3D Morphable Models

Idea and basic model

- Geometry of face is represented with shape vector S = $(X_1,Y_1,Z_1,X_2,Y_2,Z_2,\ldots,X_n,Y_n,Z_n)^T$ in R^{3n} which contains the X,Y,Z coordinates of its n vertices.
- The texture of the face is represented by the texture vector T = $(R_1,G_1,B_1,R_2,G_2,B_2,\ldots,R_n,G_n,B_n)^T$ in R^{3n} that contains the R,G,B values of n corresponding vertices.
- 3D morphable model is then constructed using data point of m faces. S_1, S_2, …, S_m and T_1, T_2, …, T_m.
- The main idea is that a new face (S_{mod}, T_{mod}) can be generated as a linear combination of some example face.

$$S_{mod} = \sum_{i=1}^{m} a_i S_i, T_{mod} = \sum_{i=1}^{m} b_i T_i, \sum_{i=1}^{m} a_i = 1, \sum_{i=1}^{m} b_i = 1$$

- 3D morphable model is defined as set of faces (S_{mod}(a), T_{mod}(b)) with parameters a = $(a_1,a_2,\ldots a_m)$ and b = $(b_1,b_2,..b_m)$. We can generate arbitrary faces by varying the parameters a and b that controls shape and texture.
- All possible combination of a and b may not result into face so it is important to be able to quantify the results in terms of their possibility of being faces.
- The probability distribution for the coefficients a_i and b_i are calculated form example faces. This distribution gives us an estimate the probability of appearance of generated face.
- A normal distribution is fit to the dataset of faces based on the averages of shape S and texture T and the covariance matrices C_S and C_T computed over the differences $\Delta \mathrm{S_i} = \mathrm{S_i} - \overline{\mathrm{S}} \mathrm{and} \Delta \mathrm{T_i} = \mathrm{T_i} - \overline{\mathrm{T}}$.
- We then use Principal Component Analysis (PCA) to the covariance matrices to get the eigenvalues and eigenvectors.

$$S_{model} = \overline{S} + \sum_{i=1}^{m-1} \alpha_i s_i, T_{model} = \overline{T} + \sum_{i=1}^{m-1} \beta_i t_i$$

where $p(\overrightarrow{\alpha}) \sim exp[-\frac{1}{2}\sum_{i=1}^{m-1}(\frac{\alpha_i}{\sigma_i})^2]$ and σ_i^2 is eigenvalues of covariance matrix C_s. Probability for β_i is computed similarly.

Using PCA for dimensionality reduction technique to get the fundamental dimensions of faces and construct a 3D morphable face model capable of generating possible faces is the fundamental idea of 3DMM.

Dense registration

This component is responsible for making the learned component meaningful. This can be thought of as parametrization where corresponding point represent same anatomical reference across all the scans. For this all the scans need to have same number of vertices, failure to these will lead to PCA making no sense.

Optical flow for dense correspondence

A reference face is choosen and modification of gradient based optical flow algorithm (please cite appropriate article if someone find it) is used to establish correspondence between each scan and reference face.

Each scan is represented by its RGB texture values and 3D coordinates, and based on these correspondence each mesh is re-parametrized with respect to the reference face so that corresponding points have the same index in the vector.

Matching 3DMM to Images

- 3DMM is used either to generate arbitrary faces or to reconstruct a face image. To reconstruct a face given an image the coefficients of the 3DMM are optimized along with a set of rendering parameters such that they produce an image as close as possible to the input
- In an analysis-by-synthesis loop, the algorithm creates a texture mapped 3D face from the current model parameters, renders an image, and updates the parameters according to the residual difference
- 3DMM matching requires both 3DMM parameters and rendering parameters. Rendering parameters contain the camera position, object scale, image plane rotation and translation, ambient light, directed light, color contrast, camera distance, light direction etc.
- The 3DMM parameters are restricted to the vector space spanned by the training data, thus non-face-like surfaces are avoided. Optimizing with respect to all those parameters is complex and costly

Limitations of standard approach

3DMM haven’t convincingly applied to face analysis applications where facial expressions are involved. Difficulty with coping with noise, local deformations and topology variations.

What can be done to tackle the issues?

- The techniques presented so far constitute the original 3DMM as first presented
- There is a 4 step process in the standard approach pipeline i.e
    - Training scans collections
    - Dense registration
    - Components learning
    - Matching to images

and we can modify each step of the pipeline.

References

- https://www-inst.eecs.berkeley.edu/~cs194-26/fa16/Papers/BlanzVetter99.pdf
- https://ibug.doc.ic.ac.uk/media/uploads/documents/0002.pdf
- http://www.cs.umd.edu/~djacobs/CMSC426/OpticalFlow.pdf