Region based Facial Expression Recognition using Gradient Directions

Vinola C., Vimala Devi K., Valarmathi K., Manjula V., Muthurajkumar S.

Abstract: Facial Expression Recognition (FER) has gained significant importance in the research field of Affective Computing in different extents. As a part of the different dimensional thinking, aiming at improving the accuracy of the recognition system and reducing the computational load, region based FER is proposed in this paper. The system is an emotion identifying system among the basic emotions, through subject independent template matching based on gradient directions. The model designed is tested on the Enhanced Cohn-Kanade (CK+) dataset. Another important contribution of the work is using only eye (including eyebrows and the nose portion near eyes) and mouth regions in the emotion recognition. The emotion classification result is 94.3% (CK+ dataset) for 6-class FER.

Keywords: Facial Expression Recognition, Histogram of Gradient directions, Region Of Interest(ROI), Template matching

I. INTRODUCTION
Emotion classification based on facial expressions, bodily gestures, physiological signals, audio signals, etc. has found an important place in the research of many scientists. The application areas of the research include understanding the emotions in the field of marketing, patient monitoring, driver drowsiness alert and other applications. FER is widely used in the work as it gives more promising results with less computation cost.

FER system developed with the raw image as input has some main components like pre-processing, feature extraction, feature space reduction, feature classification and finally emotion recognition. Each and every stage in the system has been under different advances targeting the recognition accuracy. Features extracted from the images act as the main factor in deciding the emotions, as they are different for each expression in the face. Different feature sets used in the research include geometrical, appearance and motion features. Facial Landmark points help in deriving geometrical features like displacement and the angle between contributing points. In comparison with the geometrical features which rely much on the contributing points, appearance features are more favorable as it takes into account the changes in texture of the image representing an expression which leads to robustness against noise and illumination constraints. They include Gabor Features [1]. Histogram of Oriented Gradients (HOG)[2], Local Binary Patterns (LBP)[3], Local Ternary Patterns (LTP) [4] and also their variants like GLTP (Gradient LTP) [5]. Motion features like Motion History Image (MHI) [6] are also well used to know the dynamics of the gestures to understand the emotion expressed. Combining different category of features and making a decision based on the features helps in performance upgradation of the system. Taking into account the local patterns of the nearby pixels for each and every pixel of the gray level input image representing the emotion, the researchers in different variations have successfully implemented FER models.

The main contributions of the work compared to other state of art approaches include 1. Lessening the feature vector space by finding the suitable feature set to classify the emotion expressed in the first stage and thus reducing the computational load. This is achieved by including only the gradient directions of the eye (including eyebrows and the nose portion near the eyes) and mouth regions alone in the feature set. 2. Person independent template matching using the reduced feature set to identify the emotion expressed.

II. PREVIOUS RELATED APPROACHES
Affective computing has been considered as the interest of recent decades by many researchers in different dimensions. Most of the recent researches are based on FER compared to other modalities as the facial emotions are the first to show the inner feelings of a person. The applications using FER extend to different domains and in particular, Human Computer Interaction (HCI) [7]. FER model consists of three important phases. They are Feature extraction and processing, Classification and finally Decision Making. Improvement in terms of accuracy, less complexity and less computation time are the prime goals of the works done. Many different types of features, both static and dynamic have been used in FER. There are also different categories of features as mentioned in section I. [8] deliberates on the different algorithms of FER in different important steps of pre-processing, feature extraction and classification where in ROI segmentation, Gabor Functions and SVM proves efficient respectively in each step. One of the interesting feature category is appearance features, which include the texture patterns like LBP (CaifengShan et.al. [9], Nazima Kauser & Jitendra Sharma [10] and LTP and its variants. Another remarkable feature used is the Histogram of oriented directional gradients (HOG) which helps in getting the idea of the shape of the edges and thus much useful in FER (Carcagni. P et.al. [11]). Identification of the contributing regions of the face and feature extraction of the active patches in the face is proposed by [12] which exhibited improved average accuracy compared to other methods.
The prime advantage that overcomes the limitations of the other works in the literature is the usage of a limited feature set of only the contributing regions (ROIs) in particular the eye and mouth regions when compared to the usage of the whole face region features.

III. PROPOSED APPROACH

The basic emotions expressed by the subjects in the CK+ database include Happiness, Sadness, Fear, Disgust, Anger, Contempt and Neutral. The first phase includes main different stages like feature set preparation, facial subregions contribution, histogram estimation and finally template matching to identify the emotion expressed.

A. Feature set preparation:

Gradient image gives more information on the edges and in general the texture of the image when compared to the raw image. Appearance feature set to be extracted and used in the system is just the orientations of the gradients as it gives the directions in which the image changes which can help in differentiating the emotions conveyed in different image frames. The direction values ranging from +180 to -180 (Fig.1) that are generated by applying the Sobel operator, contribute to the feature set preparation. The subregions in the facial region have their own gradient directions which can be used for differentiating the emotion expressed in the respective image frame as they vary for different emotions. Histogram values evaluated from the gradient directions will serve as the input feature set to identify the emotion as discussed in the next section.

Fig.1. Range of Gradient Directions (Angles)

B. Facial Sub regions Contribution and Histogram estimation:

In most of the research, the whole facial region is taken into consideration for the feature set extraction. Some of the researchers have used weights to be assigned for different facial regions. Our work aiming at reducing the feature vector length and identifying only appropriate contributing regions, has explored that the eye regions along with eyebrows and mouth regions can alone be used for decision making and both the regions can be given equal weightage for the purpose. The nose region histogram of oriented feature vector after many tests was found as additional contribution, but not the prime contribution like eye and mouth regions. The eye and mouth regions have to be detected in the first step using Viola Jones Algorithm.

Firstly detected eye and mouth portions are normalized or resized to a uniform size (it is [60 180] for eye and [90 120] for mouth in our work). Then the eye and mouth regions are split into 3x4 cells as component based histogram extraction has been proved in literature to be more effective than the whole region based processing. The ranges are fixed in equal intervals of 45° in the positive and negative orientation (Fig. 1). Therefore the total number of ranges considered is 8 and for each cell the histogram values (frequency or count of occurrence) are calculated for the 8 ranges. Then the feature vector is made ready by combining the histogram values for all the cells (size of feature vector-12x8=96) (Fig. 2). Separate feature vectors for the eye and mouth region are used in the system. They are flexible to be used by giving suitable weightage for the regions (eye and mouth) as discussed in the next section.

Fig.2. Histogram of gradient directions (Eye and Mouth region)

C. Template matching

As discussed in the previous section, histogram of gradient directions is calculated. Recently multiclass SVM is used for the purpose of classification of the emotions. Template matching is more useful in cases where the feature vector length is less and when the feature vector can directly relate to the emotion expressed. In our case, the feature vector length is 96 for each region. Separate template matching is done for the eye and mouth regions. Firstly, two template sets are prepared for each emotion class, one including less to medium intensity expression and another one including medium to peak intensity expressions. This helps in confining the classification to any of the two template sets rather than misclassification. It is done for both eye and mouth regions. Not less than 64 templates are considered for each set. Final template for each set is prepared by finding the mean of the histogram values of the gradient orientations available in the set. The Mean templates of Histogram of gradient directions for the eye and mouth region of two basic emotions for descriptive purpose are shown in the figure (Fig. 3). In the same way remaining emotion classes are also considered. The test image frame is compared with the average templates of all the emotion classes using MSE (Mean Square Error) estimate which is evaluated as follows:

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (T(i) - P(i))^2
\]
where $AEerr_{i,j}$ is the MSE between the emotion template $i$ and the test image of the sub region $j$ in the eye region. $AETemp_{i,j,k}$ is the mean template histogram values of the emotion class $i$ in the $j$th subregion and $k$th range of gradient orientations. Here number of ranges $(m)$ is 8. Next the mean error $AE_{en}$ is calculated for the whole eye region for the emotion class $i$ considering all the subregions. Here the number of subregions $(r)$ is 12. $AETest$ is the eye feature vector of the test image frame.

Similarly the mouth region error estimate is evaluated as follows:

$$AEerr_{i,j} = \frac{\sum_{k=1}^{m} (AETemp_{i,j,k} - AETest_{i,j,k})^2}{m}$$

$$AEerr_{i} = \frac{\sum_{j=1}^{r} AEerr_{i,j}}{r}$$

where $AMerr_{i,j}$ is the MSE between the emotion template $i$ and the test image of the sub region $j$ in the mouth region. $AMtemp_{i,j,k}$ is the mean template histogram values of the emotion class $i$ in the $j$th subregion and $k$th range of gradient orientations. Next the mean error $AM_{en}$ is calculated for the whole mouth region for the emotion class $i$ considering all the subregions. $AMTest$ is the mouth feature vector of the test image frame.

The weighted average of $AEerr$ and $AMerr$ is the final error estimate $Ei$ of each emotion class $i$ considered for decision making.

$$E_i = \frac{W_E x AEerr_{i} + W_M x AMerr_{i}}{W_E + W_M}$$

Here in our work, the weights WE and WM are tested with different values based on the emotion set considered. For example, Happiness and Surprise emotions expressed by the subjects can be differentiated by the mouth region more appropriately. Therefore, more weightage could be given to the mouth region. But when generally tested for all emotions, equal weightage with $WE=1$ and $WM=1$ gave promising results. The emotion expressed in the image frame is identified by the minimum value of the $E$ vector.

**Fig. 3 Mean Template Histogram of gradient directions for 2 basic emotions (Eye and mouth region)**

**IV. SYSTEM ARCHITECTURE**

Algorithm:

Module: Feature set preparation
1. Detect the eye region from the face using Viola-Jones algorithm
2. Detect the mouth region using manipulations
3. Normalize the size of the eye portion to 60 x 180 and the mouth region to 90 x 120 pixels
4. Apply Sobel operator to the eye region and mouth region and find the gradient directions
5. Divide the eye and mouth region into 3x4 subregions individually
6. Find the Histogram of gradient directions in the interval of 45° for eye and mouth region by combining the histograms of each subregion.

Training phase:

Given: A set of training samples for all emotions
Feature set preparation module is called for all training samples and the two average templates (one with less to medium intensities and other with medium to maximum intensities) is found for each class of emotion in the training sample.

Testing phase:

Given: Test image frame
1. Feature set preparation module is called for the test image
2. Comparison of the histogram found from the above step and the average templates found in the training phase is done to find the mean square error between them. The emotion template which gives the least error will indicate the emotion expressed by the test image.
V. RESULTS AND DISCUSSIONS

CK+ database is used in validating the system developed (Lucey.P, Cohn J.F (2010)). 123 subjects involving different genders (65% were female) and races (15% were African-American and 3% were Asian or Latin descent) available with 593 sequences. 327 out of 593 sequences are emotion sequences. The image sequences were captured by front facing camera ranging from neutral to peak target expression and the images are digitized to 640x480 or 640x490 pixel resolutions and stored in .png format. 75 subjects with about 1122 images are selected of which nearly 25% are less intensity frames, 40% are moderate intensity frames and 35% are peak intensity frames.

For the two templates to be prepared for each emotion class the data set is grouped based on the intensities. 6-class expression recognition (happy, sadness, anger, surprise, fear and disgust) is attempted in our work. 'Contempt' expression in CK+ database is not considered in many researches.

Emotion Recognition phase is tested with image samples of different emotions with different intensities. Many researchers do not train or test the system developed with less intensity image samples, as they are not more expressive. But in our work, to provide input for the second phase of intensity estimation less expressive image samples are also included in the training samples as well as testing samples.

The person independent testing method is followed. In this method, the whole dataset except for certain subjects selected at random are used for preparing templates and the remaining subjects are used only for testing the system. This is repeated for all the subjects and the accuracy is evaluated by the number of correctly recognized frames against the total number of frames considered.

Histogram of directional gradients which is calculated for the testing image is compared with the stored templates of average histogram values for each emotion class. LMSE calculated indicates the matching template and the emotion class is identified. The confusion matrix for the 6-class FER is shown in the table 1. The average accuracy percentages for the 6-class in our proposed approach is 94.3%.

The highest accuracy is achieved for surprise emotion in all the three classes of expression recognition and the next highest is for the happiness emotion followed by sadness and disgust emotions. Anger and Fear emotions are mostly mistaken as surprise emotions giving less accuracy. The improvement from the recent approaches in literature for facial expression recognition is in terms of achieving acceptable and improved accuracy using person independent testing strategy. The work avoids the problem of reducing the dimensions and usage of complex machine learning techniques involving over fitting and other issues. In many works, only more expressive images from the image sequences in the dataset are included, perhaps within the last four or five images. But as our work does not end with emotion classification and extends to grade the emotional intensities, the image frames from neutral to peak expressions are randomly picked to form the templates and also for testing.

<table>
<thead>
<tr>
<th>Table 1 Confusion Matrix 6-class Expression Recognition-CK+ dataset (Average Accuracy-94.3%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Happy</td>
</tr>
<tr>
<td>Sadness</td>
</tr>
<tr>
<td>Anger</td>
</tr>
<tr>
<td>Surprise</td>
</tr>
<tr>
<td>Fear</td>
</tr>
<tr>
<td>Disgust</td>
</tr>
</tbody>
</table>

The drawback of the proposed approach based on the outcome is that the anger emotion is being misclassified as most other emotions, mostly neutral and surprise emotion. As the work includes less, moderate and peak intensity frames, the misclassification occurs and it happens more with the less intensity frames, which are similar in template to the neutral frames and in particular the eye region feature vector matches with the surprise emotion template. For the same reason mentioned, i.e., including less intensity frames, the neutral frames are also misjudged as other emotions. For example, for the frames of sadness emotion misinterpretation of neutral as less intensity sadness frame occurs which could be logical, but against the labelling of emotions.

VI. COMPARISON WITH THE STATE OF ART

Many researchers have contributed using different feature extraction methods and classification strategies for FER and estimation of emotion intensity. They have carried out the work and presented results in available standard datasets using different approaches in testing. Therefore a direct comparison of the results and validation is not possible. But the outcome of the works using the same dataset with more or less the same testing strategy could be compared with a valid discussion on the methodologies followed. Based on the outcomes of the work tested on CK+ dataset for 6-class expression recognition in the recent years is considered and listed in Table 2. Previously LBP and LTP patterns were used by researchers for face and facial expression recognition. An extension of those patterns and including other features and methodologies is done to improve the accuracy of FER.

In our work, person independent template matching is performed to validate the system against the cross-validation used in many literatures as subject independent testing will aid in the second phase of intensity estimation as the working of the system is required irrespective of the subject. Other works include only most expressive frames (3 to 4) only for training and testing. But our work is developed to identify the expression, even with less intensity as the templates are prepared for those types of frames also. Therefore, even though direct comparison is not possible relative performance could be validated and discussed. Taheri.S et.al., (2013), suggests a dictionary based approach for face recognition and also FER which uses person independent testing on CK+ providing 89.21% with the least accuracy in detecting...
anger and fear expressions. Using HOG features and template matching for contributing eye and mouth regions.

Our person-independent testing result for 6-class is 94.3%. Our work has achieved it with the least number of features. Using 955 directional patterns in the work, Ryu.B et.al (2017) have excelled in performance by giving 94.2% for 7-class in person independent validation against our result of 88%. This is only due to the neutral frames inclusion, neutral images are misinterpreted as low intensity emotional image frames thereby providing accuracy of 68.7%.

K-fold cross validation results are obtained by most researchers and they are always higher than the person-independent results as shown by Holder et.al (2017), because the data set is randomly picked and grouped. But the person independent testing is stringent as certain subjects are totally left out and tested, which is more suitable to justify with real time data. In particular the template matching can use less number of data samples for preparing templates (average of HOG features for each emotion class) when compared to training dataset size used in hard core machine learning techniques. When the templates with just 20 data samples were prepared and then used in testing, the system yielded 100% accuracy in recognition rate of happiness and surprise emotions (6-class FER) and above 60% for all other emotion classes.

The K-fold cross validation results shows that Bayezid et.al. (2018) gives maximum of 99.57% using Gabor features and SVM followed by Holder et.al. (2017) giving 99.3%, using improved GLTP showing improvement on the work done by Ahmed et.al (2013) with 97.2%. But the person independent validation results by Holder et.al obtained is 86.5% for 6-class FER with very less accuracy for fear and sadness when compared to our accuracy of 94.3%. HOG features are often used in literature with SVM(Efficient Machine learning algorithm) and such an attempt is made by [11] proposing a system which excels other methods as discussed with a recognition rate of 95.8%. [13] uses the region based FER as proposed in our work, with different variants of LBP and LTP and using 10-fold cross validation testing scheme gives the maximum of 92.19% using 6 ROIs of the face classified with SVM. The advantage of using template matching against SVM in our method which has less feature space is to avoid the complex training and add the flexibility in maintaining separate templates for each emotion class. Thus the problem of over fitting and under learning is avoided in phase I of emotion recognition. Using HOG features with template matching is rarely attempted and work has proved its efficacy. As our work focusses on the eye and mouth regions, [12] also have worked with LBP of the salient patches in the facial regions and uses SVM to classify the pattern and the recognition rate is 94.09%. Following the person independent testing strategy, our approach has proved with good accuracy.

Table 2 Comparison to the state of art (6-class FER if not mentioned)- CK+ dataset

<table>
<thead>
<tr>
<th>Author</th>
<th>Feature used /Classification method</th>
<th>Recognition rate(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahmed et.al</td>
<td>GLTP/SVM</td>
<td>97.2&lt;sup&gt;CV&lt;/sup&gt;</td>
</tr>
<tr>
<td>(2013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tahteri.S. et.al</td>
<td>DCS</td>
<td>89.21&lt;sup&gt;PI&lt;/sup&gt;</td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Author</th>
<th>Feature used /Classification method</th>
<th>Recognition rate(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy et.al</td>
<td>LBP of salient patches/Multiclass SVM</td>
<td>94.09&lt;sup&gt;CV&lt;/sup&gt;</td>
</tr>
<tr>
<td>(2015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cacagni et.al</td>
<td>HOG/SVM</td>
<td>95.8&lt;sup&gt;CV&lt;/sup&gt;</td>
</tr>
<tr>
<td>(2015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holder et.al</td>
<td>Improved GLTP/SVM</td>
<td>99.3&lt;sup&gt;CV&lt;/sup&gt;, 99.7&lt;sup&gt;LOO&lt;/sup&gt; 86.5&lt;sup&gt;PI&lt;/sup&gt;</td>
</tr>
<tr>
<td>(2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ryu.B et.al</td>
<td>LDTP/two grids</td>
<td>94.2(7-class)&lt;sup&gt;PI&lt;/sup&gt;</td>
</tr>
<tr>
<td>(2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lekdoui.K et.al</td>
<td>Variants of LBP and LTP in face regions</td>
<td>92.19&lt;sup&gt;CV&lt;/sup&gt;</td>
</tr>
<tr>
<td>(2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayezid et.al</td>
<td>2D Gabor features/SVM</td>
<td>99.57&lt;sup&gt;CV&lt;/sup&gt;</td>
</tr>
<tr>
<td>(2018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Our approach</td>
<td>HOG/Template matching</td>
<td>94.3&lt;sup&gt;CV&lt;/sup&gt;, 88&lt;sup&gt;PI&lt;/sup&gt;(7-class), 87.12&lt;sup&gt;CV&lt;/sup&gt;(8-class)</td>
</tr>
</tbody>
</table>

CV-Cross Validation PI-Person Independent
LOO-Leave One Out
DCS-Dictionary based Component Separation Algorithm

VII. CONCLUSION AND FUTURE WORK

This paper presents a work for less computational and accurate Facial Expression Recognition (FER) system to detect the mood of a person so as to promote HCI, which is suitable for various applications. It uses Histogram of Oriented Gradients of only the contributing regions like eyes, surrounding regions of the eyes and mouth regions, avoiding unnecessary and redundant information in the feature set, which is capable of distinguishing the emotion classes with the help of template matching. Template matching also avoids the problem of complex machine learning techniques and also increases the flexibility in preparing templates for each emotion class without the influence of the other class, which is not possible in other machine learning methodologies. Decision making about the emotion expressed is obtained by giving equal weightage for the eye and mouth regions. FER system developed is tested using person independent method achieving better accuracy compared to the state of art methods using the same testing strategy.

REFERENCES

Region based Facial Expression Recognition using Gradient Directions


AUTHORS PROFILE

Vinola C., has completed her M.E degree in Computer Science and Engineering and pursuing PhD as part time scholar in Anna University, Chennai and has good number of publication record. Her area of interest includes image processing and artificial intelligence techniques.

Vimala Devi, K., is working as Associate Professor in the School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India. She received her M.E (2003) and Ph.D (2008) degrees from Anna University Chennai and MCA degree from University of Madras in the year 1997. She has more than 20 years of experience in Research and Development, Programming and Teaching. Her interest in research areas are like Computer Networks, Network Management, Network Security, Distributed Networks, Active Networks Wireless Sensor Networks and Object Oriented Programming and Network Security. She has published 90 technical research papers in reputed journals and conferences, which includes 8 journal papers indexed in ACM portal & SCI, 15 journal papers indexed in Scopus and 15 IEEE conferences.

Valarmathi K., obtained her Ph.D degree from Anna University, Chennai in 2008. She is a member of IEEE. She is a reviewer for various reputed journals. Her research interest includes Process Control, System Identification, Biomedical Instrumentation, Image Processing, Machine Learning, Wireless Network, Cloud Computing. Currently, she is working as a professor and head of the department of Electronics and Communication Engineering, at P.S. R Engineering College. She has published 23 papers in SCI journals and 58 papers in conferences.


Muthurajkumar S., is working as Assistant Professor in Department of Computer Technology, Anna University, MIT Campus from 2014. His areas of interest are Cloud Computing, DBMS, Data Structures, Computer Architecture and Digital Systems. He has published 25 papers in various International and National Journals and also presented 4 Research Papers in International Conferences and one in National Conference.

Retrieval Number: L34271081219/2020©BEIESP
DOI: 10.35940/ijitee.L3427.029420

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

2776