A PRECISE RECOGNITION MODEL FOR HUMAN FACIAL EXPRESSION RECOGNITION SYSTEM

Muhammad Hameed Siddiqi[1], *Adil Mehmood Khan*[2], *Tae Choong Chung*[1], *Sungyoung Lee*[1]

[1]Department of Computer Engineering, Kyung Hee University (Global Campus), Rep. of Korea.
[2]Division of Information and Computer Engineering, Ajou University, Suwon, Rep. of Korea
e-mail: siddiqi@oslab.khu.ac.kr, amtareen@ajou.ac.kr, tcchung@khu.ac.kr, sylee@oslab.khu.ac.kr

ABSTRACT

In this paper, a new recognition model, based on Hidden Conditional Random Fields (HCRFs), is proposed for human Facial Expression Recognition (FER) systems. At the first stage of our FER system, some well-known statistical techniques are used for both global and local feature extraction. These features are then fed to the HDRF-based recognition model at the final stage to recognize six human facial expressions. The proposed recognition model is tested on Cohn-Kanade database of facial expression videos. The training and testing is performed using one-leave-one-out cross-validation rule, which means that every facial expression is used for both training and testing. By means of a confusion matrix, it is shown that the proposed scheme has yielded improved recognition rate, with a mean recognition rate of 93%, when compared with some of the existing techniques. This shows the feasibility of using our HCRF-based recognition model for accurate human facial expression recognition.

Index Terms— PCA, ICA, LDA, HCRF

1. INTRODUCTION

Recognition module plays a significant role in determining the overall accuracy of a Facial Expression Recognition (FER) system. Many of the existing systems have focused on implementing new improved algorithms for face detection and feature extraction; however, most of them have failed or faced difficulties in recognition stage.

The authors in [1] used Artificial Neural Networks (ANNs) to detect different types of facial expression, and achieved an accuracy of 73%. However, neural networks are considered as black box and have incomplete capability to explicitly categorize possible fundamental relationships [2].

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2013 – 0030823).

This work (Grants No. 00048272) was also supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2011.

Gaussian Mixture Model (GMM) is another well-known recognition model used in many previous works to improve the accuracy of the recognition part of an FER system [3] and [4], but with very little success. A GMM model operates along similar principles to a Bayesian classifier. However, the likelihood function is not assumed to be a single Gaussian probability density. Instead, it is assumed to be of unknown shape and functional form and thus approximated by a weighted mixture of Gaussian functions. The weights and the parameters (centers and covariance) of the mixture components are calculated using the expectation-maximization (EM) algorithm.

Support Vector Machines (SVMs) was employed in FER systems by [5]–[9], with claims of achieving higher recognition accuracies. However, one drawback of SVMs is that they do not ensure good recognition accuracy if the numbers of samples are less than the number of the selected features. Moreover, there is no direct method for probability estimation in SVMs. Furthermore, SVMs simply disregard the temporal addictions among video frames thus each frame is supposed to be statistically independent from the rest. This confines their prediction ability, particularly when uncertainties exist in some video sequences, where it is complicated to recognize a suitable expression label for one frame until we have information of its temporal circumstance [10].

Another well-known classifier used in numerous previous FER systems, for sequential data, is Hidden Markov Model (HMM). It is known for producing high recognition accuracy. However, HMM has shown a higher degree of complexity in incorporating long-range dependencies between the states and observations. This is because, HMM is restrictive to model observations in a uniform way [11]. Furthermore, HMM is generative in nature and assumes the states and observations to be independent of each other. A limitation of generative model is that latent variables are assumed to be independently provided by the observations. Since, we want to incorporate long range dependencies in the model and allow hidden variables to depend on several local features, so specifying such a generative model is a challenging task [12]. Maximum entropy Markov model (MEMM) is a non-generative model. It was developed to overcome the limitations of HMM and showed good results [21]. Nevertheless, MEMM has a commonly known weakness called *label bias problem* [13]. In other words, it uses per-state normalization of transition scores, implying score conservation at each state.

Conditional Random Fields (CRFs) [13] and Hidden Conditional Random Fields (HCRFs) [12] are the generalizations of MEMM that are proposed to take advantages of MEMM and to solve the *label bias problem*. HCRF extends the capability of CRF with hidden states, making it able to learn hidden structure of sequential data. Both of them use global normalization instead of per-state normalization. Thus, they allow weighted scores, making the parameter spaces larger than those of MEMM and HMM. In order to inherit the advantages of HCRF model and completely tackle the limitations of the existing work, in this paper, we proposed the use of HCRF which is able to explicitly utilize mixture of full-covariance Gaussian distributions. We applied the proposed model to the publicly available facial expression data and compared the results with that obtained by HMM.

2. MATERIALS AND METHODS

2.1. Feature Extraction using Statistical Techniques

There lots of methods have been developed and validated for the purpose of feature extraction for FER systems, among them Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are widely used, and their performance has already been validated in [14]. Therefore, we decided to use PCA and ICA for feature extraction to extract both the global and local features respectively.

2.3. Linear Discriminant Analysis (LDA)

LDA maximizes the ratio of between-class variance to within-class variance in any particular data set, thereby guaranteeing maximal separability. LDA produces an optimal linear discriminant function that maps the input into the classification space on which the class identification of the samples is decided. LDA easily handles the case in which the within-class frequencies are unequal and their performances have been examined on randomly generated test data. For more details on LDA, please refer to [15].

2.4. Proposed Recognition Model

As mentioned earlier, the current HMM and Gaussian mixture HCRF models are not competent to utilize full distributions. These models do not guarantee the convergence of their parameters to some specific values, at which the conditional probability is modeled as a mixture of the normal density functions. To overcome these limitations, we explicitly included a mixture of Gaussian distributions in the feature functions, thus our feature functions could be described in the following forms

$$f_s^{Prior}(Y,\bar{S},X)=\delta(s_1=s)\forall s \tag{1}$$

$$f_{s,s'}^{Transation}(Y,\bar{S},X)=\sum_{t=1}^{T}\delta(s_{t-1}=s)\,\delta(s_t=s')\forall s,s' \tag{2}$$

$$f_s^{Observation}(Y,\bar{S},X)=\sum_{t=1}^{T}\log\left(\sum_{m=1}^{M}\Gamma_{s,m}^{Obs}N\left(x_t,\mu_{s,m},\Sigma_{s,m}\right)\right)\delta(s_t=s) \tag{3}$$

$$N\left(x,\mu_{s,m},\Sigma_{s,m}\right)=\frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_{s,m}|^{\frac{1}{2}}}\exp\left(-\tfrac{1}{2}(x-\mu_{s,m})'\Sigma_{s,m}^{-1}(x-\mu_{s,m})\right) \tag{4}$$

where M is the number of density function, D is the dimension of the observation and $\Gamma_{s,m}^{Obs}$ is the mixing weight of the m^{th} component with mean $\mu_{s,m}$ and covariance matrix $\sum_{s,m}$. As we can see in (3), Γ, μ, and $\sum$ can be updated during the training phase, hence we can set

$$\Lambda_{s,m}^{Obs}=1\forall s \tag{5}$$

As a result, the conditional probability can be rewritten as:

$$p\left(Y\mid X;\Lambda,\Gamma,\mu,\Sigma\right)$$

$$=\frac{\sum_{S}\exp\begin{pmatrix}\sum_{s}\Lambda_s^{Prior}f_s^{Prior}(Y,\bar{S},X)+\\ \sum_{s\,s'}\Lambda_{s\,s'}^{Transition}f_{s\,s'}^{Transition}(Y,\bar{S},X)+\\ \sum_{s}f_s^{Observation}(Y,\bar{S},X)\end{pmatrix}}{z(X;\Lambda,\Gamma,\mu,\Sigma)} \tag{6}$$

$$=\frac{\sum_{\bar{S}=S_1,S_2,\ldots\ldots,S_t}\exp\left(\Lambda_{S_1}^{Prior}+\sum_{t=1}^{T}\begin{pmatrix}\Lambda_{s_{t-1},s_t}^{Transition}+\\ \log\left(\sum_{m=1}^{M}\Gamma_{s_t,m}^{Obs}N\left(x_t,\mu_{s_t,m},\Sigma_{s_t,m}\right)\right)\end{pmatrix}\right)}{z(X;\Lambda,\Gamma,\mu,\Sigma)} \tag{7}$$

$$=\frac{Score\left(Y|X;\Lambda,\Gamma,\mu,\Sigma\right)}{z(X;\Lambda,\Gamma,\mu,\Sigma)} \tag{8}$$

Based on equations (7) and (8), we can compute the conditional probability by using the well-known forward and backward algorithm as:

$$\alpha_\tau(s)=\sum_{\bar{S}=S_1,S_2,\ldots\ldots,\{S_\tau=S\}}\exp\begin{pmatrix}\Lambda_{S_1}^{Prior}\\ +\sum_{t=1}^{T}\begin{pmatrix}\Lambda_{s_{t-1},s_t}^{Transition}+\\ \log\left(\sum_{m=1}^{M}\Gamma_{s_t,m}^{Obs}N\left(x_t,\mu_{s_t,m},\Sigma_{s_t,m}\right)\right)\end{pmatrix}\end{pmatrix}$$

$$=\sum_{s'}\alpha_{\tau-1}(s')\exp\left(\Lambda_{s's}^{Transition}\log\left(\sum_{m=1}^{M}\Gamma_{s,m}^{Obs}N\left(x_\tau,\mu_{s,m},\Sigma_{s,m}\right)\right)\right) \tag{9}$$

$$\beta_\tau(s)=\sum_{\bar{S}=\{S_\tau=S\},S_{\tau+1},\ldots\ldots,S_T}\exp\begin{pmatrix}\Lambda_{S_1}^{Prior}\\ +\sum_{t=1}^{T}\begin{pmatrix}\Lambda_{s_{t-1},s_t}^{Transition}+\\ \log\left(\sum_{m=1}^{M}\Gamma_{s_t,m}^{Obs}N\left(x_t,\mu_{s_t,m},\Sigma_{s_t,m}\right)\right)\end{pmatrix}\end{pmatrix}$$

$$= \sum_{s'} \beta_{\tau+1}(s') \exp\left(\Lambda^{Transition}_{s's} \log\left(\sum_{m=1}^{M} \Gamma^{Obs}_{s,m} N\left(x_\tau, \mu_{s,m}, \Sigma_{s,m}\right)\right)\right) \quad (10)$$

$$Score(Y \mid X; \Lambda, \Gamma, \mu, \Sigma) = \sum_{S} \alpha_T(s)$$

$$= \sum_{S} \beta_1(s) \quad (11)$$

In the training phase our goal has been to find the parameters (Λ, Γ, μ, and ∑) to maximize the conditional probability of the training data due to which best accuracy of classification was achieved.

3. RESULTS AND DISCUSSION

In this research we have tested the idea of employing HCRFs for human facial expressions recognition. The tests are found to be successful and we have achieved significant improvements in the recognition rate.

In order to evaluate and validate the proposed model, a publicly available dataset [16] of facial expressions has been utilized. Six different types of expressions were used from this dataset, namely: “Happy”, “Sad”, “Surprise”, “Disgust”, “Anger”, and “Fear”. Each expression was performed by a different subject/person. Mostly, the image data in this dataset of facial expressions display the frontal view of the face and each expression is composed of several sequences of expression frames.

The performance of the proposed recognition model has been validated by comparing it against the existing systems. In order to demonstrate the recognition rate of six universal facial expressions, the proposed model has been trained and tested on the abovementioned dataset of facial expression based on one-leave-one-out cross validation rule. The size of frame in the experiments was set to 60 x 60.

As mentioned earlier, PCA and ICA were employed in order to calculate both the global and local features. However, most of these feature were merged with each other in the feature space, therefore, a linear classifier, LDA, was used for dimension reduction and better classification.

After feature extraction, Mel-frequency cepstral coefficients (MFCCs) were extracted at the recognition stage. Then the training and testing were performed based on the one-leave-one-out cross-validation rule, and then the classification experiments were performed using the proposed recognition model (HCRF).

At the initial step, the experiments were run using HMM with different number of states and Gaussian mixtures. The states of HMM that produced the most accurate classification were applied for training and evaluating the proposed model. The average classification rates based on the one-leave-one-out cross-validation rule are shown in Table I. Furthermore, the results of the proposed recognition model on four facial expressions are also shown in Figure 1.

TABLE I. CONFUSION MATRIX OF THE PROPOSED MODEL USING COHN-KANADE DATABASE OF FACIAL EXPRESSIONS (UNIT: %)

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**92.6**	0	1	0	6.4	0
Sad	0	**93.6**	5	1.4	0	0
Anger	0	3.7	**94.3**	2	0	0
Disgust	0	0	8	**92**	0	0
Surprise	0	0	0	5	**95**	0
Fear	0	2	7	0	0	**91**
Average	**93.08**					

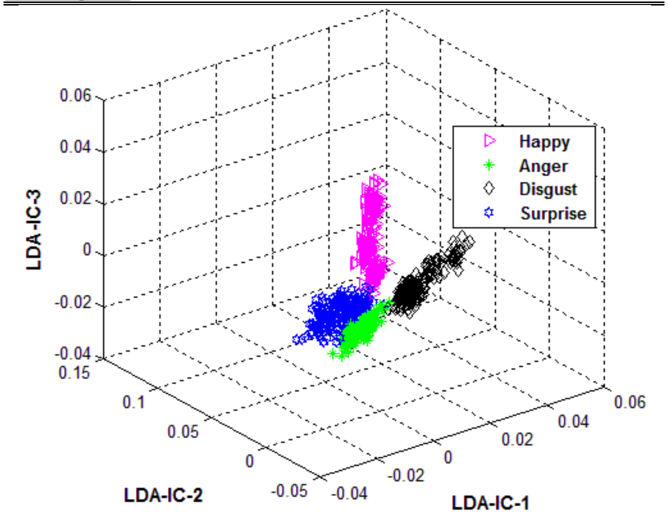

Figure 1. 3D-feature plot of the proposed model for four different types of facial expressions after LDA.

It is to be noted from Figure 1 that proposed model succeeded in achieving a high class separation in the feature space which helped the system in achieving a high recognition rate. The proposed system was further compared with some of the existing works that used the same database of facial expressions for their experiments. The comparison results of the proposed model against these existing works [17]–[20] are shown in Figure 2.

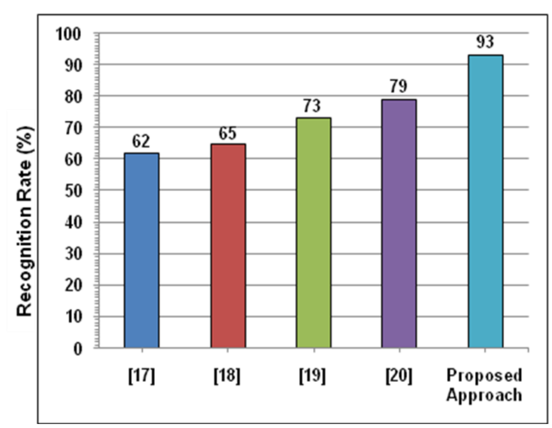

Figure 2. Comparison of the proposed recognition model with some of the existing works.

Figure 2 indicates that the proposed model is much better than those of the existing works. This improvement in accuracy could be attributed to the use of HCRFs that solved the limitations of the CRFs and HMM. The proposed recognition model works well for frontal view of facial expressions and also extracts the hidden details between the two transition states to calculate the likelihoods.

As said earlier, the training and testing of the proposed model is performed using the one-leave-one-out cross-validation rule, which means that every facial expression was used for both training and testing. The current implementation was done in Matlab, on a dual-core Pentium processor 2.5 GHz with 3 GB RAM, it takes very small amount of time for classification of human facial expressions.

4. CONCLUSION

The analysis of the human facial expressions has become a dynamic and challenging research area for applications like communication and psychology over the past few years.

In this research, a new recognition model has been developed for the sake of highly accurate facial expression recognition. Well know statistical methods, such as PCA and ICA were employed for feature extraction. Furthermore, a linear classifier, LDA, was used for dimension reduction and class separation. Finally, the proposed model was utilized for recognition. The proposed FER system achieved 93% recognition rate when applied on Cohn-Kanade dataset of facial expressions. The results of the proposed recognition model showed a significant improvement in recognition when compared with some of the existing techniques applied on the same dataset.

Most of the expressions in Cohn-Kanade dataset are in frontal views. In future, we are looking forward to work on expression recognition based on side views which are not covered in this research. So, further research work is required to modify the proposed model and test it for side views of facial expressions for better performance.

5. REFERENCES

[1] M. Gargesha and P. Kuchi, “Facial expression recognition using artificial neural networks,” *Artif. Neural Comput. Syst.*, pp. 1– 6, 2002.

[2] J.V. Tu, “Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes,” *J. Clin. Epidemiol.*, vol. 49, pp. 1225 – 1231, 1996.

[3] W. Liu, J. Lu, Z. Wang, and H. Song, “An expression space model for facial expression analysis,” Proc. *Congress on Image and Signal Processing,* vol. 2, pp. 680 – 684, 2008.

[4] M. Schels, and F. Schwenker, “A multiple classifier system approach for facial expressions in image sequences utilizing GMM supervectors,” *Proc. International Conference on Pattern Recognition (ICPR),* pp. 4251 – 4254, 2010.

[5] M. Kabir, T. Jabid, and O. Chae, “Local Directional Pattern Variance (LDPv): A Robust Feature Descriptor for Facial Expression Recognition,” *Int. Arab J. Info. Tech*., vol. 9, no. 4, pp. 382 – 391, 2012.

[6] T. Jabeed, M.H. Kabir, and O. Chae, “Robust Facial Expression Recognition Based on Local Directional Pattern,” *ETRI Journal*, vol. 32, no. 5, pp. 784−794, 2010.

[7] 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.

[8] G.U. Kharat and S.V. Dudul, “Human emotion recognition system using optimally designed SVM with different facial feature extraction techniques,” *WSEAS Transactions on Computers*, vol. 7, no. 6, pp. 650 – 659, 2008.

[9] C. Shan., S. Gong, and P. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” *Image and Vision Computing*, vol. 27, no. 6, pp. 803 – 816, 2009.

[10] Q. Shi., L.Wang, L. Cheng., and A. J. Smola. “Discriminative human action segmentation and recognition using semi-Markov model.” *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1 – 8, 2008.

[11] A. Gunawardana, M. Mahajan, A. Acero, and J. C. Platt. Hidden conditional random fields for phone classification. In *INTERSPEECH*, 2005.

[12] A. Quattoni, S. Wang, L. Morency, M. Collins, and T. Darrell, “Hidden Conditional Random Fields*,” IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1848 – 1852, 2007.

[13] J. Lafferty, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data”, *Proc. of the 18th International Conference on Machine Learning*, pp. 282 – 289, 2001.

[14] M. Z. Uddin, J. J. Lee, and T.-S. Kim, “An enhanced independent component-based human facial expression recognition from video,” *IEEE Transactions on Consumer Electronics*, vol. 55, no. 4, pp. 2216 – 2224, 2009.

[15] P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711 – 720, 1997.

[16] J. F. Cohn, A. Zlochower, J. Lien, and T. Kanade, “Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding,” *Psychophysiology*, vol. 36, pp. 35 – 43, 1999.

[17] S. M. Lajevardi, and Z. M. Hussain, “Facial expression recognition: Gabor filters versus higher-order correlators,” *Proc. Int. Conf. on Communication, Computer and Power (ICCCP)*, pp. 354 – 358, 2009.

[18] S. M. Lajevardi, and Z. M. Hussain, “Higher Order Orthogonal Moments for Invariant Facial Expression Recognition,” *Digital Signal Processing*, vol. 20, no. 6, pp. 1771 – 1779, 2010.

[19] S. S. Kulkarni, N. P. Reddy, and S. I. Hariharan, “Facial Expression (Mood) recognition from facial images using Committee Neural Networks,” *Biomedical Engineering online*, vol. 8, no. 16, pp. 1 – 12, 2009.

[20] S. M. Lajevardi, and Z. M. Hussain, “Feature Extraction for Facial Expression Recognition based on Hybrid Face Regions,” *Advances in Electrical and Computer Engineering*, vol. 9, no. 3, pp. 63 – 67, 2009.

[21] Yohan Kim. Application of maximum entropy markov models on the protein secondary structure predictions. Technical report, *University of California*, 2001.