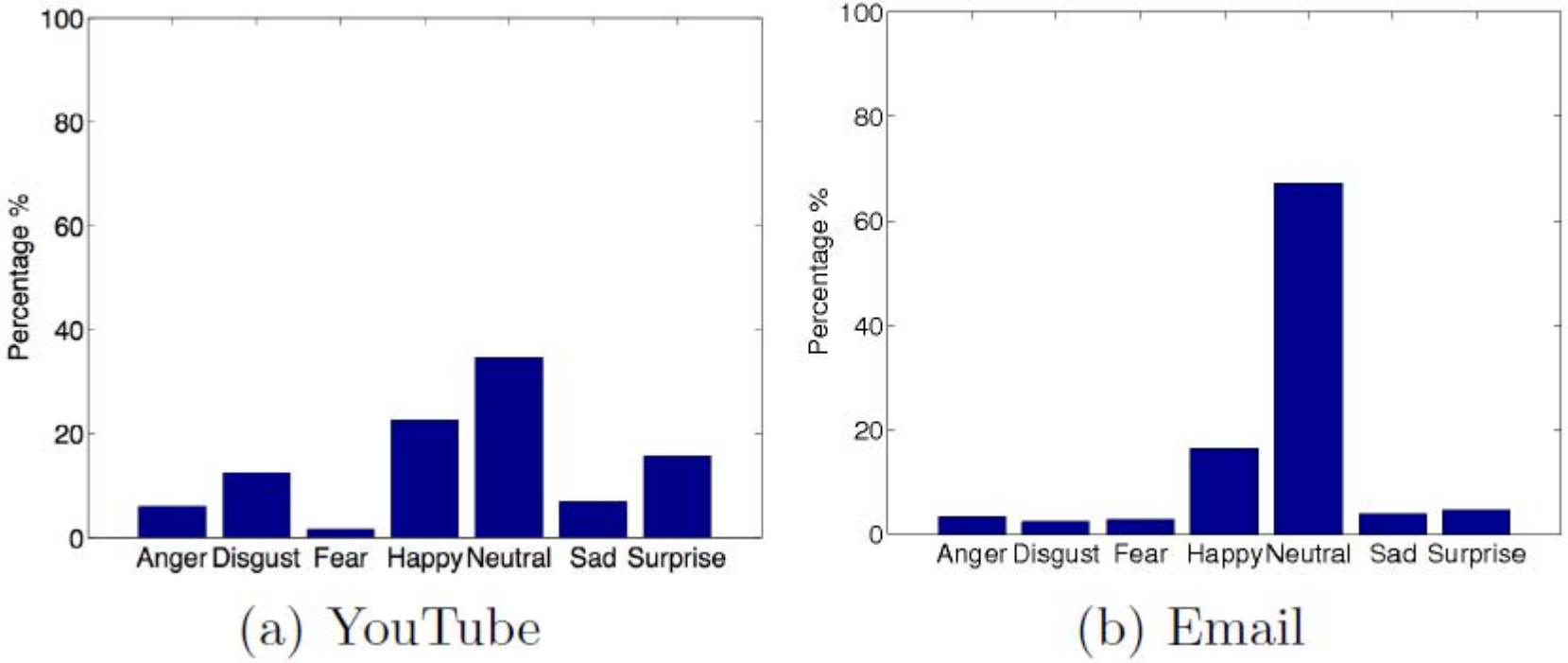
(b) Head facing front

(c) Head facing left

(d) Head facing right

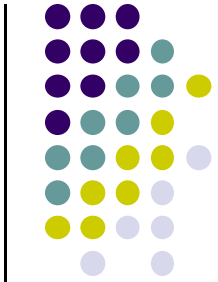
Streetview+ (based on Google Streetview) application automatically changes the view based on the rotation of head

Using Facial Expression – Mood Profiler



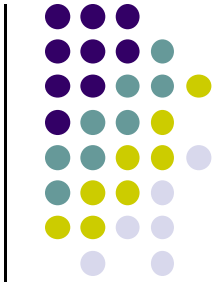
Shows a user's expression while (a) watching YouTube and (b) reading email – depends on accuracy of facial classification

Conclusion



- Using Phone's Camera As a Sensor
- Possible to do Facial Recognition in Realtime
- Compensate for Contextual Factors
- Experiment Results show robustness
- Use Camera to Build Integrated Apps
  - Head motion can be used in Apps like Streetview
  - Facial expressions can be used …
    - Provide feedback
    - Or even change mood (not in paper)

Critique/Thoughts …



- The Good …
  - Use of camera as a sensor
  - Myriad of experiments show robustness
  - Great Potential …
    - Play “happy” music if anger is detected
    - Notify friends if sadness detected
- The Not so Good …
  - Applications/Examples aren’t practical
  - Little discussion on Battery Usage
  - No experiments different skin tones

References

- *http://www.cs.dartmouth.edu/~campbell/visage.pdf*
- *http://copterix.perso.rezel.net/?page_id=58*
- *http://www.aforgenet.com/articles/posit/*
- *http://en.wikipedia.org/wiki/Project_Glass*