Project Proposal

Face Reconstruction in Fisheye Images/Videos

1 Abstract

3D face reconstruction has been a long-standing task in the computer vision and computer graphics community. 3D Morphable Models (3DMM) [1] have played a significant role in 3D face modelling. In 3DMMs, face attributes are represented by $3m$ dimensional vectors where m is the number of vertices in a triangular mesh. The most commonly used attributes are shape and texture. A parametric face model is a multi-linear PCA model. New faces are generated from the model as linear combinations of the principal components

$$\boldsymbol{s}(\boldsymbol{a}) = \boldsymbol{u}_s + \boldsymbol{U}_s diag(\boldsymbol{\sigma_s})\boldsymbol{a} \qquad \boldsymbol{t}(\boldsymbol{b}) = \boldsymbol{u}_t + \boldsymbol{U}_t diag(\boldsymbol{\sigma_t})\boldsymbol{b} \tag{1}$$

where $\boldsymbol{u}_{s,t} \in \mathbb{R}^{3m}$ are the mean, $\boldsymbol{\sigma}_{s,t} \in \mathbb{R}^{n-1}$ are the standard deviations and $\boldsymbol{U}_{s,t} = [\boldsymbol{u_1}, \boldsymbol{u_2}...\boldsymbol{u_m}] \in \mathbb{R}^{3m \times n-1}$ are an orthonormal basis of principle components of shape and texture components.

3DMMs have also been widely used for image-based reconstruction. Reconstructing a 3D face from an observed image(s) involves optimization of the 3DMM coefficients along with a set of rendering parameters in an analysis-by-synthesis loop that can best explain the observation. The synthesis is dependent on face model parameters $\boldsymbol{a}$, $\boldsymbol{b}$, the illumination parameters $\boldsymbol{l}$, the rigid transformation $\boldsymbol{R}$, $\boldsymbol{t}$, the camera parameters $\boldsymbol{K}$ [4]. Given a video sequence, we can jointly optimize all the parameters (Θ) using the Levenberg-Marquardt algorithm. We formulate the optimization problem as an energy minimization problem

$$E(\Theta) = w_{col}E_{col}(\Theta) + w_{reg}E_{reg}(\Theta), \tag{2}$$

where $E_{col}(\Theta)$ is the photo consistency error between the synthesized image and the input image, $E_{reg}(\Theta)$ is a prior of parameters and w_{col}, w_{reg} are the weights.

The image formation models to synthesize the 3DMMs are much simpler than those used in graphics and many other areas of computer vision. To the best of our knowledge, we are not aware of any work that accounts for non-linear distortion in the image formation process. Moreover, the effect of this assumption is not understood. However, in other fields like visual SLAM and SfM, it is standard to consider non-linear distortion in the image formation process. To this end, we want to investigate the 3DMMs-based reconstruction of faces captured using the wide-angle lens. Figure 1 shows the overview of our method.

2 Requirements

- Basel face model [3] (3DMM model)

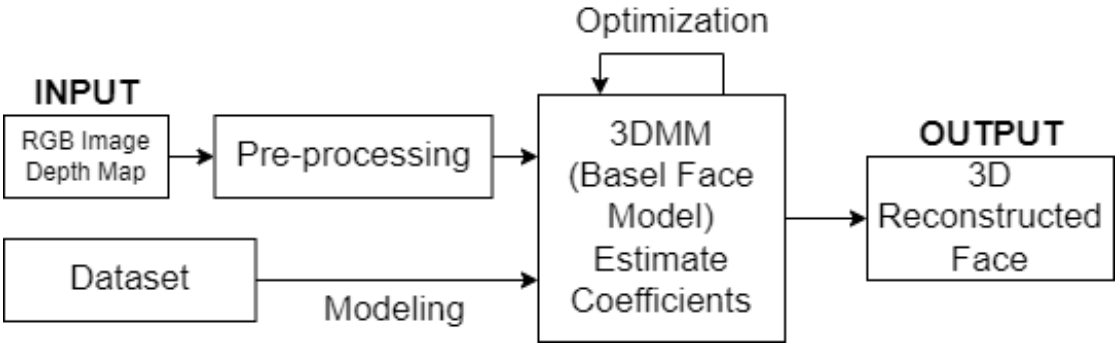

Figure 1: Method overview.

- FDDB-360 [2] (Face Detection Data Set and Benchmark (FDDB) that have been processed to look like fisheye images coming from a typical 360-degree camera.)
- C++ (Pipeline)
- Ceres (Optimization)
- OpenCV (Rendering)

3 Team

- Ananta Bhattarai - 03742706
- Magdy Mahmoud - 03763991
- Maqarios Saleh - 03758594

4 Milestones

- Week 1: Acquire Dataset, Add Github Repository, Implement C++ Project Skeleton.
- Week 2: Acquire Basel Face Model, C++ Conversion and Ceres Residual Block.
- Week 3: Benchmarking and Review Papers related to Fisheye.
- Week 4: Estimate 3DMM Coefficients for Fisheye Scans.
- Week 5: Optimisation and Finalizing Final Report.
- Week 6: (Optional) Implement real-time Application.

References

[1] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, page 187194, USA, 1999. ACM Press/Addison-Wesley Publishing Co.

[2] Jianglin Fu, Saeed Ranjbar Alvar, Ivan V. Bajic, and Rodney G. Vaughan. Fddb-360: Face detection in 360-degree fisheye images, 2019.

[3] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. A 3d face model for pose and illumination invariant face recognition. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 296–301, 2009.

[4] Justus Thies, Michael Zollhöfer, Matthias Nießner, Levi Valgaerts, Marc Stamminger, and Christian Theobalt. Real-time expression transfer for facial reenactment. 34(6), 2015.