Recognizing Fine Facial Micro-Expressions using Two-Dimensional Landmark Feature

Dong Yoon Choi*, Dae Ha Kim, Byung Cheol Song
{pride0723, kdhht5022} @ gmail.com, bcsong @inha.ac.kr
Department of Electronic Engineering, Inha University, Republic of Korea

Introduction

- Micro-expression recognition in computer vision
  - Most of the existing emotion recognition methods are studied only on large facial expression.
  - If we pay attention to the micro expression of human face, it makes a difference even though there is emotion.

- Motivation and our approach
  - Landmark is not influenced by personal characteristics and environment, and is similar in movement pattern on each emotion
  - Convert to 2-D feature of Image type and apply to convolutional neural network (CNN)

Related Works

- Deep temporal appearance-geometry network (DTAGN)[1]
  - A neural network-based algorithm that uses image and landmark simultaneously
  - Perform emotion recognition through joint fine-tuning number of each network
- Preprocessing method using video magnification[2]
  - Apply the existing recognition algorithm after applying the video motion magnification

2-D Landmark Feature

- Generation of frame-based landmark feature (LMF)
  - The variation of the distance between each landmark is represented by a two-dimensional matrix.
    - Unique pattern can be generated according to emotion
    - Perform normalization using the maximum value of distance variation
  - Normal pattern can be generated robust to the size of expression

Landmark Feature-Based Face Expression Recognition algorithm

- CNN-LSTM-based classifier
  - Encode 2D feature of each frame through VGG16 based CNN
  - Perform classification through stacked LSTM using encoded feature sequence as input

- Jointly FER scheme using image and LMK
  - A simple facial expression metric intensity (SFEM): Measure the size of the expression using the sum of the position change of the landmark.
  - Algorithm can be selectively processed according to expression size

Experimental Results

- Dataset
  - CK+ dataset (327 sequences), seven emotion categories
  - In the case of small motion data, synthesis is performed using only the first three frames of the sequence Create nine-frame small motion sequence through video frame interpolation[3]
  - Landmark information is extracted using activate appearance model (AAM)

- Visual analysis of landmark feature (LMF)
  - Example of 2D landmark feature based on seven emotions

Conclusion

- Recognition accuracy
  - CK+ training process, used general expression data
  - In test process, used synthesized micro expression data and mixed dataset
- Recognition difficulty
  - Unique pattern can be generated robust to the size of expression
  - Through SFEM and joint framework, general and micro motion can be processed


\[ \text{SFEM} = \left( \frac{N}{n} \sum_i \sum_j | LMF_i(j)| \right) - \left( \frac{n}{N} \sum_i \sum_j | LMF_i(j)| \right) \]