Abstract—Automatic analysis of human facial expression is one of the challenging problems in machine vision systems. The most expressive ways humans display emotion is through facial expression. In this paper, we extend texture based facial expression recognition, with a method of 2D image processing implemented for extraction of features and new neural network based decision trees. The proposed algorithm applied set of pre-processing and divided image into two main parts (eyes and lips parts), Then, the Discrete Cosine Transform (DCT) implemented on each part to reduce image data size in different parts of the face. Different decision tree models have examined in order to find the best recognition rate. Experimental results have shown that the combination of decision tree with neural network to identify different facial expressions could improve the recognition rate significantly.

Keywords—Facial Expression Recognition; DCT; Neural Network; Decision Tree

I. INTRODUCTION

The use of facial expression for measuring people’s emotions has dominated psychologically since 1960s [1]. The research studied in [2] has indicated that 7% of communication information between people transferred by linguistic (Verbal part), 38% by paralanguage (Tone of voice), and 55% by facial expression in human face to face communications. Therefore, this shown that a large amount of information is hidden in the human facial expression.

Automatic Facial Expression recognition plays a significant role in human computer interaction systems such as robotics, machine vision, monitoring of stress, and etc. [3]-[5]. Recently, significant advances have been achieved in the area of face recognition and facial expression recognition [6]-[9]. However there are still many challenges remaining. For example, face recognition in uncontrolled environments and conditions still bears limitations due to light illumination, head pose and person identity [10].

Human faces are extremely similar, thus the extraction of facial features characteristics and selection of appropriate classifier are two key ways to solve the facial expression recognition problems. Two main methods in the current research on facial expression recognition are texture information (e.g. Pixel intensity) and geometrical based information (e.g. Muscle action detection). The most frequently used texture features are Gabor filter output [11]-[13], Pixel intensity [14]-[15], discrete cosine transform features [16]-[18] and skin color information [19]. Furthermore, in order to enhance texture information, feature extraction methods based on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) had been used [20]. Different classifiers were applied on texture information, using neural networks, support vector machine, empirical classification rules and Bayes classifier [13]-[19].

It is found that facial expression is usually correlated with identity [21], and variations in identity (extra-personal features) dominate over those in expression (intra-personal features). The unsolved problem might be expression recognition of a novel person, but not in the database. Therefore, most of existing algorithms, which seem to work well on person dependent (PD) expression recognition, are substantially less efficient on person independent (PI) algorithms.

In this study, we proposed an approach combining the most effective texture feature extraction and neural network based decision trees for gray scale face image. The recognized facial expressions are 6 basic ones (anger, happiness, sadness, surprise, fear, disgust) which include main facial gestures. All the simulations have been done on both person dependent (PD) and person independent (PI) database. In this research, firstly face position is detected in whole image and local features are extracted from two separated parts of the image. There are also local decision trees to recognize and vote for one of six facial expressions. In order to study the systems performance on PD and PI samples, there is also a face recognition system implemented beside other networks to identify human face.

The rest of paper is organized as follows. Section 2, introduces image pre-processing steps and proposed feature extraction method. Section 3, represents and explains different decision trees studied in this research. Section 4, illustrates experimental results and best recognition rates achieved and Section 5 concludes the paper.
II. TEXTURE FEATURE EXTRACTION METHOD

In the following, it is assumed that images from a face-image database are used for facial expression recognition. Relevant details on face-image database is given later in section 4. The first steps of feature extraction would be preprocessing of database images. In the second step, the partitioning of detected face is applied to output images from step 1. Using discrete cosine transform (DCT), preprocessed images are transformed in frequency domain and compressed to form the feature vectors (step 3). Finally, low frequency information of the transformed images are used as input vectors for the classifier. Fig. 1 shows these steps for image pre-processing and feature extraction method. In the following, we describe each of these steps.

A. Preprocessing and partitioning

With the view to limit the processing area of an image only to required parts of the face. We first detect face with a CERT face detector, which is a Viola and Jones type [22]. Then, we crop the image to remove extra borders in the picture, such as hairs, chin and etc. Finally, all the images are resized to the uniform size of 240 × 260 pixels.

Fig. 1 shows how preprocessing algorithm departs each image into two upper and lower regions. Upper part contains eyes and eyebrows and lower part contains mouth and lips. The reason why face analysis is done in two separated parts is behind the data hid in different facial components. Human faces upper facial components have the ability to move and rotate (like eyebrows) but lower facial components could be reshaped (like mouth and cheeks) and each takes totally different changes in various facial expressions. We found that defining facial component into two main groups gives satisfactory results in feature extraction.

B. Discrete Cosine Transform

Obviously, it is difficult for the classifier to deal with preprocessed images of expression models, as they are large sets of data. To facilitate the recognition process, we need to compress all the train and test images to reduce data in a proper way, without losing key features. The two dimensional discrete cosine transform (2D-DCT) is a powerful way for image frequency transform that can either compress the image in frequency domain or use images feature. The DCT is a real domain transform which represents entire image as coefficients of different frequencies of cosines.

In frequency domain where the lower frequencies present high magnitudes and higher frequencies relatively indicate smaller magnitudes. Thus, the higher frequency components could be ignored without damaging the original image. The 2D-DCT of an image gives the result matrix such low frequency components stay in the top left corner, while high frequency components gather in the bottom right corner of DCT matrix, Fig. 1. The 1D discrete cosine transform is defined as (1):

\[
C(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos \left( \frac{(2x+1)u\pi}{2N} \right)
\]  

(1)

And the corresponding 2D transform is defined as (2):

\[
C(u, v) = \alpha(u) \alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \left( \frac{(2x+1)u\pi}{2N} \right) \cos \left( \frac{(2y+1)v\pi}{2N} \right)
\]  

(2)
Where:

\[
\alpha(u) = \begin{cases} 
\sqrt{\frac{1}{N}} & \text{for } u = 0 \\
\sqrt{\frac{2}{N}} & \text{for } u = 1, 2, \ldots, N - 1
\end{cases} \quad (3)
\]

And

\[
\alpha(v) = \begin{cases} 
\sqrt{\frac{1}{N}} & \text{for } v = 0 \\
\sqrt{\frac{2}{N}} & \text{for } v = 1, 2, \ldots, N - 1
\end{cases} \quad (4)
\]

A small \(L_1 \times L_2\) for low frequency part of the \(N \times N\) original image would be enough as input vector of NNs for training and simulating.

III. PROPOSED RECOGNITION TECHNIQUE

Neural Network based Decision Tree is selected as a classification function. One distinctive advantage of this classifier for facial expression recognition is the step by step detection format in tree nodes. Obviously, different facial expression has some similarities which makes harder for classifiers to decide. Thus, for some facial expressions are it’s easier to be recognized correctly than others. For example, for a testing sample like sad expression, it’s much easier to be misrecognized with happy expression rather than surprised. From our point of view, it could be productive to separate different expressions with a priority, using the similarity of expressions in both upper and lower facial components.

Motivated by the above facts, in the proposed decision tree, former nodes use the facial similarities to classify similar expressions in main groups and later nodes try to separate new groups into smaller classes \[17\]. Hence, the difference of similar classes would get more attention in training process of classifier from the classification viewpoint.

Fig. 2 shows different NN based decision trees used in this research where each node is implemented with a back propagation feed forward neural network. In Fig. 2a there are three nodes defining the faces class, Fig. 2b consists of four nodes and Fig. 2c illustrates five nodes in the decision tree. Experimental results show that the result of this step by step method will be way more accurate than single neural network \[14\] and implementing more nodes to define facial expressions gives more accuracy to the system. It is shown in Fig. 2c, node 1 is trained and placed to divide “Anger-Sadness-Disgust-Happy” category from “Surprise-Fear” category, in the next step, node 2 divides “Anger-Sadness-Disgust” category from “Happiness” and node 3 separates “Surprise” from “Fear”. Next, node 4 divides “Anger-Sadness” category from “Disgust” and at last “Anger” will separate from “Sadness” in node 5.

In the proposed method, each face has two feature vectors. It’s because distinct changes in upper-face facial components and lower-face facial components. Results of both classifiers are then used to make an accumulated decision. Dimension of all NNs input vector would be \(L_1 \times L_2\) which is the same for all NNs, Fig 1.

IV. EXPERIMENTAL STEUP

All the results are experimented in MATLAB\textsuperscript{TM} software. We first study the performance of the system on a specific database and the next database used to verify the experimental results.

A. Data Preparation

Two below facial expression databases are used to verify the proposed method. (1) Japanese Female Facial Expression (JAFFE) database (2) Taiwanese Facial Expression Image (TFEID) Database. JAFFE database contains 213 images of 10 Japanese female models, including six main facial expressions, Fig. 3. In each database, we divide the database into two parts, train set and test set. Since the dividing rule was according to each person’s photos, the same person will not appear in both training and learning, it is also person independent(PI). Thus, the system deals with expression of
a novel person and the resulting recognition accuracy is substantially lower than person dependent (PD) samples.

B. Experimental Setting

In the experiments, the $L_1 \times L_2$ size of feature vectors are 8x8 blocks, which are low frequency DCT components. Bigger blocks would significantly result in higher recognition accuracy but greater computational cost. Therefore, for all the following experiments we assume the feature vectors are 8x8 length elements. For classification, 240 images of 20 men and 20 women having 6 different facial expressions are taken. One-hidden-layer neural networks are considered in this work, which are trained with the “TRAINGDX” program in MATLAB toolbox. The training process is terminated within 200 epochs and mean square error (MSE) reaching pre-specified goal $10^{-5}$. The nodes of decision tree shows different performance on various NNs and the mean of recognition rates are subject to transcribe.

C. Experimental Results

As we expect an increase in accuracy is achieved. Table 1 shows the mean achieved classification accuracy rate for each facial expression on JAFFE database with different proposed decision trees. Table 2 implies on the same recognition accuracies on TFEID database.

Table 1. Recognition Rate (%) on JAFFE Database with Different Decision Trees

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Surprise</th>
<th>Happy</th>
<th>Disgust</th>
<th>Fear</th>
<th>Angry</th>
<th>Sad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree a</td>
<td>77.9</td>
<td>60.1</td>
<td>54.8</td>
<td>61</td>
<td>58.7</td>
<td>60</td>
</tr>
<tr>
<td>Tree b</td>
<td>75.6</td>
<td>63.3</td>
<td>63.1</td>
<td>59.4</td>
<td>68.1</td>
<td>70.5</td>
</tr>
<tr>
<td>Proposed</td>
<td>79.5</td>
<td>75.2</td>
<td>66</td>
<td>62.8</td>
<td>73.9</td>
<td>78.4</td>
</tr>
</tbody>
</table>

Table 2. Recognition Accuracy (%) on TFEID Database with Different Decision Trees

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Surprise</th>
<th>Happy</th>
<th>Disgust</th>
<th>Fear</th>
<th>Angry</th>
<th>Sad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree a</td>
<td>95.7</td>
<td>77.4</td>
<td>71</td>
<td>77.2</td>
<td>79.5</td>
<td>75.1</td>
</tr>
<tr>
<td>Tree b</td>
<td>98.2</td>
<td>72.2</td>
<td>74.5</td>
<td>74.3</td>
<td>84.8</td>
<td>84.8</td>
</tr>
<tr>
<td>Proposed</td>
<td>98.1</td>
<td>81.8</td>
<td>83.4</td>
<td>79</td>
<td>84</td>
<td>80.2</td>
</tr>
</tbody>
</table>

Table 3. Confusion Matrix (%) for the Best Result of Proposed System on TFEID

<table>
<thead>
<tr>
<th></th>
<th>Surprise</th>
<th>Happy</th>
<th>Disgust</th>
<th>Fear</th>
<th>Angry</th>
<th>Sad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surprise</td>
<td>98.1</td>
<td>0.8</td>
<td>0</td>
<td>1.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Happy</td>
<td>4</td>
<td>81.8</td>
<td>4.1</td>
<td>0.5</td>
<td>2.1</td>
<td>7.5</td>
</tr>
<tr>
<td>Disgust</td>
<td>0</td>
<td>1.3</td>
<td>83.4</td>
<td>9.3</td>
<td>3.3</td>
<td>2.7</td>
</tr>
<tr>
<td>Fear</td>
<td>1.1</td>
<td>0</td>
<td>16.9</td>
<td>79</td>
<td>1.9</td>
<td>1.1</td>
</tr>
<tr>
<td>Angry</td>
<td>0.5</td>
<td>0</td>
<td>2.5</td>
<td>0</td>
<td>84</td>
<td>13</td>
</tr>
<tr>
<td>Sad</td>
<td>1.9</td>
<td>0</td>
<td>4.9</td>
<td>0</td>
<td>13</td>
<td>80.2</td>
</tr>
</tbody>
</table>

Fig. 4. Best recognition results (%) for JAFFE and TFEID databases
Table 4 compares the results of proposed method with three other similar works. As we expect an increase in accuracy is achieved. It should be noted that, the results are not directly comparable due to different databases, preprocessing methods and experimental setups.

V. CONCLUSION

In this paper, a new approach is proposed for facial expression classification. The similarities between facial components in different facial expressions are used to improve recognition rates. In the proposed method, a decision tree with five neural network based nodes designed to separate six different facial expressions in four steps. The preprocessing is used to improve the training data to increase the accuracy rate substantially. Experimental results on two widely used databases are presented to demonstrate the efficiency of the proposed method. Working on hard to recognize nodes in this decision tree gives an interesting topic for future investigations.

REFERENCES