Automated Restyling of Human Portrait Based on Facial Expression Recognition and 3D Reconstruction
Cheng-Han(Dennis) Wu, Hsin Chen
Department of Electrical Engineering, Department of Mechanical Engineering, Stanford University

Motivation
Inspired by the movie Inside Out, the project showcases the capabilities of AI in visual effects automation in photography or film making. Furthermore, through convolutional neural network (CNN) with enhanced classification algorithm, we are able to achieve high expression recognition accuracy.

Related Work
1. Depth Based Visual Effect
Matthias Ziegler, Andreas Engelhardt, Stefan Miller, Joachim Keinert, Frederik Zilly, Siegfried Foessel. Multi-camera system for depth-based visual effects and compositing. CVMP, 2015.
2. Caffe Model: GoogleNet
3. Caffe Model: DeXpression

Classification
1. Classification Algorithm Benchmark
- Cross-validation accuracy
- Max. Accuracy: GoogleNet (~99%) > DeXpression (~94%)

2. KNN and Subspace KNN
KNN:
Given a set of N points in an r-dimensional feature space, calculate proximity of the test point with the closest k neighbors using their relative Euclidean distances. Assign the test point to the class that has the most frequent occurrence.
Subspace KNN:
Given a set of N points in an r-dimensional feature space, randomly choose m-dimensional subspace (m < n), in which find k nearest neighbors for a test point. Repeat such procedure for c_n^k times, then assign the test point to the class that has the most frequent occurrence.
For this model, n = 8, m = 4, k = 30.

System Architecture
1. RGB & Depth Preprocessing
2. CNN Model Training
3. CNN Model Comparison
4. Classification Algo. Optimization
5. Expression Recognition
6. 3D Face Model Calculation
7. Shading/Relighting/Restyling

Misclassification Analysis
1. DeXpression:
- prone to contempt → neutral misclassification
- Confusion Matrix (subspace KNN)
2. GoogleNet:
- prone to angry → disgust misclassification

Image Processing
1. Kinect for Windows V2 Calibration
2. Cropping ROI and Image Masking Using Depth
- Crop with OpenCV tracked face ROI.
- Mask with dilated, hole-filled, and threshold depthmap.
3. Depth Denoising
- Calculate directional lighting intensity and color for normal map.

Feature Extraction
1. Dataset(CK+)
- Extended Cohn-Kanade dataset
- 5876 images, 8 Emotions
- Trained Size:
  - Small: 981 images
  - Large: 1635 images
2. Training Platform: Caffe
3. Model
- DeXpression(DeX):
  - 11 main layers
  - 400000 Iterations (loss ~ 0.0008)
- GoogleNet(GIN):
  - 27 main layers
  - 240000 Iterations (loss ~ 0.0004)

Results
1. 8 Emotion Effects
2. Unseen Dataset Test accuracy
- Using 746 labeled Cohn-Kanade(CK) images
- Fine KNN has a 73.5% accuracy over unseen data. KNN/subKNN robust toward two models.
- Softmax performance heavily depends on model and feature quality.
Automated Restyling of Human Portrait Based on Facial Expression Recognition and 3D Reconstruction

Cheng Han (Dennis) Wu
Department of Electrical Engineering, Stanford University
1. Cropping ROI and Image Masking
   - Crop with OpenCV tracked face ROI
   - Mask with dilated, hole-filled, and thresholded depthmap.

2. Shading/Relighting
   - Calculate surface normal from depth.
   - Calculate directional lighting intensity and color for the face.

2. Restyling
   - Apply mood lighting and background, processed using above mask applied on both images.