Estimation of Human Emotions Using Thermal Facial Information

Hung Nguyen¹, Kazunori Kotani¹, Fan Chen¹, Bac Le²

¹Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, Japan
²University of Science, 227 Nguyen Van Cu, Ho Chi Minh city, Vietnam

ABSTRACT

In recent years, research on human emotion estimation using thermal infrared (IR) imagery has appealed to many researchers due to its invariance to visible illumination changes. Although infrared imagery is superior to visible imagery in its invariance to illumination changes and appearance differences, it has difficulties in handling transparent glasses in the thermal infrared spectrum. As a result, when using infrared imagery for the analysis of human facial information, the regions of eyeglasses are dark and eyes’ thermal information is not given. We propose a temperature space method to correct eyeglasses’ effect using the thermal facial information in the neighboring facial regions, and then use Principal Component Analysis (PCA), Eigen-space Method based on class-features (EMC), and PCA-EMC method to classify human emotions from the corrected thermal images. We collected the Kotani Thermal Facial Emotion (KTFE) database and performed the experiments, which show the improved accuracy rate in estimating human emotions.

Keywords: Thermal facial information, estimation human emotion, principal component analysis, eigen-space method based on class-feature, KTFE database.

1. INTRODUCTION

Nowadays, Human – Computer Interaction (HCI) is a very attractive research area in computer vision. One of the key researches in HCI is to detect the inner emotions through human faces by performing automatic analysis of facial expressions. Many previous works have proposed towards developing facial emotion estimation [1]. However, we still lack an accurate and robust facial emotion estimation method to be deployed in uncontrolled environments. Several factors that affect facial emotion estimation include pose variations, occlusions, and most importantly, illumination changes [1]. Therefore, it is a new and imaginative way to use IR imagery, which is not sensitive to light condition, to fill the gap in the human emotion estimation field. Besides, human emotions could be manifested by changing temperature of face skin which is obtained by IR camera. Consequently, thermal infrared imagery gives us more information to help us robustly estimate the human emotions.

Recently, a number of studies have demonstrated that thermal infrared imagery offers a promising alternative to visible imagery in facial emotion estimation problems by better handling the visible illumination changes. Y.Yoshitomi et al. used two dimensional detection of temperature distribution on the face using infrared rays [2]. Based on studies in the field of psychology, several blocks on the face are chosen for measuring the local temperature difference. With Back Propagation Neutral Network, the facial expression is recognized. The recognition accuracy reaches 90% with neutral, happy, surprising and sad expressions. However, the testing database is obtained from only one female frontal view. Y. Yoshitomi generated feature vectors by using a two-dimensional Discrete Cosine Transformation (2D-DCT) to transform the grayscale values of each block in the facial area of an image into their frequency components, and used them to recognize five expressions, including “angry”, “happy”, “neutral”, “sad”, and “surprise”. The mean expression accuracy is 80% with four test subjects [3]. Y.Koda et al. used the idea from [3] and added a proposed method for efficiently updating of training data, by only updating the training data with “happy” and “neutral” facial expression after an interval [4]. The expression accuracy increased from 80% to 87% with this new approach. Sophie Jarlier et al. extracted the features as representative temperature maps of nice action unit (AUs) and used K-nearest neighbor to classify seven expressions [5]. The database for testing has four persons and the accuracy rate is 56.4%. M.M.Khan et al. suggested using Facial Thermal Feature Points (FTFPs), which are defined as facial points that undergo significant thermal changes in presenting an expression, and used Linear Discriminant Analysis (LDA) to classify intentional facial expressions based on Thermal Intensity Values (TIVs) recorded at the Facial Thermal Feature Points (FTFPs) [6]. The database has 16 persons with 5 expressions and the accuracy rate ranges from 66.3% to 83.8%. L.Trujillo et al. proposed using a local and global automatic feature localization procedure to perform facial expression in thermal images. They
used PCA to reduce the dimension and interest point clustering to estimate facial feature localization and Support Vector Machine (SVM) to classify three expressions [7]. B.Hernandez et al. used SVM to classify the expressions “surprise”, “happy”, “neutral” from two inputs. First input consists of selection of a set of suitable regions where the feature extraction is performed, second input is the Gray Level Co-occurrence Matrix used to compute region descriptors of the IR images [8]. B.R.Nhan et al. extracted time, frequency and time-frequency features from thermal infrared data to classify the natural responses in terms of subject-indicated levels of arousal and valence stimulated by the International Affective Picture System [9]. All these studies with thermal infrared imagery have shown that human emotion states are related to the facial skin temperature property. However, to our knowledge, most approaches that use the extracted features from a single infrared thermal image may lose some useful information which could be contained in the sequences.

Although there are many significant advantages when we use thermal imagery, IR have several drawbacks. One of drawbacks is eyeglasses. Glass is opaque to IR and the sensitivity of thermal IR to facial occlusions is decreased by eyeglasses. This is because objects made of glasses act as a temperature screen, completely occluding the parts located behind them. To eliminate the effect of presence of eyeglasses, we propose a temperature space method for thermal infrared data. We will also reduce the impact of ambient temperature by normalizing it between frames. To estimate five emotions, we use PCA, PCA-EMC, and EMC to extract feature vectors and then find the similarity between the testing data and training data.

2. METHOD

In this section, we propose a method to reduce the effect of eyeglasses in thermal data and then use three conventional methods PCA, EMC, and PCA-EMC to analyze emotions. Before applying temperature spaces, to avoid the temperature change of ambience from frame to frame, we calculate the mean of ambient point temperatures of each frame. Then we find the difference of the mean of each frame to that of its previous frame and update the temperature of all points of each frame by subtracting those temperatures with the difference.

2.1 Temperature spaces

We propose using temperature spaces to support the temperature analysis of human face area. Based on temperature data and observation, we classify images into two main spaces, i.e., non-face space and face space. Non-face space is for the background area, which further includes two sub-spaces. In face space, there are three sub-spaces because the eyebrows and nose are usually cold whereas the cheek is warm [10], and the forehead is usually the warmest. Using the spaces, we find eyeglasses area and replace each point of them by the mean of image temperature.

Let h and g be maps from image to temperature space and image to glass space respectively, as the following:

\[ h: I \rightarrow T \quad ; g: I \rightarrow G \quad ; \quad g \circ h(i, j) = \begin{cases} 0 & \text{if } (i, j) \in \text{glass} \\ h(i, j) & \text{if } (i, j) \not\in \text{glass} \end{cases} \]

where I is image space, T is temperature space, and G is glass space.

We calculate the temperatures for image and face as following:

\[ \Delta T_i = T_{i,\text{Max}}^I - T_{i,\text{Min}}^I; \bar{\sigma}T_i = \Delta T_i / 5 \]

where \( T_{i,\text{Max}}, T_{i,\text{Min}} \) are maximum and minimum of temperature of each image, respectively.

\[ I'_k = \left\{ \frac{f(i, j)}{T_{i,\text{Min}}} + \bar{\sigma}T_i \ast (k - 1) \leq f(i, j) < \frac{T_{i,\text{Max}}}{T_{i,\text{Max}}} - \bar{\sigma}T_i \ast (5 - k), (i, j) \in I \right\} \]

where \( k \in (1, 5) \)

\[ \Delta T_F = T_{F,\text{Max}} - T_{F,\text{Min}}; \bar{\sigma}T_F = \Delta T_F / 5 \]

where \( T_{F,\text{Max}}, T_{F,\text{Min}} \) are maximum and minimum of temperature of each human face area, respectively.

\[ I'_F = \left\{ \frac{f(i, j)}{T_{i,\text{Min}}} + \bar{\sigma}T_F \ast (i - 1) \leq f(i, j) < \frac{T_{i,\text{Max}}}{T_{i,\text{Max}}} - \bar{\sigma}T_F \ast (5 - i), (i, j) \in F \right\} \]

where \( i \in (1, 5) \)
Based on (3), (5), (6), we find the glass area as followings:
\[
glass = \{ f(i, j) / f(i, j) < T_{min}^f, (i, j) \in F \}
\]  
(6)

After using our proposed method to find eyeglasses’ area in thermal data, to reduce the effect of eyeglasses, we replace the temperature of the areas by the mean of image temperature.

### 2.2 Estimate emotion using PCA, EMC, and PCA-EMC

After reducing the effect of eyeglasses, we use three conventional methods PCA, EMC, and PCA-EMC to estimate the human emotion using thermal information of emotion.

With PCA, the aim is to build a face space, including the basis vectors called principal components, which better describes the face images [11]. To estimate emotions using PCA, we divide the training set into five classes and compute the eigen-space of each class as following:

Step 1: Each thermal data frame is shaped from 2D matrix to 1D vector. Given $M_t$ frames of thermal data as training data, we convert these datum to corresponding column vectors.

Step 2: Find covariance matrix that represents the scatter over the mean of training data
\[
S = \frac{1}{M_f} \sum_{k=1}^{M_f} (x_k - \bar{x}) (x_k - \bar{x})^T ; \quad \bar{x} = \frac{1}{M_f} \sum_{k=1}^{M_f} x_k
\]  
(8)

where $x_k$ is an N-dimension vector, $M_f$ is total number of frames of each class and S is covariance matrix.

Step 3: Solve the eigenvalues problem $S \phi=\lambda \phi$ and obtain the eigenvalues and corresponding eigenvectors.

Step 4: Choose the largest H eigenvalues and eigenvectors. Eigen-space includes H eigenvectors.

For each test thermal data, we project it to the eigen-space of each class and derive the reconstruction thermal data from each eigen-space. Using mean square error, measuring the similarity, between input thermal data and reconstruction thermal data, we can choose a suitable class for input thermal data which is a minimum of the mean square errors [11].

In [12] and [13], the authors suggest using Eigen-space Method based on Class-features (EMC) to analyze the facial expressions. The difference between PCA and EMC is that PCA finds the eigenvector to maximize the total variance of the projection to line, while EMC obtains eigenvector to maximize the difference between the within-class and between-class variance. The difference between the within-class and between-class variance is calculated as following:
\[
S = S_B - S_W ; \quad S_B = \frac{1}{M} \sum_{f \in F} M_f (x_f - \bar{x})(x_f - \bar{x})^T ; \quad S_W = \frac{1}{M} \sum_{f \in F} \sum_{m=1}^{M_f} (x_{fm} - \bar{x})(x_{fm} - \bar{x})^T
\]  
\[
\bar{x}_{fm} = \frac{1}{M_f} \sum_{m=1}^{M_f} x_{fm} ; \quad \bar{x} = \frac{1}{M} \sum_{f \in F} \sum_{m=1}^{M_f} x_{fm}
\]  
(9)

where $F$ is a set of expression classes, $M_f$ facial-patterns are given for each class $f \in F$ and $x_{fm}$ is an N-dimension vector of the m-th facial pattern, $m = 1, M_f$.

To estimate emotion using EMC, we divide the training set into five classes and compute the eigen-space including eigen-vectors of difference matrix. For each test thermal data, we project it into the eigen-space of each class and an emotion is chosen if it gets maximum of cosine of angle between obtained vector after projection and eigenvector of each class. With PCA-EMC, we use PCA to reduce the dimension and apply EMC to the obtained data.

### 3. EXPERIMENTS

For database, we use KTFE (Kotani Thermal Facial Emotion) database [14]. This database contains 26 subjects who are Vietnamese, Japanese, Thai from 11 year-olds to 32 year-olds with seven emotions. The example of thermal images
is shown in Fig. 1. In our experiments, we use only sequence thermal data to estimate human emotions. From 130 GB thermal data, we extracted 2.4 GB of sequence thermal data for five emotions.

![Examples thermal images of five expressions (neutral, fear, anger, happiness, sadness)](image)

In our experiments, we use PCA, EMC, and PCA-EMC with before and after removal of eyeglasses. From the obtained database, we separate the training and testing data as 60% and 40% of the total thermal data.

Table 1 shows the results with EMC. The results show that accuracy increases with happy, sad, and neutral emotions but remains constant with angry and neutral emotions.

Table 2 shows the results with PCA. The accuracy of angry emotion increases to a large extent. However, with neutral emotion, the accuracy decreases. In total, the mean accuracy rate increases.

Table 3 shows the results with PCA-EMC. Although the mean accuracy rate increases, the accuracy of angry and happy emotion decreases. In totality, the mean accuracy rate is only slightly increased because some information is lost when we reduce the dimension by using PCA.

### TABLE I. EMC WITH BEFORE AND AFTER REMOVAL OF EYEGLASSES

<table>
<thead>
<tr>
<th></th>
<th>Ag</th>
<th>Ha</th>
<th>Fe</th>
<th>Ne</th>
<th>Sa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag</td>
<td>81.33</td>
<td>-</td>
<td>14.67</td>
<td>4.00</td>
<td>-</td>
</tr>
<tr>
<td>Ha</td>
<td>2.71</td>
<td>64.64</td>
<td>12.03</td>
<td>4.12</td>
<td>16.50</td>
</tr>
<tr>
<td>Fe</td>
<td>5.88</td>
<td>12.28</td>
<td>55.60</td>
<td>1.18</td>
<td>25.06</td>
</tr>
<tr>
<td>Ne</td>
<td>5.17</td>
<td>10.67</td>
<td>9.97</td>
<td>58.29</td>
<td>15.90</td>
</tr>
<tr>
<td>Sa</td>
<td>1.36</td>
<td>11.26</td>
<td>12.83</td>
<td>10.26</td>
<td>64.29</td>
</tr>
<tr>
<td>Av.</td>
<td>64.83</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### TABLE II. PCA WITH BEFORE AND AFTER REMOVAL OF EYEGLASSES

<table>
<thead>
<tr>
<th></th>
<th>Ag</th>
<th>Ha</th>
<th>Fe</th>
<th>Ne</th>
<th>Sa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag</td>
<td>93.33</td>
<td>-</td>
<td>7.00</td>
<td>6.77</td>
<td>4.00</td>
</tr>
<tr>
<td>Ha</td>
<td>1.46</td>
<td>67.80</td>
<td>8.07</td>
<td>2.89</td>
<td>19.78</td>
</tr>
<tr>
<td>Fe</td>
<td>11.37</td>
<td>9.46</td>
<td>57.13</td>
<td>8.59</td>
<td>13.45</td>
</tr>
<tr>
<td>Ne</td>
<td>8.15</td>
<td>9.09</td>
<td>12.99</td>
<td>58.29</td>
<td>11.48</td>
</tr>
<tr>
<td>Sa</td>
<td>1.18</td>
<td>10.14</td>
<td>15.39</td>
<td>5.60</td>
<td>67.70</td>
</tr>
<tr>
<td>Av.</td>
<td>66.45</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### TABLE III. PCA-EMC WITH BEFORE AND AFTER REMOVAL OF EYEGLASSES

<table>
<thead>
<tr>
<th></th>
<th>Ag</th>
<th>Ha</th>
<th>Fe</th>
<th>Ne</th>
<th>Sa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag</td>
<td>81.33</td>
<td>-</td>
<td>14.67</td>
<td>4.00</td>
<td>-</td>
</tr>
<tr>
<td>Ha</td>
<td>5.10</td>
<td>65.46</td>
<td>6.59</td>
<td>3.97</td>
<td>18.88</td>
</tr>
<tr>
<td>Fe</td>
<td>9.42</td>
<td>14.47</td>
<td>55.60</td>
<td>1.18</td>
<td>19.33</td>
</tr>
<tr>
<td>Ne</td>
<td>9.71</td>
<td>18.98</td>
<td>5.15</td>
<td>59.62</td>
<td>6.54</td>
</tr>
<tr>
<td>Sa</td>
<td>2.71</td>
<td>11.73</td>
<td>14.01</td>
<td>2.59</td>
<td>68.96</td>
</tr>
<tr>
<td>Av.</td>
<td>66.19</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### 4. CONCLUSION

This paper describes a novel way to reduce the effect of changing ambient temperature which occurs during the data acquisition and the proposed method to reduce the effect of eyeglasses using temperature space in thermal data.
estimate human emotion and prove the efficiency of the proposed method, three convention methods, namely PCA, EMC, PCA-EMC, are used with KTFE database.

Since the eyeglasses areas are replaced with the averaged-ambient temperature, these is no more effect of ambient temperature to these areas. The experiment with after removing glasses and before removing glasses shows the increasing of accuracy rate. Specially, with PCA, the accuracy of angry emotion increases from 59.17% to 93.33%. In general, the accuracy of each emotion and each method has been improved with the correction of eyeglass areas. However, in some case, the loss of information causes the decrease of accuracy rate. In the future, to improve the efficiency of the proposed method, we intend to combine visible sequence image and thermal sequence image to have fused-model.

REFERENCES


