THE EXTRACTION AND THE RECOGNITION OF FACIAL FEATURE STATE TO EMOTION RECOGNITION BASED ON CERTAINTY FACTOR

1 IGDE ARIS GUNADI, 2 AGUS HARJOKO, 3 RETANTYO WARDYO, 4 NEILA RAMDHANI

1 Student Doctoral Program in Computer Science UGM
2, 3 Doctoral Program in Computer Science UGM
4 Doctoral Program Psychology UGM
E-mail: 1 aria_sukaat@yahoo.com, 2 aharjoko@ugm.ac.id, 3 rw@ugm.ac.id, 4 neila_psi@ugm.ac.id

ABSTRACT

Psychologically, emotion is related to someone’s feeling in particular condition. Some fields like: health, psychology, and police investigation need the information of emotion recognition. Human’s emotion can be classified into six types include happy, sad, angry, fear, disgusted, and normal. Psychologically, there are some methods that can be used to emotion recognition, like self-report analysis, automatic measure, startle response magnitude, FMRI analysis (Functional Magnetic Response Imaging), and behavior response. Each of those methods has their own advantage and disadvantage.

The aim of this research is to determine someone’s emotion in a video scene. The video was decomposed into image frame and in each image frame was extracted into feature (component) face, which include mouth, eyes, nose, and forehead. The feature extraction was done by combining two methods based on the color and the facial geometric figure. The state of each face feature related to AU’s (Action Unit) face that used to emotion recognition. State recognition of mouth and eyes can be seen based on the feature elongation, state on the forehead and nose are known based on the wrinkle density. In the last part of emotion, recognition is done with certainty factor method to determine the quality of each emotion, the classification of actor’s emotion is determined based on the quality level of maximum emotion. The results showed recognition of emotion in a single image, the average accuracy of 77%, while in the video, the average accuracy of 76.6%.

Key Words: emotion recognition, facial feature extraction, eyes figure, mouth figure, Certainty Factor.

1. INTRODUCTION

In general, emotion is related to someone’s mind and feeling in particular condition. Most of the expert stated emotion in many different definitions. In a research publication, it stated 92 definitions of emotion, then in general it can be concluded that emotion is the group of interactions between subjective and objective factor through neural system and hormonal, that it can give some changes of feeling, builds cognitive process into perception that appear, and the physiological adaptation causes certain behavior both reactively and adaptively [1].

Furthermore, Ekman stated that emotions in some classifications are normal, happy, sad, fear, angry, and disgusted. Based on Ekman’s opinion, emotion is universal, it does not depend on the racial, culture, and nation [2]. The knowledge and recognition about emotion has important role in some fields like health, character building, communication, social psychology, etc. One factor to the success of communication is knowing the emotional condition of the audience. In the health field, it stated that someone who feels negative emotion, he has two more possibility to get some serious diseases like: indigestion, headache, cancer, and heart attack [3].

Emotion recognition can be done with some methods. Scherer stated some methods that can use to emotion recognition, like: subjective experience, peripheral physiology, affect modulate started, central physiology, and behavior [4]. A kind of emotion recognition based on behavior is face expression [4]. Face expression can be stated in a
group of emotion rule, with each of rule is related to the coming of AU’s (Action units) in the particular face. For the example, happy emotion is related to the coming of AU’s6 and AU’s12 [5]. Patic et al stated that a method to classify emotion based on the coming of AU’s in each emotion rule, for the example disgusted emotion is stated with the formula, AU’s9+AU’s17+AU’s26, as in the image is only found Au's9 and Au's17, so the image size of the disgusted emotion is 66% [6].

In this research, emotion recognition is done in some stages. The first stage is feature extraction (component): face, mouth, eyes, forehead, and nose. The extraction of face component is based on color and face geometrical figure. The eyes and mouth feature is extracted based on color, whereas the nose and forehead feature is extracted with face geometrical figure that based on the position of mouth and eyes spot that gotten previously. The second stage is state recognition of face feature, eyes and mouth are used elongation feature, on the forehead and nose are used the wrinkle density. The third stage to introduce emotion based on the certainty factor in each emotion rule.

2. THE EMOTION FORMULA BASED ON AU’s FACE

Visually, someone’s emotion can be seen from the face expression. The face expression of someone who gets angry will be different with someone who is happy. In some research, the scoring emotion on the face is related to the coming of AU’s (Action Unit’s) in certain face like formula [7], [8], [9] called FACS (Facial Action Code System). Table 1 shows the formula list for each emotional class.

<table>
<thead>
<tr>
<th>No</th>
<th>Emotion</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Happy</td>
<td>AU's 6 + AU's 12</td>
</tr>
<tr>
<td>2</td>
<td>Fear</td>
<td>AU's 1 + AU's 2 + AU's 4 + AU's 5 + AU's 7 + AU's 20 + AU's 26</td>
</tr>
<tr>
<td>3</td>
<td>Sad</td>
<td>AU's1 + AU's 4 + AU's 15</td>
</tr>
<tr>
<td>4</td>
<td>Angry</td>
<td>AU's 4 + AU's 5 + AU's 7 + AU's 23</td>
</tr>
<tr>
<td>5</td>
<td>Disgust</td>
<td>AU79 + AU15 + AU16</td>
</tr>
</tbody>
</table>

3. THE FACIAL FEATURE EXTRACTION

Facial feature (component) extraction is done in each image, face extracted components are mouth, eyes, forehead, and nose. The extraction method uses geometrical approach, face figure, and color.
room HSV [12]. Equation 1 stated the mouth figure
equation that used in [10], [13], [14], [15], whereas
equation 2 stated the repairing mouth figure [16].

\[
\text{Mouth Map} = C \sum \left( C_r^2 - \eta \frac{C_i}{C_r^2} \right) \tag{1}
\]

\[
\text{Mouth Map} = S \left( \sum \left( C_r^2 - \eta \frac{C_i}{C_r^2} \right) \right) \tag{2}
\]

\[
\eta = 0.95 \times \frac{\sum C_i}{\sum C_r^2} \tag{3}
\]

In this research, it developed mouth figure
cromatic like in the equation 1 with block system.
Generally, both mouth figure that based on equation
1 and 2 have some weaknesses because of the
lighting problem, the color of the lips and the face
skin is similar on the particular people with the
image background. It causes the detection of other
RoI (Region of Interest) not the mouth RoI. Picture
2 shows the result of mouth detection using mouth
map. It showed the detected mouth RoI which
followed RoI that is not mouth.

Fig 2: RoI Not Mouth

The block system suggested in this research
based on mouth position, which is on the face. In
the detection result of, it will be taken the block or
area with absis score 0.65 until 0.92 from the face
height. After that, in that area will be found RoI
with the biggest area that probably is mouth.

3.3 The Eyes Extraction

For the eyes feature extraction, it also uses the
approach of chromatic and luminance color. Some
previous research also uses an approach that called
the eyes figure [13], [17], [18], and [19]. The eyes
figure is stated as the multiplication of chromatic
eyes figure with luminance figure, each of them
stated in equation 4, 5, and 6.

\[
\text{Eye Map} = \text{Eye Map C} \times \text{Eye Map L} \tag{4}
\]

\[
\text{Eye Map C} = \frac{C_r^2 + C_i + C_b}{3} \tag{5}
\]

\[
\text{Eye Map L} = \frac{Y + f(x,y)}{Y + f(x,y)} \tag{6}
\]

f(x, y) is hemmisibleperical function, to determine
the coordinate of eyes center point, it also used
block system like done in detecting mouth. The
eyes probably on the \( \frac{1}{4} \) to \( \frac{1}{2} \) from face. Then, in
that, area we found two items of the biggest RoI
determined the center point of that RoI is the
eyes area. The using of block system is to solve the
noise problem, in some cases found the coming of
noise that is RoI with the bigger size than the eyes
area. Picture 3 shows the eyes RoI extraction based
on the eyes figure and the block system.

Fig 3: Eye Map and Block System

3.4 The Extraction of Forehead and Nose
Feature

The location of forehead and nose is determined
based on the mouth or eyes position, in this research
eyes point as the reference is used. Caused by the
center point of eyes are two, the right and left eyes,
so the eyes center point is taken from the average of
the left and right eyes. The problem is when the
position of the face in oblique position, in this case
previously it is determined by the direction and
oblique of the face. After that, based on the center
point of eyes, the oblique direction and the
geometrical figure of Indonesian face, it will
probably as the position of forehead and nose. At
whole, the face extraction result is gotten simultaneously, showed in Picture 4.

![Face Feature Segmentation](image)

**Fig 4. Result Face Feature Segmentation**

4. THE RECOGNITION OF FACIAL FEATURE STATE

After each face component, which can be segmented, the next stage is the state recognition from each face component.

4.1. The Recognition of Forehead and Nose State

In the research of Navarro J and Karlin M, they stated that the wrinkle sign was one of the sign which made someone discomfort or negative emotion [20]. Some previous research revealed the wrinkle detection method like [21], citing the research [22], so that in the research [23] stated the way to calculate the wrinkle density with equation 7.

$$D_a = \frac{P_w}{P_a}$$  (7)

*Wa* is the group of pixel with wrinkle, whereas *Pa* is the group of all pixels. Da is around 0 to 1. In this research, to determine the existence of wrinkle, there will be some stages. The first stage is normalization of the segmented result in the size of 100 x 150 windows. The second stage is the side detection using LoG (Laplacian of Gaussian). The third stage is tresholding to strengthen the wrinkle line and to disappear, which is not wrinkle. Picture five shows the comparison between the wrinkle detection result on the forehead with *canny*, LoG, and LoG thresholding. Picture 6 shows the wrinkle detection on the nose.

![Wrinkles Detection on Forehead](image)

**Fig 5. Wrinkles Detection on Forehead (a) Test Image (b) Edge Detection by canny (c) Edge Detection LoG, (d) Edge Detection LoG and Thresholding.**

4.2. The Recognition of Mouth State

One feature that can be used to analyze the mouth state is elongation feature, which states the comparison between the mouth height and width, it is shown in Picture 7 and Equation 8.

$$\text{Elongation} = \frac{\text{Height of Mouth}}{\text{Width of Mouth}}$$  (8)

The elongation score is determined based on the mouth RoI, for extracting the mouth RoI, it uses equation of the mouth figure, Equation 1. Though the analyzed mouth state is related to some face AU's (*Action Unit*). Picture 8 shows extraction of the mouth ROI.

![Mouth RoI Extraction](image)

**Fig 8. The Mouth RoI Extraction**

4.3. The Recognition of Eyes State

There are some state or the eyes condition that used in this research like normal view (Au's1), closed eyes (Au's4), become narrowed (Au's6), and glared eyes (Au's5). The analysis of recognition of that eyes state is also based on elongation geometrical feature, with gotten eyes RoI firstly. Based on the eyes figure, Equation 6 is determined by the eyes position, and then *crooping* is done to get the right and left eyes area. The eyes RoI is determined by looking the area that is not skin on the *crooping* result of the eyes area. Skin tresholding is based on the YCbCr color room, with the limitation, Y > 80, 102<Ch<121, and 139<Cr<160. This limitation score is appropriate with the research [24]. Picture 9 shows extraction of the eyes RoI.
4.4. The Feature’s Score of Each Face Component

At whole, the state of face component is summarized in Table 2 and 3. For the eyes and mouth, it is as the mean score of elongation. Whereas, on the forehead and nose, it is considered as wrinkle if the wrinkle density more than 0.3.

Table 2. The Mouth Elongation Feature and Wrinkles Density on Forehead Nose

<table>
<thead>
<tr>
<th>AU</th>
<th>Illustration</th>
<th>Elongation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU12</td>
<td></td>
<td>0.3887 ± 0.0758</td>
</tr>
<tr>
<td>Au15</td>
<td></td>
<td>0.2796 ± 0.1102</td>
</tr>
<tr>
<td>Au20</td>
<td></td>
<td>0.4207 ± 0.1175</td>
</tr>
<tr>
<td>Au26</td>
<td></td>
<td>0.6638 ± 0.208</td>
</tr>
<tr>
<td>Au23</td>
<td></td>
<td>0.3595 ± 0.1103</td>
</tr>
<tr>
<td>Au16</td>
<td></td>
<td>0.4794 ± 0.1025</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component</th>
<th>Wrinkles Density</th>
<th>Elongation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forehead</td>
<td>Wrinkles/No</td>
<td></td>
</tr>
<tr>
<td>Nose</td>
<td>Wrinkles/No</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Eyes Elongation Feature

<table>
<thead>
<tr>
<th>State</th>
<th>Illustration</th>
<th>Elongation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU's 6</td>
<td></td>
<td>0.349 ± 0.113</td>
</tr>
<tr>
<td>AU's 1</td>
<td></td>
<td>0.516 ± 0.0736</td>
</tr>
<tr>
<td>AU's 4</td>
<td></td>
<td>0.349 ± 0.113</td>
</tr>
<tr>
<td>AU's 5</td>
<td></td>
<td>0.791 ± 0.2296</td>
</tr>
<tr>
<td>Au's 41</td>
<td></td>
<td>0.263 ± 0.0236</td>
</tr>
</tbody>
</table>

5. THE EMOTION CLASSIFICATION

At the first time, the certainty factor method was introduced by Shortliffe and Buchanan in 1970’s, as the expert system to diagnose a disease, then it was known as MYCIN [25]. Certainty Factor (CF) states the trustworthy degree toward an event (fact or hypothesize) based on evidence or expert assessment. This method is generally used to solve the uncertainty of managing system, which contains a lot of rule base [26]. In that research, Certainty Factor (CF) method is used to recognize emotion in the static image. A number of rules are arranged based on the distance feature among the face component, the distance feature is determined based on the face point feature on the frontal face model. If the CF becomes bigger, it shows the trustworthy level toward a fact becomes higher. CF will be similar with the probable conditional, however it has some differences like: (1) CF more than only states the fact probability level, CF more states trustworthy size of the fact. (2) Probability is stated in the range of 0 to 1, whereas CF is stated around -1 and 1, CF score around 0 and -1 states the untrustworthy size, whereas CF score around 0 and 1 states the trustworthy size. Certainty Factor notation is written in the Equation 9.

\[ CF(H, e) = CF(e) \cdot CF(H) \]  

Equation 9 states the trustworthy size on H hypothesize if the fact evidence e happened, generally CF (H) is CF rule determined by the expert, whereas CF (e) determined by the user. CF evidence is related to the trustworthy score of the face component state, that trustworthy score is calculated based on the probability score of the fact evident evidence e, stated P(e). The probability calculation of evidence in each facial state is stated with Equation 10.

\[ F(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]  

\( \mu \) and \( \sigma \) score is the average and the deviation standard refers to Table 1 and 2 with \( \pi = 3, 14159 \) is constant. The basic reference in calculating the fact probability (evidence) in AU’s eyes and mouth component is the average score of elongation feature score, whereas on the forehead is the existence of wrinkle, score 0 if there is no wrinkle and oppositely score 1 with wrinkle. Based on the CF (En) score, the quality of each emotion type is determined based on the formula of Table 4. The formula of Table 4 is gotten by the emotion rule of Table 1.
Table 4. Certainty Factor Emotion Rule

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Certainty Factor Emotion Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>Max(Au's6</td>
</tr>
<tr>
<td>Fear</td>
<td>Max(Au's1</td>
</tr>
<tr>
<td>Sad</td>
<td>Max(CF AU's1</td>
</tr>
<tr>
<td>Angry</td>
<td>Max (AU's 4</td>
</tr>
<tr>
<td>Disgust</td>
<td>Max (AU's9</td>
</tr>
</tbody>
</table>

CF rule is the score that shows the trustworthy level from each rule, in the research is stated with 0.8 score.

6. THE EXPERIMENT AND RESULT

The research’s experiment was done in two stages, the first stage in that experiment was done in the singular image, whereas in the second stage was done in the video scene. In the first stage, 20 images were used singularly in each emotion character. In the second stage, 25 video scenes were used; the emotion experiment of image video emotion was based on set image frame from the video decomposition.

Validation was done by comparing the classification result system with that visual emotion video. Table 5 shows the experiment result in the single image.

Table 5. The Experiment Classification on Single Image

<table>
<thead>
<tr>
<th>Emotion</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>AC (%)</th>
<th>ER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>17</td>
<td>2</td>
<td>1</td>
<td>85</td>
<td>15</td>
</tr>
<tr>
<td>Angry</td>
<td>16</td>
<td>1</td>
<td>3</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>Sad</td>
<td>15</td>
<td>2</td>
<td>3</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Fear</td>
<td>15</td>
<td>1</td>
<td>4</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>Disgust</td>
<td>14</td>
<td>2</td>
<td>4</td>
<td>70</td>
<td>30</td>
</tr>
</tbody>
</table>

Note: TP (True Positive), FP (False Positive), FN (False Negative), AC (Accuracy), ER (Error Rate)

The experiment in the video was done frame by frame, the emotion classification at whole was determined based on the average of the emotion in each frame. The experiment on the video was done by taking the video cut in the average of 3 seconds, which average contains of 75 image frames. The experiment result on the video cut is shown in Table 6.

Table 6. The Experiment on Video Cut

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Average True Positive frame/Video</th>
<th>Average Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>60 frames</td>
<td>80</td>
</tr>
<tr>
<td>Angry</td>
<td>63 frames</td>
<td>83</td>
</tr>
<tr>
<td>Sad</td>
<td>61 frames</td>
<td>81</td>
</tr>
<tr>
<td>Fear</td>
<td>49 frames</td>
<td>65</td>
</tr>
<tr>
<td>Disgust</td>
<td>55 frames</td>
<td>74</td>
</tr>
</tbody>
</table>

Picture 10 shows the experiment sample for image sequence frame from happy emotion video, and picture 11 shows the changing of elongation feature score of the mouth and eyes at the sequence image frame in a video with the happy and neutral classification.

Fig 10. The sequence images frame for emotions happy
From the Picture 11, it can be seen that the happy emotion in the average around 0.5 closes to Au’s16 score that shows the smiling mouth condition. On the wrinkle, on the happy emotion, from Picture 11 shows the lowest score under 0.35. This case shows appropriateness with the theory that wrinkle on the forehead or nose shows negative emotion (high stress) [20]. Based on the observation from the error frame, it was found two cause problems, the first thing was the error of identifying the mouth shape. This case was found mostly on the angry emotion, based on the elongation feature and it was very difficult to differentiate between the mouth shape Au’20 for identifying angry with Au’23 and also Au’s16. The extraction of feature point to determine elongation, based on RoI of the lips skin was still imperfect, it was caused by the lips color which imperfectly red. Found 21% RoI segmentation imperfect mouth, which causes an error determining the state of the mouth, shown in picture 12.

The second problem happened because the influence of the light changing, this case was caused of some cases like the moving object and the changing of camera focus. Because of that changing, there was error recognition of the eyes shape. Like what have been stated in the part 3, the eyes state is known based on the RoI shape, whereas RoI is segmented based on the area, which is not skin. The difference of the lighting influence will cause the range changing on the color of Cr skin. Found 17% RoI segmentation imperfect eyes, which causes an error determining the state of the eyes, shown in picture 13.

In Figure 13, the eyes state is actually closed, but based on segmentation RoI eye, eye elongation value obtained 0.47, as though a normal open eyes
Comparison with other risets related to the emotion recognition, patic el al [6] which used FACS coder have accuracy of 85%. Ratliff et al [28] used AAM (Active Appreance Model) on single image have accuracy, fear 90%, angry 63.9%, disgust 93.3%, sad 63.9%, and happy 93.3%

CONCLUSION

On the average, emotion recognition accuracy on a single image by 77%, while in the video, in the form set image frame sequence, the average accuracy of 76.6%. There are two main problems, the first difficulty determining the state mouths based on the value of elongation. This is due to the segmentation of RoI lip is less than perfect, to as much as 21%. The second problem, the problem of determining the eyes state, based on elongation RoI. Eyes RoI segmentation on the basis of skin color negation, prone to lighting problems, the percentage found 17%

REFERENCES


