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

- Sensing Stage
  - Camera
  - Motion sensor

- Preprocessing Stage
  - Adaptive exposure
  - Face detection
  - Phone posture

- Tracking Stage
  - Feature points tracking

- Inference Stage
  - AAM face fitting
  - Pose inference
  - Expression classification
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

The graph shows the relationship between the minimum detection window size and the processing time (ms). The graph indicates that a window size of 128 x 128 results in a processing time of 80 ms.
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 2x 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

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Avg. CPU usage</th>
<th>Avg. memory usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUI only</td>
<td>&lt; 1%</td>
<td>3.18MB</td>
</tr>
<tr>
<td>Pose estimation</td>
<td>58%</td>
<td>6.07MB</td>
</tr>
<tr>
<td>Expression inference</td>
<td>29%</td>
<td>4.57MB</td>
</tr>
<tr>
<td>Pose estimation &amp; expression inference</td>
<td>68%</td>
<td>6.28MB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component</th>
<th>Average processing time(ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face detection</td>
<td>53</td>
</tr>
<tr>
<td>Feature points tracking</td>
<td>32</td>
</tr>
<tr>
<td>AAM fitting</td>
<td>92</td>
</tr>
<tr>
<td>Facial expression classification</td>
<td>3</td>
</tr>
</tbody>
</table>
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

(a) Without motion-based reinitialization

(b) 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

<table>
<thead>
<tr>
<th>Expressions</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Neutral</th>
<th>Sadness</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>93.33</td>
<td>6.67</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Disgust</td>
<td>6.90</td>
<td>75.86</td>
<td>17.24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
<td>7.41</td>
<td>92.54</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.23</td>
</tr>
<tr>
<td>Happy</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>87.10</td>
<td>6.45</td>
<td>3.23</td>
<td>0</td>
</tr>
<tr>
<td>Neutral</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>90.00</td>
<td>10.00</td>
<td>0</td>
</tr>
<tr>
<td>Sadness</td>
<td>0</td>
<td>6.45</td>
<td>9.68</td>
<td>3.23</td>
<td>9.68</td>
<td>70.97</td>
<td>0</td>
</tr>
<tr>
<td>Surprise</td>
<td>0</td>
<td>0</td>
<td>3.33</td>
<td>3.33</td>
<td>0</td>
<td>0</td>
<td>93.33</td>
</tr>
</tbody>
</table>
Using Head Rotation – Streetview+

Streetview+ (based on Google Streetview) application automatically changes the view based on the rotation of head.
Using Facial Expression – Mood Profiler

(a) YouTube

(b) Email

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://copterix.perso.rezel.net/?page_id=58
- http://www.aforgenet.com/articles/posit/