Facial Motion Retargeting

Bindita Chaudhuri

Paul G. Allen School of Computer Science & Engineering, UW
(Linda Shapiro, Alex Colburn, Barbara Mones, Gary Faigin)

Visual Intelligence Group, Microsoft AI&R, Redmond
Introduction

• **Goal**: Predict and transfer facial motion from 2D images to 3D models
Introduction

• **Goal**: Predict and transfer facial motion from 2D images to 3D models

  - bounding box detection and tracking
  - per-frame expression and pose

• Recent applications include:

  - video games
  - motion capture films
  - social VR experience
Methodology

- **Blendshape based approach** *(better generalizability to multiple characters)*:

- **Example based approach** *(better generalizability to out-of-space expressions)*:
Blendshape based Approach

3DMM tensor

user-specific blendshapes

3D vertex coordinates

2D landmarks

\[ P_{lm} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \end{bmatrix} \begin{bmatrix} R \ast (\hat{M} \ast w_{id} \ast w_{exp}) + t \end{bmatrix} \]

\[ T = V \times B_{exp} \times B_{id} \]

identity coefficients \( w_{id} \)

expression coefficients \( w_{exp} \)

pose (R, t) and projection (f) parameters
Single Face Network

pose from global features

identity and expression from local + global features

B. Chaudhuri, N. Vesdapunt, B. Wang, Joint Face Detection and Facial Motion Retargeting for Multiple Faces, CVPR 2019
Multi Face Network

- two sequential networks; memory inefficient
- runtime increases linearly with number of faces

B. Chaudhuri, N. Vesdapunt, B. Wang, Joint Face Detection and Facial Motion Retargeting for Multiple Faces, CVPR 2019
Multi Face Network

- YOLO loss function; bounding box and 3D face prediction help each other

\[ b_{lm_x} = b_x + b_w \cdot b_{lm_x} ; \quad b_{lm_y} = b_y + b_h \cdot b_{lm_y} \]

B. Chaudhuri, N. Vesdapunt, B. Wang, Joint Face Detection and Facial Motion Retargeting for Multiple Faces, CVPR 2019
Network Performance for Test Images
Results for Single Face Based Application

Hardware: Google Pixel 2
Live Performance Capture for Multiple Faces
Example based Approach (semi-supervised)

Online triplet generation based on distance metric

\[ \phi_d = \alpha |\text{JS Distance}| + \beta |\text{Geometric Distance}| \]

Expression feature vectors  Geometry feature vectors

D. Aneja, B. Chaudhuri, A. Colburn, G. Faigin, L. Shapiro, B. Mones, Learning to Generate 3D Stylized Character Expressions from Humans, WACV 2018
Example based Approach (semi-supervised)

HCNN → SCNN → fused-CNN → 3D-CNN

Online triplet generation based on distance metric

(E, E+, E-)

Binary similarity score

matching pairs

256 × 256

100 × 1

D. Aneja, B. Chaudhuri, A. Colburn, G. Faigin, L. Shapiro, B. Mones, Learning to Generate 3D Stylized Character Expressions from Humans, WACV 2018
Results for Videos

Frame-by-frame transfer; jitter removed by temporal smoothening using Savitzky-Golay filter
Example based Approach (unsupervised)

Aim:

• Use single network that directly regresses 3D vertices of character

• Generalize to a broader range of expressions
Example based Approach (unsupervised)

• Compute facial landmarks:
  ![Facial landmarks](image)

• Convert 3D model to 2D position map:
  ![3D to 2D conversion](image)

• Train CycleGAN:
  ![CycleGAN](image)
  - Landmark loss
  - generator A→B
  - generator B→A
  - 2D discriminator
  - 3D discriminator
Results

Input

Blendshape based

Semi-supervised Example based

Unsupervised Example based
Results

Input

Blendshape based

Semi-supervised Example based

Unsupervised Example based
Thank you