COLOR CHANNEL-WISE RECURRENT LEARNING FOR FACIAL EXPRESSION RECOGNITION

*Jinhyeok Jang, Dae Hoe Kim, Hyung-Il Kim, and Yong Man Ro**

Image and Video Systems Lab, School of Electrical Engineering, KAIST, Republic of Korea

ABSTRACT

Facial expression recognition is increasingly gaining importance in emerging affective computing applications. In practice, achieving accurate facial expression recognition is still challenging due to environmental variations. In this paper, we propose a color channel-wise recurrent facial feature learning. The proposed method adopts recurrent neural network to learn expression features sequentially along color channels. The proposed network preserves discriminative expression feature through a long short-term memory for the sequence of color spatial features. Comprehensive experiments have been conducted on the publically available CMU Multi-PIE dataset under illumination variations. Experimental results showed that the proposed method achieved higher recognition rates compared to the state-of-the-art methods.

Index Terms— Facial expression recognition (FER), facial expression features, deep learning, color channel sequencing

1. INTORDUCTION

Automatic facial expression recognition (FER) has been an interesting research topic in the area of the affective computing including human emotion analysis and human-computer interaction (HCI) [1-3]. Nonetheless, achieving accurate FER is still a challenging problem. It is mainly attributed to the illumination variation which causes degradations in acquired image [4,5].

While many research efforts have been devoted to explore FER, most of previous works have adopted hand-crafted texture features extracted from local patches [6-9]. Meanwhile, color information has been used to improve recognition rate: [10] used color local texture features. [11] proposed the fusion of color channels in global image and local patches. [12] and [13] showed that optimal color channels would be different according to applications. [14] proposed optimally combined features from multiple color channels to improve recognition performances. In combining multiple color channels, most previous works adopted weighted fusion or color channel selection based on discriminative power of each color channel [14, 15]. In those deterministic methods like weighted fusion and color channel selection, latent features from low weight or not selected color channels could be ignored.

In this paper, we propose a color channel-wise recurrent feature learning for FER. The proposed method is devised to extract latent information resided in all color channels. For this purpose, long short-term memory (LSTM) for color channel sequence is proposed. Color channels are sequenced in descending order along the discriminability of color channel measured by convolutional neural network (CNN) [16] and t-distributed stochastic neighbor embedding (t-SNE) [17]. Then, CNN features obtained from the sequenced color channels are fed to the LSTM. From the sequence of spatial features of color channels, the LSTM stores facial color-texture information to the memory cell controlled by three gates (i.e., input, forget, and output gates [18]). The three gates enable the LSTM to preserve expression information in discriminative color channels through the sequence of CNN features. To the best of our knowledge, this work is the first attempt to devise an effective latent feature representation using recurrent neural network to learn color channel patterns.

Comprehensive experiments have been conducted on the publically available CMU Multi-PIE expression dataset [5]. Experimental results showed that the proposed method achieved higher recognition rates compared to the state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 details the proposed method. In Section 3, the comparative experimental results of the proposed method are presented. Finally, conclusions are drawn in Section 4.

2. PROPOSED FACIAL EXPRESSION RECOGNITION WITH COLOR CHANNEL-WISE LSTM

Fig. 1 represents the proposed FER framework, which is mainly composed of three parts: 1) channel-wise spatial coding with CNN, 2) color channel-spatial feature sequencing, and 3) LSTMs to learn expression features along color channels. The details of each step of the proposed method are described in the following subsections.

* Corresponding author (ymro@kaist.ac.kr)

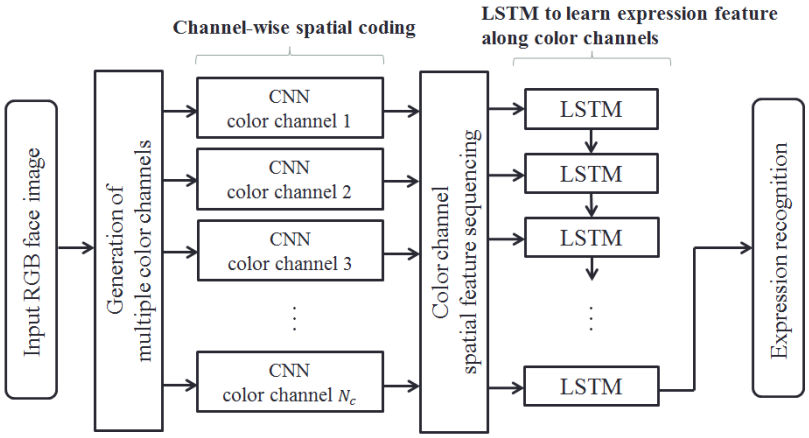


Fig. 1. A block diagram of proposed method.

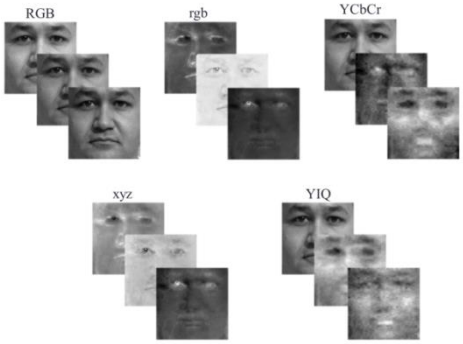
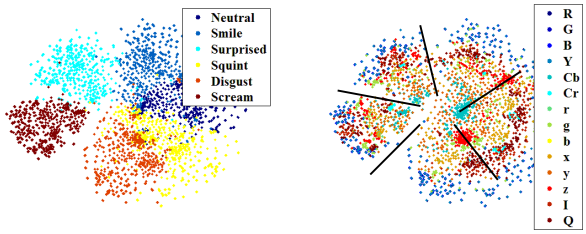


Fig. 2. Examples of color channels of five color spaces.



(a) (b)

Fig. 3. t-SNE visualization of spatial features extracted from all channel images. Color in (a) represents the expression and (b) represents the color channels.

2.1. Channel-wise spatial coding

Given an input RGB face image, multiple color channels are generated. In this paper, five different color spaces are used (i.e., RGB, YCbCr, xyz, rgb and YIQ) as shown in Fig. 2. To encode spatial information (i.e., texture) of each color channel, CNN is utilized. The CNN is comprised of an input layer of size 64x64 pixels, three convolutional layers of kernel size 3x3 and the number of kernel is 32, 64, and 64, respectively. Each convolutional layer is followed by a max pooling layer of kernel size 3x3 with stride 2. Lastly, two fully connected layers with 1024 units are used. In this paper, a single CNN model is learned, which shares trainable weights across all color channel images.

2.2. Color channel-spatial feature sequencing

It is known that order of input in LSTM is important for resultant performance [19]. In this paper, we propose discriminability-ordered sequencing. The color channel-spatial features are sequenced based on their discriminability with respective to facial expression recognition. To discriminate the features, t-SNE is employed which is a dimension reduction method based on Euclidian distance between each feature pair.

Fig. 3 shows the t-SNE visualization of CNN spatial features extracted from all channel images. Colors in the figure represent the facial expressions in Fig. 3(a), and the color channels in Fig. 3(b), respectively. As shown in the figure, different color channels have different discriminability with respective to facial expressions (Fig. 3(b)). Specifically, the color channels whose samples are located on inner regions in Fig. 3(b) have low discriminability because distances between different expressions are close to each other. On the contrary, color channels whose samples are located on outer regions have high discriminability.

Based on the aforementioned observation, color channel-spatial feature sequencing is proposed. The detail is shown in the Algorithm 1. The spatial features, which are extracted from CNNs with color channels, are used for the sequencing.

In the first step, dimension-reduced features are calculated by t-SNE. Then, color channel distances from the global mean are calculated. Finally, the sequencing of color channels is found by sorting the color channel distances in descending order.

Algorithm 1. Proposed color-sequencing with t-SNE analysis

Input :

- Number of input channels N_c
- Input feature set, $\boldsymbol{X} = \{x_1^1, x_2^1, \dots, x_j^1, x_1^2, x_2^2, \dots, x_j^2, \dots, x_j^{N_c}\}$ where j is the number of face images
- Number of iterations T, learning rate η, momentum α

Output :

- Index of sequenced color channels O

Begin:

-

Calculate dimension reduced feature set $Y = \text{tSNE}(\boldsymbol{X}, \text{T}, \eta, \alpha)$ where $\boldsymbol{Y} = \left\{y_1^1, y_2^1, \dots, y_j^1, y_1^2, y_2^2, \dots, y_j^2, \dots, y_j^{N_c}\right\}$

- Calculate color channel distance from global mean $D = \{d_1, \dots, d_k, \dots, d_{Nc}\}$ where d_k is calculated as follows:

$$d_k = \frac{\sum_{l=1}^{len(y^k)} dist(y_l^k, mean(Y))}{len(y^k)}$$

- Initialize index of sequenced color channels $\mathbf{O} = \{\phi\}$

For i=1 **to** N_c do

- $o_i = \arg\max_k d_k$
- $\mathbf{D} = \mathbf{D} - \{d_{o_i}\}$

End

End

In this paper, the sequencing of color channels found by the Algorithm 1 is [B-G-Y-R-I-Cr-Q-g-r-z-b-y-Cb] from the most discriminative color channel to the least discriminative color channel. Fig. 4(b) shows t-SNE visualization of spatial discriminative color channel. The figure represents that the proposed method successfully orders color channels based on the discriminability of color channels. Fig. 5 shows examples of color channel images sequenced by the proposed method.

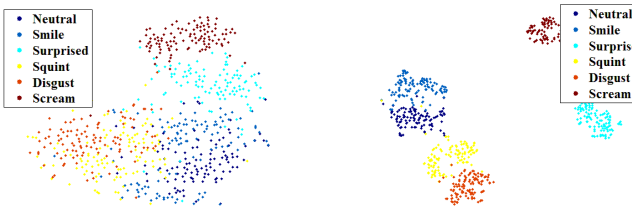
2.3. LSTMs to learn expression features along color channels

Color-texture information that resides along the sequenced color channel-wise spatial features are consequently learned by using LSTM. From the sequenced color channel-wise spatial features, the LSTM stores facial color-texture information to the memory cell controlled by three gates (i.e., input, forget, output gates [18]). The three gates enable the LSTM to preserve information in discriminative color channel through the sequence of features. The depth of LSTM structure in this paper has two layers, where each layer contains 512 hidden units.

3. EXPERIMENTS AND DISCUSSION

3.1 Experimental conditions

We performed experiments to verify the proposed method. In experiments, a subset of CMU Multi-PIE dataset [5] composed of 961 images under illumination variations was used, where frontal facial images of randomly selected 100 subjects were used. Each facial image was labeled with one of the six expressions (i.e., neutral, smile, surprise, squint, disgust, and scream). The face images were cropped and resized to 64x64 pixels.



(a) (b)

Fig. 4. t-SNE visualization of spatial features extracted from (a) Cb channel images (least discriminative color channel) and (b) B channel images (most discriminative color channel) in training set. Color represents the expression of samples.

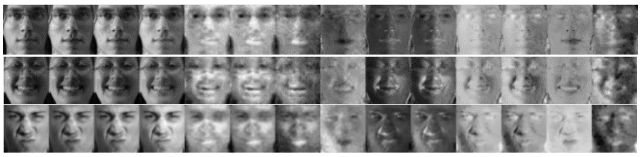


Fig. 5. Color channel images sequenced by the proposed color-sequencing. The channels are sequenced from the most discriminative channel to the least discriminative channel. Each row shows images of different subject.

The evaluations in experimental results were conducted with a five-fold cross validation, such that the facial images of test subject were excluded from the training set. For each fold, CNN and LSTM were trained on the training set. Then the independent testing set was tested by the trained models. This process was repeated for each fold so that testing was performed at least once for each fold. After the five-fold cross validation, the recognition rate was calculated by averaging recognition rates of five folds.

3.2. Effectiveness of the proposed FER method

Table 1 shows the recognition rate of the proposed FER on the CMU Multi-PIE dataset. Comparative results are shown with the state-of-the-art methods as well. For the comparison, hand-crafted features and sparse representation based classifier (SRC) [20] were used. In addition, conventional CNN with RGB image was also evaluated. For this purpose, a softmax classification layer was attached to the spatial coding network in CNN and input layer was RGB image. As shown in Table 1, the proposed method clearly outperformed existing state-of-the-art methods and the conventional CNN methods with the same dataset. The result indicates that the proposed method could improve the recognition rate by effectively encoding patterns with multiple color channels.

Table 1. Recognition rate comparison with existing state-of-the-art methods.

Method	Recognition rate (%)						
	Neutral	Smile	Surprise	Squint	Disgust	Scream	Overall
Proposed method	**81.25**	**93.13**	**93.75**	**78.88**	**69.37**	98.12	**85.74**
CNN+RGB image	77.50	86.88	89.38	59.01	67.50	**98.75**	79.81
SRC+Intra-class variation image [3]	70.67	85.98	89.98	61.66	66.36	97.49	78.72
SRC+Raw pixel [6]	48.15	58.06	63.47	39.14	48.15	85.09	57.49
SRC+LBP [6]	54.66	75.38	79.88	51.25	64.57	89.79	69.15
SRC+RawLBP [7]	54.59	77.30	83.24	58.92	67.03	94.59	72.61
SRC+Gabor [8]	59.93	79.75	84.25	50.02	63.53	94.16	71.74
SRC+LPQ [9]	58.32	79.04	81.10	49.41	67.44	93.58	71.48

Table 2. Recognition rate with different color channel sequencing.

Used color-spaces	Recognition rate (%)		
	Proposed sequencing	Random order 1	Random order 2
RGB	80.33	-	-
+YCbCr	80.54	79.71	79.60
+xyz	83.76	81.38	83.04
+rgb	83.87	83.35	83.66
+IQ	85.74	84.81	84.60

3.3. Effectiveness of the proposed color channel-spatial feature sequencing

In this experiment, the effectiveness of the proposed color channel-spatial feature sequencing was evaluated. In the evaluation, different color spaces were incrementally added for the color channel sequencing. For given color channels, spatial features of color channels sequenced by the proposed method were fed to the LSTM. For the comparison, two sets of randomly sequenced color channels were generated. Table 2 shows the recognition rates with different color channel orders. As shown in Table 2, the proposed color channel sequencing outperformed the random sequencing. The result indicates that the order of color channels makes an effect on recognition rate. In addition, overall recognition rates were improved when more color channels were used.

3.4. Effectiveness of the proposed LSTM along color channels

We evaluated the effectiveness of the LSTMs along color channels. In order to focus the effect of the LSTM, we replaced CNN features to hand-crafted features (i.e., LBP [6] and LPQ [9]). Namely, LBP and LPQ feature extractors were used in the channel-wise spatial coding. Hand-crafted features of each channel sequenced by color-order were fed into LSTM. As shown in Table 3, the recognition rates of hand-crafted features with the proposed method were improved compared to hand-crafted feature with SRC. This result indicates that the proposed LSTMs along color channels channel can be used with any feature extractor.

Table 3. Recognition rates of hand-crafted features with different FER method.

Method	Recognition rate (%)
Proposed+LBP	**74.30**
SRC+LBP [6]	69.15
Proposed+LPQ	**77.73**
SRC+LPQ [9]	71.48

4. CONCLUSION

In this paper, we proposed a color channel-wise recurrent feature learning for FER. In the proposed method, color channel sequencing based on discriminability of channel-wise spatial features was devised. The CNN features from the sequenced color channels were fed to the LSTM where expression features were learned sequentially along color channels. The experimental results showed that the proposed method was robust to illumination variations and outperformed existing state-of-the-art methods in terms of recognition rate.

5. ACKNOWLEDGEMENT

The authors would like to thank Wissam J. Baddar for his valuable discussion and contribution on this work. This work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2015R1A2A2A01005724).

6. REFERENCES

[1]. I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines,'' *IEEE Trans. Image Processing*, vol. 16, no. 1, pp. 172-187, 2007.

[2] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Trans. Systems, Man, and Cybernetics Part B,* vol. 36, no. 1, pp. 96-105, 2006.

[3] S. H. Lee, K. N. Plataniotis, and Y. M. Ro. "Intra-class variation reduction using training expression images for sparse representation based facial expression recognition," *IEEE Trans. Affective Computing,* vol. 5, no. 3, pp. 340-351, 2014.

[4] G. Stratou, A. Ghosh, P. Debevec, L.-P. Morency. "Effect of illumination on automatic expression recognition: a novel 3D relightable facial database," *IEEE Int'l Conf. Automatic Face and Gesture Recognition (FG)*, pp. 611-618, 2011.

[5] R. Gross, I. Matthews, J. Cohn., T. Kanade, and S. Baker. "Multi-pie," *Image and Vision Computing,* vol. 28, no. 5, pp. 807-813, 2010.

[6] M. Huang, Z. Wang, and Z. Ying, "A New Method for Facial Expression Recognition Based on Sparse Representation Plus LBP," *IEEE Int'l Congress on Image and Signal Processing (CISP)*, vol. 4, pp. 1750-1754, 2010.

[7] Z. L. Ying, Z. W. Wang, and M. W. Huang "Facial expression recognition based on fusion of sparse representation," *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, vol. 6216, pp. 457-464, 2010.

[8] S. Zhang, X. Zhao, and B. Lei, "Robust facial expression recognition via compressive sensing," *Sensors,* vol. 12, no. 3, pp. 3747-3761, 2012.

[9] W. Zhen and Y. Zilu, "Facial expression recognition based on local phase quantization and sparse representation," *IEEE Int'l Conf. Natural Computation (ICNC),* pp. 222-225, 2012.

[10] J. Y. Choi, Y. M. Ro, and K. N Plataniotis, "Color face recognition for degraded face images," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 5, pp. 1217–1230, 2009.

[11] Z. Liu and C. Liu. "Fusion of color, local spatial and global frequency information for face recognition," *Pattern Recognition,* vol. 43, no. 8, pp. 2882-2890, 2010.

[12] W. Shangxuan, Y. Chen X. Li, and A. Wu, "An enhanced deep feature representation for person re-identification," *IEEE Winter Conf. Applications of Computer Vision (WACV)*. pp. 1-8, 2016.

[13] Z. Lu, J. Xudong, and K. Alex, "An effective color space for face recognition," *IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP)*. pp. 2019-2023, 2016.

[14] T. Dang and X. T. Thi. "Weighted feature-level fusion of color local texture features for face recognition," *Int'l Journal of Information and Electronics Engineering,* vol. 5, no. 5, pp. 346, 2015.

[15] M. M. Oghaz, M. A. Maarof, A. Zainal, M. F. Rohani, and S. H. Yaghoubyan, "A hybrid color space for skin detection using genetic algorithm heuristic search and principal component analysis technique", *Plos One*, vol. 10, no. 8, 2015.

[16] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," *Proc. ACM Int'l Conf. Multimodal Interaction*, pp. 435–442, 2015.

[17] L. van der Maaten and G.E. Hinton, "Visualizing high-dimensional data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579-2605, 2008.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[19] O. Vinyals, S. Bengio, and M. Kudlur. "Order matters: Sequence to sequence for sets." *Int'l Conf. Learning Representation (ICLR)*, 2016.

[20] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 210–227, 2009.