9.913 Pattern Recognition for Vision

Class 9 – Object Detection and Recognition

Bernd Heisele
Overview

1. Detection
2. Recognition
3. Demo
General Problems of Detection and Recognition

Rotation/illumination invariance

Applicable to many classes of objects
Object Detection

Task

Given an input image, determine if there are objects of a given class (e.g. faces, people, cars..) in the image and where they are located.

Photograph by MIT OCW.
Detection—Problems

1. Classifier must generalize over all exemplars of one class.

2. Negative class consists of everything else.

3. High accuracy (small FP rate) required for most applications.

Images removed due to copyright considerations.
Face Detection

Feature vector \((x_1, x_2, \ldots, x_n)\)

Off-line training

Feature Extraction

Classification Result

Photograph by MIT OCW.

Search for faces at different resolutions and locations

Face examples

Non-face examples

Pixel pattern
Training and Testing

Training Set

Train Classifier

Labeled Test Set

Classify

Photograph by MIT OCW.

Sensitivity

False Positive

Correct

Photograph by MIT OCW.
Image Features

Gray
- 19x19 Histogram Equalization
- Masking

Gradient
- 17x17 [0, 1] x-y-Sobel Filtering

Wavelets
- Haar Wavelet Transform
- Normalization

283 [0, 1]
17x17 [0, 1]
1740 [0, 1]
ROC Image Features

(CMU Testset 1, 127 images, 479 faces, 56.774.966 windows, res 19x19, pos 2429, neg 19932)
Positive Training Data

Real
2900 faces + 2900 mirrored faces

Synthetic
(T. Vetter, Univ. of Freiburg)

Illumination

Faces

Rotation

3D Morphing: + =

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Real vs. Synthetic

(CMU Testset 1, 127 images, 479 faces, 58,774,866 windows, res 19x19, real pos 2429, pos synth 4536, neg 19932)
Negative Training Data

Problem: 1 face in 116,440 examined windows

Initial set of 25,000 non-faces

Add to training set

Retrain Classifier

Determine false-positives on large set of non-face images

Bootstrapping

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Bootstrapping

(CMU Testset 1, 127 images, 479 faces, 56,774.666 windows, res 19x19, pos 5,762, neg no boot 23,300, neg boot)

Correct

0.9 FP/image   6.7 FP/image

FP / inspected window
### Performance of Global Face Detectors

<table>
<thead>
<tr>
<th>System</th>
<th>Subset of test set 1 23 images, 155 faces</th>
<th>Test set 1 130 images, 507 faces</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Det. Rate</td>
<td>FPs</td>
</tr>
<tr>
<td>[Sung 96] Neural Network</td>
<td>84.6%</td>
<td>13</td>
</tr>
<tr>
<td>[Osuna 98] SVM</td>
<td>74.2%</td>
<td>20</td>
</tr>
<tr>
<td>[Rowley et al. 98] Single neural network</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>[Rowley et al. 98] Multiple neural networks</td>
<td>84.5%</td>
<td>8</td>
</tr>
<tr>
<td>[Schneiderman &amp; Kanade 98]³ Naïve Bayes</td>
<td>91.1%</td>
<td>12</td>
</tr>
<tr>
<td>[Yang et al. 99]⁴ SNoW, multi-scale</td>
<td>94.1%</td>
<td>3</td>
</tr>
<tr>
<td><strong>Our system</strong>⁵</td>
<td>84.7%</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>90.4%</td>
<td>26</td>
</tr>
</tbody>
</table>
Rotation

Rotation in the image plane

- Rotation invariant features
- Apply 2D rotation to image

Rotation out of image plane

- Component-based classification
- Train on rotated faces

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Global vs. Components

Single template

Component templates

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Component-based Detection

1st Level: Components

2nd Level: Geometrical relation between components

maximum response of each component classifier + x, y location

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Learning Components

Components:
• discriminatory
• robust against changes in pose and illumination

Synthetic faces:
• 7 different 3-D head models
• 2,500 faces
  Rotation: -30° to + 30°
• 3-D correspondences for automatic location of components

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Learning Components—One Way To Do It

1. Start with small initial regions
2. Expand into one of four directions
3. Extract new components from images
4. Train SVM classifiers
5. Choose best expansion according to error bound of SVMs

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Margin, Radius and Expected Error

Bound on error
\[ E < c(R / M)^2 \]

Feature Space

Cross Validation might be better
Some Examples

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Test on CMU PIE Database

Faces have been manually labeled (only $-45^\circ$ to $45^\circ$ of rotation)

- About 40,000 faces
- 68 people
- 13 poses
- 43 illumination conditions
- 4 different expressions

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
ROC Component vs. Global

Pedestrian Detection

- Representation: dictionary of Haar wavelets; high dimensional feature space (>1300 features)
- SVM classifier

Examples

Components

- Haar wavelets
- 5 components
- Can deal with partial occlusions

Components are small, and prone to false detection, even within the face.

Training on Faces

Use the remainder of the face in the negative training set

Positive

Negative

Training on Faces Only

Red: Trained only with faces.

Blue: Trained on faces and non-faces.


Courtesy of Stan Bileschi. Used with permission.
Errors

Often, many components classify correctly, with only a few errors.


Courtesy of Stan Bileschi. Used with permission.
Using Models of Pair-wise Positions


Courtesy of Stan Bileschi. Used with permission.
Pair-wise Biasing Leads to Tightened Result Images


Courtesy of Stan Bileschi. Used with permission.
Pair-wise Biasing

Using Greedy Optimization (Dotted Line)


 Courtesy of Stan Bileschi. Used with permission.
Application: Eye Detection


Courtesy of Stan Bileschi. Used with permission.
Task:

Given an image of an object of a particular class (e.g. face) identify which exemplar it is.

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Recognition—Problems

1. Multi-class problem

2. Classifier must distinguish between exemplars that might look very similar.

3. Classifier has to reject exemplars that were not in the training database.
System Architecture

- Training Data
- Feature extraction
- Classifier
- Classification Result
  - Support Vector Machine,
  - ….
  - Gray, Gradient,
  - Wavelets, …
- Face Image

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Multi-class Classification with SVM

Bottom-Up 1vs1

A or B or C or D

A or B

C or D

Training: \( L \cdot \frac{(L-1)}{2} \)
Classification: \( L-1 \)

1 vs. All

A / B,C,D

B / A,C,D

C / A,B,D

D / A,B,C

Training: \( L \cdot \frac{(L-1)}{2} \)
Classification: \( L \)
Global Approach

Detect and extract face

Feed gray values into $N$ SVMs

Classify based on maximum output

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Global Approach with Clustering

Partition training images of each person into viewpoint-specific clusters

Train a linear SVM on each cluster

Take maximum over all SVM outputs

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Component-based Approach

Detect and extract components

Feed gray values of components to $N$ SVMs

Take max. over all SVM outputs

Photographs courtesy of CMU/VASC Image Database at http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
Why Components for Recognition?

Images removed due to copyright considerations. See: Huang, Jennifer. "Face Recognition Using Component-Based SVM Classification and Morphable Models."
More ROC Curves

Morphable Models for Face Recognition, Jennifer Huang

Images removed due to copyright considerations.

Images removed due to copyright considerations.
Some Training Images

Preliminary Results on Synthetic Images

Current Work – Testing on Real Images

Problems Encountered:

Detection
Inaccurate Component detection

Recognition
Accuracy of 3D models
Choice of Illumination and Pose

Jennifer Huang

See also CBCL Web page