Human to 3D Avatar Retargeting (Face Only; Facial Expressions and Deformations)

Anusha Bongale (01FE21BCS153), Bhimashankar Malgi(01FE21BCS178), Karthik Jodangi (01FE21BCS354)

Under the guidance of
Uma Mudenagudi, Ramesh Ashok Tabib

KLE Technological University, Vidyanagar, Hubballi-580031, Karnataka, India

April 15, 2024

Overview

- Introduction: Human to 3D Avatar retargetting(Face Only)
- Motivation: : Human to 3D Avatar retargetting(Face Only)
- Literature Survey: Human to 3D Avatar retargetting(Face Only)
- Problem Statement and Objectives
- Design Alternative: Delaunay Triangulation
- Methodology
  - Final Architecture: Block Diagram
  - Mediapipe Canonical Face Model
  - Basel Face Model (BFM)
- Observations

Introduction: Human to 3D Avatar retargeting(Face Only)

- A 3D Avatar is a computer-generated virtual character that represents a real person in a virtual world.
- Transferring facial expressions from one subject to another is a long-standing problem in computer animation, which is also known as expression cloning and retargeting.

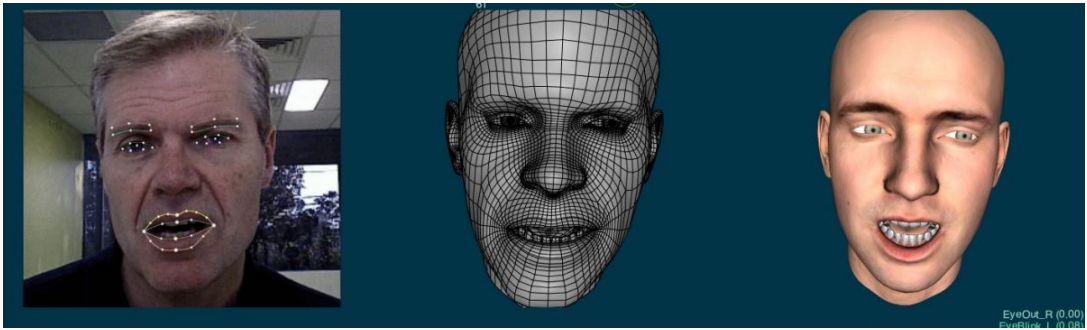


Figure: 1. Human face retargeting to 3D Avatar

Motivation: Human to 3D Avatar retargeting(Face Only)

- 3D avatars are beneficial for visual storytelling.
- It enhances user experience in video games by extensive avatar expressions.
- It boosts social VR experience in virtual universe.
- Helpful in enriching the field of Facial Puppetry.

Figure: 2. Real-time facial remapping in movies

Figure: 3. Virtual avatar in virtual universe

Literature Survey

"Learning to generate 3D stylized character expressions from humans." (Aneja, Deepali, et al. WACV 2018)

Takes images of human faces and generates the character rig parameters that best match the human's facial expression.Also generalizes to multiple characters

- **Dataset Used:**
  - Human Expression Database (HED): (a) SFEW (b) CK+(Extended) (c) MMI database (d) KDEF (e) DISFA
  - Character Expression Database(CED): FERG-DB + 3 additional characters for validation
- **Architecture Used:** 3D-CNN,C-MLP

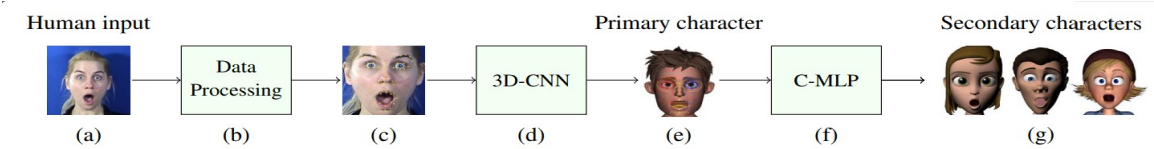


Figure: Multi-stage expression transfer system ExprGen

Literature Survey

"Synergy between 3dmm and 3d landmarks for accurate 3d facial geometry"
(Wu, Cho-Ying, et al. 3DV. 2023)

Overview: Synergy between 3dmm and 3d landmarks for accurate 3d facial geometry
Architecture: 3DMM, 3D Facial Landmarks
Limitations:
1. Inaccurate results for low-resolution, blurry inputs.
2. Since ICP(iterative closet point) is not used, pose estimation would affect performance.

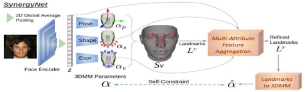


Figure: Architecture Framework of SynergyNet

"EMOCA: Emotion driven monocular face capture and animation."
(Daněček, Radek, Michael J. Black,et al. CVPR 2022)

Overview: Introduces a deep perceptual emotion consistency loss during training, ensuring the reconstructed 3D expression the input image.
Architecture: DECA
Limitations:
1. Emotion consistency loss difficult to optimize.
2. DECA sometimes predicts 3D faces and expressions that are slightly misaligned with the input due to fixed coarse shape encoder.

Figure: EMOCA regresses 3D faces from images with facial geometry that captures the original emotional content

"Collaborative regression of expressive bodies using moderation."
(Feng, Yao, et al. 3DV 2021)

Overview: Introduces PIXIE, which produces animatable, whole-body 3D avatars with realistic facial detail, from a single image.
Architecture: SMPL-X
Limitations:
1. Mesh-to-image misalignment of the image, losing local information.
2. Due to photometric term the model prefers to explain image evidence using lighting, rather than albedo, leading to wrong skin tone predictions..

Figure: Estimation of fine facial details by PIXIE

[0]

Literature Survey

"Learning to generate 3D stylized character expressions from humans." (Aneja, Deepali, et al. WACV 2018)

Overview: Generates character rig parameters that best match the human's facial expression. Also generalizes to multiple characters
Architecture: 3D-CNN,C-MLP
Limitations:
1. When the new secondary character expression is perceptually or geometrically ambiguous, model tries to find the closest match based within the wrong expression classes leading to wrong training

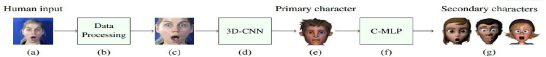


Figure: Multi-stage expression transfer system ExprGen

"3d face reconstruction with dense landmarks." (Wood, Erroll, et al. European Conference on Computer Vision 2022.)

Overview: Accurately predicts facial deformations with 10 times as many landmarks as usual, covering eyes and teeth, using synthetic training data
Architecture: CNN, Minimization of Energy function
Limitations:
1. Unavailability of synthetic data.
2.If landmarks are poorly predicted, the resulting model fit suffers.
3.Tongue movements cannot be recovered due to not inclusion of

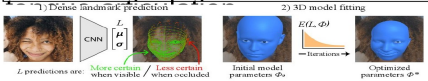


Figure: Architecture framework of dense landmark reconstruction

Problem Statement and Objectives

Problem Statement

Developing an efficient system for Human to 3D Avatar Retargetting, specifically focusing on facial expressions and deformations.

Objectives

- Collection of facial input data.
- Detection and extraction of facial landmarks from the input
- Fitting a 3D face model using the extracted landmarks.
- Mapping texture to a 3D facial model to create a 3D avatar.

Design Alternative- Flowchart

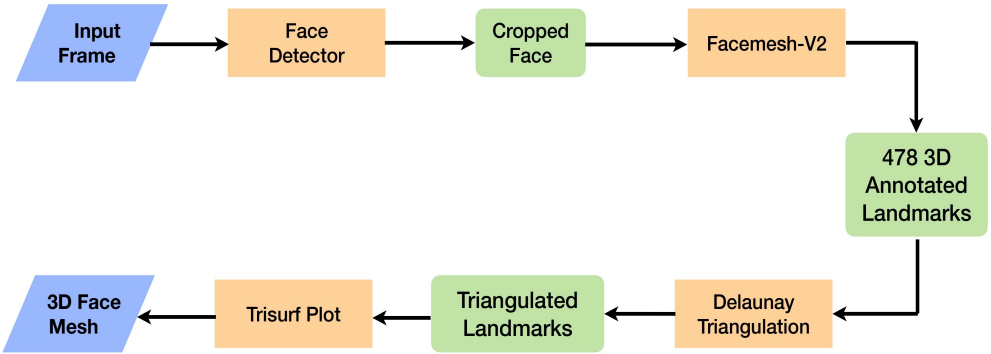
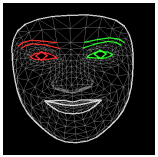


Figure: Flowchart for Face Mesh using Delaunay's Triangulation

Pipeline for Delaunay's Triangulated Mesh

Face Detection and cropping

Face keypoint extraction

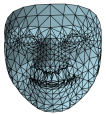


Input-frame (Image/Video/Webcam)

Cropped (ROI)

Landmark Annotated Face Mesh

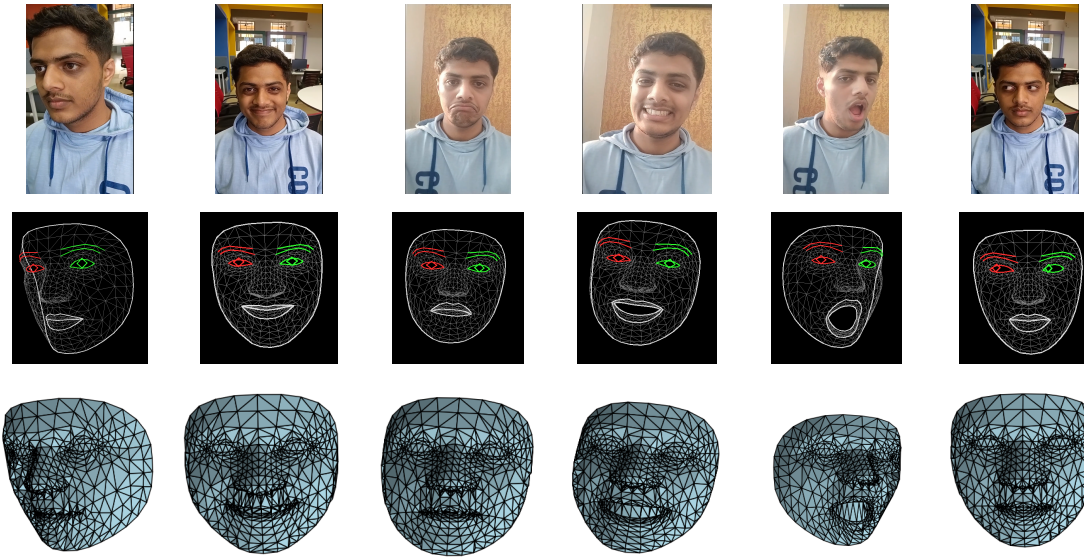
Trisurf plot



Face Mesh

Figure: Pipeline describing the process of Delaunay's Triangulation

Experimental Results: Delaunay's Triangulation



Final Architecture- Block Diagram

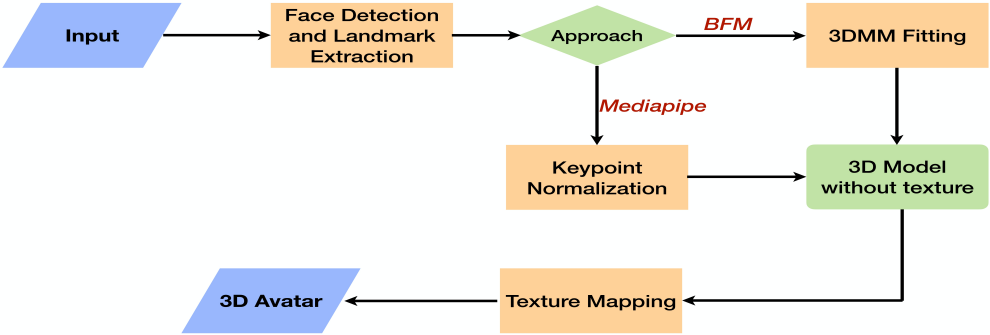


Figure: Block Diagram for the two final architectures: Mediapipe Canonical Face Model and Basel Face Model

Mediapipe Canonical Face Model-Flowchart

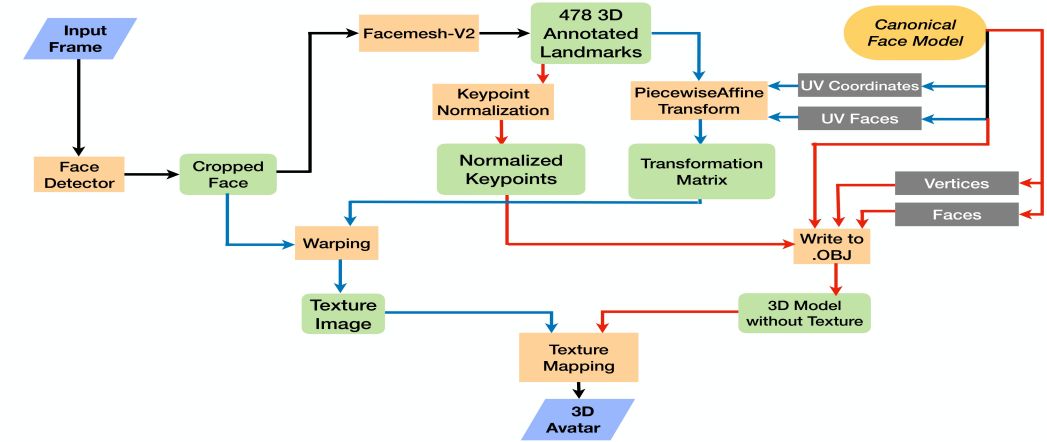
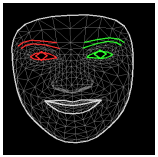


Figure: Flowchart describing Mediapipe Canonical Face Model

Pipeline for Canonical Face model

Face Detection and cropping

Face keypoint extraction



Input-frame (Image/Video/Webcam)

Cropped (ROI)

Landmark Annotated Face Mesh

Keypoint normalization

Texture mapping

3D model

3D avatar

Figure: Pipeline describing the process of Mediapipe Face Model

Experimental Results: Mediapipe (Canonical Face Model)

Basel Face Model- Flowchart

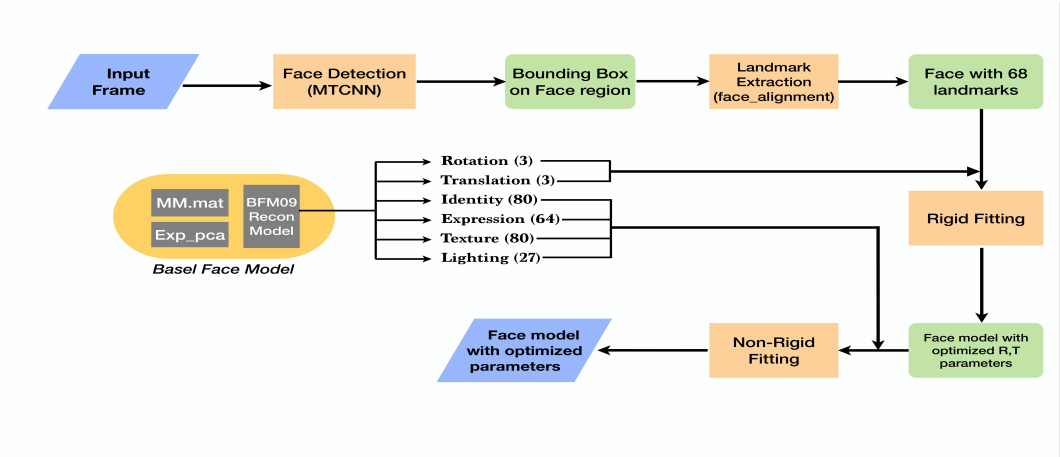
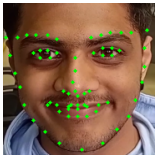


Figure: Flowchart for Basel Face Model

Pipeline for BFM

Face Detection and cropping

Face keypoint extraction



Input-frame (Image/Video/Webcam)

Cropped (ROI)

Landmarks overlaid on Face

3DMM Fitting

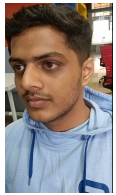
Texture mapping

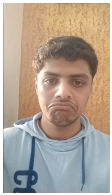
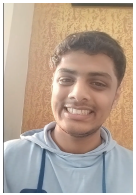
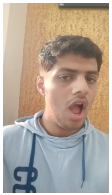
3D model

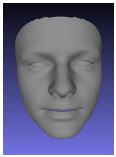
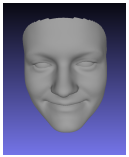
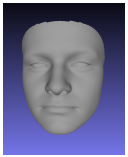
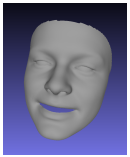
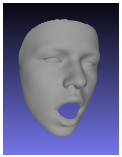
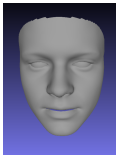
3D avatar

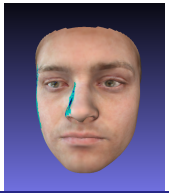
Figure: Pipeline describing the process of Basel Face Model

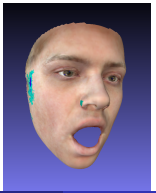
Experimental Results: BFM - Basel Face Model






Observations

Attributes	BFM	Mediapipe
No of vertices	35709	1834
Facial landmarks	68	478
Texture	Principal Component Analysis	Warping (UV wrapping-unwrapping)
Type of fitting	Rigid and Non-rigid fitting by minimizing the loss	Normalization of keypoints
Time taken to retarget for realtime	Delay of 2-5 minutes per frame	Delay of 2-5 seconds.

Conclusion

- Basel Face Model (BFM) and Mediapipe with the Canonical Face Model were the primary approaches studied until the output stage.
- BFM approach excelled in precision and high-quality retargeting, making it suitable for scenarios prioritizing accuracy, even at the expense of real-time processing.
- Mediapipe approach stood out in situations requiring swift and real-time retargeting, where compromises in result quality were acceptable for the sake of speed.
- The study emphasized the trade-offs in real-world applications, contributing to a broader understanding of when to prioritize precision or real-time processing in facial retargeting.

References

1. Wu, Cho-Ying, Qiangeng Xu, and Ulrich Neumann. *Synergy between 3DMM and 3D Landmarks for Accurate 3D Facial Geometry.* In *2021 International Conference on 3D Vision (3DV)*, IEEE, 2021.
2. Daněček, Radek, Michael J. Black, and Timo Bolkart. *EMOCA: Emotion Driven Monocular Face Capture and Animation.* In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
3. Feng, Yao, et al. *Collaborative Regression of Expressive Bodies using Moderation.* In *2021 International Conference on 3D Vision (3DV)*, IEEE, 2021
4. Aneja, Deepali, et al. *Learning to Generate 3D Stylized Character Expressions from Humans.* In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2018.
5. Wood, Erroll, et al. *3D Face Reconstruction with Dense Landmarks.* In *European Conference on Computer Vision*, Springer Nature Switzerland, 2022.

Thank You