SVM Based Recognition of Facial Expressions Used In Indian Sign Language

Daleesha M Viswanathan *daleesha_mv@address.com*
Department of Computer Science
Cochin University of Science and Technology
Kochi-22, Kerala, India

Sumam Mary Idicula *sumam@cusat.ac.in*
Department of Computer Science
Cochin University of Science and Technology
Kochi-22, Kerala, India

Abstract

In sign language systems, facial expressions are an intrinsic component that usually accompanies hand gestures. The facial expressions would modify or change the meaning of hand gesture into a statement, a question or improve the meaning and understanding of hand gestures. The scientific literature available in Indian Sign Language (ISL) on facial expression recognition is scanty. Contrary to American Sign Language (ASL), head movements are less conspicuous in ISL and the answers to questions such as yes or no are signed by hand. Purpose of this paper is to present our work in recognizing facial expression changes in isolated ISL sentences. Facial gesture pattern results in the change of skin textures by forming wrinkles and furrows. Gabor wavelet method is well-known for capturing subtle textural changes on surfaces. Therefore, a unique approach was developed to model facial expression changes with Gabor wavelet parameters that were chosen from partitioned face areas. These parameters were incorporated with Euclidian distance measure. Multi class SVM classifier was used in this recognition system to identify facial expressions in an isolated facial expression sequences in ISL. An accuracy of 92.12 % was achieved by our proposed system.

Keywords: Indian Sign Language, Facial Expression, Gabor Wavelet, Euclidian Distance, SVM.

1. INTRODUCTION

Facial expressions convey important synaptic information in sign language. For example, facial expression would modify the meaning of hand gesture sign for “you eat” in to “please eat”, “did you eat?” and “can you eat?” In ASL, the facial expressions and head movements are generally classified into expression of emotions, conversation regulator, expressions that modify the quality and quantity of a sign [1]. Similar to ASL, there are many occasions in ISL where facial expressions are used in sign language recognition. Whether a sentence is interrogative or simple sentence is decided by the associated facial expression [1]. As an example, “where are you going?” in ISL following “subject object verb” pattern is “you going where”, of which, ‘you’ and ‘going’ are signed by hand gesture but the question “where” is conveyed by facial expression [2]. In addition, the emotions such as anger, happiness, surprise and sadness are signed by facial expression in ISL [3]. In ASL, facial expression and head movement recognition systems have been standardized (Google Scholar and Web of Science citations) [1]. However, these works would not be directly applicable in ISL due to many reasons. Inherent difference exist between ASL and ISL [2]. Contrary to ASL, head movements are less conspicuous in ISL and there is clear variation in facial expressions between the two systems [3]. In addition, the answers to questions, such as yes or no are signed by hand in ISL

This paper deals with a system which can recognize facial changes in an isolated facial expression ISL sentences. The facial feature changes were detected by the appearance based approach [5]. Prior to analysis and feature extraction, the region of interest on which features has to be extracted would be identified and cropped using image processing methods. The facial expressions were modeled with the help of Gabor wavelet parameters and these parameters were incorporated by Euclidean distance measure. Multi class SVM classifier was used in this recognition system to identify facial expression in ISL. This paper is organized as follows: After a brief survey of related works in Section 2, facial expression recognition framework is presented in Section 3. Section 4 deals with result and discussion and Section 5 summarize the paper.

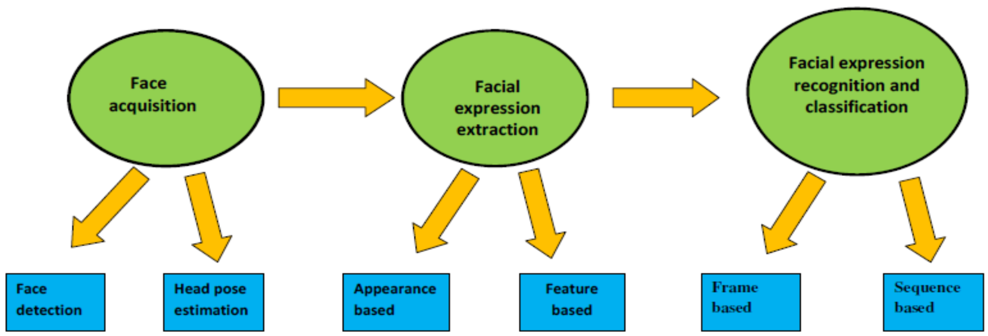


FIGURE 1: Basic steps in facial expression analysis.

2. RELATED WORKS

The accurate analysis of facial expression along with hand gesture is highly crucial in deciding the overall success of a recognition system. Accurately detecting the changes due to facial expression, its representation and classification are the major tasks of a recognition system (Figure 1). There are many challenges in facial expression features due to the dynamic nature of the signals that transmit information over time. Even the best classifiers fail to achieve accurate recognition rate due to inadequate features [4]. In facial feature extraction, for expression analysis, there are mainly three types of approaches. The holistic approach based on appearance, the geometric feature based method, and hybrid method based on the appearance and geometry [5].

In the Geometric method, the location of key facial components such as mouth, eyes, eyebrow and nose are being tracked and any variation due to expression on these parts are being targeted [4]. Subsequently, the feature vector transmits the extracted facial components at these key geometric regions on face [6]. This analysis has wide applicability in exploiting facial representation. In an approach, Candida wire frame model and active appearance algorithm for tracking and SVM for classification are used on image sequences [7].

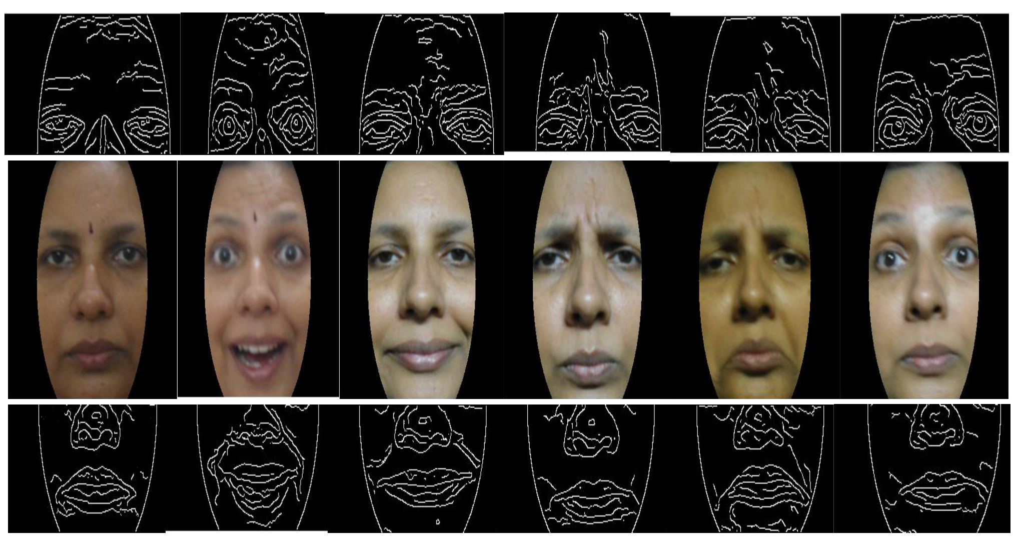
The appearance based method, on the other hand, focuses on the whole face or specific region on face to frame the feature vector [4] [5]. The appearance based features method targets the textural changes on face such as wrinkles and furrows on face. Well known Holistic approaches are based on Principal component analysis [8]. Linear Discriminate analysis (LDA) [9], Independent component analysis (ICA) [10] and Gabor wavelet analysis [11] which were applied to either the whole face or specific face regions to extract the facial appearance changes [5]. In a study, Affine Moment Invariant was used as feature vector to compute the changes in eyebrows, nose, lips and mouth area.[12] Hybrid feature were also used to recognize facial expressions. In an approach, the face was partitioned and on the upper part, the Gabor wavelet method was applied and Active Appearance Model to the whole face [13].

One of the major applications of facial expression recognition system is the Sign language recognition process [5]. Without facial expression, signer cannot convey the complete intended meaning of the signs. In these situations, transition of expressions from neutral stage to the peak expressive state would be the major change takes place [14]. These expressions and changes can vary with the signer. Very few works were done in this area accounting this complexity. The percentage of works on facial expression over the total works in sign language computing is very low. In ISL, a thorough literature search in Google scholar and web of science (Technology) revealed lack of published research in this area. Therefore research efforts are required in this area to improve the ISL based recognition systems. In ASL, head movements along with facial expression and hand gestures are a major component in sign language representation.

Gabor wavelet representation has been successfully adopted in facial expression analysis [5]. In our work, we adopt, appearance based feature expression using Gabor Wavelet as the basic method in the facial expression analysis. Facial expressions are key components of ISL and facial gestures consistently lead to change in skin textures by forming wrinkles and furrows [4]. Gabor wavelet is well-known for capturing subtle textural changes on surface [4]. Therefore, in this analysis, Gabor wavelet parameters with Euclidian distance measure and Multi class SVM classifier were used to identify facial expressions in ISL.

3. RECOGNITION FRAMEWORK

Facial expressions in ISL are described using facial feature changes. A simplified description of six main expressional changes considered in this paper are represented in (Figure 2).Our recognition system uses the changes that appears in the upper and lower face areas of the signer to classify the expressions during the information exchange (Fig.2). The four main phases in this recognition system are presented in the block diagram (Fig.3).



(a)

(b)

(c)

FIGURE 2: Row (b) shows the expressions neutral, surprise, happy, angry, sad and WH grammatical Marker (question). Row (a) and (c) shows the wrinkle changes in the upper and lower area of the face during signing.

3.1 Image Acquisition

Videos of isolated expression sequences, representing a short ISL sentence were recorded. In all the videos, signers' facial expression starts from neutral expression state and evolved into apex or peak point of expression. The length of the sequences varies depending on the facial expression and the signer/subject. The capturing is done using camera with frontal view of the person.

3.2 Pre-processing

This phase of the system is to extract the expressed faces from the frames of the video for recognition process. The two sub tasks involved in this pre-processing phase are the detection of the face from the frames and the extraction of frames with apex expressions. The cropped face area alone is used in the later phases of our recognition system.

3.2.1 Frame Extraction

In a video, there are two phases of interest. The neutral phase and the apex phase where maximum variation from the neutral phase occurs. The most relevant information is present in the final phase or apex phase and is crucial in facial expression recognition [14]. Following this methodology, the first frame representing neutral expression and last frame with peak expression are extracted for this study.

3.2.2 Face Detection and Extraction

Aim of the module is to extract the face area from extracted frames. Viola and Jones method based on Haar-like features and the AdaBoost learning algorithm [15] was used to detect the face part from the frames. Detected face part was cropped as a square, resized and masked into an elliptical shape as shown in Figure 4 to get the exact area of facial expression. Elliptically shaped cropped face would be partitioned horizontally along the elliptical centre into upper and lower face regions. The RGB segmented face areas were then converted into a normalized gray scale image. Figure 3 shows the example of the facial extraction process.

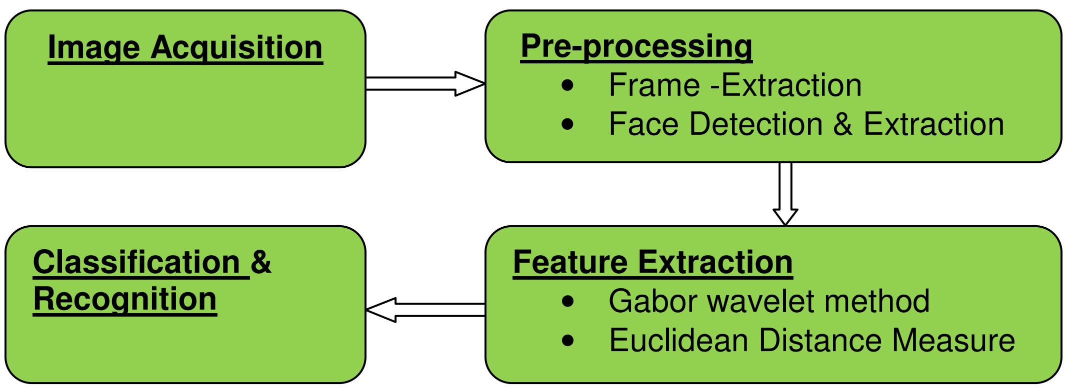


FIGURE 3: The four main phases in recognition system.

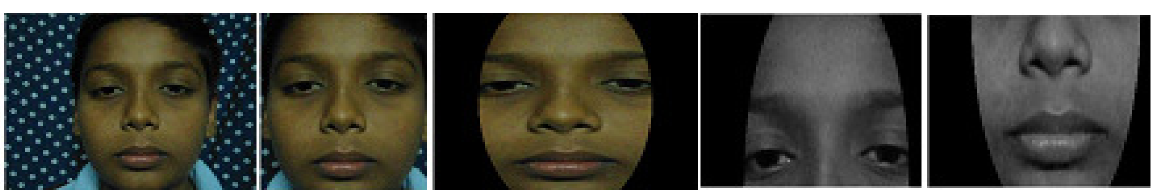


FIGURE 4: Steps in facial region extraction.

Function of facial expression system as shown in block diagram (Fig. 3) is summarized as follows.

3.3 Feature Extraction

The Figure 2 clearly indicates the importance of wrinkles on facial expression recognition. This motivates us to apply the well known textural analysis feature extraction method to the recognition process. Gabor wavelet method, which is well known for textural analysis and facial feature recognition, was chosen in our recognition system to represent the facial expressions. The signers' expression intensity and duration variations are not unique. Therefore, some common measuring criteria are required for modelling facial expression changes in an image sequence. Facial expressions were described with the help of Gabor wavelet parameters chosen from the partitioned facial areas. These feature parameters were incorporated with Euclidean distance measure to represent the facial expression changes in an image sequence.

3.3.1 Gabor Wavelet Feature Representation

Gabor features were calculated by convolution of input image with Gabor filter bank [16] [17]. Gabor filter works as a band pass filter for the local spatial frequency distribution thereby achieving an optimal resolution in both the spatial and frequency domain. The 2D Gabor filter $\psi(x,y,f,\theta)$ can be represented as a complex sinusoidal signal, modulated by a Gaussian kernel function as in Eq (1).

$$\psi(x,y,f,\theta) = [1/2\pi\sigma^2]\,[\exp\{-(x_1^2 + y_1^2) / 2\sigma^2\}]\,[\exp(2\pi f x_1)] \qquad (1)$$

where $x_1 = x \cos\theta + y \sin\theta$
$y_1 = -x \sin\theta + y \cos\theta$

σ is the standard deviation of Gaussian envelop along the x, y dimension, f is the central frequency of the sinusoidal plane wave, θ is the orientation of gabor filter.

Feature extraction procedure can then be written as the convolution of gray scale facial expression image $I(x,y)$, with the Gabor filter $\psi(x,y,f,\theta)$ as in Eq (2).

$$G_{(u,v)}(x,y) = I(x,y) * \psi(x,y,f,\theta) \qquad (2)$$

In Eq (2), $G_{(u,v)}(x,y)$ represent the complex convolution output which can decomposed into real and imaginary part as follows:

$E_{(u,v)}(x,y) = Re[G_{(u,v)}(x,y)]$ and $O_{(u,v)}(x,y) = I_m[G_{(u,v)}(x,y)]$.

Based on these result, both the phase as well as the magnitude response of the filter can be computed. In our work Gabor feature representation was based only on the magnitude response of the Gabor filter by neglecting the phase information. Small spatial displacement causes significant variation in phase value. Due to this variation, the two Gabor features could not be directly compared. Magnitude response $A_{(u,v)}(x,y)$ of the filter can be computed as in Eq (3).

$$A_{(u,v)}(x,y) = \sqrt{E^2 + O^2} \quad \text{where } E = E_{(u,v)}(x,y) \text{ and } O = O_{(u,v)}(x,y) \qquad (3)$$

A Gabor filter bank with 5 frequencies and 8 orientations was used to extract Gabor features in our work. Down sampling was done on all magnitude response, which were then normalized and concatenated into Gabor Feature Vector.

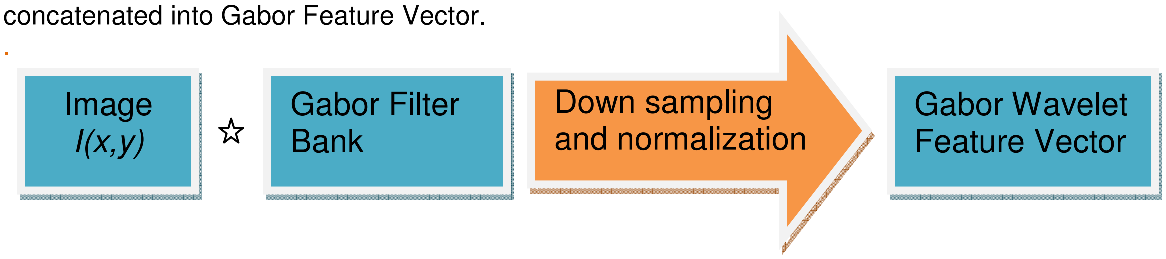


FIGURE 5: Gabor wavelet feature extraction sequence.

3.3.2 Distance Measure On Gabor Feature Vector

Gabor wavelet feature vector representing magnitude information in 2D real matrix form was converted to one dimensional matrix without any loss in information. For a facial expression corresponding to a frame, feature vectors were generated for upper and lower face regions which was partitioned horizontally alone the elliptical centre. In our work we concentrated only on first and last frames in the videos corresponding to isolated expression sentences. By this processing method, four feature vectors were extracted from a video where two feature vectors were representing the upper part of the faces in the first and last frame and another two vectors representing the lower parts. Later, the Euclidean distance measure was applied between feature vectors corresponding to the upper parts and lower part feature vectors. In addition, the percentage change that occurs in the upper and lower face areas corresponding to the total change between the neutral and peak expression vectors were calculated. All these four measures later act as the feature descriptors for the facial expression changes.

3.4 Expression Classification using SVM

Support Vector Machine (SVM) classifier exhibit high classification accuracy for small training sets and good generalization performance on difficult to separate data set. Because of the superior classification performance, SVM was chosen for this study for facial expression changes. Displacement measure vectors corresponding to the facial expression changes, extracted in the previous phase were fed to the SVM for classification training.

SVM originally developed as a linear binary classifier [18] based on optimal separating hyper plane [Eq (4)] in the feature space and decision function [Eq (5)]. For a given training data pair (x_i , y_i), y_i € {+1, _1}.

$$(w \cdot x) + b = 0,\ w \in R^n, b \in R, \text{ weight vector } w, \text{ bias } b \quad (4)$$

$$f(x) = (\Sigma_i \alpha_i y_i .\ k(x_i \cdot x) + b) \quad (5)$$

The SVM map the original input space into a higher dimensional feature space in order to achieve a linear solution [19] [20]. This mapping is done using kernel function. In Eq (5) k (x_i . x) is the kernel transformation function. This can be a linear or non-linear kernel. Lagrange co-efficient α_i chosen from training samples are called support vectors and the decision function is defined by these support vectors. SVM with Radial Basis Function (RBF), a non –linear kernel function was used in this work for the analysis of facial expression changes.

4. RESULT AND DISCUSSION

The accuracy and performance of the proposed system were evaluated by using a set of videos that involve some common facial expression patterns used in the construction of ISL sentences.

4.1 Data Set

Ten video frames of isolated facial expression for happiness, sadness, angry, surprise and question were taken from five different persons at different time and location. The testing data set was chosen from different signers (person independent) and also from the available video sequences from internet. All the extracted frames for training and test were converted into normalized gray scale image of resolution 320 x 240 before processing.

4.2. Performance Analysis

For performance analysis, the facial expression was partitioned into upper and lower halves. The Gabor wavelet methodology was applied to both halves at the neutral frame and peak frame. The variation between these frames was computed by using Euclidian distance measures. In addition, the percent variation of each half between the neutral and peak frame was computed over the whole face variation. The student paired t test was performed to test any significant changes between the upper and lower halves for each facial expression category. The results indicated significant difference between both the halves ($p < 0.001$). This clearly justifies the partitioning approach, as this proposed scheme improve the recognition accuracy compared to facial gestures as whole face. Subsequently, a training data set is created by this approach. Displacement measure vectors corresponding to the facial expression changes taken by training were fed to SVM for classification of each expression variation.

The testing phase, output were tested with statistical measures such as sensitivity (recall), specificity, precision, f-measure and accuracy. Accuracy which shows the overall correctness of a model alone is not sufficient measure to make decision on the performance of a recognition system. Along with accuracy, we need to calculate sensitivity (recall) which gives how good a test is at detecting the correct expression, whereas specificity gives an estimate of how good a method in identifying negative expressions correctly, precision is a measure of exactness and F-measure is the harmonic mean of precision and recall. Score of all these measure reaches its best values at 1 and worst score at 0. A concomitant increase or decrease in values of specificity and sensitivity is required to consider a method to be superior or inferior. Similarly, a higher value for precision along with recall is required to rate a method as superior.

In this study lowest value for all the statistical measures were more than 0.7 (Fig 6 & 7). Generally values more than 0.5 were considered to be an acceptable and successful system for recognition [21]. Confusion matrix for recognizing facial expression changes is presented (Table 1). Table 2 shows the recognition accuracy of each expressional change from neutral to peak phases. The result indicated an average accuracy of 92.12 %. In addition, the performances of other statistical measures were superior (Figure 6 and 7).

TABLE 1: Confusion matrix for recognizing facial expression changes.

Expr.	Happy	Angry	Sad	Surprise	WH_expr	Neutral
Happy	**0.9**	0.0	0.0	0.0	0.0	0.0
Angry	0.0	**0.9**	0.0	0.0	0.1	0.0
Sad	0.0	0.0	**1.0**	0.0	0.0	0.0
Surprise	0.0	0.0	0.1	**0.8**	0.1	0.0
WH_expr	0.0	0.1	0.1	0.1	**0.7**	0.0
Neutral	0.0	0.0	0.0	0.0	0.0	**1.0**

TABLE 2: Expression Recognition Accuracy.

Expression	Happy	Angry	Sad	Surprise	WH_expr	Neutral
Accuracy	96%	93%	90%	93%	87%	100%

5. CONCLUSION AND FUTURE WORKS

We proposed a method to recognize the facial expressional changes in isolated ISL sentences. Six common facial expression categories of ISL were tested with Gabor wavelet methodology. Isolated facial expression sentences in ISL were taken for training and testing. Gabor wavelet parameters from the partitioned face areas, Euclidian distance measure and Multi class SVM classifier with RBF kernel function were used in this recognition system. There was significant difference between the upper and lower halves for their variability. Statistical measures were computed and the results indicated an overall accuracy of 92.12 % for the proposed system. Also found that the values for all the statistical measures were more than 0.7. The performance analysis revealed that the method followed in this study as highly promising and this could be used for the further up gradation of facial expression recognition systems in ISL. Since there is no commonly used dataset for Indian Sign language recognition, it is very difficult to compare different methods quantitatively. The scientific literature available in Indian Sign Language (ISL) on facial expression recognition is also lacking, which further complicates the comparison. The proposed method can be further extended for the recognition of multiple facial expressional changes in ISL sentences.

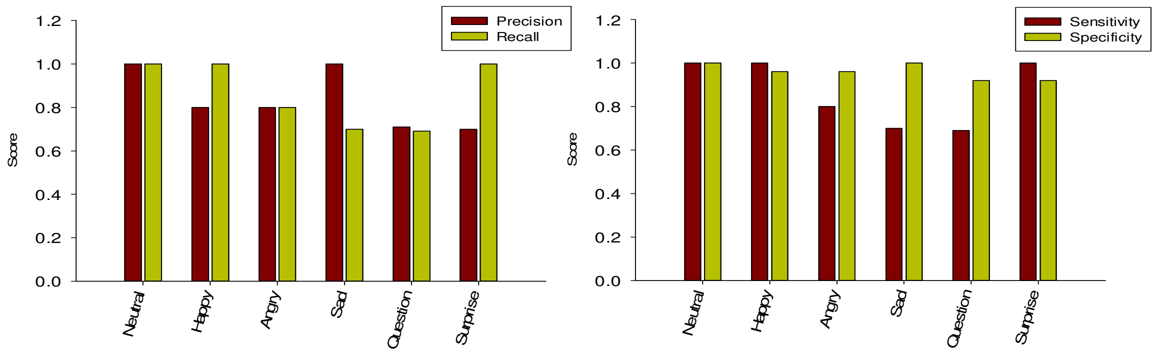


FIGURE: 6 The statistical measures precision, recall, specificity and sensitivity for 6 facial expression categories.

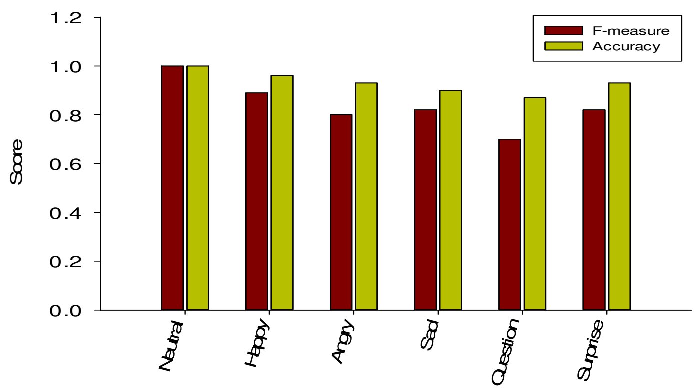


FIGURE 7: The F measure and accuracy for six facial expression categories.

6. REFERENCES

1. N. Tan Dat and S. Ranganath, "Facial expressions in American sign language: Tracking and recognition", Pattern Recognition, vol 45, pp. 1877-1891, 2012.

2. M.W. Morgan,"Topology of Indian Sign Language verbs from a comparative perspective", An annual review of South Asian languages and linguistics,vol 222, pp. 103-131, 2009.

3. B. Bridges and M. Metzger," Deaf tend your", Calliope press. Silver Spring, 1996, pp 27-38 .

4. Y.L. Tiam, T. Kanade and J.F. Cohn, " Facial expression analysis", Springer New York, 2005, pp. 247-275.

5. C. Shan, S. Gong and P.W. McOwan," Facial expression recognition based on local binary patterns: A comprehensive study", Image and Vision Computing, vol. 27, pp. 803-816, 2009.

6. D. Das,"Human's facil parts extraction to recognize facial expression", International journal on information theory, vol. 3, pp. 65-72. 2014.

7. R.A, Patil, V. Sahula and A.S. Mandal,"Facial expression recognition in image sequences using active shape model and SVM", Uksim Fifth European Modelling Symposium on Computer Modelling and Simulation,, 2011, pp. 168-173.

8. M. Turk and A.P. Pentlanad,"Face recognition using Eigen face", IEEE Conference on computer vision and pattern recognition, 1991.

9. P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman,"Eigen faces versus Fisher faces: Recognition using class specific linear projection", IEEE Transactions on pattern analysis and machine intelligence, pp. 711-720, 1997.

10. Bartlett, M.S., J.R. Movellan and T.J. Sejnowski," Face recognition by independent component analysis", IEEE Transactions on Neural Networks, pp. 1450-1464, 2002.

11. M.J. Lyions, J. Budynek and S. Akametsu,"Automatic classification of single facial image", IEEE Transactions on pattern analysis and machine intelligence, pp. 1357-1362, 1999.

12. R. Londhe and V. Pawar,"Facial expression recognition based on affine moment invariants", International Journal of Computer Science, vol. 9, pp. 382-392, 2012.

13. Y. Zhan and G. Zhou,"Facial expression recognition based on hybrid features and fusing discrete HMMs", Virtual reality-Book chapter, Springer 2007, pp. 408-417.

14. J.M. Gold,"Efficiency of dynamic and static facial expression",Journal of vision, pp. 1-12, 2013.

15. P. Viola and M. Jones,"Rapid object detection using a bossted cascade of simple features", Computer visison and pattern recognition, 2001.

16. C. Liu and H. Wechsler," Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition", IEEE Transactions on image processing, 2002, pp. 467-476.

17. V. Struc and N. Vavesic,"From Gabor magnitude to Gabor face features: Tackling the problem of face recognition under severe illumination changes",Face recognition-Chaper 12. www.intechopen.com, 2011, p. 215-238.

18. V. Vapnik,"The nature of statistical learning theory",Springer, 2000.

19. S. Sindhumol, K. Anil and B. Kannan,"Abnormality detection from Multispectoral brain MRI using multiresolution independent component analysis", International journal of signal processing, image processing and pattern recognition, pp. 177-190, 2013.

20. I. Ari and L. Akarun,"Facial Feature Tracking and Expression Recognition for Sign Language",2009 IEEE 17th Signal processing and communications applications conference, 2009, pp. 479-482.

21. S.S.K, Nair, N.V.S. Reddy and K.S. Hareesha," Explointing heterogenous features to improve insillico prediction of peptide status of amyloidogenic or non-amyloidogenic", BMC bioinformatics, pp. 2-9. 2011.