Face detection, face alignment, and face image parsing

Brandon M. Smith
Guest Lecturer, CS 534
Monday, October 21, 2013

Lecture overview

- Brief introduction to local features
- Face detection
- Face alignment and landmark localization
- Face image parsing

Local features: broad goal

- What are local features trying to capture?
  The local appearance in a region of the image

Local features: motivation

What are their uses?
  - Matching

Local features: motivation

What are their uses?
- Matching
- Image indexing and retrieval
- Aligning images, e.g., for panorama stitching
- Video stabilization

Shen et al., CVPR 2012

http://www.leet.it/home/lale/files/Garda-pano.jpg

Local features: motivation

What are their uses?
- Matching
- Image indexing and retrieval
- Aligning images, e.g., for panorama stitching
- Video stabilization
- 3D reconstruction

Local features: motivation

What are their uses?
- Matching
- Image indexing and retrieval
- Aligning images, e.g., for panorama stitching
- Video stabilization
- 3D reconstruction
- Object recognition, including face recognition

Local features: types

Types of features and feature descriptors
- Image intensity or gradient patches
- Shift Invariance Feature Transform (SIFT) – very popular!
- DAISY
- SURF
- Many more…

Face detection: goal

Automatically detect the presence and location of faces in images.

Face detection: motivation

- Automatic camera focus
Face detection: motivation

- Automatic camera focus
- Easier photo tagging

Face detection: challenges

- Large face shape and appearance variation
- Scale and rotation (yaw, roll, pitch) variation
- Background clutter
- Occlusions
- Image noise
- Efficiency
- False positives

Face detection: Viola-Jones*

  - Feature type?
  - Which features are important?
  - Decide: face or not a face

* Next few slides are based on a presentation by Kostantina Pall & Alfredo Kalaitzis, available at http://www1.cs.columbia.edu/~belhumeur/courses/biometric/2010/violajones.ppt
Face detection: Viola-Jones

Feature type?
- Useful domain knowledge:
  - The eye region is darker than the forehead or the upper cheeks
  - The nose bridge region is brighter than the eyes
  - The mouth is darker than the chin
- Encoding
  - Location and size: eyes, nose bridge, mouth, etc.
  - Value: darker vs. brighter

Face detection: Viola-Jones

Feature type?
- Rectangle features
  - Value = \( \sum \) (pixels in black) - \( \sum \) (pixels in white)
  - Three types: 2, 3, 4 rectangles
  - Very fast: integral image

Face detection: Viola-Jones

Which features are important?
- Tens of thousands of features to choose from
- AdaBoost (Singer and Schapire, 1997)
  - Given a set of weak classifiers: \( h_t(x) \in \{-1,1\} \)
  - Iteratively combine classifiers to form a strong classifier:
    \[
    H(x) = \begin{cases} 
    1 & \text{if } \left( \sum \alpha_i h_t(x) \right) > \text{threshold} \\ 
    0 & \text{otherwise}
    \end{cases}
    \]

Final decision: face or not a face
- Cascade of classifiers
  1. Two-feature classifier: >99% recall, >60% precision
  2. Five-feature classifier
  3. 10-feature classifier
  4. 
  5. 20-feature classifier

Face detection: Viola-Jones

http://vimeo.com/12774628#

Face detection: recent approaches


Face detection: recent approaches

Shen et al., Detecting and Aligning Faces by Image Retrieval, CVPR 2013.

(a) Validation result of a true positive.

(b) Validation result of a false positive.
Face alignment and landmark localization: goal

Goal of face alignment: automatically align a face (usually non-rigidly) to a canonical reference

Goal of face landmark localization: automatically locate face landmarks of interests

Face alignment and landmark localization: motivation

- Preprocess for:
  - Face recognition
  - Portrait editing wizards
  - Face image retrieval
  - …
- Face tracking
- Expression recognition
- Facial pose recognition

Face alignment and landmark localization: challenges

- Pose
- Expression
- Identity variation
- Occlusions
- Image noise

Face alignment and landmark localization: approaches

Parametric appearance models
- Cootes, Edwards, and Taylor, Active Appearance Models, ECCV 1998
Face alignment and landmark localization: approaches

Parametric appearance models
- Cootes, Edwards, and Taylor, Active Appearance Models, ECCV 1998

Part-based deformable models
- Saragih et al., Face Alignment through Subspace Constrained Mean-Shifts, ICCV 2009

Supervised descent
- Xiong and De la Torre, Supervised Descent Method and Its Applications to Face Alignment, CVPR 2013
Face alignment and landmark localization: approaches

Exemplar-based/non-parametric methods
- Shen et al., Detecting and Aligning Faces by Image Retrieval, CVPR 2013.

Face image parsing

Smith, Zhang, Brandt, Lin, and Yang, Exemplar-Based Face Parsing, CVPR 2013.

Face image parsing: goal

Given an input face image, automatically segment the face into its constituent parts.

Face image parsing: motivation

- Like face alignment, can be used as a preprocess for face recognition, automated portrait editing, etc.
- Encodes ambiguity
- Generalizes to hair, teeth, ears etc., across datasets
Face image parsing: our approach

Database

2K exemplar images
11 landmarks
~150 SIFT features

Exemplar labels

Step 0: Rough alignment &
Top exemplar selection

Step 1: Nonrigid alignment

Input

Step 2: Exemplar label aggregation

12/6/2013
Face image parsing: our approach

**Database**
- 11 landmarks
- ~150 SIFT features

**Exemplar labels**

**Output**

**Step 0: Rough alignment & Top exemplar selection**

**Step 1: Nonrigid alignment**

**Step 2: Exemplar label aggregation**

**Step 3: Pixel-wise label selection**

\[ w_1 \ast \text{Label 1} + w_2 \ast \text{Label 2} + \cdots + w_9 \ast \text{Label 9} \]

Face image parsing: quantitative results

<table>
<thead>
<tr>
<th>Method</th>
<th>Eyes</th>
<th>Brows</th>
<th>Nose</th>
<th>In Mouth</th>
<th>Upper Lip</th>
<th>Lower Lip</th>
<th>Mouth/All</th>
<th>Face Skin</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhu &amp; Romer [20]</td>
<td>0.533</td>
<td>m/a</td>
<td>m/a</td>
<td>0.425</td>
<td>0.472</td>
<td>0.455</td>
<td>m/a</td>
<td>m/a</td>
<td>m/a</td>
</tr>
<tr>
<td>Saragih et al. [13]</td>
<td>0.679</td>
<td>0.598</td>
<td>0.899</td>
<td>0.600</td>
<td>0.576</td>
<td>0.700</td>
<td>0.749</td>
<td>m/a</td>
<td>0.773</td>
</tr>
<tr>
<td>Liu et al. [12]</td>
<td>0.770</td>
<td>0.640</td>
<td>0.843</td>
<td>0.600</td>
<td>0.576</td>
<td>0.700</td>
<td>0.749</td>
<td>m/a</td>
<td>0.773</td>
</tr>
<tr>
<td>Gu &amp; Kanade [1]</td>
<td>0.743</td>
<td>0.681</td>
<td>0.889</td>
<td>0.545</td>
<td>0.568</td>
<td>0.700</td>
<td>0.749</td>
<td>m/a</td>
<td>0.773</td>
</tr>
<tr>
<td>Ours, Steps 1 &amp; 3 united</td>
<td>0.766</td>
<td>0.687</td>
<td>0.896</td>
<td>0.628</td>
<td>0.637</td>
<td>0.703</td>
<td>0.853</td>
<td>0.861</td>
<td>0.779</td>
</tr>
<tr>
<td>Ours, Step 3 united</td>
<td>0.772</td>
<td>0.708</td>
<td>0.914</td>
<td>0.659</td>
<td>0.629</td>
<td>0.697</td>
<td>0.850</td>
<td>0.872</td>
<td>0.779</td>
</tr>
<tr>
<td>Ours, full pipeline</td>
<td>0.750</td>
<td>0.672</td>
<td>0.922</td>
<td>0.713</td>
<td>0.651</td>
<td>0.700</td>
<td>0.857</td>
<td>0.882</td>
<td>0.804</td>
</tr>
</tbody>
</table>

Face image parsing: qualitative results

**Input**

**Soft segments**

**Hard segments**

**Ground truth**

*CS 534: Computation Photography* 12/6/2013 #43

*CS 534: Computation Photography* 12/6/2013 #44

*CS 534: Computation Photography* 12/6/2013 #45

*CS 534: Computation Photography* 12/6/2013 #46