3D Face Recognition

Anil K. Jain

Dept. of Computer Science & Engineering
Michigan State University

http://biometrics.cse.msu.edu
Face Recognition

Face detection using OpenCV Viola-Jones algorithm
Face Detection in Video

- Cluttered background
- Frontal face is not always visible
Surveillance Video
Why Face?

- Most common biometric used by humans
- Allows identification at a distance
- Easy to capture from low-cost cameras
- Non-contact data acquisition
- Covert data acquisition (surveillance cameras)
- Legacy database (passport, driver license)

http://www.infotech.oulu.fi/Annual/2004
Applications

• Access control
  – Entrance door

• Surveillance
  – Inspecting public areas

• Civil applications
  – E-passports
  – Mobile phones

• Forensic application
  – Criminal/victim identification
  – Missing children

www.3g.co.uk/PR/March2005/1109.htm
Challenges

- Automatically locate the face
- Discriminate “similar” faces (inter-class similarity)
- Accommodate intra-class variability due to:
  - head pose
  - illumination condition
  - expression
  - facial accessories
  - aging effect
- Face recognition in video & at a distance
Face Detection

Intra-class Variability

- Faces with intra-subject variations in pose, illumination, expression, accessories, color, occlusions, and brightness
Inter-class Similarity

- Different persons often have very similar appearance

Twins

Father and son
Representation

- Humans recognize caricatures and cartoons
- How do we learn salient facial features?
- Discriminative vs. descriptive approaches
State-of-the-Art (FRVT 2002)

- Still, frontal & neutral images
- Eight vendors
- Best rank-1 performance = 73%

121,589 images from 37,437 subjects (US visa database)
“Left/right” and “up/down” show identification rates for the non-frontal images on a database of 87 individuals.
FRGC 2.0 (2005 – 2006)

Goal: Improve verification rate from 80% (FRVT) to 98% at FAR=0.1%; order of magnitude improvement

Exp 1: Controlled indoor still versus indoor still

Exp 2: Multiple still versus multiple still

Exp 3: 3d versus 3D
   3t - Texture only
   3s - Shape only

Exp 4: Uncontrolled still versus indoor still

• 16,000 Controlled and 8,000 Uncontrolled Stills (~400 Subjects)
• Mostly frontal and neutral images
Large Scale Deployment

• At NY airports, ~300,000 passengers/day
• Suppose a face recognition system operating at 2% FRR & 0.1% FAR is deployed
• 6,000 genuine passengers may be subjected to secondary search
Sensing

- Passive vs. Active
- Static (single frame) vs. Video
- Visible vs. Non-visible spectrum
- Unimodal vs. Multimodal
Change in Ambient Lighting

2D Color Images under intensity variations

Registered Range images under intensity variations

Lighting and Pose are effectively handled by range (surface or depth) images
Face Recognition Scenarios

<table>
<thead>
<tr>
<th>Template</th>
<th>Query/Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D still/sequence</td>
<td>2D still/sequence</td>
</tr>
<tr>
<td>3D/2.5D</td>
<td>3D/2.5D</td>
</tr>
</tbody>
</table>

Using 3D models to retrieve 2D images!
3D Face Modeling

- SfM (Structure from Motion)
- Stereography
- Direct Range sensing
Structure from Motion

- 72 landmarks detected and tracked using Active Shape Model (ASM) on a Video with 60 frames

Landmark detection **without** temporal coherency

Landmark detection **with** temporal coherency (estimated feature points at current frame are used as the initial state for the next frame)
Reconstruction with SfM

- Reconstruction based on 72 landmarks in 60 frames

Dense reconstruction with texture mapping
Direct 3D range sensing
3D Face Image Acquisition

Minolta Vivid 910 scanner
Acquisition time ~ 1sec.
320x240 image size
Depth resolution < 0.1 mm
Match multiview 2.5D non-neutral scans to 3D models or 2.5D frontal scans in gallery
3D Model Construction

- Geomagic Studio software used to stitch 5 scans
- Remove noise and fill holes
- Decimate the model to 50,000 polygons
3D Face Recognition (Version 1)

Utilize both 3D surface geometry and appearance

Automatic Feature Extraction

Feature Location Model
Feature Extraction Results

Results on the multiview MSU database

Results on the UND database
Iterative Closest Point Algorithm

Given: \( N_p \) and \( N_m \) points from the test scan \( P \) & template \( M \)

\( T_0 = \) Coarse Alignment Transformation

\( P_0 = T_0(P) \) and \( k = 0 \)

1. Compute the closest points:
   \( M_k = C(P_k, M) \)

2. Compute the transformation \( T_k \) and error \( e_k \):
   \( (T_k, e_k) = Q(P, M_k) \)

1. Apply the transformation:
   \( P_{k+1} = T_k(P) \)

2. Iterate Until:
   \( |e_k - e_{k+1}| < t \) or \( k > N_I \)

Besl and McKay, PAMI, 1992; Chen and Medioni, IVC, 1992
Control Point Selection

- Cover the entire scan
- “Invariant” w.r.t. expression
Surface Matching

Query scan

Matching with true model

Matching with impostor model

matching distance = 8.0444

matching distance = 2.659
LDA with Synthesized Images

(a) Intensity image of test scan
(b) Intensity image of the aligned 3D model
(c-f) synthesized images from the model (b) with shift displacement
(g-j) synthesized images with lighting changes
Matching Score

- Shape matching distance

\[ MD_{ICP} = \sqrt{\frac{1}{N_c} \sum_{i=1}^{N_c} d^2(\psi(p_i), S_i)} \]

- Appearance-based matching

\[ MS_{LDA} = \frac{<v_1^*, v_2^*>}{\|v_1^*\| \cdot \|v_2^*\|} \]

- Fusion

\[ MD_{comb} = MD_{ICP} + \alpha \cdot MD_{LDA} \]
Database

3D model examples: MSU database (left), USF database (right)

Query scan examples

- Multi-view facial scans
- Expression changes, e.g., smiling
- 598 2.5D independent query scans
- Gallery: 200 3D models (100 each from MSU & USF)
### Matching Results

598 2.5D independent test scans and 200 3D models

<table>
<thead>
<tr>
<th>Test scan category</th>
<th>Surface matching + LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal &amp; Neutral (99)</td>
<td>99% (1)</td>
</tr>
<tr>
<td>Profile &amp; Neutral (213)</td>
<td>98% (5)</td>
</tr>
<tr>
<td>Frontal &amp; Smiling (98)</td>
<td>77% (23)</td>
</tr>
<tr>
<td>Profile &amp; Smiling (188)</td>
<td>84% (31)</td>
</tr>
</tbody>
</table>

Almost all the matching errors are due to expression changes.

Number of scans in each category and the number of misclassified test scans are shown in the parentheses.
Grins Banned From Passport Pics

• Travelers have been ordered not to look too happy in their passport photographs to avoid confusing facial recognition scanners.

• A Home Office spokesman said: "When the mouth is open it can make it difficult for facial recognition technology to work effectively."

BBC News, 6 August, 2004, 13:05 GMT 14:05 UK
Expression Modeling

- Collecting and storing multiple templates for each subject with different expressions is not practical
  - Synthesis
- The same type of expression can appear with different intensity
  - Deformable model

Same expression with different degree
Deformation Modeling

Landmark Extraction

- Identify 20 landmarks with significant surface or intensity changes
- Obtain additional landmarks in “smooth” regions by interpolating between landmarks
- Geodesic path (set of points along the shortest path along the surface connecting two vertices) is “invariant” to expression
Geodesic-based Landmark Extraction
Hierarchical Surface Resampling

In total, 94 landmarks are obtained
Deformation Transfer & Synthesis

Gallery neutral model

Control group

Synthesized model
Deformable Model

$S = S_{ne} + \sum_{i=1}^{M} \alpha_i (S_i - S_{ne})$

$M$ is the number of synthesized models, $\alpha_i$'s are the weights, $S_{ne}$ denotes the neutral model and $S_i$ represents the synthesized model.
Deformable Model

Examples of the deformable model with varying weights ($\alpha_i$)
Deformable Model Fitting

Minimize the cost function

\[ E = \left\| S - \psi(S_t, (R | T)) \right\|^2 \]

\[ = \left\| S_{ne} + \sum_{i=1}^{M} \alpha_i \cdot (S_i - S_{ne}) - \psi(S_t, (R | T)) \right\|^2 \]

\( S \): deformable model

\( S_t \): test scan

\((R | T)\): rotation matrix and translation vector

\( S_{ne} \): neutral model

\( S_i \): model with synthesized expressions
Deformable Model Fitting

(a) Query scan
(b) 3D neutral model
(c) Deformed model after fitting to (a)

Match dist.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>2.7</td>
</tr>
<tr>
<td>b</td>
<td>1.3</td>
</tr>
</tbody>
</table>
Control Group

10 subjects, 7 expressions (neutral, happy, angry, smile, surprise, deflated, inflated)
Frontal View & Non-Neutral

- A subset of the FRGC Version 2.0 database
- 100 subjects in the gallery database (100 neutral scans)
- 877 independent query scans with multiple expressions
- Each query is matched to all the gallery templates)
- Fully automatic

Query scan examples
CMC Curves (FRGC)
Matching Examples in FRGC

(a) Test scans, (b) Genuine neutral templates, (c) Genuine deformable templates after successful match
Matching Errors in FRGC

(a) Test scans
(b) Mis-matched deformable templates after convergence

Reasons for errors: (i) Local minima, (ii) Small number of subjects and expressions in the control group
Multiview & Non-Neutral

- 90 subjects in the gallery database (90 3D neutral models)
- 533 independent query scans
  - Two expressions, namely, neutral and smiling
  - Large head pose changes (45+ degrees from frontal)

Query scan examples
CMC Curves (MSU)
ROC Curves (MSU)
Summary

- Face recognition continues to be a challenging problem with many applications
- We have described an automatic system
  - that integrates range and intensity modalities
  - automatically extracts feature points across large head pose variations
  - utilizes deformation synthesis and modeling scheme to accommodate facial expression
- Future work: Size of control group, automatic landmark labeling, efficient matching
Future Directions

• Robust and efficient feature extraction, especially in the presence of large pose and expression variations

• Automatic landmark labeling. Although this is conducted in the offline training stage, it is desirable to make it a fully automatic process

• More efficient matching algorithms to achieve real-time performance