FACIAL ATTRACTIVENESS PREDICTION USING PSYCHOLOGICALLY INSPIRED CONVOLUTIONAL NEURAL NETWORK (PI-CNN)

Jie Xu[1], *Lianwen Jin*[1,∗], *Lingyu Liang*[1,2,∗], *Ziyong Feng*[1], *Duorui Xie*[1], *Huiyun Mao*[1]

[1] South China University of Technology, Guangzhou, China
[2] The Chinese University of Hong Kong, Shatin, Hong Kong

ABSTRACT

This paper proposes a psychologically inspired convolutional neural network (PI-CNN) to achieve automatic facial beauty prediction. Different from the previous methods, the PI-CNN is a hierarchical model that facilitates both the facial beauty representation learning and predictor training. Inspired by the recent psychological studies, significant appearance features of facial detail, lighting and color were used to optimize the PI-CNN facial beauty predictor using a new cascaded fine-tuning method. Experiments indicate that the cascaded fine-tuned PI-CNN predictor is robust to facial appearance variances, and obtains the highest correlation of 0.87 in the SCUT-FBP benchmark database, which is superior to the related hand-designed feature and related deep learning methods.

***Index Terms*—** Facial attractiveness prediction, facial beauty analysis, convolutional neural network, deep learning

1. INTRODUCTION

Facial attractiveness has allured humans for centuries. Psychology research revealed that facial attractiveness perception of a human is influenced by several factors, like facial averageness, symmetry, lighting, smoothness and color [1–3]. Although a universal definition of facial beauty remains elusive, studies indicate that facial attractiveness can be learned by a machine using data-driven methods [4, 10].

Recently, facial attractiveness prediction has become a significant problem of facial beauty analysis [11], and has led to many interesting studies in computer vision and machine learning communities [4–12]. Based on the data-driven learning methods, automatic facial attractiveness predictors can be build for various useful applications, such as recommendation system [15], content-based image retrieval [16], face beautification [17] and face editing [18].

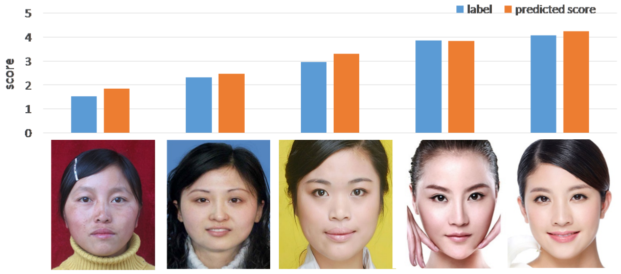

Fig. 1. Input faces and their facial beauty score predicted by our deep model. The score ranges between [1, 5], where the large score means the face is more attractive. The results indicate the PI-CNN facial beauty predictor is consistent to human perception.

However, it is not trivial to achieve automatic facial attractiveness prediction that is consistent to human perception. The challenge is twofold. Firstly, large facial appearance variance and the complexity of human perception make it difficult to construct an effective and robust beauty assessment model. Secondly, most of face benchmark databases were originally designed for face recognition problems, and they may not be suitable for attractiveness prediction. To address the latter problem, Xie et al. proposed a benchmark database, called SCUT-FBP, for facial beauty assessment and analysis [14]. Based on the SCUT-FBP database, this paper aims to propose an learning-based method to achieve automatic facial attractiveness prediction.

Facial attractiveness prediction can be formulated from a supervised learning perspective, and previous studies addressed the problem using different learning models with hand-designed facial feature [4, 7–10]. Kagian et al. experimented with the geometric and appearance facial feature with several models for attractiveness prediction, like linear regression and LS-SVM [4]. Chiang et al. extracted 3D facial features using a 3dMD scanner to train an adjusted fuzzy neural network, and high accuracy was achieved [6]. Kalayci et al. used dynamic features obtained from video clips with static facial features to build the attractiveness predictor [8]. Yan proposed a cost-sensitive ordinal regression model with

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61472144 and Grant 61502176; in part by GDSTP under Grant 2015B010101004, Grant 2015B010130003, Grant 2015B010131004; and in part by the National Key Research & Development Plan of China under Grant 2016YFB1001405. ∗ Corresponding authors: L. Jin (lianwen.jin@gmail.com); L. Liang (lianglysky@gmail.com)

978-1-5090-4117-6/17/$31.00 ©2017 IEEE

ranked labels for beauty assessment [10]. Previous studies indicate that both the facial representation and the prediction model are critical for facial beauty assessment. Since the heuristic hand-designed feature is difficult to optimize for the facial beauty predictors, it would be significant to learn the facial attractiveness representation adapted to the predictor.

Due to the recent success of deep neural network (DNN) for many high-level recognition problems [19, 20], we introduce the DNN models to address the facial beauty representation learning and prediction problem. Gan et al. trained a SVM regression predictor using the facial feature extracted by a deep self-taught learning method with two-layer convolutional deep belief networks (CBDN) [12]. Wang et al. built pairs of auto-encoders for attractive and non-attractive faces with different types of visual descriptors, and used the rank-minimized late fusion scheme to perform beauty/not-beauty prediction [13]. However, previous methods took the facial beauty representation learning and predictor training in two separated processes, which may fail to achieve the optimal performance. It is significant to obtain the facial beauty representation and the predictor under a whole learning and fine-tuning process.

In this paper, we propose a psychologically inspired convolutional neural networks (PI-CNN) to achieve automatic facial beauty prediction that is consistent to human perception. The beauty scores predicted by a human and our PI-CNN system are shown in Fig. 1. Different from the previous DNN, the PI-CNN is a hierarchical model that facilitates both the facial beauty representation learning and predictor training. Inspired by the recent psychological studies [1–3], significant appearance features of facial detail, lighting and color were extracted to build the PI-CNN facial beauty predictor. We used the facial detail feature obtained by a adaptive WLS edge-preserved smoothing filter [22] to pre-train the model, and propose a cascaded fine-tuning method with lighting and color feature to optimize the whole performance. Experiments illustrate that our PI-CNN predictor with cascaded fine tuning obtains the highest correlation of 0.87 in the SCUT-FBP database [14], and robust to facial appearances variances in JAFFE [24] and GT [25] face database.

In summary, the contributions of this paper include: (i) a deep CNN-based facial attractiveness predictor, which is consistent to human beauty perception and robust to different facial variances; (ii) a cascaded tuning method, which effectively improves the performance of the deep CNN model with psychologically inspired facial feature.

2. PI-CNN FOR FACIAL BEAUTY PREDICTION

The goal of the PI-CNN is to achieve representation learning and model training within a whole deep CNN model for facial beauty prediction. To optimize the performance of the deep model, the facial features that psychologically effect the perception of facial beauty are extracted to adjust the model parameters following a cascaded fine-tuning scheme.

2.1. The Basic CNN Architecture of Beauty Prediction

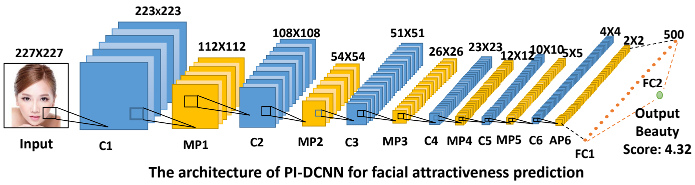

Fig. 2. The basic CNN architecture of facial attractiveness predictor, where ’C’ is the convolution layer; ’MP’ is the max-pooling layer; ’AP’ is the average pooling layer; and ’FC’ is the full connection layer.

The construction of PI-CNN is based on the convolution neural networks (CNN), which have obtained the state-of-the-art in many high level image recognition problems [19, 20].

The basic CNN of beauty prediction consists of three typical layers: convolution layer (C-layer), pooling layer (P-layer) and full connection layer (FC-layer), as shown in Fig. 2. In C-layers, several convolutions perform in parallel to produce a set of pre-synaptic activations, and then each pre-synaptic activation is run through a nonlinear activation function. In P-layers, a pooling function (like max-pooling) is used to modify the output of the C-layer. The C-layers and P-layers help to make the facial representation become invariant to small translations of the input, which is inspired by the visual cortex structure of human vision [19]. Finally, the FC-layers are on the top of the network to produce the output score of facial beauty prediction.

Specifically, let $\mathbf{c}_k$ and $\mathbf{h}_k$ are the outputs of k-th C-layer and P-layer, respectively; $\mathbf{h}_{FC}$ is the output of FC-layer. The input-output functions of the PI-CNN beauty predictor $f_{\boldsymbol{\theta}}(\mathbf{x})$ with multiple C-, P-, FC-layers can be formulated as:

$$\begin{cases} \mathbf{c}_k = \Phi(\mathbf{b}_k + \mathbf{W}_k \mathbf{h}_{k-1}), \;\; k = 1, 2, ..., K, \\ \mathbf{h}_k = P_k(\mathbf{c}_k), \\ \mathbf{h}_{FC2} = \mathbf{b}_{FC1} + \mathbf{W}_{FC1} \mathbf{h}_K, \\ f_{\boldsymbol{\theta}}(\mathbf{x}) = \mathbf{b}_{FC2} + \mathbf{W}_{FC2} \mathbf{h}_{FC2}(\mathbf{x}). \end{cases} \tag{1}$$

where $\mathbf{h}_0 = \mathbf{x}$ is the input of the basic CNN model; $\Phi(a)$ is the non-linear activation functions of C-layer, and here we use the rectified linear unit (ReLU) in PI-CNN, i.e. $\Phi(a) = max(0, a)$; P_k is the pooling operations of k-th P-layer, where PI-CNN performs average pooling in the K-th P-layer and max-pooling in the others. The whole PI-CNN architecture is illustrated in Fig. 2. In this paper, the depth of the PI-CNN for the C-, P-layers is set to $K = 6$.

The model parameters are $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\omega}\}$, where $\boldsymbol{\beta} = \{\mathbf{b}_k, \mathbf{b}_{FC1}, \mathbf{b}_{FC2}\}$ contain the offset parameters and $\boldsymbol{\omega} = \{\mathbf{W}_k, \mathbf{W}_{FC1}, \mathbf{W}_{FC2}\}$ contain the weight matrixs. During training, we minimize the cost function:

$$J(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}||\mathbf{y}^{(n)} - f_{\boldsymbol{\theta}}(\mathbf{x}^{(n)})||^2 + \lambda||\boldsymbol{\omega}||^2, \qquad (2)$$

where $(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})$ is the n-th training $(feature, score)$ pair. The scalar λ controls the weight decay that penalizes the squared norm of the weights $\boldsymbol{\omega}$, and $\lambda = 0.0005$ is set in this paper. Since the non-linear hierarchical model has many local minimums [20], we used stochastic gradient descent to train the model with a dropout technique [21] to avoid over-fitting.

2.2. PI-CNN with Cascaded Fine-Tuning

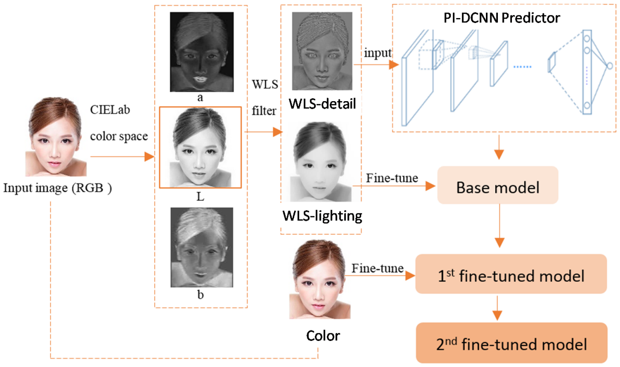

Fig. 3. The cascaded fine-tuning process of PI-CNN using the facial feature extracted by the adaptive WLS filter [21].

Recent psychology studies indicates that the skin color, smoothness, and lighting are three significant factors influencing the perception of facial beauty [1–3]. It inspires us to construct and improve the PI-CNN using a cascaded fine-tuning method with these facial features.

In the PI-CNN, we used RGB channels as the color feature, and extracted the detail and lighting features by an edge-aware filter. Specifically, we used the adaptive WLS filter of Liang et al. [22] to decompose the input into separated layers of facial lighting and detail feature, i.e. WLS-lighting and WLS-detail. Then, we pre-train the PI-CNN predictor using WLS-detail, and sequentially fine-tune the pre-trained model with WLS-lighting and RGB, as shown in Fig. 3.

3. EXPERIMENTS AND ANALYSIS

We evaluated the performance of PI-CNN beauty predictor on the recently proposed SCUT-FBP benchmark database [14], which is specifically designed for facial beauty perception. The SCUT-FBP database contains 500 Asian female faces with beauty scores ($score \in [1, 5]$) labeled by 75 raters. In the experiments of this paper, 400 samples were randomly selected as the training set, while the rest is the testing set. The ground-truth is the average of the 75 scores. Following the previous studies [7, 10, 11], the Pearson Correlation [26] is used to evaluate the performance between the ground-truth and the predicted result. To reduce the sample variances, all the experiments were performed in 5-folds cross validation [27].

3.1. Evaluations of PI-CNN

We evaluate the effectiveness of PI-CNN for facial beauty prediction with different model architecture and different fine-tuning schemes. Experiments indicate that PI-CNN with deeper structure and cascaded fine-tuning obtained better performance. The results also show that PI-CNN is robust to facial variance, like pose and expression.

Evaluations of Basic CNN Architecture. We constructed three CNN beauty predictor (without tuning, RGB image as input) according to Eq. (1) with network depth of $K = 3$ (CNN-3), $K = 5$ (CNN-5) and $K = 6$ (CNN-6), respectively. In the current settings, we found that the CNN tends to overfit when $K > 6$. The results of 5-folds cross validations are shown in Table 1, which indicate that properly increasing the numbers of C-,P- layers can improve the facial beauty prediction.

Table 1. Comparisons of three basic CNN architectures with different depths. The results are the Pearson Correlation between the predicted and the ground-truth facial beauty score.

Test Set	1	2	3	4	5	Average
CNN-3	0.77	0.77	0.75	0.77	0.70	0.75
CNN-5	0.84	0.80	0.76	0.79	0.75	0.79
CNN-6	0.86	0.85	0.85	0.79	0.80	0.83

Evaluations of PI-CNN Cascaded Fine-Tuning. To evaluate the effectiveness of our cascaded fine-tuning with different facial feature for PI-CNN, we perform different fine-tuning schemes with different input feature, including eigenface, LBP, Gabor and WLS-detail, as shown in Fig. 4. The results illustrate that cascaded fine-tuning can effectively improve the accuracy of the deep model without fine-tuning for beauty prediction.

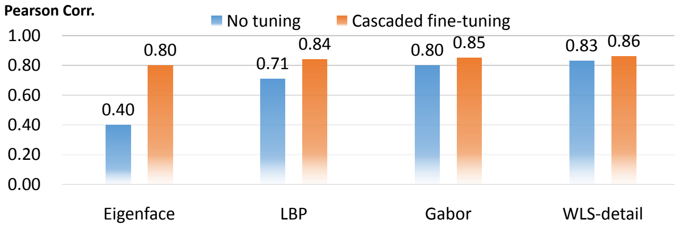

Fig. 4. Evaluations of PI-CNN cascaded fine-tuning with different input features, including Eigenface, LBP, Gabor, and WLS-detail.

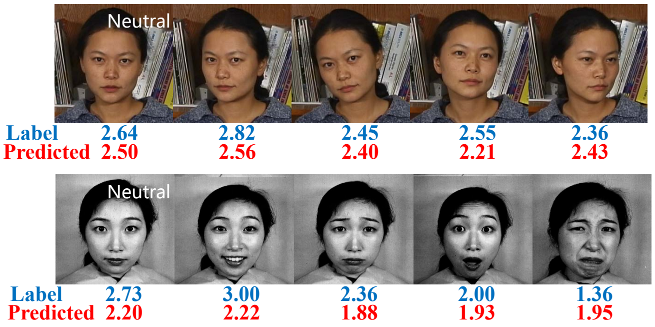
Neutral
Label 2.64 2.82 2.45 2.55 2.36
Predicted 2.50 2.56 2.40 2.21 2.43
Neutral
Label 2.73 3.00 2.36 2.00 1.36
Predicted 2.20 2.22 1.88 1.93 1.95

Fig. 5. Robustness of PI-CNN beauty predictor to the facial variances of pose and expression. Blue: the ground true beauty label by human. Red: the score predicted by PI-CNN.

Robustness to Different Facial Variances. Since the C-, P-layers of PI-CNN help to obtain the facial representation that is invariant to small translations of the input, we evaluate the sensitiveness of the beauty predictor under facial variance of pose and expression, as shown in Fig. 5. The faces with pose variances were taken from the Japanese Female Facial Expression (JAFFE) database, while those with expression variances were taken from the Georgia Tech (GT) face database [25]. The beauty score of the test faces were labelled according to the same setting as the SCUT-FBP benchmark database [14].

The results show that the predicted beauty scores for faces with pose and expression variances are not only consistent to human label, but also consistent to the faces with neutral pose and expression, which indicate the robustness of PI-CNN.

3.2. Comparisons with Related Methods

We compared our method with the traditional shallow regression model and the related deep learning models, respectively. The results show that the PI-CNN method obtain the highest correlation 0.87 among all the methods, which illustrates the effectiveness of the cascaded fine-tuned PI-CNN predictor.

Comparisons with Shallow Learning Methods. We compared the methods using hand-crafted feature with traditional shallow regression and our PI-CNN method, as shown in Table 2. The results illustrate that PI-CNN is superior to the other regression model with the same input feature.

Comparisons with Related Deep Learning Methods. We also compared the PI-CNN with the related deep learning methods, including multi-layer perception (MLP), neuron networks with radial basis function kernels (RBFnet), and the recently proposed PCAnet [23]. The results of 5-folds cross validation in each test set are shown in Fig. 6, which indicate the effectiveness of PI-CNN for the representation learning and predictor training.

We also compared our PI-CNN model with Gan et al. [12]. According to the setting of [12], CDBN model was used to learn the facial feature with Gabor as input, and the SVM beauty predictor was trained by the CDBN-based feature. The results show that Gan's CDBN-based feature obtain 0.83 Pearson correlation in the training set, but its performance drops to 0.49 in the testing set of SCUT-FBP. It seems that the feature learned by the separated CDBN-based model fails to adapt to the predictor well in some cases. In contrast, our PI-CNN simultaneously optimizes the facial beauty representation and predictor.

Table 2. Comparisons between shallow regression models and PI-CNN with different input feature. 'LR' is Linear Regression; 'GPR' is Gaussian Process Regression; 'SVR' is Support Vector Regression. Note that the deep models (CNN and PI-CNN) here were trained without fine-tuning to indicate the effectiveness of the deep structure.

Model	LR	GPR	SVR	CNN	PI-CNN
Gabor	0.72	0.69	0.75	0.80	0.85
Eigenface	0.11	0.13	0.16	0.40	0.80
LBP	0.30	0.29	0.28	0.71	0.84
WLS-detail	0.63	0.64	0.57	0.83	0.86

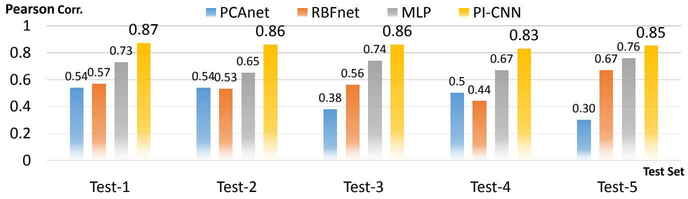

Fig. 6. Comparisons between the related deep learning methods (PCAnet, RBFnet and MLP) and the fine-tuned PI-CNN in 5-folds cross validation.

4. CONCLUSIONS

This paper addresses the facial attractiveness prediction problem using deep learning. A psychologically inspired deep convolutional neural networks (PI-CNN) is proposed, which facilitates both the facial beauty representation learning and predictor training. We used a new cascaded fine-tuning method to further improve the performance of PI-CNN facial beauty predictor with facial features of detail, lighting and color. Experiments indicate that the cascaded fine-tuned PI-CNN predictor obtains the highest correlation of 0.87 in the benchmark database, which is superior to the related hand-designed feature with shallow regressors and related deep learning methods.

Experiments in this paper were based on a relatively small SCUT-FBP database [14]. To tackle with the benchmark evaluation problem [11], we would design a large scale benchmark database for facial beauty analysis in the future work.

5. REFERENCES

[1] R. Russell, “Sex, beauty, and the relative luminance of gacial features,” *Perception*, vol. 32, no. 9, pp. 1093-1108, 2003.

[2] G. Rhodes, “The evolutionary psychology of facial beauty,” *Annu. Rev. Psychol.*, vol. 57, pp. 199-226, 2006.

[3] I. Stephen, M. Law Smith, M. Stirrat, and D. Perrett, “Facial skin coloration affects perceived health of human faces,” *International Journal of Primatology*, vol. 30, no. 6, pp. 845-857, 2009.

[4] A. Kagian, G. Dror, T. Leyvand, D. Cohen-Or, and E. Ruppin, “A Humanlike Predictor of Facial Attractiveness,” *Proc. of NIPS*, pp. 649-656, 2006.

[5] H. Mao, L. Jin and M. Du, “Automatic classification of Chinese female facial beauty using Support Vector Machine,” *Proc. of IEEE SMC*, pp. 4842-4846, 2009.

[6] D. Zhang, Q. Zhao and F. Chen, “Quantitative analysis of human facial beauty using geometric features,” *Pattern Recognition*, vol. 44, no. 4, pp. 940-950, 2011.

[7] W. Chiang, H. Lin, C. Huang, L. Lo, and S. Wan, “The cluster assessment of facial attractiveness using fuzzy neural network classifier based on 3D Moir features,” *Pattern Recognition*, vol. 47, no. 3, pp. 1249-1260, 2014.

[8] S. Kalayci, H. K. Ekenel, and H. Gunes, “Automatic analysis of facial attractiveness from video,” *Proc. of ICIP*, pp. 4191-4195, 2014.

[9] J. Fan, K. P. Chau, X. Wan, L. Zhai and E. Lau, “Prediction of facial attractiveness from facial proportions,” *Pattern Recognition*, vol. 45, pp. 2326-2334, 2012.

[10] H. Yan, “Cost-sensitive ordinal regression for fully automatic facial beauty assessment,” *Neurocomputing*, no. 129, pp. 334-342, 2014.

[11] D. Zhang, F. Chen and Y. Xu, *Computer Models for Facial Beauty Analysis*, Springer International Publishing Switzerland, 2016.

[12] J. Gan, L. Li, Y. Zhai and Y. Liu, “Deep self-taught learning for facial beauty prediction,” *Neurocomputing*, no. 144, pp. 295-303, 2014.

[13] S. Wang, M. Shao and Y. Fu, “Attractive or not? Beauty prediction with attractiveness-aware encoders and robust late fusion,” *ACM Multimedia*, pp. 805-808, 2014.

[14] D. Xie, L. Liang, L. Jin, J. Xu and M. Li, “SCUT-FBP: A Benchmark Dataset for Facial Beauty Perception,” *Proc. of IEEE SMC*, pp. 1821-1826, 2015.

[15] L. Liu, J. Xing, S. Liu, H. Xu, X. Zhou, and S. Yan, “Wow! you are so beautiful today!,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 11, no. 1s, p. 20, 2014.

[16] N. Murray, L. Marchesotti L, and F. Perronnin, “AVA: A large-scale database for aesthetic visual analysis,” *Proc. of CVPR*, pp. 2408-2415, 2012.

[17] L. Liang, L. Jin, and X. Li, “Facial skin beautification using adaptive region-aware mask,” *IEEE Trans. on Cybernetics*, vol. 44, no. 12, pp. 2600-2612, 2014.

[18] L. Liang, L. Jin, X. Zhang, and Y. Xu, “Multiple facial image editing using edge-aware PDE learning,” *Comput. Graph. Forum*, vol. 34, no. 7, pp. 203–212, Oct. 2015.

[19] Y. LeCun, Y. Bengio and G. Hinton, “Deep learning,” *Nature*, vol. 251, pp. 436-444, 2015.

[20] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.

[21] S. Nitish, C. G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929-1958, 2014.

[22] L. Liang and L. Jin, “A New Face Relighting Method Based on Edge-Preserving Filter,” *IEICE Trans. Information and Systems*, vol. E96-D, no. 12, pp. 2904-2907, 2013.

[23] T-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng and Y. Ma, “PCANet: A Simple Deep Learning Baseline for Image Classification?,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017-5032, 2014.

[24] M. Lyons, S. Akemastu, M. Kamachi and J. Gyoba, “Coding Facial Expressions with Gabor Wavelets,” *3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200-205, 1998.

[25] L. Chen, H. Man and A. Nefian, “Face recognition based multiclass mapping of Fisher scores,” *Pattern Recognition*, vol. 38, no. 6, pp. 799-811, 2005.

[26] B.H. Cohen, *Explaining psychological statistics*, John Wiley & Sons, 2008.

[27] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” *Proc. of IJCAI*, pp. 1137-1145, 1995.