1-1-2010

Feature selection for facial expression recognition

Abdesselam Bouzerdoum
University of Wollongong, bouzer@uow.edu.au

Son Lam Phung
University of Wollongong, phung@uow.edu.au

Fok Hing Chi Tivive
University of Wollongong, tivive@uow.edu.au

Peiyao Li
University of Wollongong, pli@uow.edu.au

Abstract
In daily interactions, humans convey their emotions through facial expression and other means. There are several facial expressions that reflect distinctive psychological activities such as happiness, surprise or anger. Accurate recognition of these activities via facial image analysis will play a vital role in natural human-computer interfaces, robotics and mimetic games. This paper focuses on the extraction and selection of salient features for facial expression recognition. We introduce a cascade of fixed filters and trainable non-linear 2-D filters, which are based on the biological mechanism of shunting inhibition. The fixed filters are used to extract primitive features, whereas the adaptive filters are trained to extract more complex facial features for classification by SVMs. This paper investigates a feature selection approach that is based on the reduction of mutual information among the selected features. The proposed approach is evaluated on the JAFFE database with seven types of facial expressions: anger, disgust, fear, happiness, neutral, sadness and surprise. Using only two-thirds of the total features, our approach achieves a classification rate (CR) of 96.7%, which is higher than the CR obtained using all features. Our system also outperforms several existing methods, evaluated on the same JAFFE database.

Keywords
feature, facial, expression, recognition, selection

Disciplines
Physical Sciences and Mathematics

Publication Details
Li, P., Phung, S., Bouzerdoum, A. & Tivive, F. (2010). Feature selection for facial expression recognition. 2nd European Workshop on Visual Information Processing (pp. 35-39). USA: IEEE.

FEATURE SELECTION FOR FACIAL EXPRESSION RECOGNITION

P. Li, S. L. Phung, A. Bouzerdom, and F. H. C. Tivive

School of Electrical, Computer and Telecommunication Engineering,
University of Wollongong, Wollongong, NSW 2522, Australia

ABSTRACT

In daily interactions, humans convey their emotions through facial expression and other means. There are several facial expressions that reflect distinctive psychological activities such as happiness, surprise or anger. Accurate recognition of these activities via facial image analysis will play a vital role in natural human-computer interfaces, robotics and mimetic games. This paper focuses on the extraction and selection of salient features for facial expression recognition. We introduce a cascade of fixed filters and trainable non-linear 2-D filters, which are based on the biological mechanism of shunting inhibition. The fixed filters are used to extract primitive features, whereas the adaptive filters are trained to extract more complex facial features for classification by SVMs. This paper investigates a feature selection approach that is based on the reduction of mutual information among the selected features. The proposed approach is evaluated on the JAFFE database with seven types of facial expressions: anger, disgust, fear, happiness, neutral, sadness and surprise. Using only two-thirds of the total features, our approach achieves a classification rate (CR) of 96.7%, which is higher than the CR obtained using all features. Our system also outperforms several existing methods, evaluated on the same JAFFE database.

***Index Terms*—** facial expression recognition, adaptive filter, feature selection, mutual information, support vector machine.

1. INTRODUCTION

Human facial expression, controlled by a complex mesh of nerves and muscles beneath the face skin, enables people to convey their emotions and perform nonverbal communications. Accurate recognition of facial expression is essential in many fields, including human-machine interaction, affective computing, robotics, computer games and psychology studies. There are seven basic facial expressions that reflect distinctive psychological activities: anger, disgust, fear, happiness, neutral, sadness and surprise. Examples of these facial expressions are shown in Figure 1.

In this paper, we propose a novel approach to recognize facial expressions from static images, via extraction and selection of salient appearance features. In our approach, fixed and adaptive nonlinear 2-D filters are combined in a hierarchical structure. The fixed filters are used to extract primitive features such as edges, whereas the adaptive filters are trained to extract more complex facial features for classification. The SVM and feature selection are combined to improve classification performance.

The paper is organized as follows. Section 2 reviews related work on facial expression recognition. Section 3 presents the proposed method for 2-D feature extraction, whereas Section 4 describes the feature selection approaches. Section 5 analyzes the performance of the proposed method on a standard database and compares it with several existing techniques. Section 6 gives concluding remarks.

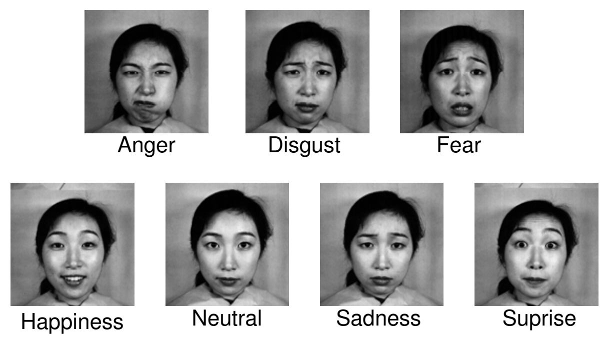

Fig. 1. Examples of facial expressions in JAFFE database. The faces are cropped from these images before facial expression recognition is performed.

2. RELATED WORK

Based on the scheme on how features are extracted from an image for classification, existing approaches for facial expression recognition can be divided into three categories: geometric-based, appearance-based, and hybrid-based approaches.

2.1. Geometric-based approaches

A face image is represented geometrically via fiducial points or the shape of facial regions [1]. Classification is done by analyzing the distances between fiducial points and the relative sizes of the facial components. Pantic *et al*. [1] proposed a method for detecting facial actions by analyzing the contours of facial components, including the eyes and the mouth. A multi-detector technique is used to spatially sample the contours and detect all facial features. A rule-based classifier is then used to recognize the individual facial muscle action units (AUs). Geometric-based methods cope well with variations in skin patterns or dermatoglyphics. However, they require accurate detection of facial fiducial points, which is difficult when the image has a complex background or a low quality.

2.2. Appearance-based approaches

Many appearance-based methods process the entire image by applying a set of filters to extract facial features. Zhen *et al*. [2] used Gabor wavelets to represent appearance changes as a set of multi-scale and multi-orientation coefficients. They proposed a ratio-image based feature that is independent of the face albedos. Their method can cope with different people and illumination conditions. Feng [3] used Local Binary Patterns (LBP) to extract facial texture features and combined different local histograms to recover the shape of the face. He developed a coarse-to-fine scheme for classification, where seven templates were designed to represent the corresponding seven basic facial expressions. Firstly, two expression classes are selected based on the distance from the test image to the seven templates. The final classification is then done via a K-nearest neighbor classifier with weighted Chi-square statistic.

2.3. Hybrid-based approaches

Facial expression recognition can be improved by combining appearance and geometric features. Zhang and Ji [4] proposed a multi-feature technique that is based on the detection of facial points, nasolabial folds, and edges in the forehead area. In their method, facial features are extracted by associating each AU with a set of movements, and then classified using a Bayesian network model.

Appearance-based methods typically use all features extracted from the face image. These features may contain redundant information which influences the classification accuracy. Feature selection aims to remove irrelevant features to build robust training models and improve the system performance. Dubuisson *et al*. [5] proposed a feature selection method that sorts the principal components, generated by principal components analysis (PCA), in the order of their importance. A forward stepwise selection method is used to select the K most discriminant components. Then the linear discriminant analysis (LDA) is applied to process the sorted eigenspace and produce a low-dimensional subset of features for classification.

3. IMAGE FEATURE EXTRACTION

The proposed system is designed to recognize the seven major facial expressions. It consists of three processing stages as shown in Figure 2. The first and second stages consist of nonlinear filters, which are used for extracting visual features. The third stage performs classification.

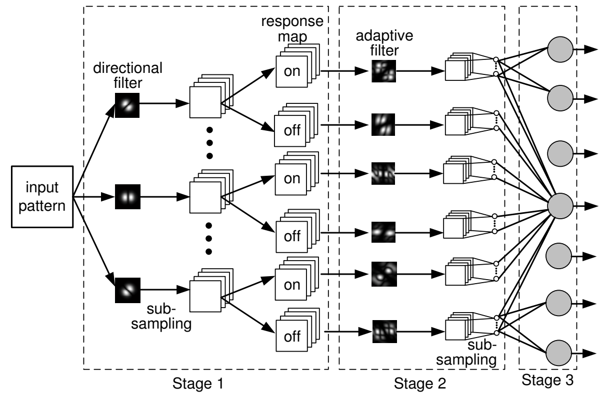

Fig. 2. Block diagram of the proposed system.

3.1. Stage 1 - Directional Filters

Stage 1 is designed to extract features at different orientations. It consists of a set of nonlinear filters that are based on a biological mechanism known as *shunting inhibition*. This mechanism, found in the early visual system [6], has been applied to improve image contrast [7]. The output response of the proposed directional nonlinear filter is computed as

$$\mathbf{Z}_{1,i} = \frac{\mathbf{D}_i * \mathbf{I}}{\mathbf{G} * \mathbf{I}}, \quad (1)$$

where $\mathbf{I}$ is a 2-D input face pattern, $\mathbf{Z}_{1,i}$ is the output of the i-th filter, $\mathbf{D}_i$ and $\mathbf{G}$ are the filter coefficients, "$*$" denotes 2-D convolution, and the division is done pixel-wise. In this paper, the subscripts 1 and 2 in $\mathbf{Z}_{1,i}$ and $\mathbf{Z}_{2,i}$ indicate the outputs of the first and second processing steps, respectively. The kernel $\mathbf{G}$ is chosen as an isotropic Gaussian kernel:

$$\mathbf{G}(x, y) \quad = \quad \frac{1}{2\pi\sigma^2} \exp(-\frac{x^2 + y^2}{2\sigma^2}). \quad (2)$$

To extract elementary facial features at different directions, the kernel $\mathbf{D}_i$ is formulated as the M-th order derivative Gaussian. Its coefficients is defined as

$$\mathbf{D}_i(x, y) = \sum_{k=0}^{M} \frac{M!}{k!(M-k)!} s_x^k s_y^{M-k} \frac{\partial^M \mathbf{G}(x, y)}{\partial x^k \partial y^{M-k}}, \quad (3)$$

where

- M is the derivative order, $M = 1, 2, ...,$

- θ_i is the angle of rotation, $\theta_i = \frac{(i-1)\pi}{N_1}$ for $i = 1, 2, ..., N_1$,
- $s_x = \sin\theta_i$ and $s_y = \cos\theta_i$.

The partial derivative of the Gaussian with respect to dimension x or y can be computed as the product of the Hermite polynomial and the Gaussian function,

$$\frac{\partial^k \mathbf{G}(x,y)}{\partial x^k} = \frac{(-1)^k}{(\sqrt{2}\sigma)^k} H_k(\frac{x}{\sqrt{2}\sigma})\mathbf{G}(x,y), \qquad (4)$$

where $H_k()$ is the Hermite polynomial of order k. Figure 3 shows the outputs of directional, derivative Gaussian filters when $N_1 = 4$ and $M = 2$.

Fig. 3. Outputs of the directional, second-order derivative Gaussian filters for input image 5 of Fig. 1. The parameters are $N_1 = 4$ and $\theta_i = 0°, 45°, 90°$ and $135°$.

Robust image classification requires visual features that are tolerant to small translations and geometric distortions in the input image. To achieve this, we perform a sub-sampling operation and decompose each filter output $\mathbf{Z}_{1,i}$ into four smaller maps:

$$\mathbf{Z}_{1,i} \rightarrow \{\mathbf{Z}_{2,4i-3}, \mathbf{Z}_{2,4i-2}, \mathbf{Z}_{2,4i-1}, \mathbf{Z}_{2,4i}\}. \qquad (5)$$

The first map $\mathbf{Z}_{2,4i-3}$ is formed from the odd rows and odd columns in $\mathbf{Z}_{1,i}$; the second map $\mathbf{Z}_{2,4i-2}$ is formed from the odd rows and even columns, and so on.

The next processing step is motivated by the center-surround receptive fields and the two configurations are on-center and off-center. Herein, we separate each sub-sampled map $\mathbf{Z}_{2,i}$, where $i = 1, 2, ..., 4N_1$, into an on-response map and an off-response map, using zero as a threshold:

$$\mathbf{Z}_{2,i} \rightarrow \begin{cases} \text{on} \ : \ \mathbf{Z}_{3,2i-1} = \max(\mathbf{Z}_{2,i}, 0) \\ \text{off} : \ \mathbf{Z}_{3,2i} = -\min(\mathbf{Z}_{2,i}, 0) \end{cases} . \qquad (6)$$

Essentially, for the on-response map, all negative entries are set to 0, whereas for the off-response map, positive entries are set to 0 and the entire map is then negated. Each map is contrast-normalized using the transformation equation:

$$\mathbf{Z}_{4,i} = \mathbf{Z}_{3,i}/(\mathbf{Z}_{3,i} + \mu),$$

where μ is the mean value of the absolute response of the output map of the directional filter and the division operation is performed pixel-wise.

3.2. Stage 2 - Trainable Filters

Stage 2 aims to detect more complex features for classification. The output maps produced by each filter in Stage 1 are processed by exactly two filters in Stage 2: one filter for the on-response and the other filter for the off-response. Hence, the number of filters, N_2, in Stage 2 is twice the number of filters in Stage 1: $N_2 = 2N_1$.

Stage 2 is also based on the shunting inhibition mechanism. Consider an input map $\mathbf{Z}_{4,i}$ to Stage 2. Suppose that $\mathbf{P}_k$ and $\mathbf{Q}_k$ are two adaptive kernels for the filter that corresponds to this input map. The filter output is calculated as

$$\mathbf{Z}_{5,i} = \frac{g\Big(\mathbf{P}_k * \mathbf{Z}_{4,i} + b_k\Big) + c_k}{a_k + f\Big(\mathbf{Q}_k * \mathbf{Z}_{4,i} + d_k\Big)}, \qquad (7)$$

where a_k, b_k, c_k and d_k are adjustable bias terms, and f and g are two activation functions. A sub-sampling operation is performed across each set of four output maps generated from the adaptive filter by averaging each non-overlapping block of size $(2 \times 2$ pixels$) \times (4$ maps$)$ into a single output signal:

$$\{\mathbf{Z}_{5,4i-3}, \mathbf{Z}_{5,4i-2}, \mathbf{Z}_{5,4i-1}, \mathbf{Z}_{5,4i}\} \rightarrow \mathbf{Z}_{6,i}. \qquad (8)$$

This sub-sampling process is repeated for each adaptive filter to generate a feature vector.

3.3. Stage 3 - Classification

The extracted features are sent to Stage 3 for classification. Stage 3 may use any type of classifiers. Previously, we used a linear classifier whose output y_j is given as

$$y_j = \sum_{i=1}^{N_3} w_{i,j}\, \mathbf{Z}_{6,i} + b_j, \qquad j = 1, 2, ..., N_4 \qquad (9)$$

where $w_{i,j}$'s are adjustable weights, b_j is an adjustable bias term, $\mathbf{Z}_{6,i}$'s are input features to Stage 3, N_3 is the number of input features, and N_4 is the number of output nodes. The output $\mathbf{y} = [y_1, y_2, ..., y_{N_4}]^T$ indicates the class or the label of the input pattern $\mathbf{I}$.

To improve classification accuracy, in this paper we use support vector machine (SVM) for Stage 3. SVM is an important tool in pattern classification. It has been developed initially for two-class problems, and has been shown to achieve good generalization by maximizing the margin between two classes. To solve multi-class problems, we can construct several SVMs to differentiate each pair of classes. For example, for seven facial expressions, we need 21 pair-wise SVMs.

To implement the new approach, we adopt a two-step process. First, we assume that a linear classifier is used in Stage 3, and calculate the coefficients of filters in Stage 2 and the weights of the linear classifier, using the Levenberg-Marquardt optimization algorithm [8]. Second, once the filters in Stage 2 are found and N_3 features are extracted,

Table 1. Classification rates for different facial expression categories. The entry at (row r, column c) is the percentage of facial expression r that is classified as facial expression c.

%	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	**96.67**	0.00	0.00	0.00	3.33	0.00	0.00
Disgust	0.00	**96.55**	0.00	0.00	0.00	3.45	0.00
Fear	0.00	0.00	**96.88**	0.00	0.00	0.00	3.12
Happiness	0.00	0.00	0.00	**100.00**	0.00	0.00	0.00
Neutral	0.00	0.00	0.00	0.00	**96.67**	3.33	0.00
Sadness	0.00	0.00	0.00	3.23	0.00	**96.77**	0.00
Surprise	0.00	0.00	0.00	6.67	0.00	0.00	**93.33**

we train the multi-class SVM to classify a selected subset of these features.

4. FEATURE SELECTION

The purpose of feature selection is to find a subset of features that jointly lead to the best separation of the target classes. The steps in our approach can be described as follows. Let $\mathbf{S}_t$ be the set of selected features at round t. Let D_{train}, D_{valid}, and D_{test} be the training, validation and test set, respectively.

- Step 1: Calculate the class separation score (CSS) for each feature in the feature pool. Let f^* be the feature with the highest CSS. Initialize $\mathbf{S}_0 = \{f^*\}$.
- Step 2: At round t, consider a feature f in the remaining feature pool. Train the classifier with features $\{\mathbf{S}_{t-1}, f\}$ on D_{train}, and evaluate it on D_{valid}. Calculate also the mutual information score for $\{\mathbf{S}_{t-1}, f\}$ on D_{train}.
- Step 3: Repeat Step 2 for all features in the remaining feature pool.
- Step 4: Select a features that correspond to the highest CR when added to $\mathbf{S}_{t-1}$. If several features have the same CR, select features that have the lower mutual information scores.
- Step 5: Increment t and go to Step 2 until a defined number of features are selected.
- Step 6: Train final classifier on D_{train} and evaluate its performance on D_{test}.

Next, we explain how the class separation score (CSS) and the mutual information are calculated. Let N be the number of classes, $N = N_4$ in the proposed architecture. For feature f, let $p_{f,i}(x)$ be the class-conditional probability density function (pdf) for class i. The class separation score for feature f is computed as

$$C(f) = \sum_{i \neq j} \{ \int p_{f,i}(x) \log p_{f,j}(x) dx \}. \tag{10}$$

A higher $C(f)$ means a better separation between the classes by feature f. In our work, the class-conditional pdfs are estimated via Gaussian kernels (i.e. Parzen window method).

The mutual information of a feature set is the sum of mutual information between all feature pairs:

$$M = \sum_{m<n} M(f_m, f_n). \tag{11}$$

In this paper, we analyze two methods of calculating the mutual information between two features. Consider two features f_m and f_n. Let $p_m(x)$ and $p_n(x)$ be probability density functions of the two features, calculated on the entire training set. Let $p_{m,n}(x, y)$ be the joint pdf of the two features.

Method 1: The mutual information is defined based on the symmetric Kullback-Leibler divergence:

$$M(f_m, f_n) = -\int \{p_m(x) \log p_n(x) + p_n(x) \log p_m(x)\} dx, \tag{12}$$

Method 2: The mutual information is defined based on the joint pdf:

$$M(f_m, f_n) = -\int\int p_{m,n}(x, y) \log \frac{p_{m,n}(x, y)}{p_m(x)p_n(y)} dxdy \tag{13}$$

5. RESULTS AND ANALYSIS

In this section, we analyze the performance of the proposed method on a benchmark facial expression data set. We also compare the proposed method and other existing methods for facial expression recognition.

5.1. Database and experimental steps

The proposed system is evaluated on the Japanese Female Facial Expression (JAFFE) database [9], which is commonly used in research on facial expression recognition. This database consists of 213 images from 10 Japanese actresses. They were instructed to produce seven types of facial expressions (see Figure 1). For each person, two to four images were recorded for each facial expression.

We applied the 10-fold cross validation on the JAFFE database, as in [10]. All images were divided into ten groups. For each validation fold, nine groups were used to train the classifier while the remaining group was used for testing. This step was repeated 10 times, and the classification rates of the ten folds were averaged to form the final estimate of the classification rate.

The proposed system uses an input image size of 44×32 pixels. The filter sizes for Stages 1 and 2 are 7-by-7 and 3-by-3 pixels, respectively. The order of the Gaussian derivative filters M can vary. To determine a suitable value for M, we conducted preliminary experiment for M equal to 1, 2, 3, 4, and 5. The classification rate for one trial is shown in Table 2. Based on this result, the order of Gaussian derivative filter is selected to be $M = 2$. That is, Stage 1 uses the second-order Gaussian derivative ($M = 2$) and four directions ($N_1 = 4$). Our experiments also used the LIB-SVM package, developed by Chang and Lin at National Taiwan University [11].

Table 2. Comparison of different values for M - the order of Gaussian derivative filters.

Order	1	2	3	4	5
CR %	94.4	96.3	94.5	94.9	94.5

5.2. Results of using all features

We first evaluated the classification performance when the SVM and all extracted features were used. The classification rates for this system are shown in Table 1 for different categories of facial expressions. In this table, the entry (at row r, column c) is the percentage of facial expression r that is classified as facial expression c. For example, 96.67% of anger expressions are correctly classified as anger, whereas 3.33% of anger expression are misclassified as sadness.

The classification rates for the seven facial expressions are: anger 96.67%, disgust 96.55%, fear 96.88%, happiness 100.0%, neutral 96.67%, sadness 96.77% and surprise 93.33%. The system can recognize happiness and neutral expressions well. It can recognize anger, disgust, and sadness expressions better than fear and surprise expressions.

5.3. Results of applying feature selection

We applied the two methods, described in Section 4, for feature selection.

- When all 560 features produced by Stage 2 were used, the CR was 96.2%.
- When feature selection method based on individual pdfs was used, the system achieved a CR of 96.2% using only 375 features.
- When feature selection method based on joint pdf was used, the system achieved a CR of 96.7% using only 373 features.

These results indicate that feature selection can lead to better classification performance with significantly fewer features.

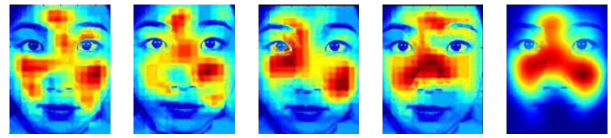

Fig. 4. Locations of selected features superimposed on a face image as yellow-red patches. The first four images correspond to features in the four directions, $\theta = 0°, 45°, 90°$, and $135°$. The last image shows the selected features, combined from all directions.

Figure 4 shows the face areas where the selected features are located. It seems features used for facial expression recognition are located near the cheek and the area between the two eyes. Surprisingly, the mouth area plays a less significant role in FER.

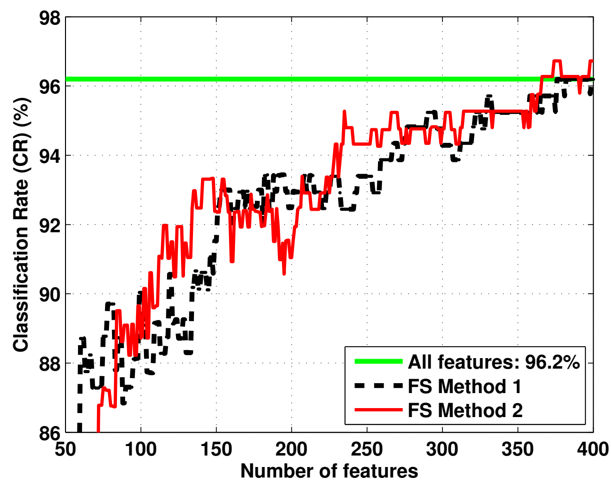

Fig. 5. System performance with different number of features selected.

Figure 5 shows the classification rate on the test set versus the number of features that are selected via the training and validation set. The solid horizontal line indicates the performance when all features are used (CR = 96.2%). The figure shows that feature selection based on joint pdf is better than feature selection based on individual pdf.

5.4. Comparison with other methods

Table 3 shows the classification rates of several FER methods, tested on the JAFFE database using ten-fold validation. Guo and Dyer [10] compared several feature selection schemes: using all features, feature selection via linear programming (FSLP), feature selection via adaptive boosting (AdaBoost). Busiu *et al*. [12] used Gabor wavelets to extract image features and the linear SVM as a classifier. Zhang *et al*. [13] used 34 manually defined fiducial points for feature extraction, and two-layer feedforward neural network for classification. Koutlas and Fotiadis [14] used 20 automatically defined fiducial points and feed-forward neural networks (MLP). The proposed system, which uses hybrid filters, SVM classifier and feature selection based on joint pdf, had a classification rate of 96.7%. It performed better than the seven methods tested on the JAFFE database.

Table 3. Classification rates of FER methods on JAFFE database.

Method	**CR** (%)
Hybrid filters + SVM + FS method 2	96.7
Hybrid filters + SVM + FS method 1	96.2
Hybrid filters + SVM + all features	96.2
Hybrid filters + Linear classifier	95.3
Gabor + Linear SVM [12]	95.2
Fiducial points + FSLP [10]	91.0
Gabor + MLP [14]	90.2
Fiducial points + two-layer MLP [13]	90.1
LBP + Coarse-to-Fine [3]	77.0
Fiducial points + AdaBoost [10]	71.9
Fiducial points + Bayes rule [10]	71.0

6. CONCLUSION

We presented an approach for facial expression recognition that is based on fixed filters and adaptive filters connected in a cascading structure. The fixed, directional filters extract primitive edge features, whereas the adaptive filters are trained to extract more complex features, which are then classified by SVMs. We also implemented and compared two feature selection methods to construct a FER system from a reduced number of features. Evaluated on the JAFFE database, the proposed system has a classification rate of 96.7%, which is higher than existing methods. The experiment results also demonstrate that this classification rate can be achieved by using only two-thirds of the features extracted by the adaptive filters in Stage 2. For future research, we plan to develop optimization algorithms that allow the SVM in Stage 3 and the adaptive filters in Stage 2 to be trained simultaneously.

7. REFERENCES

[1] M. Pantic and L. J. M. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Trans. SMC, Part B: Cybernetics*, vol. 34, no. 3, pp. 1449–1461, 2004.

[2] W. Zhen and T. S. Huang, "Capturing subtle facial motions in 3d face tracking," in *Proc. ICCV*, 2003, pp. 1343–1350.

[3] X. Feng, "Facial expression recognition based on local binary patterns and coarse-to-fine classification," in *Proc. ICCIT*, 2004, pp. 178–183.

[4] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Trans. PAMI*, vol. 27, no. 5, pp. 699–714, 2005.

[5] S. Dubuisson, F. Davoine, and M. Masson, "A solution for facial expression representation and recognition," *Signal Processing: Image Communication*, vol. 17, no. 9, pp. 657–673, 2002.

[6] L. J. Borg-Graham, C. Monier, and Y. Fregnac, "Visual input evokes transient and strong shunting inhibition in visual cortical neurons," *Nature*, vol. 393, no. 6683, pp. 369–373, 1998.

[7] T. Hammadou and A. Bouzerdoum, "Novel image enhancement technique using shunting inhibitory cellular neural networks," *IEEE Trans. Consumer Electronics*, vol. 47, no. 4, pp. 934–940, 2001.

[8] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Trans. Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[9] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proc. IEEE FG*, 1998, pp. 200–205.

[10] G. Guo and C. R. Dyer, "Learning from examples in the small sample case: face expression recognition," *IEEE Trans. SMC, Part B: Cybernetics*, vol. 35, no. 3, pp. 477–488, 2005.

[11] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: a library for support vector machines," 2001.

[12] I. Buciu, C. Kotropoulos, and I. Pitas, "Ica and gabor representation for facial expression recognition," in *Proc. ICIP*, 2003, vol. 2, pp. II–855–8.

[13] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Proc. IEEE FG*, 1998, pp. 454–459.

[14] A. Koutlas and D. I. Fotiadis, "An automatic region based methodology for facial expression recognition," in *Proc. IEEE SMC*, 2008, pp. 662–666.