Generate 3D Stylized Character Expressions from Humans

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

June 14, 2024

Overview

Introduction

- Clear facial expressions are essential for effective human interaction and communication, making them vital in animated movies and illustrations.
- Animator-created character expressions can be expressive and clear but require expertise and hours of work. In order to speed up the process, using human actors to control and animate a 3D stylized character is ideal.

Figure: Expression transfer to human and non-human characters.

Motivation

- 3D Stylized character enhances user experience in video games by extensive avatar expressions.
- It is beneficial for visual storytelling.
- It boosts social VR experience in virtual universe.
- Stylization enhances the virtuality of the character.

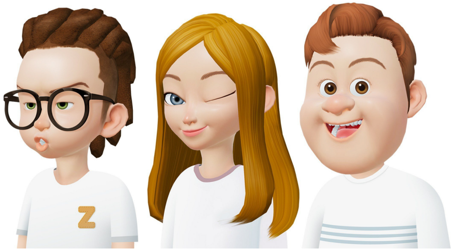

Figure: Expressive avatars in video games

Figure: Visual storytelling

Facial Expression Retargeting from Human to Avatar Made Easy [5] (Juyong, Zhang, et al. IEEE 2020)

Transferring facial expressions from a human face to a virtual avatar in a more user-friendly way compared to traditional techniques.

Methodology:

- Variational autoencoders for expression embedding.
- User-friendly triplet-based correspondence learning.
- Automatic triplet generation and network training.

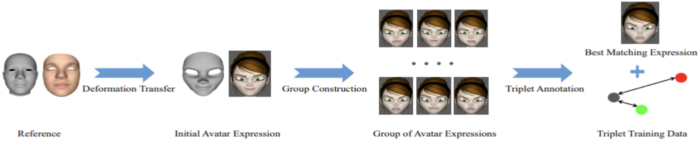

Figure: Semantic correspondence construction

Learning to generate 3D stylized character expressions from humans. [2] (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

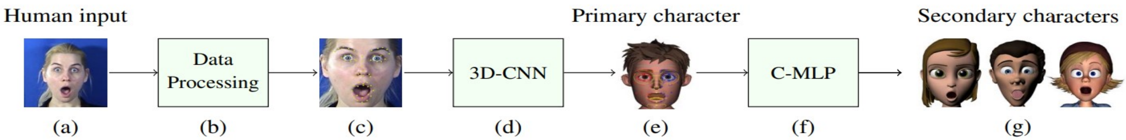

Figure: Multi-stage expression transfer system ExprGen

Modeling Stylized Character Expressions via Deep Learning [1] (Aneja, Deepali, et al. ACCV 2016)

A novel approach which accurately recognizes human expressions and transfers them to a stylized character without relying on explicit geometric markers.

Architecture: CNN and transfer learning.

Methodology:

- Multiple convolutional layers followed by max-pooling layers and fully connected layers with softmax activation
- Transfer learning between human dataset and character dataset
- Retrieval of character closest expression match to human using Jensen—Shannon divergence and Geometric distance

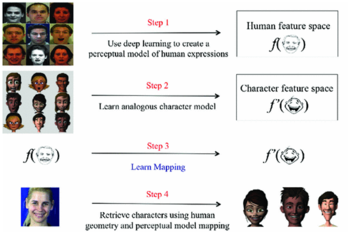

Figure: Pipeline proposed in the paper

Literature Survey

Real-time Facial Animation for 3D Stylized Character with Emotion Dynamics [4] (Pan, Ye, et al. ACM 2023)

- 3D emotion transfer network: Convert 2D images to generate a 3D stylized rig parameter.
- Blendshape adaptation network: Blendshape weights as input and generates expressive controller values.

Methodology:

- Interpolation: Takes the human facial images as input and predicts the character's controller values.
- Retrieve matched character image with closest emotional and geometric distance.

Limitations:

- 3D modeling, texturing and lighting.
- Unsatisfactory rendering results.

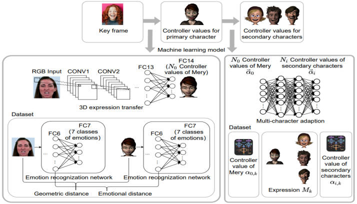

Figure: Overview of first solution (Interpolation)

Problem Statement and Objectives

Problem Statement

Generate 3D Stylized Character (Human/Non-Human) Expressions from Humans

Objectives

- Find correspondence between a human expression and the desired character expression, in 2D.
- Face 3D landmarks extraction from the character image obtained from mapping of human expression to character expression.
- Generation of an expressive face mesh for the desired character.

Dataset Description

- FER Dataset is used for human expressions.
- Resolution = 48x48, Training samples = 27,000, Testing samples = 3,000
- All the images are in greyscale and are re-labelled into classes.
- Images are divided into 7 classes: Happy, Sad, Anger, Fear, Disgust, Surprise, and Neutral

Figure: Sample images from FER-2013 dataset

- FERG DB Dataset for character expressions.
- Resolution = 256x256, Training samples = 50,190, Testing samples = 5,576
- Each character is divided into 7 classes: Happy, Sad, Anger, Fear, Disgust, Surprise, and Neutral

Figure: Sample images from FERG-DB dataset

Approach-Training

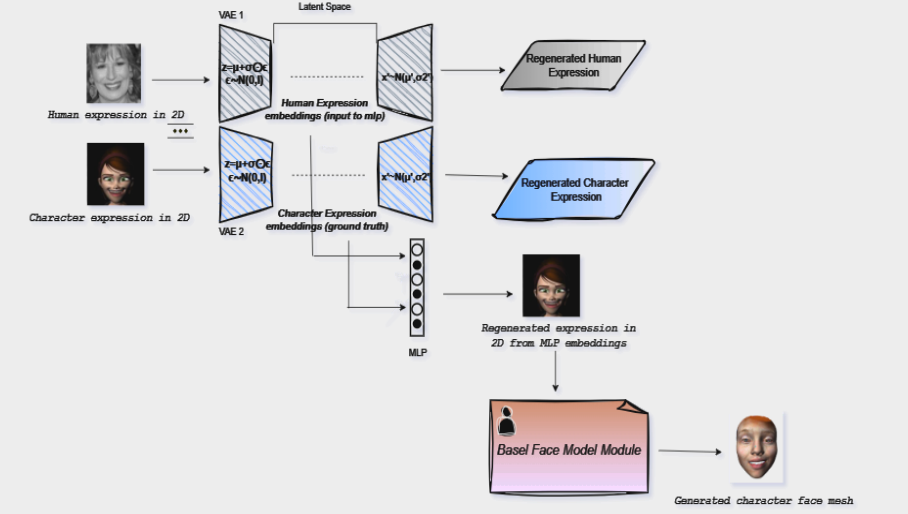

Figure: Figure representing the training pipeline of proposed model

Basel Face Model Module

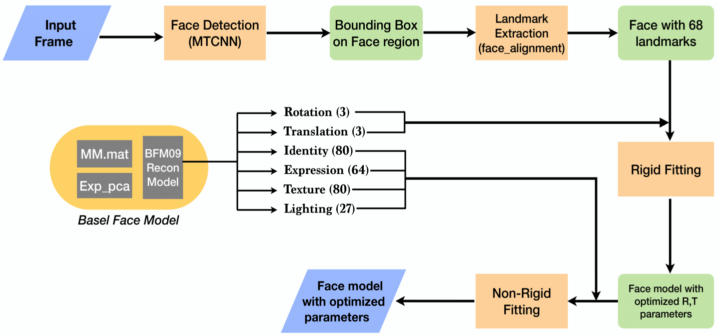

Figure: Pipeline of Basel face model module

Approach-Training

- VAE for human expression: Generates a latent space for human expressions.
- VAE for character expression: Generates a latent space for character expressions.
- MLP: Trains the model for latent space of human expression (input) to latent space for character expression (ground truth)
- Basel Face Model: Takes the character image as input and generates a 3D character face mesh as output

Approach-Testing

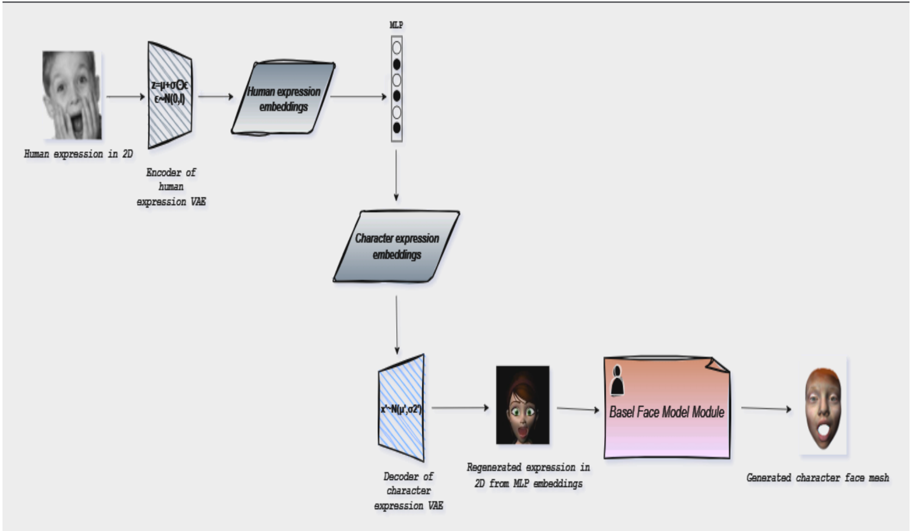

Figure: Figure representing the testing pipeline of proposed model

Approach-Testing

- Human expression is passed as input to the encoder of human VAE and a latent space is generated.
- The latent space is mapped to the latent space of a character through MLP.
- Character expression is reconstructed by passing the mapped latent space through the decoder of character VAE.
- Basel Face Model: Takes the character image as input and generates a 3D character face mesh as output

Initial Results-Human to character correspondence

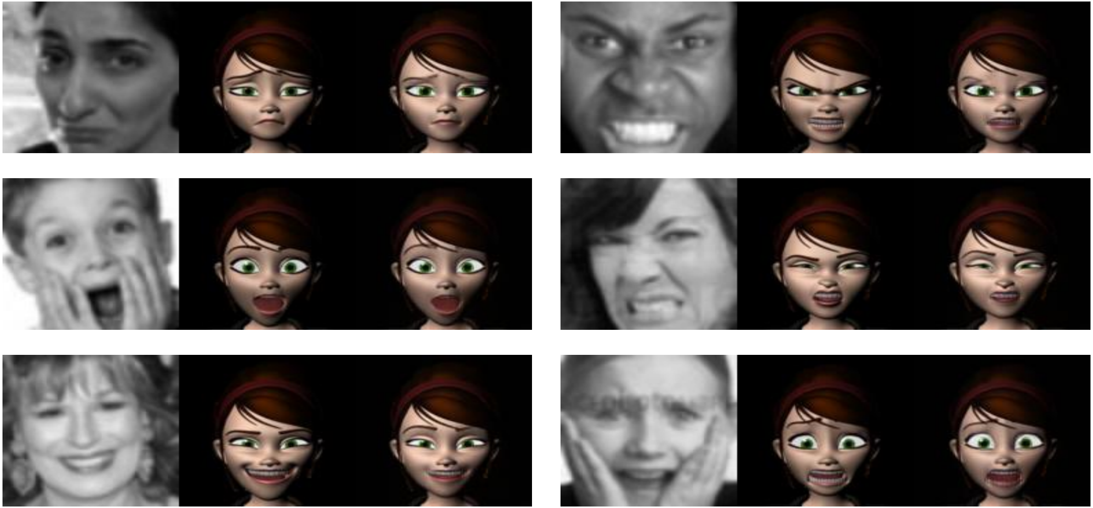

Figure: Results indicating the reconstructed character expression from human expression

Generated Results-Textured character face mesh for Mery

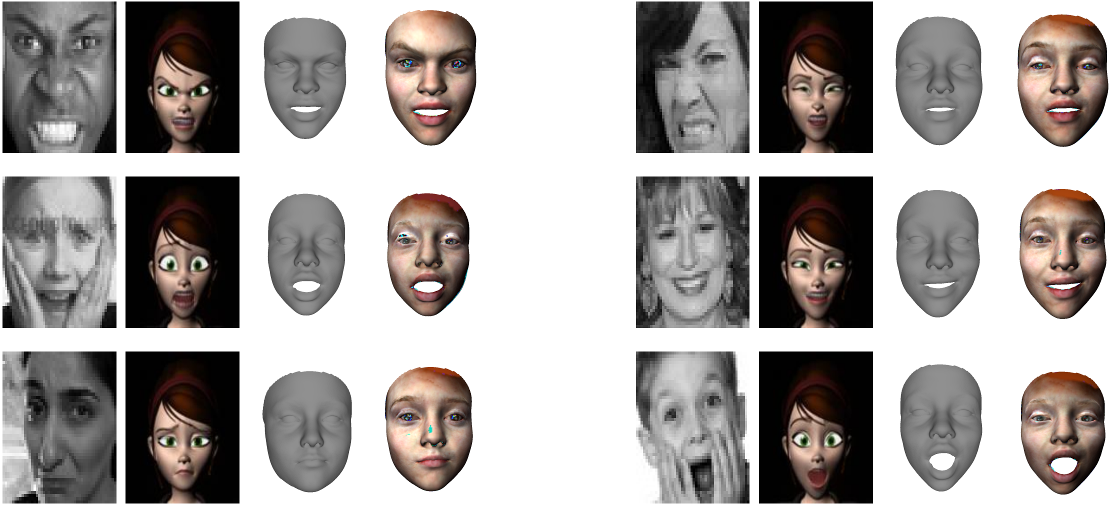

Figure: Human expression transforms to its character equivalent image, and to its 3D character mesh

Generated Results-Textured character face mesh for Malcom

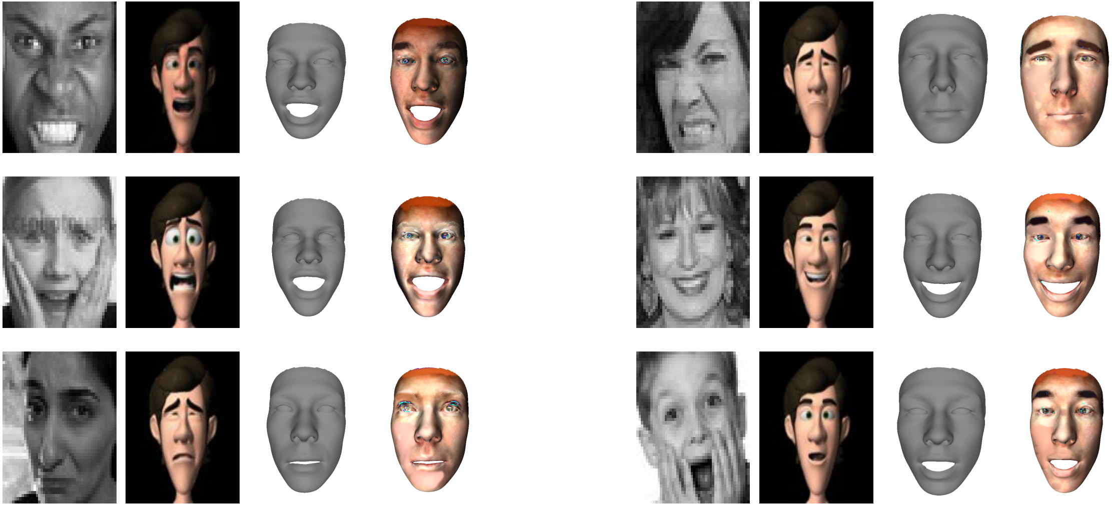

Figure: Human expression transforms to its character equivalent image, and to its 3D character mesh

Conclusion

Our model proposes an innovative and efficient methodology to generate stylised character meshes from 2D human input images and 2D character images. This method is highly cost effective and can be interpolated to any character, if a 2-dimensional appearance of the character is provided.

This approach can be used in animations and gaming applications, in order to transfer expressions from a human to a character.

Reference I

Deepali Aneja and et al.
Modeling stylized character expressions via deep learning.
Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part II 13. Springer International Publishing, 2017.

Deepali Aneja and et al.
Learning to generate 3d stylized character expressions from humans.
2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018.

Jiangke Lin, Yuan Yi, and Zou Zhengxia.
Meingame: Create a game character face from a single portrait.
Proceedings of the AAAI conference on artificial intelligence. Vol. 35. No. 1. 2021., 2021.

Ye Pan and et al.
Real-time facial animation for 3d stylized character with emotion dynamics.
Proceedings of the 31st ACM International Conference on Multimedia. 2023, 2023.

Juyong Zhang, Keyu Chen, and Jianmin Zheng.
Facial expression retargeting from human to avatar made easy.
IEEE Transactions on Visualization and Computer Graphics 28.2 (2020): 1274-1287., 2020.

Thank You