Towards a Video Annotation System using Face Recognition

Lucas Lindström

Umeå University

lucasl@cs.umu.se

January 14, 2014
Introduction: Background

- Applications of face recognition
  - Biometrics
  - Crime prevention
  - Web indexing
- Codemill AB
- Vidispine
Introduction: Goals

- Extract Vidispine face recognition plugin into standalone application.
- Improve and evaluate recognition speed and accuracy.
- Integrate face recognition and object tracking.
- Integrate frontal face recognition and profile recognition.
  - Not addressed.
Main problem: Identify or verify the identity of one or more individuals in a static image or video sequence (*probe*) by comparison to a known set of images or videos (*gallery*).

Three steps:
- Detection
- Normalization
- Identification

Most common model:
- Feature extraction
- Similarity measure
Theory: Studied approaches

- Detection
  - Cascade classification with Haar-like features

- Identification
  - Eigenfaces
  - Fisherfaces
  - Local binary pattern histograms
  - Wawo
Theory: Face recognition in video

- Multiple observations
- Temporal/continuity dynamics
- 3D model
Problem: Locate object(s) in video sequences, track their movement from frame to frame and/or analyze object tracks to recognize behavior

Studied approach: CAMSHIFT
Libraries

- OpenCV
  - Huge open source computer vision library.
- Wawo SDK
  - Small closed-source library.
  - Unpolished.
Algorithmic extensions: Face recognition/object tracking integration

- **Concept**
  - Frame-by-frame recognition in video disregards continuity.
  - Most recognition algorithms are view-dependent.
  - CAMSHIFT tracking is based on color probability histograms.
  - Tracking provides continuity and color probability histograms are view-insensitive.

- **Algorithm**
  1. Detect faces using arbitrary face detection algorithm in each frame of the video.
  2. Track faces across the video if the detected regions do not intersect with an existing track.
  3. In the search region of each track, in each frame, first apply face detection and then face recognition.
  4. When the entire video has been processed, compute the mode of recognized identities for each track, and assign it as the identity of the entire track.
Algorithmic extensions: Example

Identifications: 

Result: 

\begin{align*}
\text{a} & \quad \text{b} & \quad \text{c} & \quad \text{d} \\
1 & \quad & 1 & \quad \\
\end{align*}
Algorithmic extensions: Rotating face detection

- **Concept**
  - Most face detectors are pose-dependent.
  - If the input image is rotated about the depth axis, a wider range of poses can be detected.

- **Algorithm**
  - Rotate the input image away from the original orientation in a given number of steps, for a given angle step size.
  - For each orientation,
    - apply face detection to the rotated image.
    - If one or more faces are detected, rotate the image back to the original orientation and compute an axis-aligned bounding box (AABB) for each face.
  - Find each set of overlapping AABBs from the previous step.
  - Compute the mean rectangle of each set of AABBs from the previous step.
System description: Overview

- Gallery
- Video file

**Technique**
- Gallery detector
- Probe detector
- Normalizer
- Recognizer
- Annotation
System description: Detectors and recognizers

- **Detectors**
  - CascadeDetector
  - RotatingCascadeDetector
- **Recognizers**
  - EigenFaceRecognizer
  - FisherFaceRecognizer
  - LBPHRecognizer
  - WawoRecognizer
  - EnsembleRecognizer
System description: Normalizers and techniques

- Normalizers
  - GrayNormalizer
  - ResizeNormalizer
  - EqHistNormalizer
  - AggregateNormalizer

- Techniques
  - SimpleTechnique
  - TrackingTechnique
System description: Other modules

- Annotation
- Gallery
- Renderer
System description: Command-line interface

Performance evaluation: Datasets

- **NRC-IIT**
  - Single subject in each video, present for the whole duration, nearly always facing the camera.
  - Static background, minimal clutter.
  - Variety of structural features, such as beards, glasses, etc.
  - Subjects express a variety of facial expressions and turn their heads slightly.

- **News**
  - Video clips of news reports.
  - One or two subjects in each video, always facing straight into the camera, speaking with neutral expressions.
  - Dynamic background, changing to illustrate news stories, occasionally containing unknown faces.

- **NR**
  - Outtakes from the TV show The Newsroom.
  - Multiple subjects, multiple unknown individuals, facing in multiple directions and frequently changing pose.
  - Dynamic, highly cluttered background.
  - Variable illumination conditions.
Performance evaluation: Experimental setup

- Regular versus tracking recognizers
  - Purpose: Evaluate the performance of the tracking extension.
  - NRC-IIT dataset.
  - Subset accuracy over gallery size.
  - Real-time factor over gallery size.

- Regular detector versus rotating detector
  - Purpose: Evaluate the performance of the rotating detector.
  - NRC-IIT dataset.
  - Subset accuracy over gallery size.
  - Real-time factor over gallery size.

- Algorithm accuracy in cases of multiple variable conditions
  - Purpose: Evaluate the impact of the variability of face, scene and imaging conditions.
  - All datasets.
  - Various metrics for largest possible gallery size.
Performance evaluation: Regular versus tracking recognizers
Performance evaluation: Regular versus tracking recognizers
Performance evaluation: Regular detector versus rotating detector

![Graph showing subset accuracy vs. gallery size per subject for different methods: Wawo, Eigenfaces, Fisherfaces, LBPH, Ensemble, Wawo (rotating), Eigenfaces (rotating), Fisherfaces (rotating), LBPH (rotating), Ensemble (rotating).](image)
Performance evaluation: Regular detector versus rotating detector

![Graph showing the performance of different face detection methods. The x-axis represents the gallery size per subject, and the y-axis represents the real time factor. Different methods are color-coded, including Wawo, Eigenfaces, Fisherfaces, LBPH, Ensemble, Wawo (rotating), Eigenfaces (rotating), Fisherfaces (rotating), LBPH (rotating), and Ensemble (rotating).]
Performance evaluation: Algorithm performance in cases of multiple variable conditions

Table: NRC-IIT

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Hamming loss</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Subset accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenfaces</td>
<td>0.104112</td>
<td>0.531498</td>
<td>0.531498</td>
<td>0.531498</td>
<td>0.531498</td>
<td>0.531498</td>
</tr>
<tr>
<td>Fisherfaces</td>
<td>0.0875582</td>
<td>0.605988</td>
<td>0.605988</td>
<td>0.605988</td>
<td>0.605988</td>
<td>0.605988</td>
</tr>
<tr>
<td>LBPH</td>
<td>0.0975996</td>
<td>0.560802</td>
<td>0.560802</td>
<td>0.560802</td>
<td>0.560802</td>
<td>0.560802</td>
</tr>
<tr>
<td>Wawo</td>
<td>0.0908961</td>
<td>0.590968</td>
<td>0.590968</td>
<td>0.590968</td>
<td>0.590968</td>
<td>0.590968</td>
</tr>
<tr>
<td>Ensemble</td>
<td>0.0933599</td>
<td>0.57988</td>
<td>0.57988</td>
<td>0.57988</td>
<td>0.57988</td>
<td>0.57988</td>
</tr>
</tbody>
</table>

- Subset accuracy at 50-60%.
- Error mainly derived from pose variation, face distortion and/or occlusion.
- Issues almost entirely overcome by tracking as shown earlier.
Performance evaluation: Algorithm performance in cases of multiple variable conditions

Table: News

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Hamming loss</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Subset accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenfaces</td>
<td>0.261373</td>
<td>0.484974</td>
<td>0.484974</td>
<td>0.605459</td>
<td>0.524676</td>
<td>0.367246</td>
</tr>
<tr>
<td>Fisherfaces</td>
<td>0.34381</td>
<td>0.398677</td>
<td>0.398677</td>
<td>0.520265</td>
<td>0.438737</td>
<td>0.27957</td>
</tr>
<tr>
<td>LBPH</td>
<td>0.309898</td>
<td>0.444169</td>
<td>0.444169</td>
<td>0.622002</td>
<td>0.50284</td>
<td>0.269644</td>
</tr>
<tr>
<td>Wawo</td>
<td>0.351944</td>
<td>0.340433</td>
<td>0.340433</td>
<td>0.463193</td>
<td>0.381086</td>
<td>0.219189</td>
</tr>
<tr>
<td>Ensemble</td>
<td>0.368211</td>
<td>0.301213</td>
<td>0.301213</td>
<td>0.438379</td>
<td>0.34654</td>
<td>0.166253</td>
</tr>
</tbody>
</table>

- About the same fraction of true positives identified as for NRC-IIT.
- Larger number of false positives, due to dynamic, cluttered background.
  - Non-face elements classified as faces.
  - Unknown faces classified as known.
Performance evaluation: Algorithm performance in cases of multiple variable conditions

Table: NR

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Hamming loss</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Subset accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenfaces</td>
<td>0.288492</td>
<td>0.210648</td>
<td>0.210648</td>
<td>0.308333</td>
<td>0.24213</td>
<td>0.119444</td>
</tr>
<tr>
<td>Fisherfaces</td>
<td>0.21746</td>
<td>0.340046</td>
<td>0.340046</td>
<td>0.55</td>
<td>0.406111</td>
<td>0.152778</td>
</tr>
<tr>
<td>LBPH</td>
<td>0.263492</td>
<td>0.244444</td>
<td>0.244444</td>
<td>0.388889</td>
<td>0.291667</td>
<td>0.105556</td>
</tr>
<tr>
<td>Wawo</td>
<td>0.194444</td>
<td>0.389583</td>
<td>0.389583</td>
<td>0.625</td>
<td>0.465463</td>
<td>0.169444</td>
</tr>
<tr>
<td>Ensemble</td>
<td>0.21746</td>
<td>0.343981</td>
<td>0.343981</td>
<td>0.55</td>
<td>0.40963</td>
<td>0.155556</td>
</tr>
</tbody>
</table>

- Wawo and Fisherfaces performed on par with the News test.
- Eigenfaces and LBPH performed significantly worse.
- All methods identified large numbers of false positives.
  - Non-face elements and unknown individuals.
  - To a greater extent than for the News test, known individuals identified as other known individuals.
    - Probably due to lower-quality training data and greater variability in pose and illumination.
Conclusion: Summary

- Wawo generally performs best, but processing time scales linearly with gallery size.
- Eigenfaces outperforms Wawo for small gallery sizes.
- Fisherfaces almost on par with Wawo for large gallery sizes.
  - Processing time doesn't scale with gallery size.
- Face recognition/CAMSHIFT integration able to improve accuracy by approximately 40 percentage points with small processing time sacrifice.
- Rotating cascade detector provides minor accuracy improvement at relatively great processing time increase.
  - Processing time scales linearly with number of orientations tested.
  - May be possible to find special cases where few additional orientations are required.
Conclusion: Limitations

- Frontal/profile integration not attempted due to lack of profile face data available.
- Results mainly acquired from the NRC-IIT dataset, which has limited variability in terms of face and image conditions.
- Lack of good test data affects field as a whole.
Conclusion: Future work

- Restrict application area.
- Gather more data.
- Add more algorithms.
- Distinguish between known/unknown subjects.
- Study normalization techniques.