Ubiquitous and Mobile Computing
CS 528: *Visage: A Face Interpretation Engine for Smartphone Applications*

Qiwen Chen

Electrical and Computer Engineering Dept.

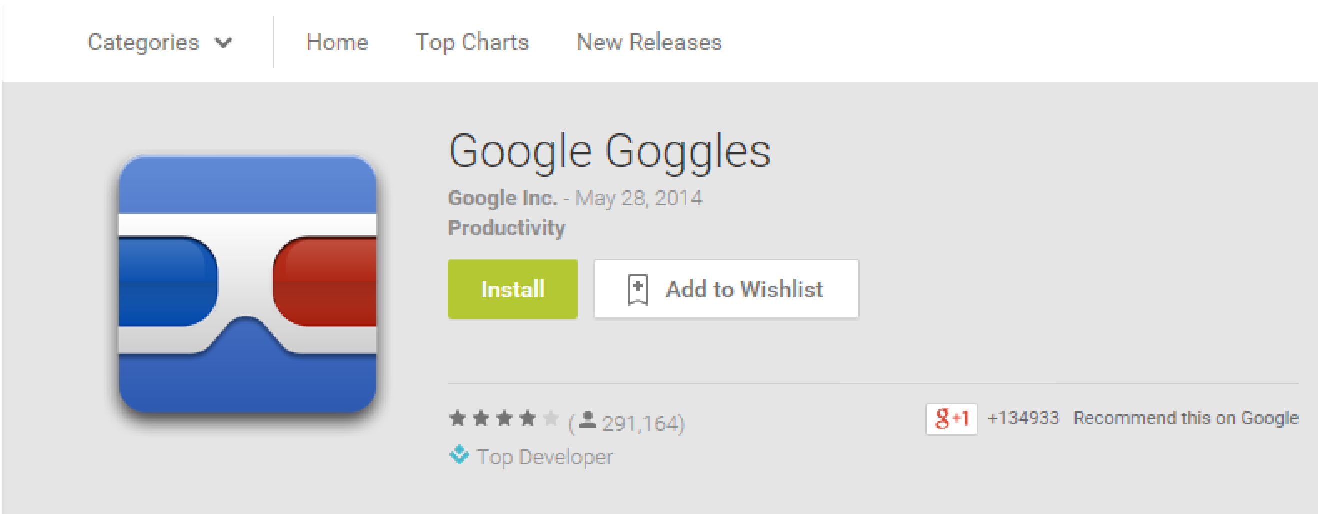
Worcester Polytechnic Institute (WPI)

Introduction

- Visage: A robust, real-time face interpretation engine for smart phones
- Tracking user’s 3D head poses & facial expression
- Fuse data from front-facing camera & motion sensor

Related Work

- Google Goggles



Related Work (Cont.)

- Recognizr

  Video Here

  Limited local image processing

- Mobile UI: PEYE

  Tracking 2D face representations

Methodology

Challenges:

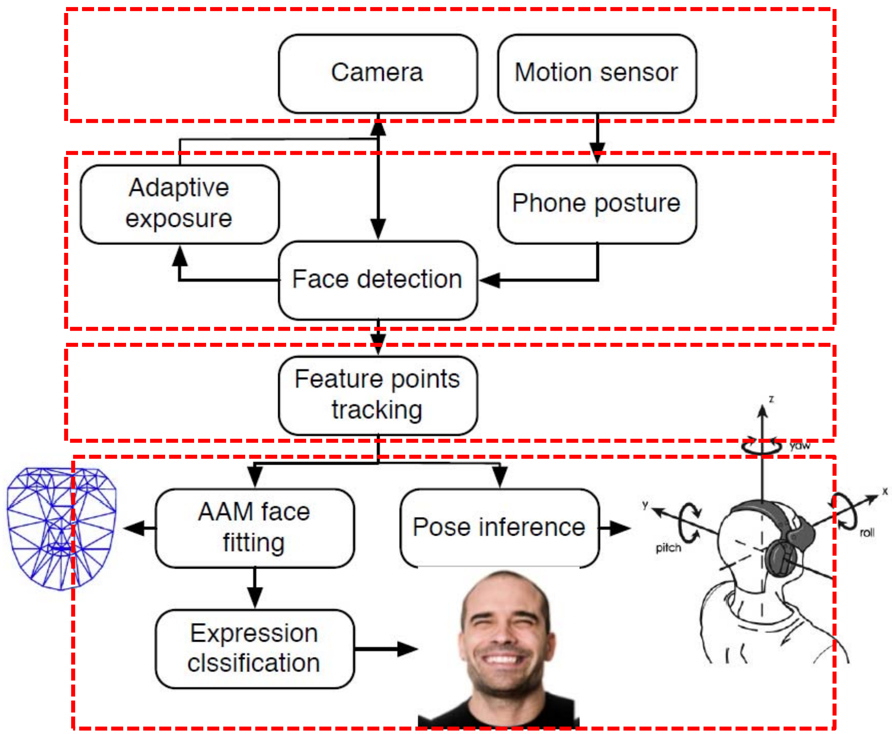
- User Mobility

Movement of the phone cause low image quality	Accelerometer & gyroscope sensor
Varying light condition	Analyze exposure level of face region

- Limited Phone Resources

  Operate in real-time

Methodology (Cont.)



Visage System Architecture

Sensing Stage

Preprocessing Stage

Tracking Stage

Inference Stage

Methodology (Cont.)
Preprocessing Stage

- Phone Posture Component

Gravity Direction: Mean of accelerometer

Motion intensity:

Variance of accelerometer & gyroscope

Methodology (Cont.)
Preprocessing Stage

- Face Detection with Tilt Compensation

AdaBoost object detector with tilt correction

$$\theta_g = \frac{180}{\pi} arctan \frac{a_x}{a_y}$$

Then the image is rotated by:

$$I_r = \begin{bmatrix} cos\theta_g & -sin\theta_g \\ sin\theta_g & cos\theta_g \end{bmatrix} I_i$$

Methodology (Cont.)

Preprocessing Stage

- Adaptive Exposure Component

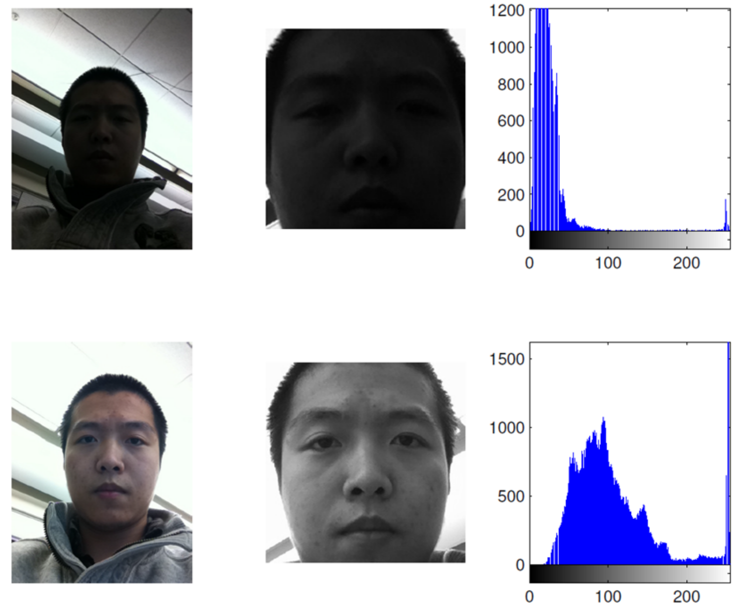
$$C_{H_{face}} = \frac{\sum_{i=0}^{255} i H_{face}(i)}{\sum_{i=0}^{255} i}$$

$C_{H_{face}}$ lies in lower ends:

Under-exposed

$C_{H_{face}}$ lies in higher ends:

over-exposed



Top: underexposed image, face region, and regional histogram; bottom: the image after adaptive exposure adjustment, face region, and regional histogram

Methodology (Cont.)
Tracking Stage

- Feature Points Tracking Component

  Select candidate feature point

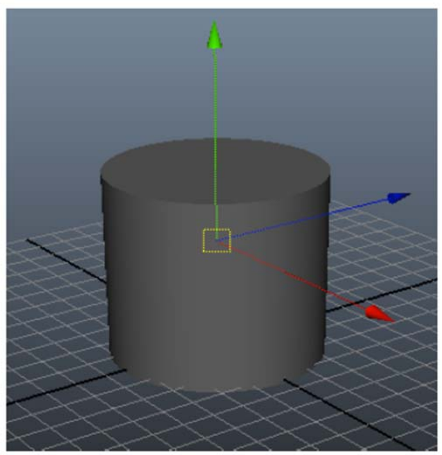
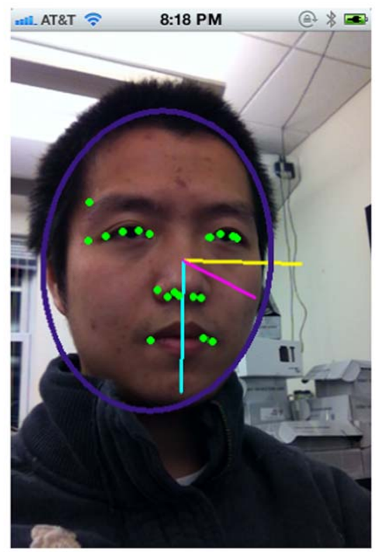
  Track points' location

Lucas-Kanade method (LK) & CAMSHIFT algorithm

Methodology (Cont.)
Tracking Stage

- Pose Estimation Component

Pose from Orthography and Scaling with Iterations

Methodology (Cont.)
Inference Stage

- Active Appearance Model

  Only when face orientation is near frontal

  Active Appearance Models: obtain shape (x) and texture (g) of user's face

$$\boldsymbol{x} = x_0 + Q_q c$$
$$\boldsymbol{g} = g_0 + Q_s c$$

Methodology (Cont.)

Inference Stage

- Expression Classification

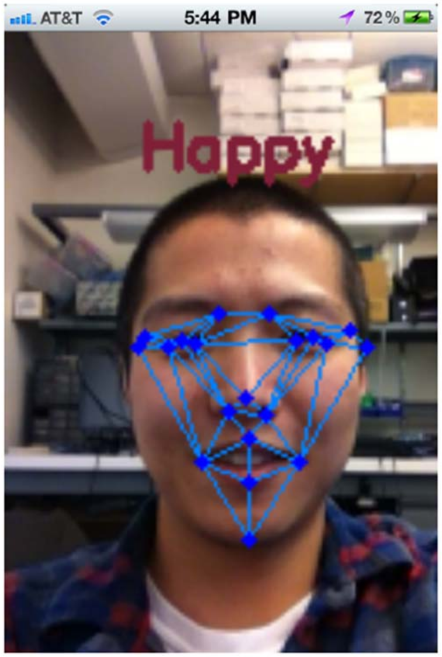
  Geometric & Appearance

Classification: Fisher Linear Discrimination (Fisherface)

$$P_{opt} = argmax_P \frac{|P^T S_B P|}{|P^T S_W P|}$$

$$S_B P = \lambda S_W P$$

$$v_{exp} = P I_{face}$$



Results
Implementation

- GUI, API: Objective C

  Core processing & inference routines: C

  Pipeline: OpenCV
- Resolution: 192 x 144 (face size 64 x 64)
- Frame skipping scheme

Results
Evaluation

Operating On Apple iPhone 4

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

CPU and memory usage under various task benchmarks

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

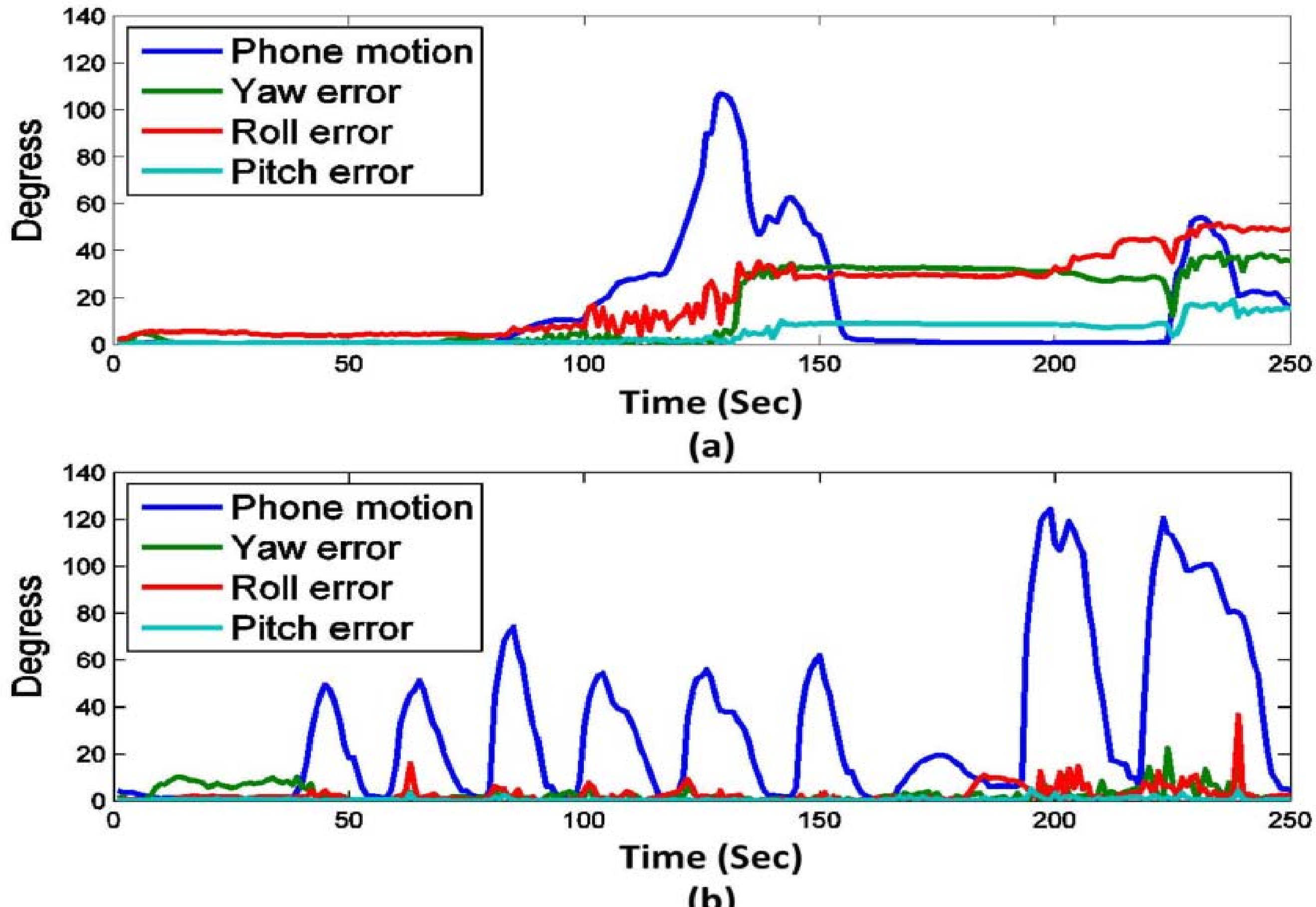
Processing time benchmarks

Results
Evaluation

Tilted angles: from -90 to 90 degrees, separated by an angle of 15 degrees.
First row : standard Adaboost face detector.
Second row is detected by Visage's detector.

Results

Evaluation

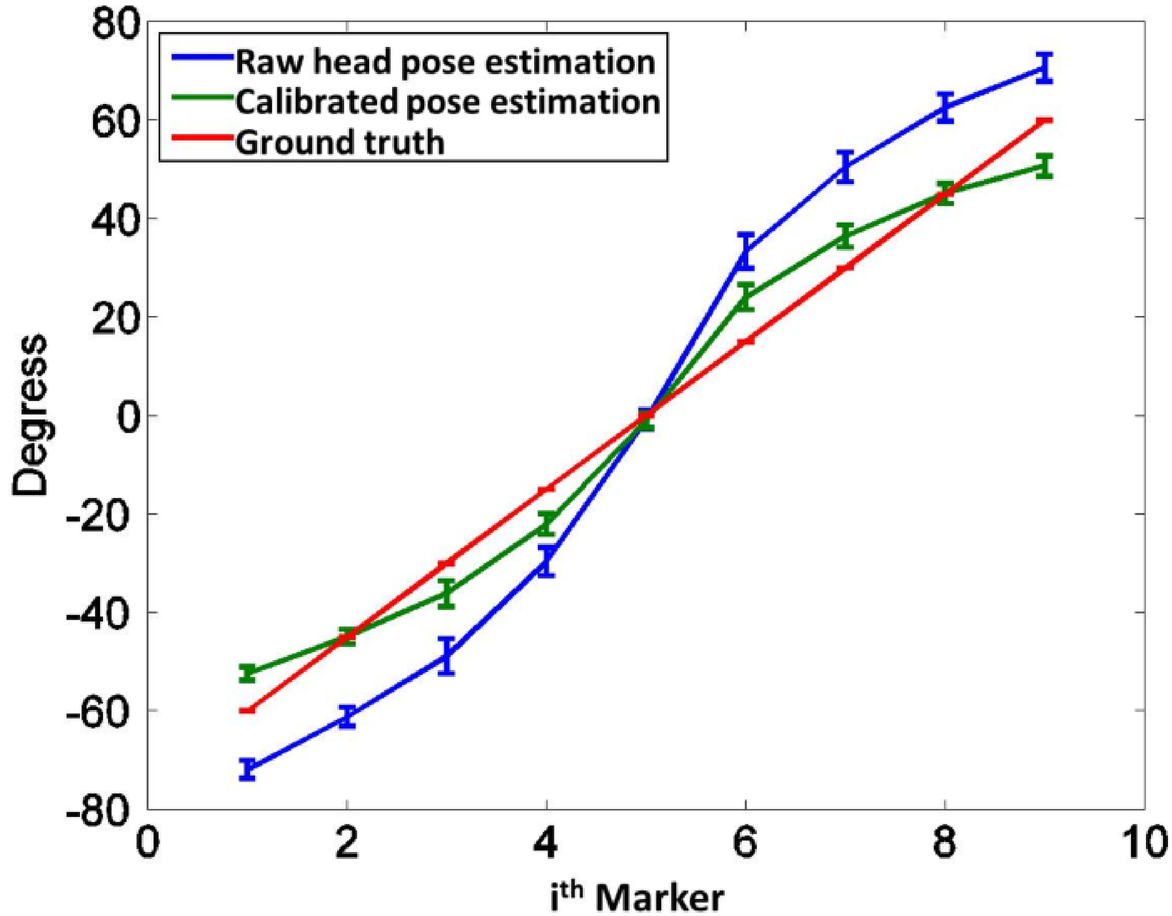


Phone motion and head pose estimation errors

(a)without motion-based reinitialization

(b)with motion-based reinitialization

Results
Evaluation



Head Pose Estimation Error, 3 volunteers, 5 samples each

Results

Evaluation

Expressions	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise
Accuracy(%)	82.16	79.68	83.57	90.30	89.93	73.24	87.52

Facial expression classification accuracy using the JAFFE dataset, 5 Volunteers. The model is personalized by user's own data

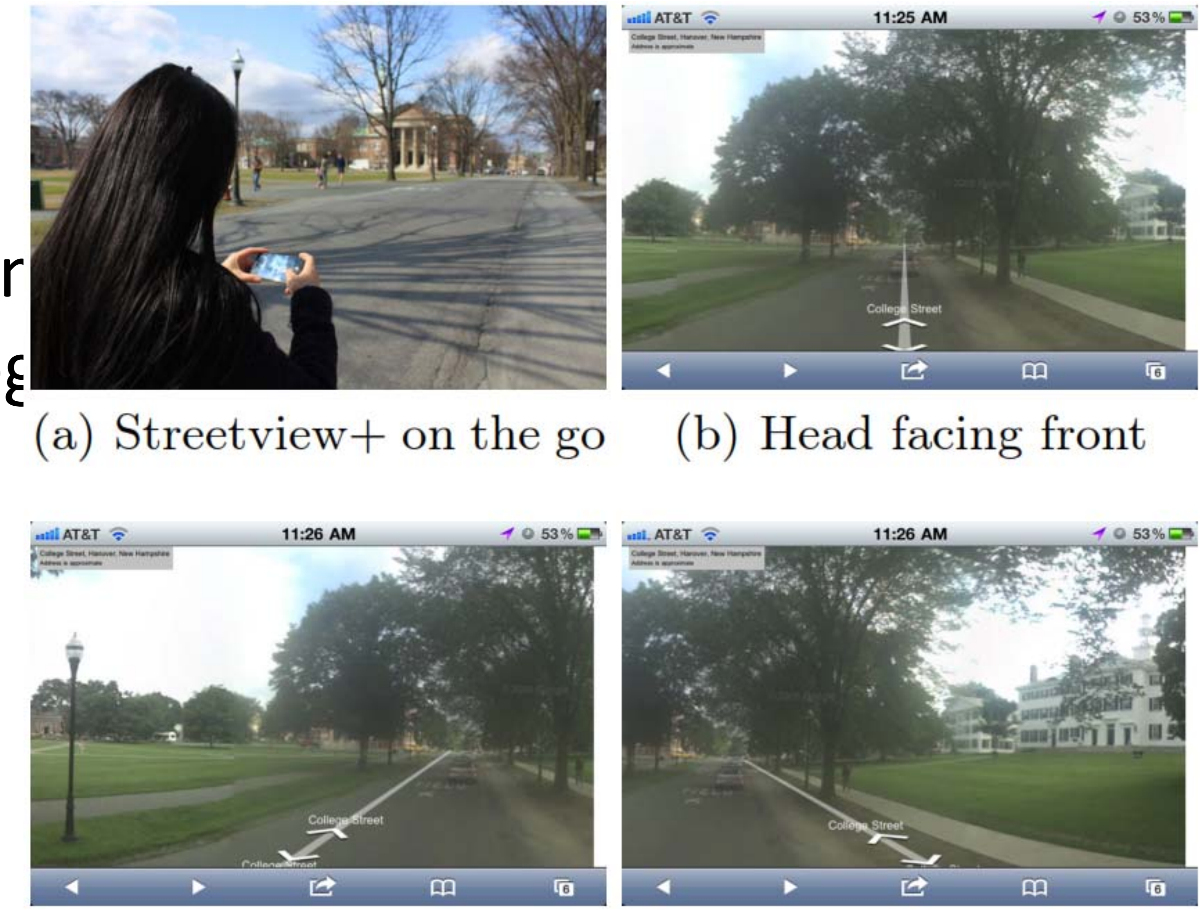
Expressions	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise
Anger	93.33	6.67	0	0	0	0	0
Disgust	6.90	75.86	17.24	0	0	0	0
Fear	0	7.41	92.54	0	0	0	3.23
Happy	0	0	0	87.10	6.45	3.23	0
Neutral	0	0	0	0	90.00	10.00	0
Sadness	0	6.45	9.68	3.23	9.68	70.97	0
Surprise	0	0	3.33	3.33	0	0	93.33

Confusion matrix of facial expression classification based on JAFFE

Application

- Streetview+

  Show the 360-degree panorar view from Goog Streetview

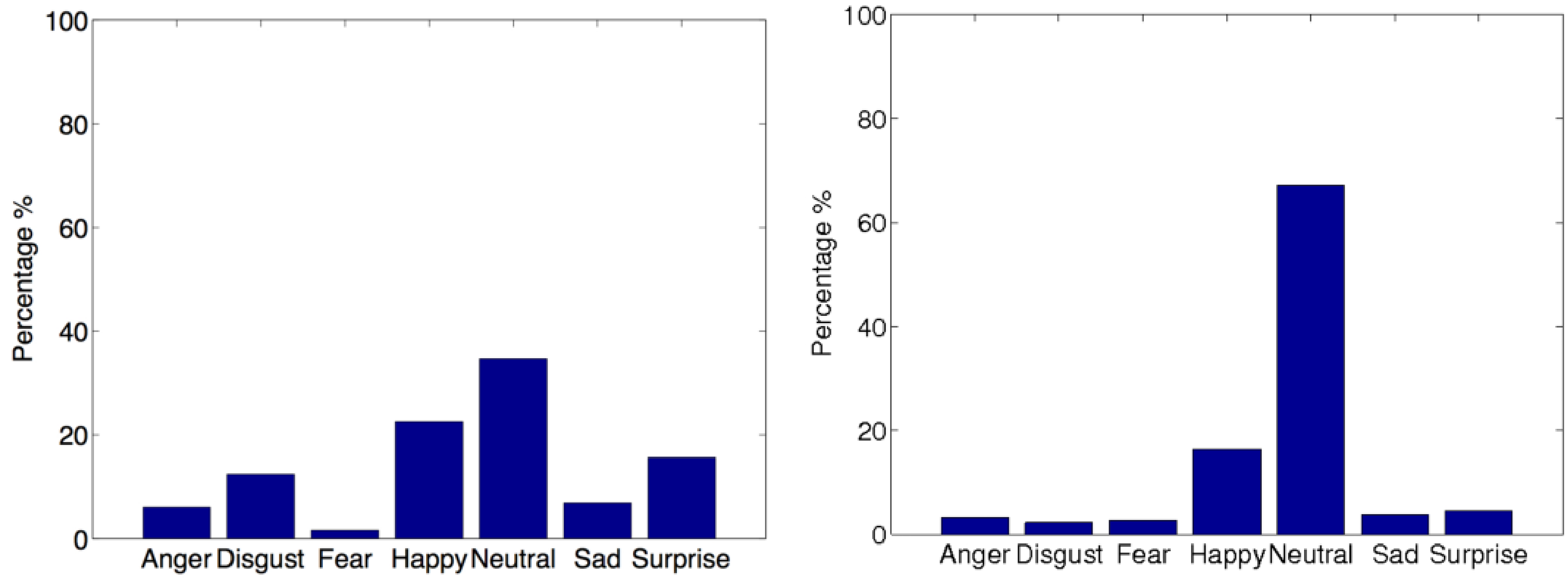


(a) Streetview+ on the go (b) Head facing front

(c) Head facing left (d) Head facing right

Application

- Mood Profiler



(a) YouTube (b) Email

References

- [1] *Recognizr, http://news.cnet.com/8301-137723-10458736-52.html*
- [2] *Hua, G., Yang, T., Vasireddy, S.: PEYE: Toward a Visual Motion Based Perceptual Interface for Mobile Devices. In: Proc. of the 2007 IEEE int'l conf. Human-computer interaction, pp. 39–48, Springer-Verlag, Berlin (2007)*
- [3] *Viola, P., Jones, M.J.: Robust Real-time Face Detection. In: Int'l J. Comput.Vision, 57, pp. 137-154 (2004)*

References

- [4] *Baker, S., Matthews, I.: Lucas-kanade 20 Years On: A Unifying Framework. In: Int'l J. Comput. Vision, 56(3),pp. 221-255 (2004)*
- [5] *Dementhon, D.F., Davis, L.S.: Model-based Object Pose in 25 Lines of Code. In: Int'l J. Comput. Vision 15, 1-2, pp. 123–141 (1995)*
- [6] *Matthews, I., Baker, S.: Active Appearance Models Revisited. In: Int'l J. Comput.Vision, 60(2), pp. 135-164 (2004)*

References

- [7] *Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection. In: Trans. Pattern Anal. Mach. Intell., 19(7), pp. 711-720 (1997)*
- [8] *Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding Facial Expressions with Gabor Wavelets. In: Proc. 3rd IEEE Int'l Conf. Automatic Face and Gesture Recognition, pp. 200-205, IEEE Computer Society, Washington, DC (1998)*