

Human Emotion Recognition from Facial Thermal Image using Histogram based Features and Multi-Class Support Vector Machine

by A. Basu*, A. Routray* and A. K. Deb*

*Department of Electrical Engineering, Indian Institute of Technology, Kharagpur, West Bengal-721302, India, ele.anushree09@gmail.com

Abstract

Emotions are one's subjective experiences which are associated with feelings, moods, affective and cognitive states. Human emotions generate unique signals which create unique patterns of sensation in our body. This paper reports a non-contact methodology to classify human emotions through thermal images of face. Histogram features obtained from facial patches are considered as the signature features and fed to a multi-class support vector machine for classification of the emotions.

1. Introduction

The advancement of technology and growing curiosity has led to the exploration of abstract and qualitative object like emotions. Specialized mathematical tools have been used for transforming the qualitative quantity into quantitative one. Research in affective computing plays an important role in Human-Computer Interaction (HCI). Literature has reported various works for estimation of human emotion using visual images and also biological signals like ECG, blood volume, EEG etc. Major contribution for emotion classification and analysis is from visual images. But the main drawback in this technique lies in the fact that genuine feeling can be easily concealed through fake expressions and also the dependency of the visual images on the surrounding illumination.

However, recent researches on emotions have seen the advent of emotion analysis from thermal images of human faces. Thermal images also known as thermograms are basically display of heat distribution of an object in form of an image. Thermal images are captured by thermal infrared camera which captures the infrared radiation emitted from an object, converts it to electrical signals and finally displays on the monitor as gray scale or pseudo-colored images. It is known from the recent literature that the change in the emotional and psychological state of a person gets reflected in blood flow rate, heart rate, temperature, pupil contraction etc. Thus it can be inferred that skin temperature pattern may be an indicator of the emotional state of the person. Wang. S et al [1] proposes a method using the temperature difference and K- Nearest Neighbour algorithm. An emotion classification using Deep Boltzmann Machine is proposed by He. S [2]. Khan M. et al. [3] in his work reports a method for emotion classification based on Facial Thermal Feature Points (FTFPs) by dividing the face into square grids and analyzing the thermal variations across the grids. Emotion classification on the basis of time-frequency analysis of temporal data of facial skin using genetic algorithm is proposed by Nhan B. et al [4].

This work presents a novel method for emotion analysis from thermal images on the basis of histogram feature extraction. The novelty here is that the classification of emotions even with partial occlusion like spectacles, facial hair etc. the method divides the face into six subsections and all the six regions are studied. Histogram features from all the subsections are extracted which contains region of interest (ROI) wise information and are used as the classifying feature. The classification is done with the help of the multi-class Support Vector Machine using appropriate kernel function for maximizing the inter class distance and minimizing the overlap between classes. *Kotani Thermal Facial Expression (KTFE)* Database has been used in this work [5]. Four emotions namely Happy, Sad, Anger and Fear have been considered here for study.

The total paper is divided into the following sections. Section 1 gives the introduction, section 2 shows the schematic diagram of the overall workflow, section 3 describes the complete methodology while section 4 forms the results and discussions part.

2. Workflow:

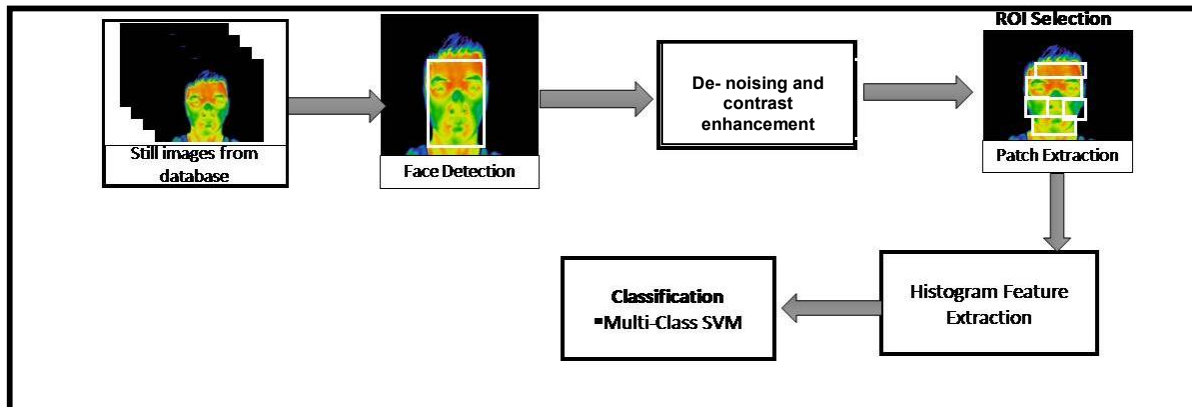


Fig. 1. Schematic diagram of the workflow

3. Methodology:

The overall procedure incorporated in this work starts with face detection from the images taken from the KTFE Database,, denoising and enhancement of facial images, selection of the region of interest (ROI) which includes facial patches extraction, histogram feature extraction and finally classification.

3.1. Face Detection:

Face detection is one of the most important step for localizing and extracting face region. The face knowledge is the most important portion for the analysis of emotions. The thermal images here are considered as color images and based on the consideration Viola-Jones algorithm is chosen in this work for thermal face detection. Viola Jones technique basically incorporates three steps which are integral image formation for feature computation, Adaboost technique which is used for feature selection and finally an attentional cascade formation for efficient computational resource allocation [6]. All these steps combined together make the Viola Jones technique highly efficient for face detection. The detected face is then de-noised and enhanced which is described in the following section.



Fig. 2. Detected face using Viola Jones

3.2. De-noising and Contrast Adjustment:

Filtering is one of the most important steps for de-noising, enhancement, smoothing and template matching of images. The facial thermal images are de-noised initially by passing them through a median filter. Median filter considers each pixel of the image and replace its value with the median of the neighbouring pixels. Median filters are robust as well as good at removing noise while preserving sharp edges. The filtered image is enhanced by using Contrast Limited

Adaptive Histogram Equalization (CLAHE) [7]. The CLAHE enhancement partitions an image into contextual regions and performs histogram equalization on each of the regions thus balancing the used gray values distributions. Thus in this process it makes the concealed features of the image properly visible. Thus the whole procedure consists of noise reduction while maintaining the high spatial content of the image by implementing the combination of median filtering, CLAHE and also edge sharpening.

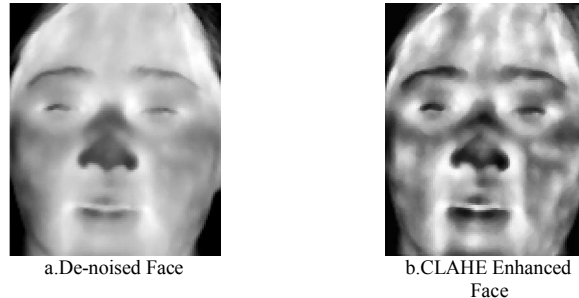


Fig. 3a. De-noised face, 3b. CLAHE enhanced face

3.2. ROI Selection:

The de-noised image face region is divided into six facial patches i.e. forehead, eyes, right cheek, left cheek, nose and mouth. Figure 4 shows the extracted facial patches from the total face. The facial patches have been extracted using the ratio based technique where the face region is segmented into the six segments based on some specific ratio considering the nose tip as the centre point of the face. Starting from the top first the forehead is segmented out, then the eyes, followed by the nose and cheek region and finally the mouth portion. The histogram features are further extracted from each of these patches and are fed to the classifier for further classification.

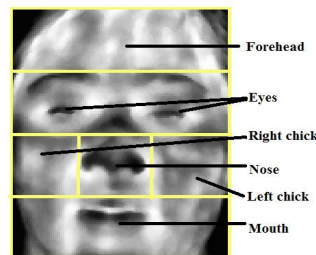


Fig. 4. Facial Patch extraction

3.3. Feature Extraction:

The image features are the main representation of the image. It helps in the proper interpretation of the image. Features may be of many kinds like corners, edges, lines etc. The process of feature extraction helps in formation of a set of feature vectors from detected features also called descriptors. Thus choice and extraction of proper features are helpful for the appropriate analysis of the images. From thermal imaging perspective feature extraction can be thought of as the procedure to convert the images into unique and distinct form to compare with the reference. Histogram features are used as the classifying features for analysis of the individual's emotions.

The histogram-based data analysis is considered to be one of the most popular techniques for analysing images [8]. Histogram preserves a lot more information than just simple averaging of raw data. The histogram features considered are basically statistics based features where the histograms are used to model the probability distributions of the intensity levels. Thus the distributions formed provide us with important information about the characteristics of the image through the intensity level distribution profile. The histogram probability of first order can be mathematically given as,

$$P(i) = \frac{A(i)}{N}$$

where, $P(i)$ is the histogram probability
 $A(i)$ is the number of pixels at gray level i
 N is the number of pixels in the image

First order histogram probability may be used for computation of mean, standard deviation, skew, energy, entropy etc. For gray level images having 256 intensity levels, the mean and the standard deviation can be given mathematically by,

$$i_m = \sum_{i=0}^{255} iP(i) = \sum_r \sum_c \frac{I(r,c)}{N}$$

$$i_{sd} = \sqrt{\sum_{i=0}^{255} (i - i_m)^2 P(i)}$$

where, i_m is the mean, i_{sd} is the standard deviation, r and c are the number of rows and columns of the image matrix respectively.

The enhanced and the de-noised face region is divided in six ROIs as already described in the previous section. Histogram features are extracted from each of the patches. All the six histogram thus extracted are concatenated together to form the feature vectors. The feature vectors of the images thus formed are fed into the multiclass Support Vector Machine (SVM) for further classification of emotions.

3.4. Classification:

The classification of emotions based on the feature vectors is performed using the multiclass SVM [9]. SVM is mainly a binary classifier that categorises both linearly as well as non-linearly separable data. For linearly separated two class data, classification is achieved by constructing a hyper plane to separate the data by maximizing the separation margin.

Let (p_i, q_i) where $1 \leq i \leq N$ be a set of training examples. A hyperplane is drawn which divides the training set such that the points belonging to the same class are on the same side of the hyperplane. Mathematical expression for the hyperplane in case of the linearly separable data can be given by,

$$q_i(\mathbf{w} \cdot \mathbf{p}_i + b) > 0 \quad i = 1, \dots, \dots, N$$

where, \mathbf{w} is normal vector to the hyperplane and b is a real quantity. Again, \mathbf{w} and b can be modified to get.

$$\min_{i \in \{1, \dots, N\}} q_i(\mathbf{w} \cdot \mathbf{p}_i + b) \geq 1 \quad i = 1, \dots, \dots, N$$

such that, the distance between the closest point and the hyperplane is $1/\|\mathbf{w}\|$. Thus finally the equation can be written into a simple form as,

$$q_i(\mathbf{w} \cdot \mathbf{p}_i + b) \geq 1 \quad i = 1, \dots, \dots, N$$

Now among all the hyperplanes, thus formed, the one for which the distance between the closest point and the hyperplane is maximum, is known as the Optimum Separating Hyperplane (OSH). For, non-linear data special functions called kernel functions are constructed for proper classification of data. Various kernel functions such as linear, quadratic, polynomial, radial basis function (RBF) can be used depending on the type of the data.

In case of multiple classes as in this work where we have multiple emotions to classify the multiclass problem is reduced into multiple binary classification problems. The one-against-one (OAO) technique is used in this case where the classifier is trained between each pair of the classes, and thus 4C2 classifiers are constructed. Appropriate choice of kernel functions very important in case of SVM classification as improper kernel can lead to poor performance. Both linear and polynomial kernel is used here between which the polynomial kernel function is found to give high amount of accuracy instead of the linear kernel function.

4. Results and discussions:

Kotani Thermal Facial Expression Database has been used in this work. This database contains subjects who are Vietnamese, Japanese, Thai from 11 year-olds to 32 year-olds with seven emotions Four emotions namely Happy, Sad, Anger and Fear have been considered here for study. 264 images of 22 subjects for 4 expressions has been considered per emotion among which about 60% have been taken for training and 40% for testing. The facial patches under consideration have been segmented into 6 sections. Histograms obtained from different patches are considered for training and testing. The histogram is normalized to make it generalized with respect to different person. Multi-class SVM approach using both linear and polynomial kernel function have been used for classification of the emotions. Obtained results have shown quite a high amount of accuracy. The figure 5 shows the emotions classified with the help of multiclass SVM using polynomial kernel.

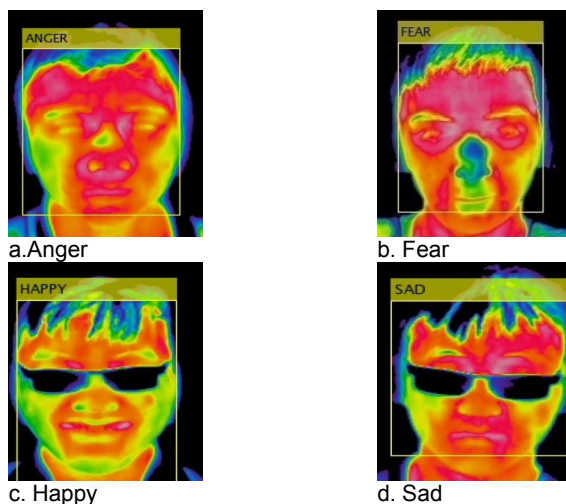


Fig. 5. Classified emotions using M-SVM

The accuracy is determined from the confusion matrix. The total images of 22 subjects are classified into four categories with 81.95% accuracy. The classified groups are given by the diagonal elements of the confusion matrix. The confusion matrix is shown in table 1. Thus from the confusion matrix it is seen that the total average accuracy of classification is 81.95% and misclassification or error rate is 18.05%.

Table 1. Confusion Matrix

Group	Predicted Group				Total
	Anger	Fear	Happy	Sad	
Anger	82.4%	4.8%	7.4%	5.4%	81.95%
Fear	7.2%	80.7%	3.4%	8.7%	
Happy	4.4%	5.1%	81.2%	9.3%	
Sad	4.3%	3.5%	8.7%	83.5%	

5. Conclusion:

In this work an efficient methodology for classification of emotions from thermal images is shown. The emotion recognition is done on the basis of image histogram features extracted from the facial patches obtained from the facial thermograms or the thermal images. These features are concatenated to form feature vectors which are further used for

classification using multiclass SVM. Four types of emotions namely happy, sad, anger and fear are classified with an accuracy of 81.95%. The work is completely based on thermal images available in KTFE database.

REFERENCES

- [1] S. Wang, P. Shen, and Z. Liu, "Facial Expression Recognition from Infrared Thermal Images Using Temperature Difference by Voting," IEEE 2nd International Conference on Cloud Computing and Intelligent Systems, pp. 94-98, 2012.
- [2] Shan He, Shangfei Wang, Wuwei Lan and Huan Fu Qiang Ji, "Facial Expression Recognition using Deep Boltzmann Machine from Thermal," Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 239-244, 2013.
- [3] M. M. Khan, M. Ingleby and R. D. Ward, "Automated Facial Expression Classification and Affect Interpretation Using Infrared Measurement of Facial Skin Temperature Variations," ACM Transactions on Autonomous and Adaptive Systems, vol. 1, no. 1, pp. 91-113, 2006.
- [4] B. R. Nhan, T. Chau, and S. Member, "Classifying Affective States Using Thermal Infrared Imaging of the Human Face," IEEE Transactions on Biomedical Engineering, vol. 57, no. 4, pp. 979-987, 2010.
- [5] H. Nguyen, K. Kotani, F. Chen, and B. Le, "Estimation of Human Emotions Using Thermal Facial Information," in proc. of Fifth International Conference on Graphic and Image Processing, vol. 9069, pp. 1-5, 2014.
- [6] P. Viola and M.J. Jones, "Robust real-time face detection," International journal of computer vision, vol. 57, no. 2, pp. 137-154, 2004.
- [7] J. Kim, L. Kim, and S. Hwang, "An advanced contrast enhancement using partially overlapped sub-block histogram equalization," IEEE Trans. Circuits Syst. Video Technol., vol. 11, no. 4, pp. 475-484, Apr. 2001.
- [8] S. Sergy, "Color Histogram Features Based Image Classification in Content-Based Image Retrieval Systems," pp. 221-224, 2008.
- [9] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support Vector Machines for Histogram-Based Image Classification," vol. 10, no. 5, pp. 1055-1064, 1999.