Face modeling (part I)
Jun-Yan Zhu
16-726 Learning-based Image Synthesis, Spring 2022
Why Human Faces?

- Face is an important subject.
  - We are humans.
  - Many commercial applications.

- Lots of useful tools
  - 3D data: geometry-based synthesis.
  - 2D/3D Computer vision works for faces.
Image Composites

Sir Francis Galton 1822-1911

The Power of Averaging
8-hour exposure

© Atta Kim
Average Images in Art

“60 passagers de 2e classe du metro, entre 9h et 11h” (1985)
Krzysztof Pruszkowski

“Spherical type gasholders” (2004)
Idris Khan
### “100 Special Moments” by Jason Salavon

<table>
<thead>
<tr>
<th>Little Leaguer</th>
<th>Kids with Santa</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Graduate</td>
<td>Newlyweds</td>
</tr>
</tbody>
</table>

**Why blurry?**
Object-Centric Averages by Torralba (2001)

Manual Annotation and Alignment

Average Image
Computing Means

Two Requirements:
• Alignment of objects
• Objects must span a subspace

Useful concepts:
• Subpopulation means
• Deviations from the mean
Images as Vectors

\[ n \times m = n^*m \]
Vector Mean: Importance of Alignment

\[ \frac{1}{2} n \] + \[ \frac{1}{2} m \] = mean image
How to align faces?

Students and staff from Technical University of Denmark
http://www2.imm.dtu.dk/~aam/datasets/datasets.html
Shape Vector

Landmark annotation

43
Appearance Vectors vs. Shape Vectors

**Appearance Vector**
- 200x150 pixels (RGB)
- Vector of 200x150x3 Dimensions
- Face landmark detection.

**Shape Vector**
- 43 coordinates (x,y)
- Vector of 43x2 Dimensions

Slide by Kevin Karsch
Average Face

1. Warp to mean shape
2. Average pixels

Students and staff from Technical University of Denmark
Objects must span a subspace
Subpopulation means

Examples:

- Male vs. female
- Happy vs. said
- Average Kids
- Happy Males
- Etc.
- http://www.faceresearch.org

Average female
Average kid
Average happy male
Average male
Average Women of the world

Several issues: 1. country ≠ race. 2. demographic diversity is lost. 3. bias in data source
Several issues: 1. country ≠ race. 2. demographic diversity is lost. 3. bias in data source
Deviations from the mean

\[ \Delta X = X - \bar{X} \]
Deviations from the mean

\[ \Delta X = X - \bar{X} \]

\[ X = 3 + 1.7 \]
Extrapolating faces

- We can imagine various meaningful directions.
Manipulating faces

- How can we make a face look younger/older, or happy/sad, etc.?
- [http://www.faceresearch.org/demos/transform](http://www.faceresearch.org/demos/transform)
Back to the Subspace
Linear Subspace: convex combinations

Any new image $X$ can be obtained as weighted sum of stored “basis” images.

$$X = \sum_{i=1}^{m} a_i X_i$$

Our old friend, change of basis! What are the new coordinates of $X$?
The Morphable Face Model

The actual structure of a face is captured in the shape vector $\mathbf{S} = (x_1, y_1, x_2, \ldots, y_n)^T$, containing the $(x, y)$ coordinates of the $n$ vertices of a face, and the appearance (texture) vector $\mathbf{T} = (R_1, G_1, B_1, R_2, \ldots, G_n, B_n)^T$, containing the color values of the mean-warped face image.
The Morphable face model

Again, assuming that we have $m$ such vector pairs in full correspondence, we can form new shapes $S_{model}$ and new appearances $T_{model}$ as:

$$S_{model} = \sum_{i=1}^{m} a_i S_i$$

$$T_{model} = \sum_{i=1}^{m} b_i T_i$$

If number of basis faces $m$ is large enough to span the face subspace then:

Any new face can be represented as a pair of vectors

$$(\alpha_1, \alpha_2, \ldots, \alpha_m)^T \text{ and } (\beta_1, \beta_2, \ldots, \beta_m)^T$$
Issues:

1. How many basis images is enough?
2. Which ones should they be?
3. What if some variations are more important than others?
   - E.g. corners of mouth carry much more information than haircut

Need a way to obtain basis images automatically, in order of importance!

But what’s important?
Principal Component Analysis

Given a point set \( \{ \vec{P}_j \}_{j=1}^P \), in an \( M \)-dim space, PCA finds a basis such that

- coefficients of the point set in that basis are uncorrelated
- first \( r < M \) basis vectors provide an approximate basis that minimizes the mean-squared-error (MSE) in the approximation (over all bases with dimension \( r \))
PCA via Singular Value Decomposition

\[ [u, s, v] = \text{svd}(A); \]
EigenFaces

First popular use of PCA on images was for modeling and recognition of faces [Kirby and Sirovich, 1990, Turk and Pentland, 1991]

- Collect a face ensemble
- Normalize for contrast, scale, & orientation.
- Remove backgrounds
- Apply PCA & choose the first $N$ eigen-images that account for most of the variance of the data.
First 3 Shape Basis

Mean appearance

Principal Component Analysis

Choosing subspace dimension $r$:

- look at decay of the eigenvalues as a function of $r$
- Larger $r$ means lower expected error in the subspace data approximation
Using 3D Geometry: Blinz & Vetter, 1999
Using 3D Geometry: Blinz & Vetter, 1999
Using 3D Geometry: Blinz & Vetter, 1999
Face + Internet Images
About 409,000,000 results (0.49 seconds)
Photobio
Challenges

Non-rigid (facial expressions, age...)

Occlusions (hair, glasses ...)

Arbitrary lighting, pose

Different cameras, exposure, focus ...

But: there are many photos!

http://vimeo.com/23561002
Image registration

Face detection
Bourdev and Brandt ‘05

Fiducial points detection
Everingham et al. ‘06

2D registration

Template 3D model

Estimate 3D pose

Kemelmacher, Shechtman, Garg, Seitz, Exploring Photobios, SIGGRAPH’11
Image registration

Face detection
Bourdev and Brandt ‘05

Fiducial points detection
Everingham et al. ‘06

3D registration

Estimate 3D pose
Template 3D model

Kemelmacher, Shechtman, Garg, Seitz, Exploring Photobios, SIGGRAPH’11
3D transformed photos

before

after
Represent the photo collection as a graph

Similarity between 2 photos

= 3D Head Pose similarity

Facial Expression similarity

Time similarity
Represent the photo collection as a graph.

- Similarity between 2 photos
- 3D Head Pose similarity
- Facial Expression similarity
- Time similarity
Represent the photo collection as a graph

Similarity between 2 photos

3D Head Pose similarity

Facial Expression similarity

Time similarity
Transfiguring Portraits

Ira Kemelmacher-Shlizerman*
Computer Science and Engineering, University of Washington

Figure 2: Illustration of our system. The system gets as input a photo and a text query. The text query is used to search a web image engine. The retrieved photos are processed to compute a variety of face features and skin and hair masks, and ranked based on how well they match to the input photo. Finally, the input face is blended into the highest ranked candidates.

https://www.youtube.com/watch?v=mILLFK1Rwhk
Me with "curly hair"
Illumination-aware Age Progression

CVPR 2014
Ira Kemelmacher-Shlizerman, Supasorn Suwajanakorn, Steven M. Seitz
Illumination-aware Age Progression
Image-Based Shaving

http://graphics.cs.cmu.edu/projects/imageshaving/
The idea

Differences

???

Beard Layer Model

[Images of different faces]
Processing steps

68 landmarks
Some results
Take-home Message

- Alignment (2D and 3D): 3D is better than 2D.
- Shape + Texture representation.
- Subpopulation mean $\bar{x}$ and deviation $\Delta x$
- 3D data and 3D shape representation helps!
  - Easy to change the viewpoint.
- Standard face pipeline:
  Given: Input Image
  Step 1: warp it to canonical pose (2D or 3D)
  Step 2: Calculate distances between faces OR apply image manipulation operations.
  Step 3: Unwarp the result back to the original image
  Step 4: Post-processing (e.g., Poisson blending)
Thank You!

16-726, Spring 2022
https://learning-image-synthesis.github.io/sp22/

Video © Kemelmacher-Shlizerman et al., SIGGRAPH 2011