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

2018 IEEE International Conference on Image Processing
October 7-10, 2018 • Athens, Greece
Imaging beyond imagination

Introduction

Micro-expression recognition in computer vision

- Most of the existing emotion recognition methods are studied only on large facial expression
- In actual situations, there are many cases where the expression does not change significantly even though there is emotion.

General-expression example(SFEM: 1.83)

Micro-expression example(SFEM: 0.20)

- In the image sequence, the amount of information change is small→ Recognition difficulty↑

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)

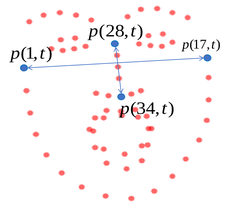
- The variation of the distance between each landmark is represented by a two-dimensional matrix.
  → Unique pattern can be generated according to emotion

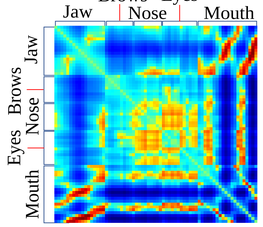
$$LMF_t(i,j) = \|p(i,t) - p(j,t)\|_2 - \|p(i,t-1) - p(j,t-1)\|_2$$

p: landmark position

- Perform normalization using the maximum value of distance variation
  → Unique pattern can be generated robust to the size of expression

$$\overline{LMF_t}(i,j) = \begin{cases} 128 \cdot \left\{ \frac{LMF_t(i,j)}{\max(LMF_t)} \right\}^{0.5} + 127 & LMF_t(i,j) > 0 \\ -128 \cdot \left\{ \frac{LMF_t(i,j)}{\max(LMF_t)} \right\}^{0.5} + 127 & LMF_t(i,j) < 0 \\ 127 & else \end{cases}$$

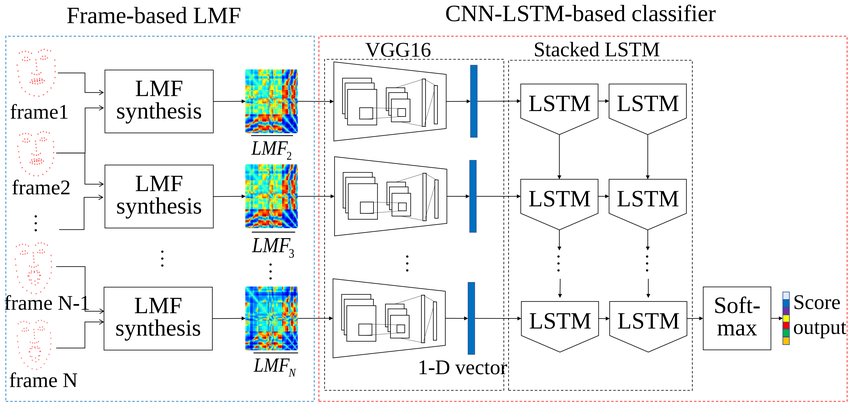




Landmark Feature-Based Face Expression Recognition algorithm

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

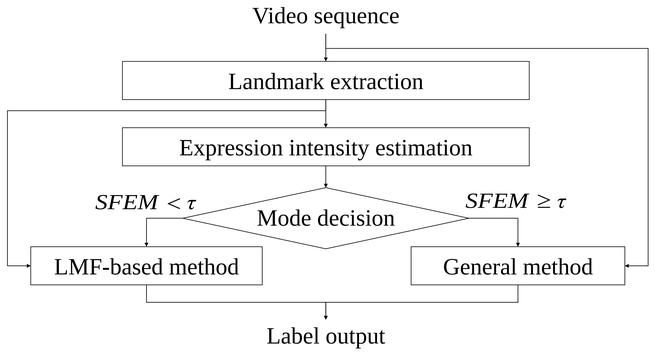


Jointly FER scheme using image and LMK

- 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

$$SFEM = \frac{\frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{M}\sum_{i=1}^{M}|LMF_t(i,j)|}{||p(1,t)-p(17,t))||_2 \cdot ||p(28,t)-p(34,t))||_2}$$

N: # of frames
M: # of landmark



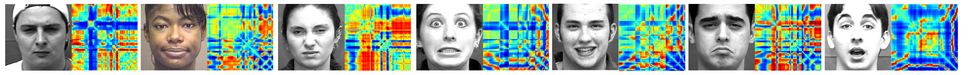
Experimental Results

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)

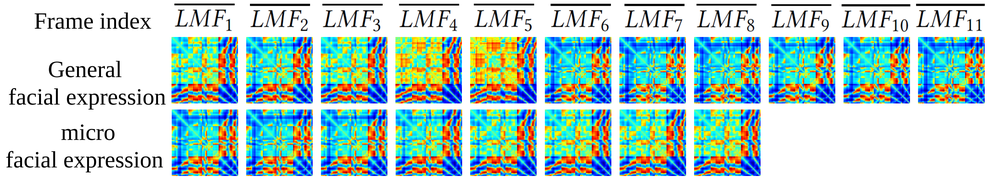
Visual analysis of landmark feature(LMF)

Example of 2D landmark feature based on seven emotions



Angry Contempt Disgust Fear Happy Sad Surprise

Example of landmark feature sequence for general case and micro case



Recognition accuracy

- In training process, used only general expression data
- In test process, used synthesized micro expression data and mixed dataset

	Test dataset[%]		
	General facial expression	Micro facial expression	Mixed
DTAGN[1]	93.88(97.25*)	43.34	70.94
LMF-FER	92.66	**77.98**	87.46
Joint framework	N/A	N/A	**88.69**

* Results stated in [1]

Conclusion

- Unique pattern can be generated robust to the size of expression
- Through SFEM and joint framework, general and micro motion can be processed

[1] H. Jung et al., "Joint fine-tuning in deep neural networks for facial expression recognition," ICCV2015
[2] A. Ngo et al., "Eulerian emotion magnification for subtle expression recognition"," ICASSP 2016
[3] S. Niklaus et al., "Video frame interpolation via adaptive separable convolution," ICCV 2017