FACIAL EMOTION RECOGNITION BY ADAPTIVE PROCESSING OF TREE STRUCTURES

Jia-Jun Wong
Division of Computing Systems
School of Computer Engineering
Nanyang Technological University
Nanyang Avenue
Singapore 639798
wong0095@ntu.edu.sg

Siu-Yeung Cho
Division of Computing Systems
School of Computer Engineering
Nanyang Technological University
Nanyang Avenue
Singapore 639798
assycho@ntu.edu.sg

ABSTRACT
We present an emotion recognition system based on a probabilistic approach to adaptive processing of Facial Emotion Tree Structures (FETS). FETS are made up of localized Gabor features related to the facial components according to the Facial Action Coding System. The proposed model is an extension of the probabilistic based recursive neural network model applying in face recognition by Cho and Wong [1]. The robustness of the model in an emotion recognition system is evaluated by testing with known and unknown subjects with different emotions. The experiment results shows that the proposed model significantly improved the recognition rate in terms of generalization.

Categories and Subject Descriptors
I.2.10 [Artificial Intelligence]: Vision and Scene Understanding, Learning.
I.5.4 [Pattern Recognition]: Computer vision.
E.1 [Data Structures]: Trees.
J.4 [Social and Behavioral Sciences]: Psychology.

General Terms
Algorithms, Measurement, Human Factors, Verification.

Keywords

1. INTRODUCTION
According to psychologist, facial expression provides information about emotional states as well as cognitive activities [2]. Emotions are revealed earlier through facial expression than people verbalize or even realize their emotional state [3]. Emotions can be classified under six basic categories initiative by the renowned psychologist, Paul Ekman [2] (i.e. fear, anger, sad, surprise, disgust, happy) as shown in Figure 1. Some of the psychologists have also found that the cognitive interpretations of emotions from facial expressions were innate and universal to all humans regardless of cultures [4, 5].

Neutral Fear Anger Sad Surprise Disgust Happy

Figure 1. 6 Basic Emotions and Neutral Expression

Models and automated systems have been created to recognize the emotional states from facial expressions. The leading method Facial Action Coding System [6], for measuring facial movements in behavioral science was developed by Ekman and Friesen in 1977. Other methods such as electromyography, which directly measures the electrical signals generated by the facial muscles and deducing the facial behavior from it, are both obtrusive non-comprehensive. According to the survey [7], FACS is the leading method for measuring facial expression in behavioral science. It uses 46 defined Action Units to correspond into each independent motion of the face. However this method takes over 100 hours of training to achieve minimal competency for a human expert [8]. Faster automation approaches, such as measurement of facial motion through optic flow [9, 10] and analysis of surface textures based on principal component analysis (PCA) [11]. Newer techniques include using Gabor wavelets [12], linear discriminant analysis [13], local feature analysis [14], and independent component analysis [15]. The techniques are benchmarked [8] and best classification accuracy of about 95% for the recognition of the twelve facial actions, was obtained using Gabor filter representation. Human experts and naïve human tester were benchmarked as well; scored about 94% and 78% respectively, and experiments were supported by Zhang et. al. [16].

Most of these systems use a set of feature vectors to represent facial images, without describing the relationship between the feature vectors. In this paper, we propose a method for emotion recognition by transforming the feature vector data into tree structure representation, which encodes the feature relationship information among the face features. Sixty Localized Gabor Features (LGF) and one Global Gabor Feature are obtained as a feature vector and transforming them into a Facial Emotion Tree Structure (FETS) representation. The layers in the FETS form a
localized to holistic analysis of the facial images and recognition is based on the collective inputs at various layers. Tsot [17] proposed using tree structures to preserve and make use of these relationships and processing them by specific machine learning models [1, 18-20]. Cho and Wong proposed using Gabor features in tree structure representation for face recognition with achieving high accuracy rate [1]. Gabor Feature extraction makes use of Gabor wavelets, which capture the properties of spatial localization, orientation selectivity, spatial frequency selectivity, and quadrature phase relationship, seem to be a good approximation to filter response profiles encountered experimentally in cortical neurons. A probabilistic based recursive neural network is proposed for classification of the FETS in this paper. This method is benchmarked against Support Vector Machines (SVM) [21], K nearest neighbors (KNN) [22], Naïve Bayes algorithm [23] where the flat vector representations in this paper. We made use of the Japanese Female Facial Expression (JAFFE) [24] database to illustrate the performance of the recognition system. Our proposed emotion recognition system is illustrated in Figure 2. This system constitutes the low-level feature extraction and the high-level tree structure representation for emotion recognition. The details of the major components in the proposed system will be described in the following sections.

2. FACIAL EMOTION TREE STRUCTURE (FETS) REPRESENTATION

2.1 Gabor Feature Extraction

We used a pre-defined global filter based on the two-dimensional Gabor wavelets, which can be defined as follows [25]:

\[
g(x, y) = \frac{1}{2\pi\sigma_x \sigma_y} \exp \left[ \frac{-\left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right)}{2} + \frac{xy}{\sigma_x \sigma_y} \right] \tag{1}
\]

where parameters \( W = U_h \), \( \sigma_x = 2\sigma_u / \pi \) and \( \sigma_y = 2\sigma_v / \pi \).

\[
\sigma_u = \frac{\left( a - 1 \right) U_h}{\left( a + 1 \right) \ln 2},
\]

\[
\sigma_v = \frac{\pi}{2K} \left[ U_h - 2 \ln \left( \frac{\sigma_u^2}{U_h} \right) \right]^{\frac{1}{2}} \left[ 2 \ln 2 - \left( \frac{2 \ln 2 \sigma_v^2}{U_h^2} \right) \right]^{\frac{1}{2}}
\]

and \( K \) is the total number of orientations, \( a = (U_h/U_f) \frac{1}{s-1} \) and \( s \) is the number of scales in the multi-resolution decomposition. \( U_h \) and \( U_f \) is the lower and upper center frequencies respectively.

The mean and standard deviation of the convolution output is used as the representation for classification purpose:

\[
\mu_{mn} = \int [W_{mn}(x,y)] dx dy, \tag{2a}
\]

and \( \sigma_{mn} = \sqrt{\int [W_{mn}(x,y)]^2 dx dy - \mu_{mn}^2} \) \( dx dy \) \( (2b) \)

Four primary feature locations are located by a hierarchical component-based feature recognizer in [26], which will provide the coordinate location for the center of the left eye, center of the right eye, tip of the nose and the center of the lips as shown in Figure 3a. The location point of the left and right eye features is being derived from the location of the center of the left eye and right eye denoted by the coordinates \((x_{LE}, y_{LE})\) and \((x_{RE}, y_{RE})\) respectively. The location of the nose bridge is the middle point of the left and right eye on the X-axis. The nose feature locations are derived from the location of tip of the nose denoted by the coordinates \((x_{NS}, y_{NS})\). The locations of lips features are derived from the center of lips coordinates \((x_{LS}, y_{LS})\).

![Figure 2. The Proposed Emotion Recognition System Diagram](image)

![Figure 3. Four primary Feature Locations and 60 Extended Local Features at various levels](image)
Each of feature point $F$ is the sub-matrix of the convolution output for the image with the Gabor filter bank. Each of the extended features is relative or an extension of the known features as shown in Figure 3a to Figure 3e. Table 1 shows the Gabor filter response of two emotions with different orientation of the Gabor wavelet.

Table 1. Gabor Filter Response of various emotions

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>Anger</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td></td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td></td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
</tbody>
</table>

2.2 Gabor Features to Facial Emotion Tree Structure (FETS) Transform

Based on Ekman’s [6] facial action coding system (FACS), we used similar areas of interest. In addition, the relationship information between the features and the details feature formed the Localized Gabor Feature vector. The Facial emotion can be represented by a 5 level deep tree structure model as shown in Figure 4, the entire face region acting as a root node and localized features upper and lower face and left, right and center of the faces became the second level branch nodes. At the third level nodes, the forehead, eyes, nose, mouth and cheek area became the corresponding branch nodes. At the fourth level, the forehead, eyes, eyebrows, nose, cheeks and mouth act as the branching nodes to the third level nodes. Sub-detail features from the 4 key fiducial points form the leaves of the tree structure. Features are grouped in the images shown in Figure 4; actual tree structure would have more connecting branches and arcs. The arcs between the two nodes corresponds to the object relationship, and features that been extracted are attached to the corresponding nodes.

Figure 4. Tree Structure Representation of the Human Face

3. ADAPTIVE PROCESSING OF FACIAL EMOTION TREE STRUCTURES

In this paper, the adaptive processing of tree structure proposed by Cho and Wong [1] was used for adaptive processing of the Facial Emotion Tree Structures (FETS). The learning algorithm and neural network architecture are addressed in the context of classification of structured patterns. The research works of [1, 20] was modified and used for the encoding method by recursive neural networks. The facial emotion relationship becomes the structure domain and all tree structures are used for presenting a learning set representing the task of the adaptive processing of data structures. The FETS can be simplified and represented in Figure 5a.

Figure 5. a) Simplified/Partial Tree Structure of the Facial Emotion. b) Encoded Tree Structure Format.

Figure 6. a) Architecture of Probabilistic based recursive neural network using GMM for neural Node Representation b) Structure of a Gaussian Mixture Model [1].

Probabilistic Neural Networks (PNNs) is one of the techniques used to embed discriminative information in the classification model in [1] and are successfully used for providing clustering analysis from the input attributes. Streit and Luginbühl [27] had demonstrated that by means of the parameters of a Gaussian mixture distribution, a probabilistic neural network model can estimate the probabilistic density functions. They showed that the
general homoscedastic Gaussian mixtures to approximate the optimum classifier could be implemented using a four layer feed-forward PNN using general Gaussian kernel, or Parzen window. Roberts and Tarassenko [28] proposed a robust method for Gaussian Mixture Models (GMMs), using a GMM together with a decision threshold to reject unknown data during classification.

The architecture of the neural network tree classifier shown in Figure 6a proposed by Cho and Wong [1], is used to represent each of the neural nodes as shown in Figure 5b. At the hidden layer, each neuron is represented by a Gaussian Mixture Model (GMM). At the output layer, each of the neurons is represented by a sigmoid activation function model. Each parameter has a specific interpretation and function in this GMM. All weights and node threshold are given explicitly by mathematical expressions involving the defining parameters of the mixture Gaussian pdf estimates and the a priori class probabilities and misclassification costs.

Suppose that a maximum branch factor of $c$ has been predefined, each of the form $q_i^{-1}$, $i = 1, 2, ..., c$, denotes the input from the $i$th child into the current node. This operator is similar to the shift operator used in the time series representation. Thus, the recursive network for the structural processing is formed as:

$$x = F_n(Aq^{-1}y + Bu),$$

$$y = F_p(Cx + Du),$$

where $x$, $u$, and $y$ are the $n$-dimensional output vector of the $n$ hidden layer neurons, the $m$-dimensional inputs to the neurons, and the $p$-dimensional outputs of the neurons, respectively. $q^{-1}$ is a notation indicating the input to the node is taken from its child so that: $q^{-1}y = \begin{pmatrix} q_1^{-1}y \\ q_2^{-1}y \\ \vdots \\ q_c^{-1}y \end{pmatrix}$. The parametric matrix $A$ is defined as: $A = \begin{pmatrix} A_1 & A_2 & \cdots & A_c \end{pmatrix}$, where $c$ denotes the maximum number of children in the tree, $A$ is an $n \times (c \times p)$ matrix such that each $A_i$, $i = 1, 2, ... , c$, is an $n \times p$ matrix, which is formed by the vectors $a_i^j$, $j = 1, 2, ... , n$. The parameters $B$, $C$, and $D$ are $(n \times m)$, $(p \times n)$ and $(p \times m)$-dimensional matrices respectively.

Let $m$ be the dimension of the input attributes in the Neural Node for each node, and $k$ be the dimension of the outputs of each node. Hence the input pattern at each GMM can be expressed as:

$$z = (u \quad q^{-1}y) = \{x_i; i = 1, 2, \ldots, (m + k \times c)\}$$

where $u$ and $z$ are the $m$-dimensional input vector and the $k$-dimensional output vector respectively. The class likelihood function of structure pattern $z$ associated with class $\omega$ would be expressed as:

$$p(z|\omega) = \frac{G}{g=1} P(\theta_g|\omega) p(z|\omega, \theta_g)$$

where $p(z|\omega)$ is the class likelihood function for class $\omega$ is a mixture of $G$ components in a Gaussian distribution. $\theta_g$ denotes the parameters of the $g$th mixture component and $G$ is the total number of mixture components. $P(\theta_g|\omega)$ is the prior probability of cluster $g$, and is termed as the mixture coefficients of the $g$th component:

$$\sum_{g=1}^{G} P(\theta_g|\omega) = 1$$

$$p(z|\omega, \theta_g) = N(\mu_g, \Sigma_g)$$

is the probability density function of the $g$th component, which typically is a form of Gaussian distribution with mean $\mu_g$ and covariance $\Sigma_g$.

Equation (9) is modified for the recursive network, and expressed as the following equation:

$$y = F_k(Wp + Vu),$$

where $F_k()$ is a $k$-dimensional vector, and their elements are the nonlinear sigmoid activation function. $p = (p_1, p_2, \ldots, p_k)$ and $W$ and $V$ are the weighting parameters in $(k \times r)$ and $(k \times m)$ - dimensional matrices respectively.

The learning scheme of the proposed probabilistic based recursive model can be divided into two phases, the locally unsupervised algorithm for the GMMs and the globally structured supervised learning for recursive neural networks. Streit and Lugimihauer has shown that in the unsupervised learning phase, the Expectation-Maximization (EM) method [27] would be optimal for this type of locally unsupervised learning scheme, which requires the parameters $\theta$ to be initialized and estimated during this learning phase. There are two steps in the EM method: The first step is called the expectation (E) step and the second is called the Maximization (M) step. The E step computes the expectation of a likelihood function to obtain an auxiliary function and the M step maximizes the auxiliary function refined by the E step with respect to the parameters to be estimated. The EM algorithm can be described as follows: Using the GMM in equation (10), the goal of the EM learning is to maximize the log likelihood of input attribute set in structured pattern, $z^* = (z_1, \ldots, z_{N_T})$,

$$\mathcal{L}(z^*, \theta) = \sum_{j=1}^{N_T} \sum_{g=1}^{G} \log p(z|\theta_g) + \log p(z|\omega, \theta_g),$$

where observable attributes $z^*$ is “incomplete” data, hence an indicator $\alpha^k_j$ is defined to specify which cluster the data belonged to and include it into the likelihood function as:
\( l(\mathbf{x}^*, \theta) = \sum_{j=1}^{N_T} \log P(\theta, \mathbf{x}^*) + \log p(\mathbf{x}^*) \),

where \( \alpha_j^k \) is equal to one if structure pattern \( \mathbf{x}_j \) belongs to cluster \( k \), else the output would be equal to zero. In E step, the expectation of the observable data likelihood in the \( n \)-th iteration would be taken as:

\[
Q(\theta, \hat{\theta}(n)) = \mathbb{E}[Q(\mathbf{x}^*, \theta) | \mathbf{x}^*, \hat{\theta}(n)],
\]

\[
= \sum_{j=1}^{N_T} \sum_{g=1}^{G} E[x_j^g, \hat{\theta}(n)] \left[ \log p(x_j^g, \theta) + \log p(\mathbf{x}_j, \hat{\theta}_g(n)) \right] \tag{17}
\]

where \( p(x_j^g, \theta) = s(\hat{x}_g(n), \hat{\Sigma}_g(n)) \) and \( E[x_j^g, \hat{\theta}(n)] = P(\theta, x_j^g, \hat{\theta}(n)) \) as the conditional posterior probabilities which can be obtained by Bayes’ rule:

\[
p(\theta, x_j^g, \hat{\theta}(n)) = \frac{p(\theta, x_j^g, \hat{\theta}(n))}{\sum_{g=1}^{G} p(\theta, x_j^g, \hat{\theta}(n))},
\]

at the \( n \)-th iteration. In M step, the parameters of a GMM are estimated iteratively by maximizing \( Q(\theta, \hat{\theta}(n)) \) w.r.t. \( \theta \).

At the next phase of supervised learning, fine-tuning in the decision boundaries of the GMMs is carried out simultaneously by a penalized heuristic strategy to obtain an optimal solution. The summary of the algorithm is as follows:

**Step 1: Initialization**

a. Set \( t_L = 0 \) and \( t_G = 0 \), where \( t_L \) and \( t_G \) are the iteration index of learning phase 1 and phase 2 respectively.

b. Initialize, by random, the parameters \( \Theta \) in the proposed probabilistic based recursive network.

c. Calculate the priori probability \( p(\theta, x_j^g, \hat{\theta}(n)) \) of each cluster by means of the input attributes and initial parameters \( \Theta \).

**Propagate the structured learning patterns to obtain the root output of each DAG.**

**Step 2: Locally Unsupervised Learning for GMMs (Phase 1)**

a. Given the input attributes \( \{x_1, \ldots, x_{N_T}\} \) of each node in all structured patterns.

b. For \( t_L = 1 \ldots T_L \):

The parameters \( \Theta_g : \{\mu_g, \Sigma_g, \pi_g\} \) of each GMM are determined iteratively by:

\[
\mu_g(t_L + 1) = \frac{\sum_{j=1}^{N_T} \pi_g \mu_j^g(t_L)}{\sum_{j=1}^{N_T} \pi_g}, \quad \Sigma_g(t_L + 1) = \frac{\sum_{j=1}^{N_T} \pi_g \Sigma_j^g(t_L)}{\sum_{j=1}^{N_T} \pi_g}.
\]

**Step 3: Globally Structural Supervised Learning (Phase 2)**

For \( t_G = 1 \ldots T_G \):

a. Calculate the outputs \( p(\theta, x_j^g) \) of each GMM in each structured pattern and class.

b. Fine-tuning the decision boundaries of each GMM using

\[
\pi_g(t_G + 1) = \frac{\pi_g(t_G) \sum_{j=1}^{N_T} \pi_g \mu_j^g(t_G) + \eta_\pi \sum_{j=1}^{N_T} \pi_g \mu_j^g(t_G) - \mu_g(t_G)}{\sum_{j=1}^{N_T} \pi_g \mu_j^g(t_G) - \mu_g(t_G)}, \quad \Sigma_g(t_G + 1) = \frac{\sum_{j=1}^{N_T} \pi_g \Sigma_j^g(t_G) + \eta_\Sigma \sum_{j=1}^{N_T} \pi_g \Sigma_j^g(t_G) - \Sigma_g(t_G)}{\sum_{j=1}^{N_T} \pi_g \Sigma_j^g(t_G) - \Sigma_g(t_G)},
\]

where \( \Gamma_j(t_G) = \sum_{j=1}^{N_T} \pi_g \mu_j^g(t_G) - \mu_g(t_G) \), \( \Sigma_g(t_G) \) is the posterior probability. \( \eta_\pi \) and \( \eta_\Sigma \) are the learning rates. \( D_a \) and \( D_b \) defined as the rejection set and the acceptance set respectively.

c. Estimate the weighting parameters iteratively by using penalized optimization algorithm

\[
a_k(t_G + 1) = a_k(t_G) + \eta_a \left[ \frac{\partial J}{\partial a_k} \right],
\]

\[
\frac{\partial J}{\partial a_k} = - \left[ a_j^k(t_G) - a_k(t_G) \right] \theta_j^k \frac{\partial a_k(t_G)}{\partial a_k} + \frac{2}{\beta} \left[ a_k(t_G) - a_k^*(t_G) \right. + \left. \beta \left( a_k(t_G) - a_k^*(t_G) \right) \right],
\]

27
where η is the learning rate. The derivative $\frac{\partial q}{\partial y_k}$ is a $(r + m) \times p$ matrix with the output gradients of the child nodes with respect to the weights.

Step 4: Algorithm complete

4. PERFORMANCE EVALUATIONS

The evaluation is based on the Japanese Female Facial Expression (JAFFE) Database [24] which contains 213 images of 7 facial expressions (including neutral) posed by 10 Japanese female models. The original images were of the size 256 x 256 pixels as shown in Figure 7. The face region were cropped out from the original images and resized to 200 x 200 pixels. The location of the eyes, nose and center of mouth can be easily detected. We have tested our system for robustness by testing under known person and unknown person conditions. The distribution of the training and test set are highlighted in Table 2. For known subjects, the training set contains 143 images of the 10 models and 7 facial expressions. The test set contains the remaining 70 images for all the models and expressions. The purpose of having unknown subjects is to test for robustness of the system in recognizing an emotion for a person not found in the training set. For our training set, we used 8 subjects with all 170 images in various expressions. The test set contains the remaining 43 images.

Table 2 is tabulated the performance of the FETS model against other methods, i.e. Naïve Bayes [23], SVM [21] and KNN [22]. They were implemented in Weka data mining package [31] and the initial parameters setting were based on their default settings. These three models used the same features extracted in Section 2.1 as a flat vector instead of FETS representation. The results show that our proposed method is the most robust in terms of recognition rate. The proposed method scored about 89% for known subjects and 86% for unknown subjects.

Table 4 shows the recall and generalization performance of the FETS versus the QuadTree structure using the probabilistic recursive model. Recall rate suggested how well the system is trained for recognizing the same training set; however, it may also indicate that the system is not robust when generalization rate is low. The result shows that using FETS as a tree representation has a better generalization rate that using Quad tree representation, i.e. FETS is more robust than Quad tree. Figure 8 shows the relationship between the different depth levels of the tree structure and the accuracy of the recognition.

Table 5 shows the confusion matrix among the various emotions recognized using the FETS model. The overall error rates are about 12% and 16% for known subjects set and unknown subjects set respectively.

From these results, our proposed model performed better because of the adaptive modeling of the FETS, where the relationship information between the feature vectors is preserved. Conventional methods such as SVM, KNN and Naïve Bayes lose such relationship information, hence producing lower recall rates for unknown samples testing.

<table>
<thead>
<tr>
<th>Table 2. Distribution of Training and Test images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Known</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3. Performance of FETS model against other methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Navie Bayes</td>
</tr>
<tr>
<td>SVM</td>
</tr>
<tr>
<td>KNN</td>
</tr>
<tr>
<td>FETS + PR</td>
</tr>
</tbody>
</table>

Figure 7. Original image of 256 x 256 pixels of various persons in JAFFE database.

Figure 8. Performance of Tree structure using different number of Tree Level Node Depths
Table 4. Recall and Generalization Rates of FETS vs QuadTree using probabilistic recursive model

<table>
<thead>
<tr>
<th>Subjects</th>
<th>QuadTree</th>
<th>FETS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall Rate</td>
<td>Generalization Rate</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>Dev</td>
</tr>
<tr>
<td>Known</td>
<td>95.00%</td>
<td>0.25%</td>
</tr>
<tr>
<td>Unknown</td>
<td>95.12%</td>
<td>0.60%</td>
</tr>
</tbody>
</table>

Table 5. Confusion Matrices showing the interclass classification errors using FETS model

<table>
<thead>
<tr>
<th>Known Subjects</th>
<th>AN</th>
<th>DI</th>
<th>FE</th>
<th>HA</th>
<th>NE</th>
<th>SA</th>
<th>SU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger (AN)</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Disgust (DI)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fear (FE)</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Happy (HA)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Neutral (NE)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sad (SA)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Surprise (SU)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unknown Subjects</th>
<th>AN</th>
<th>DI</th>
<th>FE</th>
<th>HA</th>
<th>NE</th>
<th>SA</th>
<th>SU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Overall error rate: 12.8% 16.3%

5. CONCLUSIONS

This proposed approach of converting facial emotion feature vectors to facial emotion tree structure representation and using adaptive processing of tree structures method hold a strong recommendation. Previous works by Donato et al. [8] has successfully classify the individual facial action, but not the cognitive emotional states represented by the combination of such facial actions. Our proposed system achieved recognition rate of about 89% for known subjects and 82% for unknown subjects for the cognitive emotional states represented by the combination of such facial actions. Our approach only made use of a total of 854 features to represent in the tree structure, which is comparatively smaller than other feature methods.

6. REFERENCES


