A Framework of Human Emotion Recognition Using Extreme Learning Machine

Conference Paper · August 2014

DOI: 10.1109/ICAICTA.2014.7005961

3 authors:

Prasetia Utama
Yokohama National University

Widodo Widodo
Jakarta State University

Hamidillah Ajie
Jakarta State University

A Framework of Human Emotion Recognition Using Extreme Learning Machine

Prasetia Utama
Universitas Negeri Jakarta
Jakarta, Indonesia
Email: prasetiaUtama@outlook.com

Widodo
Universitas Negeri Jakarta
Jakarta, Indonesia
Email: widodo03@yahoo.com

Hamidillah Ajie
Universitas Negeri Jakarta
Jakarta, Indonesia
Email: hamidillah@yahoo.com

***Abstract*—Human emotion recognition has been challenging issue in field of human-computer interaction. In order to form an interaction that is more natural between human and computer, the computer should be able to discern and respond to human emotion. In this paper, an approach for recognizing human emotion is proposed. The proposed approach operates HAAR-classifier to detect mouth, eyes, and eyebrow on face, and, to extract features from them, it uses Gabor wavelet. Before classifying the features, PCA is performed to reduce its dimension. The proposed framework employs SLFNs with ELM as its learning algorithm to classify the features. In this experimental, the proposed framework is tested in two cases, personalize and generalize face case, with ten subjects expressing six basic emotions and neural state. The robustness of ELM is evaluated with comparing it to K-NN and SVM. Preliminary experiment shows that the proposed approach has promising performance in personalize face case.**

***Keywords*—Human Emotion Recognition, Extreme Learning Machine, Gabor Wavelet, Human Computer Interaction, SLFNs**

I. INTRODUCTION

Interaction between human and computer has been raising as high as requirement of computer in human's life. Humans need computer more than a stuff to assist them in their job, they expect a stuff that has ability to create natural interaction as if an interaction among human being. Hence, a computer should be able to response to human's emotion, since emotion is a reaction from human in adapting immediately to environment. There are six basic emotions, angry, disgust, fear, happy, sad, and surprise, in which each individual expresses in identical way [1]. According to Darwin [2] , emotion of human can be recognized through their facial expression, because it is one of expressing behavior that is directly correlated to human's response to emotion.

Researchers have undertaken research in human's emotion for decades, and in computer vision, the issue has begun to collect interest from 1990s [3]. Various methods have been proposed by experts to address this issue. Sawada and Samad [4] proposed a framework employing Gabor wavelet and Support Vector Machine to recognize emotion of human through facial expression. Sarode dan Bhatia [5] introduced a recognition system operating edge projection and geometrical relationship. Quraishi, et al. [6] presented a framework using Sobel edge detection and MLP with Back Propagation. In recent study, Uar, et al. [7] proposed an approach performing OSELM and curvelet transform; the approach has promising result tested on JAFFE and Cohn-Kanade database. One of the limitations suffered by those methods is they have unclear result when tested in the case of limited features attribute, since most of them use features from a whole of face to recognize the emotion.

In this paper, a framework employing Gabor wavelet [8] in features extraction and SLFNs with Extreme Learning Machine (ELM) in features classification is proposed to cope with the case of limited feature attribute; it should be noted that this proposed framework employs ELM for Regression and Multiclass Classification that proposed in 2012 [9]. According to Marcelja [10] and Daugman [11], Gabor function is able to represent simple cell of mammal's visual cortex . Moreover, Gabor wavelet has both multi-resolution and multi-orientation properties that are optimal to measure local spatial frequencies [12]. ELM is selected as learning algorithm for SLFNs because of its less time-consumption and generalization performance; tested with iris database in multiclass case, ELM with sigmoid additive node reaches 97.6% in the testing phase, 3% higher than SVM and 1% than LSSVM. Moreover, the training time of ELM is faster 97% than SVM and 5% than LSSVM [9]. The essential point of ELM is learning without iterative tuning, as a result, the time-consumption in training process of ELM is less-far than conventional method e.g. Back Propagation. Furthermore, the ability of ELM to produce minimum norm weights induces better generalization in learning performance.

In this experimental, eyes, eyebrow, and mouth are selected as parameters to determine human's emotion, as they produce unique cues in each emotion [13]. They are detected using HAAR-classifier [14], and Gabor wavelet is employed to extract features from them. PCA [15] is used to reduce the dimension of features, 132 features are retained to be fed into SLFNs. In the experiment, the dataset are taken from ten subjects expressing six basic emotions and neural state. The experiment divided in two cases, generalized and personalized face case. The performance of SLFNs with ELM is compared with K-NN [16] and SVM [17] to evaluate its performance in those cases.

978-1-4799-5100-0/14/$31.00 ©2014 IEEE

II. Overview

This section briefs the methods that are used in the proposed framework for extracting and classifying the features of eyes, eyebrow, and mouth. Before extracting feature from them, the proposed approach detects them using HAAR-classifier; on the pre-processing stage, face image is converted into gray-scale, and histogram equalization is applied on it. After eyes and mouth regions are detected, they are cropped automatically with uniform size, eyes are cropped in height of 35 pixels and width of 84 pixels, while mouth in height of 20 pixels and width of 48 pixels.

A. Edge-detection using Gabor wavelet

Gabor wavelet is linear filtered that mainly used in Image Processing as method for extracting feature. Marcelja [10] and Daugman [11] stated that Gabor functions can model simple cells of mammal's visual cortex. In an experiment conducted by Sawada and Samad [4], edge on eyes and mouth produced by filtering using Gabor wavelet produces unique pattern, implemented in case of human emotion recognition. A Gabor wavelet (kernels) can be definened as [18]

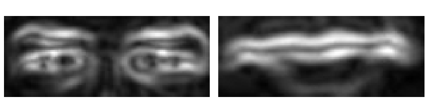
$$\psi_{x,v} = \frac{\|k_{\mu,v}\|}{\sigma^2} e^{-\frac{\|k_{\mu,v}\|^2\|z\|^2}{2\sigma^2}} \left[e^{ik_{\mu,v}} - e^{-\frac{\sigma^2}{2}}\right] \qquad (1)$$

where $z = (x, y)$, μ and v define the orientation and scale of Gabor wavelet, $\|.\|$ denotes norm operator, and the wave vector $k(\mu, v)$ defined as follow:

$$k_{\mu,v} = k_v e^{i\phi\mu} \qquad (2)$$

where k_v is k_{max}/f^v, $\phi_u = \pi\mu/8$ (if eight orientations are applied [19]), and i denotes imaginary number. To detect edge on eyes the parameters of Gabor wavelet are set as follow: $v = 2$, $\mu = \{0, \ldots, 7\}$, $\sigma = \pi/2.5$, $k_{max} = \pi/2$. In this proposed method, filtering process uses Gabor filter bank with eight orientations . This method produces noticeable edge on eyes and wrinkle around it (Fig. 1(a)).

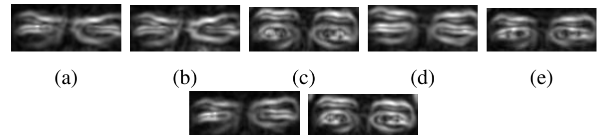
On the other hand, to detect edge on mouth, the parameters of Gabor wavelet are set as follow: $v = 2$, $\mu = \{0, \ldots, 7\}$, $\sigma = \mu/3$, $k_max = \pi/2$. The change in the value of σ produces thinner edge on mouth region (Fig. 1(b)).



(a) (b)

Fig. 1. (a) Edge on eye. (b) Edge on mouth. The images are magnitude value from filtering using Gabor wavelet.

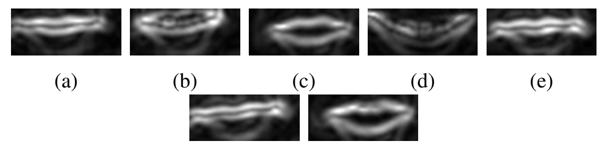
Edge detection with Gabor wavelet produces unique and noticeable pattern, implemented in six basic emotions and neutral state (Fig. 2)(Fig. 3).



(a) (b) (c) (d) (e)

(f) (g)

Fig. 2. Result of filtering using Gabor wavelet on eyes for six basic emotions and neutral state. (a) Angry. (b) Disgust, (c) Fear, (d) Happy, (e) Neutral, (f) Sad, (g) Surprise.



(a) (b) (c) (d) (e)

(f) (g)

Fig. 3. Result of filtering using Gabor wavelet on mouth for six basic emotions and neutral state. (a) Angry. (b) Disgust. (c) Fear. (d) Happy. (e) Neutral. (f) Sad. (g) Surprise.

rea

B. SLFNs Overview

SLFNs is single hidden layer feedforward neural networks, a network architecture with three layers: input, hidden, and output layers. SLFNs is different from single-layer feedforward networks, where the input nodes in the input layer of the network project directly to the output nodes in the output layer of the network. In SLFNs, the input nodes in the input layer supply input vector to the nodes in the hidden layer, and output vector from the hidden layer is fed to the nodes in the output layer.

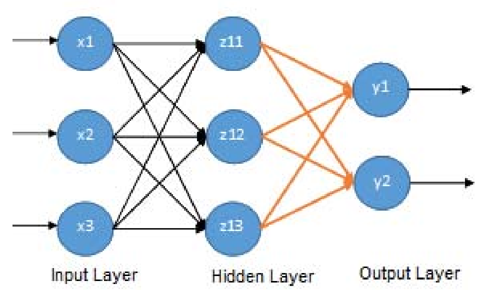


Fig. 4. Single hidden layer feedforward neural networks architecture.

Typically, the output of SLFNs can be modeled as

$$\sum_{i=1}^{\tilde{N}} \beta g(\boldsymbol{w}_i \cdot \boldsymbol{x}_j + b_i) = \boldsymbol{o}_j,\ j = 1, \ldots, N \qquad (3)$$

where β is the weight vector linking the hidden layer to the output layer, $g(x)$ is any activation function in hidden nodes, $\boldsymbol{w}_i = \{w_{i1}, \ldots, w_{in}\}$ is the weight vector connecting the input layer and the hidden layer, $\boldsymbol{x}_j = \{x_{j1}, \ldots, x_{jm}\}$ is the input vector, $b_i = \{b_{i1}, \ldots, b_{ip}\}$ is the hidden bias, and $\boldsymbol{o}_j = \{o_{j1}, \ldots, o_{jq}\}$ is the output vector network. Alike neurons, in learning process all parameters of SLFNs need to be tuned to obtain desired performance. The value of weight will adapt correspondingly with the error-correction.

C. Extreme Learning Machine

Most traditional learning algorithms of SLFNs employ gradient descent based methods. The methods are generally slow, since it needs to adjust iteratively all parameters of networks to obtain desired learning performance. In addition, the gradient descent based methods are easily trapped in local minima [20]; as a result, reaching the smallest squared error is difficult to do by the methods in some cases.

Proposed by Guang Huang-Bin et al. [21] in 2004, ELM is a new learning algorithm designed for SLFNs. ELM is different from traditional learning algorithm. ELM need not to adjust all parameters of the networks iteratively, the output weight linking the hidden layer to the output layer are determined through inverse operation of hidden layer output matrices [21], after choosing arbitrary the input weights and hidden biases. In learning process of ELM, there is no any iterative step; in addition, ELM is able to reach not only smaller training error, but also smaller norm weights. Algorithm of ELM can be defined as follow [21]:

Algorithm 1 ELM algorithm

Given training set $\aleph = \{(x_i, t_i)|x_i \in R^m, t_i \in R^m, i = 1, \ldots, N\}$, activation function $\varphi(x)$, and hidden neuron number $\tilde{N}$.

Step 1: Assign arbitrary value of input weight $\boldsymbol{w}_i$, and bias $b_i, i = 1, \ldots, N$.

Step 2: Calculate hidden layer output matrix $\mathbf{H}$.

Step 3: Calculate the output weight β:

$$\beta = \mathbf{H}^{\dagger}\mathbf{T} \tag{4}$$

where $\mathbf{H}^{\dagger}$ is Moore-Penrose generalized inverse [22] of hidden output matrices, and $\mathbf{T}$ is desired response.

Considering to huge training datasets that are used in the training phase, an optimized function to calculate the output weight of ELM (5) [9] is applied, in order to obtain better learning performance.

$$\beta = \left(\frac{\mathbf{I}}{C} + \mathbf{H}^{\mathbf{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathbf{T}}\mathbf{T} \tag{5}$$

where $\mathbf{I}$ is identity matrix, and C is constant defined by user.

Guang Huang-Bin et al. stated that ELM is a Universal Classifier that can be applied in both classification and regression. In classification cases, ELM is sufficient for binary and multi-class cases. In multi-class cases with huge training datasets, there are two optional output functions. The first is the output function for single-output node (6).

$$f(x) = sign\left(h(x)\left(\frac{\mathbf{I}}{C} + \mathbf{H}^{\mathbf{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathbf{T}}\mathbf{T}\right) \tag{6}$$

On the other hand, while multi-output nodes are applied, the predicted class label is the index number of node with the highest output value [9].

III. Proposed Framework

In this section, it is introduced the proposed facial expression algorithm. The proposed approach extracts features on eye and mouth region using Gabor-wavelet, and applies ELM to classify their pattern. The proposed framework consists of six-folds:

1) Preprocessing: convert the image to gray scale, and perform histogram equalization on it.
2) Detect eye and mouth regions using Haar-classifier [14].
3) Extract features from eye and mouth employing Gabor Wavelet [8], and convert them to binary image.
4) Compose features vector form eye and mouth.
5) Reduce the dimension of the features vector using PCA subspace [15], 132 feature elements are retained.
6) Perform ELM for Regression and Multiclass Classification [9] to classify the features.

IV. Experimental Result

In this section the performance of proposed approached is presented, the value of C and L of ELM is set in various values to analysis its performance. Moreover, to evaluate the robustness of ELM in recognizing the pattern from the extracted features, comparison among SLFNs with ELM, K-NN [16], and SVM with linear kernel [17] is undertaken. All simulations for K-NN, SVM, and ELM are implemented in OpenCV, and they are tested on Intel Core i5 @ 2.40GHz with 4GB RAM using ten subjects expressing six basic emotions and neural state, which are selected from FEEDTUM [23]. They are tested in two cases:

1) In the first case, training and testing datasets are taken from same subject, subject 1-10; it should be noted that the training and testing datasets are not same images.
2) The second cases uses 10-folds cross-validation.

In the first case, the value of C and L of ELM are set to $\{2^5, 2^{10}, 2^{20}, 2^{25}, 2^{30}, 2^{40}, 2^{45}\}$ and $\{100, 200, 400, 600, 800, 1000, 1500\}$, while in the second case the value of C and L are set to 2^{25} and 1500. And for SVM, 32 and 35 are assigned to the value of C and gamma γ in both the first and the second case. The threshold (0.25) is applied to convert the input image to binary image after extracting features with Gabor-wavelet. In order to decrease the time consumption, the features dimension is reduced using PCA subspace, 132 feature elements are retained. In this phase, it does not only analyze the recognition rate, but also calculate RMSE of actual and desired output of compared methods.

TABLE I. Detail of datasets

Case	#train	Training Subject	#test	Testing Subject
One	560	1 - 10	934	1 - 10
Two	@504	Cross	@56	Cross

TABLE II. Recognition rate (%) of proposed approach for each emotion and its miss-classification in case one for $C = 2^{25}$ and $L = 1500$

Predicted Class	Actual Class: Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	Miss-classification
Angry	**92.31%**	2.79%	2.09%	2.21%	0.00%	0.60%	0.00%	7.69%
Disgust	2.15%	**87.76%**	7.94%	0.72%	0.72%	0.00%	0.71%	12.24%
Fear	2.70%	2.21%	**88.11%**	0.00%	2.19%	1.39%	3.40%	11.89%
Happy	1.39%	2.09%	0.70%	**95.10%**	0.72%	0.00%	0.00%	4.90%
Neutral	1.12%	1.12%	2.26%	0.00%	**95.50%**	0.00%	0.00%	4.50%
Sad	0.00%	0.91%	0.00%	0.00%	1.49%	**97.60%**	0.00%	2.40%
Surprise	2.84%	0.00%	0.00%	0.00%	0.00%	0.00%	**97.16%**	2.84%

A. Case 1

In this case, 560 training datasets and 934 testing datasets are used. The training set is composed of the first and the last four images of each subject, whereas the testing set is generated from the remaining images. The recognition rate of ELM reaches 93.36%, in which the value of $L = 1500$, and $C = 2^{25}$. The average of miss-classification is only 6.64%, and RMSE is merely 0.378. Moreover, the training process does not require much time to be convergent. For this case, increasing the value of L affects the recognition rate; generally, the recognition rate increases on the increase of L. The proposed framework not only attains high recognition rate, but also achieves small squared error in this case.

TABLE III. Performance comparison in case one

Method	Testing Result: RMSE	Testing Result: Recognition Rate(%)	Training Time(s)
ELM	0.378	93.36	108.57
K-NN	0.361	87.20	-
SVM	0.343	85.8	0.191

Compared to SVM and K-NN, the recognition rate of ELM is outperform. In this experimental, the SVM with 406 support vectors runs 568.42 times faster than ELM (it should be noted that computing inverse matrix using openCV is at the expense of speed while working with large scale matrix). However, its recognition performance is slightly lower than ELM. K-NN performs better than SVM in this case, yet its recognition rate is not preferable compared with ELM.

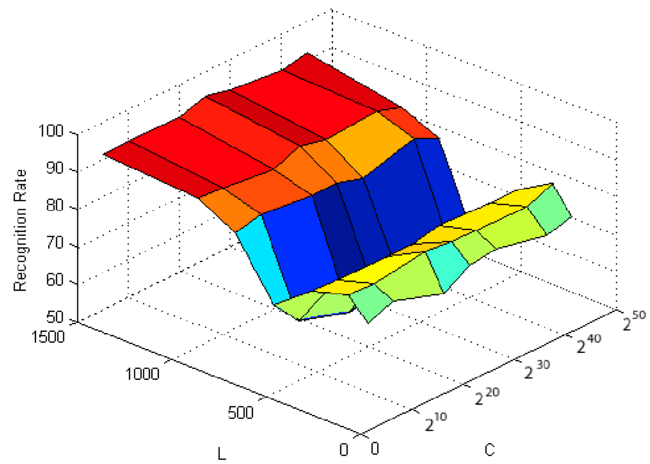


Fig. 5. The recognition rate of proposed approach in case one with different values of C and L, the recognition rate increases with the value of L.

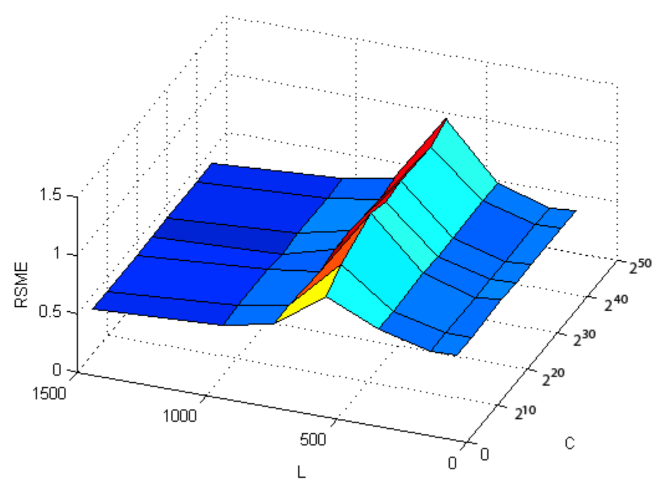


Fig. 6. The RMSE of proposed approach in case one with different values of C and L, the RMSE decreases with the value of L

B. Case 2

For the second case, the proposed approach tested with 10-folds cross-validation, the test set is composed of one subject, out of the total subject. While the train set is generated from the remaining subjects. The value of C and L are set to 2^{25}

TABLE IV. Recognition rate (%) of proposed approach for each emotion and its miss-classification in case two for $C = 2^{25}$ and $L = 1500$

		Actual Class							Miss-classification
		Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	
Predicted Class	Angry	**18.75%**	21.25%	15.00%	15.00%	7.50%	6.25%	16.25%	81.25%
	Disgust	10.00%	**23.75%**	12.50%	23.75%	3.75%	8.75%	17.50%	76.25%
	Fear	20.00%	11.25%	**17.50%**	0.0%	13.75%	11.25%	26.25%	82.50%
	Happy	2.50%	15.00%	5.00%	**57.50%**	0.00%	0.00%	20.00%	42.50%
	Neutral	17.50%	15.00%	13.75%	1.25%	**25.00%**	5.00%	22.50%	75.00%
	Sad	16.25%	20.00%	15.00%	11.25%	11.25%	**15.00%**	11.25%	85.00%
	Surprise	10.44%	9.98%	9.22%	8.03%	1.42%	5.53%	**55.38%**	44.62%

and 1500. Implemented in this case, the ELM does not have satisfying result. The ELM obtains recognition rate merely 30.41%. Not only ELM, but also the other compared methods have poor performance. Nevertheless, compared to SVM and K-NN, the recognition rate of ELM is higher.

TABLE V. Performance comparison in case two

Method	Testing Result		Training Time(s)
	RMSE	Recognition Rate(%)	
ELM	0.771	30.41	97.76
K-NN	1.026	25.14	-
SVM	1.079	14.06	0.125

V. Conclusion

The research in human emotion recognition has attracted the attention of researchers in computer science for decades. Typically, the recognition method consists of two stages, extracting and classifying features. Gabor wavelet considered as the method of extracting features in this proposed framework. The edge produced by Gabor wavelet is thin and noticeable. It shows unique pattern on each emotion. In classifying stage, the proposed approach employs SLFNs with ELM, as its machine learning, to classify the features. The approach is tested in two cases. Implemented in case one, personalized face, the proposed approach reach satisfying performance. Compared to SVM and K-NN, the robustness of ELM is outperformed, the recognition rate of ELM increases with increasing the value of L. The proposed approach not only reaches high recognition rate but also has small squared-error and less time-consumption in training process. However, the performance of proposed approach is inadequate implemented in generalize face case; moreover, in this case, none of the tested methods have satisfying result.

Although the performance of proposed approach is not satisfied enough implemented in generalization face case, the proposed framework has promising result implemented in personalization face case. Therefore, we believe that the proposed framework can be utilized to develop personal application to recognize six basic human emotions.

Acknowledgment

The authors would like to thank to Frank Wallhoff from Institute for Technical Assistive Systems Jade University for providing datasets for this research, and, to reviewer for the worth comment.

References

[1] P. Ekman and W. Friesen, *Pictures of Facial Affect*. Consulting Psychologists Press, 1976.
[2] C. Darwin, *The Expression of Emotion in Man and Animals*. London: John Murray, 1872.
[3] S. V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis, and S. D. Kollias, "Emotion recognition through facial expression analysis based on neurofuzzy network," in *Neural Network 18*, March 2005, pp. 423–435.
[4] R. Samad and H. Sawada, "Edge-based facial feature extraction using gabor wavelet and convolution filters," in *Machine Vision Applications*, Nara, Japan, June 2011, pp. 430–433.
[5] N. Sarode and S. Bhatia, "Facial expression recognition," in *International Journal on Computer Science and Engineering*, vol. 02, 2010, pp. 1552–1557.
[6] M. I. Quraishi, J. P. Choudhury, M. De, and P. Chakraborty, "A framework for the recognition of human emotion using soft computing models," in *International Journal of Computer Applications*, vol. 40, February 2012, pp. 51–55.
[7] A. Uar, Y. Demir, and C. Gzeli, "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering," *Neural Computing and Applications*, pp. 1–12, 2014. [Online]. Available: http://dx.doi.org/10.1007/s00521-014-1569-1
[8] J. Daugman, "High confidence visual recognition of persons by a test of statistical independence," in *Pattern Analysis and Machine Intelligence*, vol. 15, 1993, pp. 1148–1161.
[9] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," vol. 42, no. 2. IEEE, 2012, pp. 513–529.
[10] S. Marelja, "Mathematical description of the responses of simple cortical cells," in *Journal of the Optical Society of Americal*, vol. 70, March 1980, pp. 1297–1300.
[11] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *JOSA A*, vol. 2, no. 7, pp. 1160–1169, 1985.
[12] L. Shen and L. Bai, "A review on gabor wavelets for face recognition," *Pattern analysis and applications*, vol. 9, no. 2-3, pp. 273–292, 2006.
[13] P. Ekman and W. Friesen, "Head and body cues in the judgment of emotion: A reformulation," in *Perceptual and Motor Skills*, vol. 24, 1967, pp. 711–724.
[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition*, March 2001.
[15] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2005.
[16] T. Cover and P. Hart, "Nearest neighbor pattern classification," *Information Theory, IEEE Transactions on*, vol. 13, no. 1, pp. 21–27, 1967.
[17] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[18] C. Liu, "Gabor-based kernel pca with fractional power polynomial models for face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 5, pp. 572–581, 2004.
[19] M. Zhou and H. Wei, "Face verification using gaborwavelets and adaboost," in *Pattern Recognition*, vol. 1, Hongkong, 2006, pp. 404–407.
[20] S. Haykin, *Neural networks: A comprehensive foundation*. New Jersey: Prentice Hall, 1999.
[21] G. B. Huang, Q. n Yu Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Neural Networks*, July 2004, pp. 25–29.

[22] C. Rao and S. Mitra, *Generalized Inverse of Matrices and its Applications*. New York: Wiley, 1971.
[23] F. Wallhoff. (2006) Facial expressions and emotion database. [Online]. Available: http://www.mmk.ei.tum.de/ waf/fgnet/feedtum.html