2009 IEEE International Conference on Control and Automation
Christchurch, New Zealand, December 9-11, 2009

Modified HMAX Models For Facial Expression Recognition

Wenfei Gu, Cheng Xiang and Hai Lin

Abstract— **Three variants of HMAX model are proposed in this paper for recognizing facial expressions of novel faces. The expressions under consideration are: happy, sad, angry, surprised, disgusted, scared and neutral. The modifications are based on recent biological findings about face processing procedure in human brain. Computational results on two different facial expression databases show the efficiency of modified HMAX models compared to the original one.**

Index Terms— **facial expression recognition, HMAX, Hebbian learning, local method.**

I. INTRODUCTION

Automatic recognition of facial expressions from face (color and gray-level) images is known to be complex in view of significant variations in the physiognomy of faces with respect to head pose, environment illumination and person-identity [1]. However, the human ability to perceive expressions is known to be highly sophisticated, with the underlying biological mechanism not yet understood. Therefore, it seems to be expedient to attempt modeling the results of empirical studies on the visual cortex. Many biologically plausible models of human object recognition [2] [3] have, in fact, been proven to outperform single-template object recognition systems. One powerful computational model of object recognition in cortex is HMAX [4], which attempts to model the rapid object recognition mechanism of human ventral visual stream in visual cortex as follows:

- The hierarchical visual processing consists of a series of stages that have increasing invariance to object transformations;
- As the receptive fields of the neurons increase along the visual pathway, the complexity of their preferred stimuli increases;
- Learning is probably involved at all stages and unsupervised learning may occur at the intermediate layers while supervised learning may occur at the top-most layers of the hierarchy.

Standard HMAX performs well for paperclip-like objects. However, it lacks the capability of dealing with facial images since the system can not capture discriminating features to distinguish facial images from natural images. Therefore, HMAX with feature learning is proposed to improve the performance on detecting faces with cluttered background [5]. In this paper, we propose several variants of the HMAX model with modifications inspired from biological findings about human face processing procedure. Computational results show that our modified HMAX models can produce satisfactory results on recognizing expressions of novel faces. The rest of the paper is organized as follows. Section II illustrates the standard HMAX model and its advanced version, HMAX with feature learning, while Section III presents the modified HMAX models for facial expression recognitions. Section IV provides experimental results and detailed discussions. The final section offers concluding remarks.

W.F.Gu, C.Xiang and H.Lin are with Department of Electrical & Computer Engineering, National University of Singapore, Singapore 117576 (email: elexc@nus.edu.sg)

II. HMAX MODEL AND LIMITATIONS ON FACIAL EXPRESSION RECOGNITION

Standard HMAX model and its advanced version, HMAX with feature learning, have a lot of biologically plausible properties, such as hierarchical architecture, max operation and RBF-like feature learning strategy. Here we are going to describe these important properties in detail and then discuss the limitations of original HMAX model on facial expression recognition.

A. Standard HMAX Model

There are a number of layers of computational units in standard HMAX. Generally, the simple S units tune to their inputs using a bell-shaped function to achieve pattern matching, while the C units perform the max operation on the S level responses. For example, the first layer of HMAX, S1, imitating the simple cells found in the V1 area of the primate brain, consists of filters (i.e., Gabor filters) tuned to stimuli with different orientations and scales in the different areas of the visual field. Then, the C1 units in the next layer perform max operation over outputs of the S1 filters that have same orientation, but different scales and positions over some neighborhoods. And in the S2 layer, composite features are obtained by combining the simple features from the C1 layer (with different orientations) into 2 by 2 arrangements. Finally, every C2 layer unit pools the max response over all S2 units in different positions and scales, resulting a specific feature which is used for classification. This kind of multiple S and C levels architecture enables the HMAX to increase specificity and invariance in feature detectors.

B. HMAX Model with Feature Learning

The HMAX architecture is supported by the experimental findings on the ventral visual pathway in primate brain and the computational results are consistent with those of physiological experiments on the primate visual system. However, since the intermediate features in HMAX are manually determined, it uses the same features for all object classes. Moreover, because these features are obtained by combining 4 bar orientations into 2 by 2 forms, they may work well

978-1-4244-4707-7/09/$25.00 ©2009 IEEE

for paperclip like objects rather than natural images like faces. To address this issue, a feature learning strategy, which corresponds to selecting a set of N prototypes $\mathbf{P}_i$ (or features) for the S2 units, has been applied to the standard HMAX model to obtain class-specific features [5]. The learning process is done by extracting a set of patches with various sizes and at random positions from training set. For example, a patch $\mathbf{P}$ of size $n \times n$ contains $n \times n \times 4$ elements can be extracted at the level of the C1 layer across all 4 orientations. These prototypes replace the intermediate S2 features in the standard HMAX. Then new S2 units, acting as Gaussian RBF-units, compute the similarity scores (i.e., Euclidean distance) between an input pattern X and the stored prototype $\mathbf{P}$: $f(X) = exp(-\frac{\|X-P\|^2}{2\sigma^2})$, with σ chosen proportional to patch size. This feature learning strategy helps the HMAX model to achieve a satisfactory performance for the task of face detection.

C. Limitations of HMAX on Facial Expression Recognition

Even though the HMAX model with feature learning could produce a strong preferences to faces against natural scenes, it is difficult to deal with facial expression recognition. Some recent physiological studies show that face processing in the human brain is a dedicated machinery, which may consist of the following aspects:

1) All of neurons in a specialized region of the human brain, such as the fusiform face area (FFA), respond only to faces. And no brain region has previously been identified that is selective for a single visual form [7].
2) Human face processing system would first perform the face detection task, then deal with the face identification and finally recognize the different facial expressions. It seems that identity information is obtained simultaneously when a face is detected, while expression recognition requires further processing [6].
3) Each cell in human face processing system would act as a set of face-specific analyzers, capturing local facial information along multiple distinct dimensions. By combining the local information of all these little analyzers, it should be possible to reconstruct any face, preserving most of facial information [7].

It is obvious that HMAX can not capture these properties which are crucial for face processing. First of all, a set of special units which deal with the face processing are missing. That is, the final layer of HMAX with C2 units, modeling the cells in the IT area, responds to a series of complex visual forms. However, according to the human face processing system, facial patterns are so complicated that additional layer is needed for further processing. Secondly, the feature learning algorithm of HMAX generates a number of random patches which are then used as the prototypes of different objects. To achieve satisfactory performance on object classification using this kind of learning strategy, a large number of natural images are required to train the system. Although the trained system is able to respond to faces, it can not capture the detailed facial information. Therefore, HMAX can at most act as face detector but can not distinguish individual faces as well as different expressions. Thirdly, even the HMAX is trained using a set of face images with different identities and expressions, the strong responses of C2 units may correspond only to some local facial components due to the randomness of the learning strategy and the max operation of the C2 units. Therefore the final decisions of identities and expressions may be not reliable.

III. VARIANTS OF HMAX FOR FACIAL EXPRESSION RECOGNITION

In this section, we are going to present some variants of HMAX model that are modified to provide a satisfactory performance on facial expression recognition compared to the original HMAX.

A. HMAX with Facial Expression Processing Units

One straightforward way to modify the HMAX such that it can deal with facial expression recognition is to add the face processing layer. Although the exact biological procedure of face processing remains unknown, we can adopt some statistical approaches instead. In our implementation, output of C2 units is further processed using principal component analysis (PCA) plus Fisher linear discriminant analysis (FLD) [8] to obtain discriminating features for facial expression recognition.

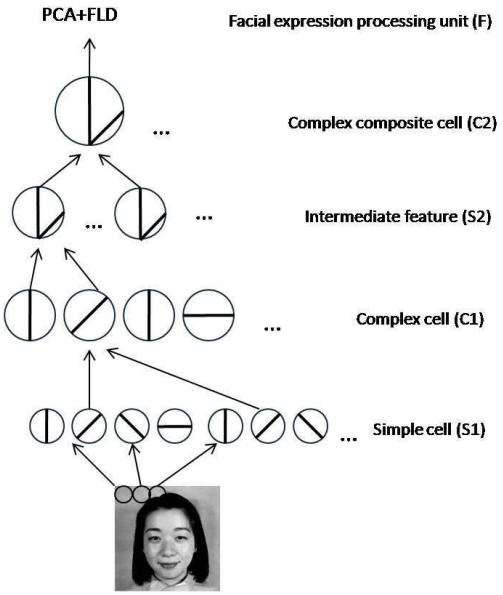


Fig. 1. Structure of HMAX with facial expression processing units.

As can be seen in Fig. 1, the architecture of the HMAX with facial expression processing units is described as follows:

1) The S1 responses are first obtained by applying a battery of Gabor filters to the input image I , which can be described by the following equation:

$$F(x,y) = \exp(-\frac{(\hat{x}^2 + \gamma^2\hat{y}^2)}{2\sigma^2}) \times \cos(\frac{2\pi}{\lambda}\hat{x}), \quad (1)$$

where

$$\hat{x} = x\cos\theta + y\sin\theta, \hat{y} = -x\sin\theta + y\cos\theta \quad (2)$$

Here, (x,y) refers to the 2D coordinate system of the input image. The five parameters (orientation θ, aspect ratio γ, effective width σ, phase ϕ and wavelength λ) determine the properties of the Gabor output. ϕ is set to be 0 and γ is set to be 0.5 since these two parameters have little influence on the final performance. Four orientations ($\theta = 0°, 45°, 90° and 135°$) are sufficient for our purpose. The remaining parameters σ and λ are determined by the following equations based on the tuning properties of cortical cells according to [5]:

$$\sigma = 0.0036 \times RFsize^2 + 0.35 \times RFsize + 0.18 \quad (3)$$

$$\lambda = \frac{\sigma}{0.8} \quad (4)$$

where $RFsize$ varies from 7 to 39 by steps of two.

2) C1 units pool responses over S1 units using max operation, and have some tolerances to certain moderate shift and scale changes.
3) S2 layer contains RBF-like units that are tuned to object-parts and compute a function of the distance between the input units and the stored prototypes.
4) C2 units perform a max operation over the whole visual field and provide the intermediate encoding of the stimulus. The difference between standard HMAX and HMAX with feature learning lies in the connectivity from C1 to S2 layer: in starndard HMAX, these connections are hard-coded to generate 256 combinations (with size of 2×2) of C1 inputs while in HMAX with feature learning, the prototypes are learned from the training set.
5) F units, (also called facial expression processing units) take the responses of C2 units as input, and perform PCA plus FLD to extract discriminating features for classifier to recognize facial expressions.

Experimental results show that F units contribute a lot to the improvement of recognizing facial expressions compared to the original HMAX (see Section. IV-A for details). Therefore, these F units are always used in other modified versions of HMAX.

B. HMAX with Hebbian Learning

Although the facial expression processing units can help to improve performance of recognizing facial expressions, the resulting improvement is still unsatisfactory. Even using the HMAX with feature learning, experimental results (see in Section. IV-B) show that when using facial images to train the HMAX, recognition accuracies on individual database test are satisfactory but recognition rates on cross database test are not stable, indicating that the instability of the RBF-like learning strategy [1]. Notice that JAFFE database contains more variations, such as pose and illumination changes, than TFEID database does. Therefore when using TFEID database as training set, the prototypes extracted by the RBF-like learning strategy may be not suitable for facial images in the JAFFE database. It indicates that the RBF-like learning strategy requires training samples to have large variations in order to achieve good generalization.

To overcome the limitations of RBF-like learning strategy described above, we now propose a Hebbian learning [8] strategy to generate prototypes from C1 to S2 layer. Let $C1_i$ denotes the outputs of C1 layer, where $i = 1,2,3,4$ stands for four orientations, so for every element of $C1_i$ in the same position (x,y), we compute the S2 response using the following formula:

$$S2(x,y) = \phi(\sum_{i=1}^{4} w_i C1_i(x,y)) \quad (5)$$

where $\phi(\cdot)$ is the Gaussian-like tuning function and w_i are the weights for 4 orientations. The weights are learned in a Hebbian learning manner:

$$\bar{\mathbf{W}}^{new} = \bar{\mathbf{W}}^{old} + \alpha\bar{\mathbf{S2}}(\bar{\mathbf{C1}} - \bar{\mathbf{W}}^{old}) \quad (6)$$

where $\bar{\mathbf{S2}}$ is the output of S2 units and $\bar{\mathbf{C1}}$ is the output of C1 units. α is the learning rate and is set to be 0.01 in our implementation. After training with facial images, the linear combination of 4 orientations can be used to represent facial expression information. In the next stage, all the S2 outputs are fed into facial expression processing units without performing max operation. This procedure is to keep as much information as possible for the subsequent processing so that the system can deal with databases with different degrees of variations.

C. HMAX with Local Method

The HMAX and modified versions described above are all holistical methods which take a whole image as input. An important limitation of these holistical methods is that some local information which contributes to facial expressions may be lost. Recently, local approaches have shown promising results not only in facial expression recognition but also in other visual recognition tasks [1]. This is consistent with the biological finding that the local facial components of the face (like the cheeks, mouth, eyes and eyebrows, and forehead) act together, i.e., globally, to compose an expression. So we now consider to improve HMAX model using local method. Fig. 2 shows the architecture of the HMAX with local method and the detailed procedure is as follows:

1) We crop the images such that the background information is removed and the size of input images is uniform. Then each image is divided into several local regions with overlap half of their size. The facial components in the local regions should be as complete as possible while the local regions should be small enough such that local features can be extracted from the facial components. To this end, we use 49 local blocks to achieve a proper trade-off between the locality and the completeness of the facial components.

[1]Please refer to Section. IV for the descriptions of test strategy and database mentioned here

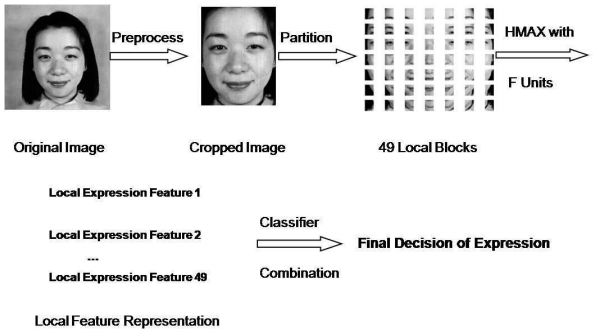


Fig. 2. Sketch of the HMAX model with local method.

2) We apply the original HMAX with facial expression processing units to every local blocks of one facial image to obtain local features. Since the local block, containing local facial information, is small, the hard-coded 2 by 2 arrangements of four orientations are sufficient for extracting local features.
3) We apply a set of local classifiers to make local decisions based on the extracted local features. The outputs of local classifiers can be used as the intermediate features and their combination can lead to a global decision for recognizing expressions.

Classical combination rules, such as Borda count [9] and decision template [10], can be used to combine local classifiers and obtain global decision. However, Borda count, based on voting, does not utilize information in training data whereas decision template, which actually is a nearest-mean classifier in the decision space, may not capture discriminating information for high dimensional decision space. Here we first concatenate outputs of all local classifiers for one facial expression image together as the intermediate feature matrix of that image. So every facial expression image is represented by an intermediate feature vector. Then PCA plus FLD are used to project the intermediate features into a discriminating low-dimensional subspace which can be effectively classified. Since the FLD can at most extract $C-1$ (where C is the number of classes) discriminating components from the input data, which may be insufficient to represent the global features with high complexity, we adopt the recursive FLD (or RFLD) [11], which uses the basic idea of FLD but extracts one feature component at each iteration, and discards the information already extracted by previous iterations from all the samples before going to next iteration. Further, in order to avoid over-fitting, we invoke the regularization method in RFLD as follows:

$$S_W \rightarrow S_W + \beta \cdot Ave(Eigv(S_W)) \cdot I, \tag{7}$$

where β is the regularization factor which controls the influence of the regularization term; S_W is the within-classes scatter-matrix (please refer to [8] for details); $Ave\,(Eigv(S_W))$ denotes the average eigenvalue of S_W, and I is the identity matrix. In this manner, we control the final performance of HMAX with local method by the two parameters, β and N_c (which is the number of extracted features), whose effect on the final recognition performance is studied in Section. IV-C.

IV. COMPUTATIONAL EXPERIMENTS AND DISCUSSIONS

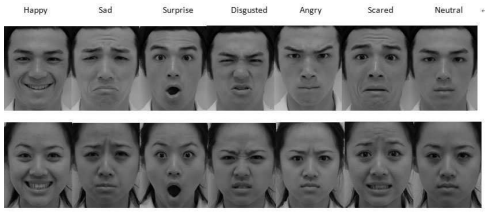
All experiments are conducted in MATLAB^{TM}, using a 2.66 GHz Intel® Core^{TM} Quad processor with 8 GB memory. We use the following facial expression databases for experiments: (1) Japanese Female Facial Expression (JAFFE) database [12];(2) Taiwanese Facial Expression Image Database (TFEID) [13].

The JAFFE database contains 213 images of 7 facial expressions of 10 Japanese female models, including 6 basic facial expressions (*happy, sad, angry, surprised, disgusted and scared* [14]) and neutral face. The TFEID database contains facial expression images of 40 persons (20 males and 20 females). Each person has 8 images corresponding to 8 expressions: happy, sad, surprise disgust, scared, neutral and contempt. In our experiment, we exclude the contempt expression, and focus on the 6 basic expressions and neutral face. Moreover, we find the downloaded TFEID database is incomplete, that is, for some particular persons in the database, some of the expression images are missing. Therefore, we use 268 images in TFEID database for our experiments. Fig. 3 shows some samples in the two databases. All the images are cropped to remove background information and normalized to uniform size (180×140).

It is well known that humans can recognize expressions of an unfamiliar person; that is, a change of identity does not seem to affect the representation of an expression in the human brain while recognizing it. However, for automatic expression recognition, it has turned out to be difficult to separate expression from identity. Here we focus on this more difficult problem of expression recognition of novel persons. There are in general two kinds of methods for testing the performance of the modified HMAX described as follows:

- Individual database test: The training images and testing images are from the same database. However, both the JAFFE and TFEID databases contain limited samples, we adopt the leaving-one-person-out cross-validation strategy. That is, we divide the database according to the number of persons in the database such that each segment contains all images belonging to only one person. After partition, one of these segments is picked out as the test set, and remaining segments are used for training. The above procedure is repeated so that all the segments are used once as the test set, and recognition accuracy is averaged over all the distinct segments.
- Cross database test: The training images are from one database while the test images are from another database in turn: 1) use JAFFE database to train and use TFEID database to test; 2) use TFEID database to train and use JAFFE database to test. The goal is to check whether the expression features, extracted by a recognition system, are representative enough such that a new facial expression image from another database can also be recognized.

(a) Cropped gray-level images in JAFFE database.



(b) Cropped gray-level images in TFEID database

Fig. 3. Samples in the two facial expression databases.

TABLE I

RECOGNITION RESULTS (%) ON INDIVIDUAL DATABASE TASK.

	Standard HMAX	HMAX with F units
JAFFE	32.39	39.44
TFEID	44.03	64.93

A. Experiments Using HMAX with Facial Expression Processing Units

HMAX with facial expression processing units is applied to both JAFFE and TFEID database to obtain feature vectors. Then nearest neighbor classifier is used to classify different expressions. We focus on recognizing expressions of novel expressers, which are considered to be very difficult for computational models. To this end, as described above, both individual database test and cross database test are performed in the simulation. Table. I shows the recognition results on individual database recognition while Table. II shows the recognition results on cross database recognition. We can see from the results that HMAX with facial expression processing units outperforms the standard HMAX. However the performance is not satisfactory which confirms that the hard-coded feature prototypes in the standard HMAX are not suitable for dealing with facial expressions.

TABLE II

RECOGNITION RESULTS (%) ON CROSS DATABASE TASK.

	Standard HMAX	HMAX with F units
JAFFE train, TFEID test	19.78	27.61
TFEID train, JAFFE test	17.37	24.41

B. Experiments Using HMAX with Hebbian Learning

HMAX with Hebbian learning is applied to both individual database test and cross database test, and the results of using nearest neighbor classifier are tabulated in Table. III. For comparison, HMAX with feature learning strategy is also applied to the facial expression recognition to perform the same tasks. Here we use facial images from JAFFE and TFEID database to train HMAX with feature learning. The recognition results are tabulated in Table. IV. We can see that the results of individual database task are slightly better than those of HMAX with RBF-like learning while the results of cross database task are more stable compare to those of HMAX with learning.

TABLE III

RECOGNITION RESULTS (%) OF HMAX WITH HEBBIAN LEARNING.

JAFFE	TFEID	JAFFE train, TFEID test	TFEID train, JAFFE test
78.87	97.01	52.99	39.83

TABLE IV

RECOGNITION RESULTS (%) OF HMAX WITH FEATURE LEARNING.

JAFFE	TFEID	JAFFE train, TFEID test	TFEID train, JAFFE test
77.46	96.27	60.19	29.80

C. Experiments Using HMAX with Local Method

HMAX with local method is applied to both individual database test and cross database test. In the computational experiments, we vary β and Nc to obtain the best results. Table. V shows the recognition accuracies of HMAX with local method on individual database task while Table. VI shows the recognition accuracies of HMAX with local method on cross database task. The results of using decision template and Borda count to combine local classifiers are also given for comparison. It is obvious that HMAX with local method can lead to satisfactory results for both individual database recognition and cross database recognition.

D. Discussions

The above three modified HMAX, which are biologically plausible, can outperform the original HMAX on facial

TABLE V

RECOGNITION RESULTS (%) OF HMAX WITH LOCAL METHOD ON INDIVIDUAL DATABASE TASK.

	Decision Template	Borda Count	PCA + RFLD
JAFFE	75.12	73.24	79.81 ($\lambda = 1.1, Nc = 16$)
TFEID	96.27	96.27	98.88 ($\lambda = 1, Nc = 6$)

TABLE VI

RECOGNITION RESULTS (%) OF HMAX WITH LOCAL METHOD ON CROSS DATABASE TASK.

	Decision Template	Borda Count	PCA + RFLD
JAFFE train, TFEID test	52.62	50.37	60.82 ($\lambda = 1.1, Nc = 30$)
TFEID train, JAFFE test	41.31	38.50	50.70 ($\lambda = 1.3, Nc = 25$)

expression recognition. However, there are still some limitations that should be improved in the future work:

1) For facial expression processing units, FLD may fail because of curse of dimensionality while PCA can not guarantee that crucial features for expressions can be preserved when reducing the dimensionality. Therefore, some advanced techniques are required for improving the performance.
2) The HMAX with local method adopts the classifier combination strategy to produce the global decision. It seems that the human brain does not process the local information in such way. Therefore, design of local feature combination strategy is needed to obtain more abstract and complex global features.

V. CONCLUSIONS

In this paper, we propose three variants of HMAX model, which are applied to facial expression recognition on two different databases. The modifications are based on recent biological findings about face processing procedure in human brain. Experimental results of both individual database test and cross database test show that our modified HMAX can produce satisfactory performance compared to the original HMAX when recognizing facial expressions of novel persons.

VI. ACKNOWLEDGMENTS

The research reported here was supported by NUS Academic Research Fund R-263-000-362-112. The authors also gratefully acknowledge the anonymous reviewers and editors for their helpful comments to substantially improve the quality of this paper.

REFERENCES

[1] B.Fasel and J.Luettin, Automatic Facial Expression Analysis: A Survey, *Pattern Recognition*, vol.36, 2003 pp.259-275.
[2] K.Fukushima, Neocognitron:A Self-Organizing Neural Network Model For A Mechanism Of Pattern Recognition Unaffected By Shift In Position,*Biological Cybernetics*, vol.36, 1980, pp 93-202.
[3] G.Wallis and E.Rolls, A Model Of Invariant Object Recognition In the Visual System, *Progress in Neurobiology*, vol.51, 1997, pp 167-194.
[4] M.Riesenhuber and T.Poggio, Hierarchical Models Of Object Recognition In Cortex, *Nature Neuroscience*, vol.2, 1999, pp 1019-1025.
[5] T.Serre, M.Kouh, C.Cadieu, U.Knoblich, G.Kreiman and T.Poggio, *A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortx, Techinical Report*, MIT, Massachusetts, USA; 2005.
[6] D.Y.Tsao, N.Schweers, S.Moeller, and W.F.Freiwald, Patches of Face-Selective Cortex in the Macaque Frontal Lobe, *Nature Neuroscience*, vol.11, 2008, pp 877-879.
[7] D.Tsao, A Dedicated System for Processing Faces, *Science*, vol. 314, 2006, pp 72-73.
[8] R.O.Duda, P.E.Hart and D.G.Stork, *Pattern Classification, Second Edition*, Wiley-Interscience, New York, USA; 2001.
[9] L.I.Kuncheva, *Combining Pattern Classifiers, Methods and Algorithms*, Wiley Interscience, New York, USA; 2005.
[10] L.I.Kuncheva, J.C.Bezdek and R.Duin, Decision Templates for Multiple Classifier Fusion: An Experimental Comparison, *Pattern Recognition*, vol.34, 2001, pp 299-314.
[11] C.Xiang, X.A.Fan and T.H.Lee, Face Recognition Using Recursive Fisher Linear Discriminant, *IEEE Tran. Image Processing*, vol.15, 2006, pp 2097-2105.
[12] M.Kamachi,M.Lyons and J.Gyoba, The Japanese Female Facial Expression (JAFFE) Database. [Online]. Available:http://www.kasrl.org/jaffe.html
[13] L.F.Chen and Y.S.Yen, 2007 Taiwanese Facial Expression Image Database. [Online]. Availble: http://bml.ym.edu.tw/˜download/html. Brain Mapping Labratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan.
[14] P.Ekman, An argument for basic emotions, *Cognition and Emotion*, vol.6, 1992, pp 169–200.