International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 12 (2015) pp. 30509-30519

http://www.ripublication.com

Facial Expression Analysis using Hybrid Transform approach

A.Punitha and M.Kalaiselvi Geetha

Assistant Professor, Dept. Of CSE ,
Annamalai University,
Tamil Nadu, India. mail-id: 12charuka17@gmail.com
Associate Professor, Dept. Of CSE ,
Annamalai University,
Tamil Nadu, India. mail-id: geesiv@gmail.com

Abstract

A novel technique for facial expression recognition from image sequences is proposed in this paper. Since human face plays a vital part in articulating emotions, the work exploits the temporal templates (Motion History Image – MHI) to capture the dynamics of face actions. Further, a Level 2 haar wavelet decomposition using Discrete Wavelet Transform (DWT) is done on the MHI image. The low frequency sub-band of the level 2 decomposed imagesis given as input to Discrete Cosine Transform (DCT) and 30 DCT coefficients are extracted using zig-zag approach. The experiments are carried out on image sequences, by considering five facial expressions viz., (happy, disgust and surprise) and two strange expressions. Support Vector Machine (SVM) and Bayesian Network are used to classify the facial expressions. The approach achieves overall recognition accuracy of 89.90% with SVM Polynomial kernel and 87.32% Bayesian Network respectively.

Keywords Motion History Image(MHI);Discrete Wavelet Transform; Discrete Cosine Transform; Zig-Zag approach; Support Vector Machine; Bayesian Network;

1. Introduction

Emotion is mostlyexposed by visual, spoken, and other physiological means. One of the important way humans display emotions is through facial expressions. Recent psychological research has shown that facial expressions are the most expressive way in which humans display emotion. The objective of facial-expression recognition techniques is to determine the expression or mental state from appropriate facial features extracted from video images automatically.Since the early 1990s, recognition

system has enthralled many researchers from numerous disciplines, such as the fields of robotics, psychology, and computer vision. Moreover, it has gained prospective applications in areas such as human-computer interaction systems, image retrieval; face modelling, patient monitoring systems for pain and depression detection, face animation and drowsy driver detection.

The paper is organized as follows. In Section 1.1, various works of the literature aiming torecognize facial expressionsis presented. A detailed description of the proposed system is described in Section 2. Section 3 describes the classifier used in this work. In Section 4, experimental results obtained with the proposed work are presented. Finally Section 5 concludes this work.

1.1 Related Work

Over the last two decades, intensive research has been carried out to develop more robust person-independent facial expression recognition systems that work in real-time and in difficult scenarios with, facial makeup, lighting changes, different ages and races, low resolution images, and facial occlusions [1].A review of existing methods on facial expression is seen in [2, 3, 4]. The authors in [5] used Local binary pattern and SVM to recognize facial expressions. In [6], moment invariants and HMM is used to recognize facial expressions. A real time face expression recognition from video using SVM is presented in [7] and a clustering based approach is seen in [8].Wavelet based facial expression recognition using a bank of SVM is presented in [9]. A distance based feature is introduced in [10] and automatic codebook generation is carried out to recognize facial expressions.In [11], a new real-time emotion retrievalscheme in video with image sequence features such as color information, keyframe extraction, video sound, and optical flow is proposed.

2. Overall Architecture of the Proposed Work

The visual data is made up of audio, video and text information. Current emotion recognition techniques focus on these features by combining two or more of them. However, it is perceived that for classification, visual information alone is adequate. This observation is based on the idea that, facial expressions have a greater influence on the listener than the spoken communication and their modulations. This paper deals with face expression recognition that aims to recognize human emotions from video sequences. To localize the face region, the method proposed by Viola and Jones [12] is used. Once the face is detected, the motion history image (MHI) of the face region is obtained. From the face MHI, low frequency sub band (LL part) of the second level decomposition of 2D Discrete Wavelet Transform is extracted. This LL region of the image is further given as input to Discrete cosine Transform and the 30 DCT coefficients are selected using zig-zag approach and given as input to classifier for expression recognition. Figure 1 depicts the overall architecture of the proposed work.

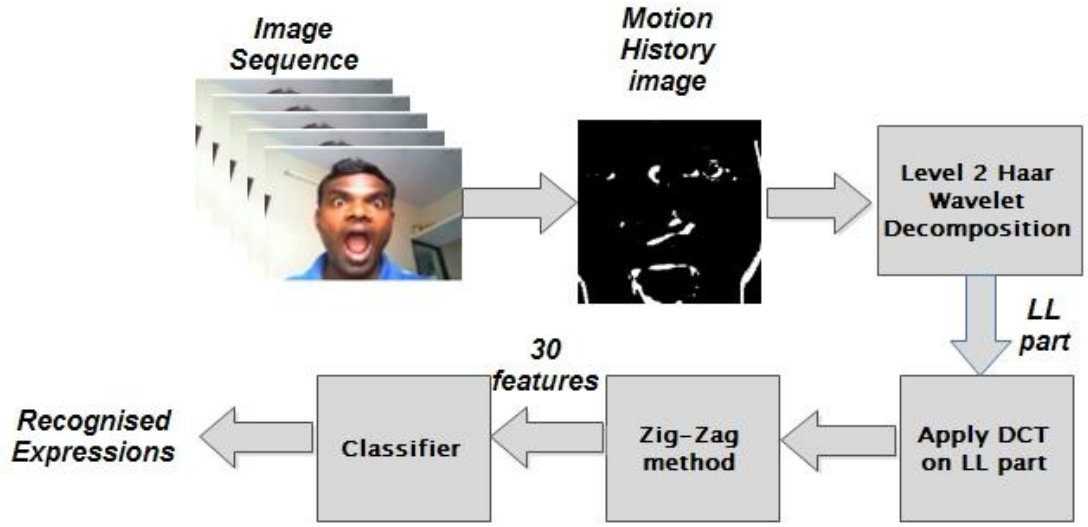

Fig. 1 Workflow of the proposed approach

2.1 Motion identification using Motion History Image(MHI)

The MHI is a cumulative gray scale image that incorporates the temporal information of the motion. Intensity of each pixel in the MHI is a function of motion density at that location. Motion history Image - its variants and applications is presented in [13] . MHI is obtained by using :

$$H_\tau(x,y,t) = \begin{cases} \tau & if \varphi(x,y,t) = 1 \\ \max(0, H_\tau(x,y,t-1) - \delta) & Otherwise \end{cases} \quad (1)$$

Here (x, y) and t show the position and time, φ (x, y, t) signals the presence of motion in the current video image, the duration τ decides the temporal extent of the movement (e.g., in terms of frames), $\delta$is the decay parameter.

The update function $\varphi(x,y,t)$ is called for every new video frame in the sequence and is defined by

$$\varphi(x,y,t) = \begin{cases} 1 & if\ D(x,y,t) \geq \in \\ 0 & Otherwise \end{cases} \quad (2)$$

where

$$D(x,y,t) = |I_k(x,y) - I_{k+1}(x,y)|, 1 \leq x \leq w, 1 \leq y \leq h \quad (3)$$

$I_k(x, y)$ is the intensity value of the pixel (x, y) in the k^{th} frame, *w* and *h* are the width and height of the frame respectively. The value of $\in$ = 40, $\delta = 10$,w=150 and h=150 is used in the experiment. Figure 2 shows the sequence of image and their corresponding MHI for various expressions.

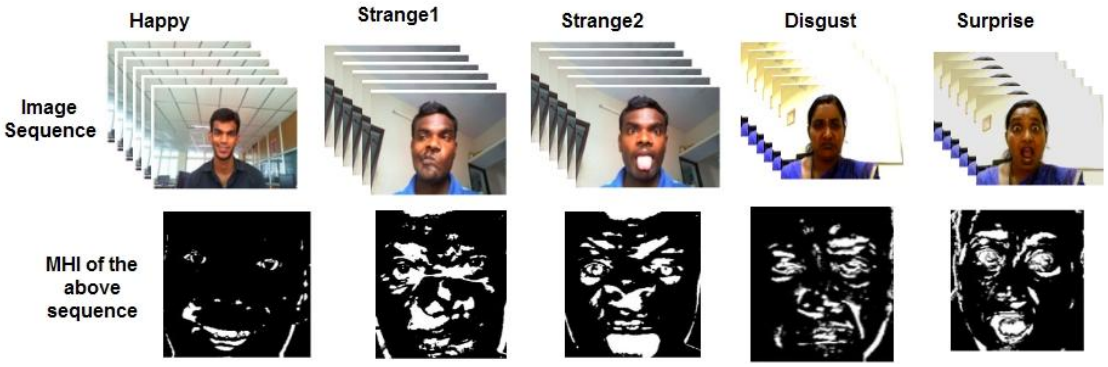

Fig 2. Image sequence & their corresponding MHI

2.2 Feature Extraction

In order to recognize facial expressions from the image sequence, a set of significant and descriptive parameters that designates a specific facial expression is essential to be extracted from the image, so that this parameter can be used to discriminatebetween different expressions. This set of parameters signifying an image is termed as the feature set of the image and the quantity of information extracted from the image to the feature set is the most essential characteristic of any efficacious feature extraction technique. The following sub section presents the feature used in this work.

2.2.1 Hybrid DWT +DCT Approach

Discrete wavelet transform is a popular method in signal processing and has been used in various research fields. Wavelet transform has advantages of multi-resolution and multi-scale decomposition. In frequency domain, when the facial image is decomposed using two dimensional wavelet transform, four sub region are obtained as shown in Figure 3.These regions are: one low-frequency region LL (approximate component), and three high-frequency regions, namely LH (horizontal component), HL (vertical component), and HH (diagonal component), respectively. The LL image is generated by two continuous low-pass filters; HL is filtered by a high-pass filter first and a low-pass filter later; LH is created using a low-pass filter followed by a high-pass filter; HH is generated by two successive high-pass filters. Subsequent levels of decomposition follow the same procedure by decomposing the 'LL' sub image of the previous level.

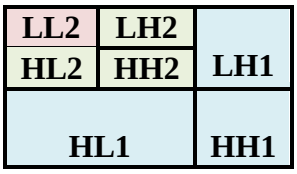

Fig.3 Level 2 wavelet decomposition

The main feature of DWT is the multi scale representation of a function. By using the wavelets, a given image can be analysed at various levels of resolution. Since the LL part contains most important information and discards the effect of noises and irrelevant parts, the LL part is adopted for further analysis in this paper. In this work,two-level 2-D discrete wavelet decomposition is performed on the MHI face images.

DCT has emerged as a widespread transformation methodcommonly used in signal and image processing. This is due to its robust energy compaction property: most of the signal information tends to be concentrated in low-frequency components of the DCT. DCT was found to be an effective method that yields high recognition rates with low computational complexity.

The LL part of the DWT image is given as input toDiscrete Cosine Transform (DCT) and a 2D DCT coefficient matrix is obtained. For most images, much of the signal energy lies at low frequencies; these appear in the upper left corner of the DCT. The zigzag scanning approach shown in Figure 4 is used for coefficient selection and the top 30 co-efficients are selected as features.

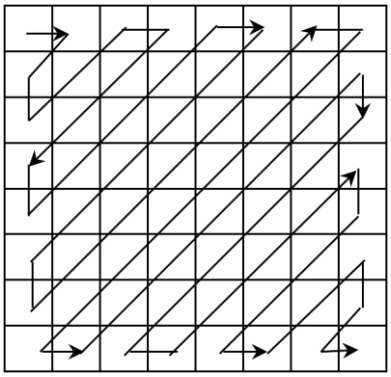

Fig. 4 Zig-Zag Scanning approach

3. Classifiers

Classification techniques explore the geometric properties of various image features and organizes them into categories.In this work, SVM classifier and Bayesian Network classifier are used in order to evaluate the effectiveness of the hybrid transform feature on the MHI of the face image.

3.1 Support Vector Machine

Support Vector Machines [14,15] are a learning method for pattern recognition problem introduced by V. Vapnik. Support Vector Machines classify data through determination of a set of support vectors, through minimization of the Structural Risk. The support vectors are members of the set of training inputs that outline a hyperplane in feature space. This k-dimensional hyperplane, where k is the number of features of the input vectors, defines the boundary between the different classes.

The classification task is simply to determinate which side of the hyperplane

the testing vectors reside in. Minimizing the structural risk reduces the average error of the inputs and their target vectors. The support vector algorithm approximately performs Structural Risk Minimization. Given a set of training examples (x_1y_1), (x_2, y_2)...(x_ky_k), if there is a hyper plane that separates the positive and negative examples, than the points x which lie on the hyper plane satisfy $w.x_i + b = 0$, where w is normal to the hyper plane and b is the distance from the origin. The margin of a separating hyper plane is defined as the shortest distance to the closest positive or negative example. The support vector algorithm looks for the separating hyper plane with the largest margin.

By applying what is known as the kernel trick, a nonlinear mapping functionmaps the original data points onto a higher dimensional feature space in which linear separability could be attained. The SVM then linearly classifies the transformed data points in the new feature space, even though the input space may not be linearly separable. The mapping function is known as the kernel function, and this work employs polynomial kernel of degree 3 and RBF kernel to carry out the classification task.

3.2 Bayesian Network

Bayesian networks [16] can represent joint distributions in an instinctive and effective way, as such, Bayesian networks are naturally suitable to classification. Bayesian network is used to compute the posterior probability of a set of labels given the observable features, and then the features are classified with the most probable label.

A Bayesian network is composed of a directed acyclic graph in which every node is connected with a variable X_i and with a conditional distribution $p(X_i|\pi_i)$, where π_i denotes the parents of X_i in the graph. The directed acyclic graph is the structure, and the distributions $p(X_i|\pi_i)$, represent the parameters of the network.

4. Experimental Setup

The experiments are carried out in MATLAB R2013 in Windows 7 operating system with Intel Xeon X3430 Processor 2.40 GHz with 4 GB RAM.

To demonstrate the effectiveness of the proposed approach, experimental studies are conducted on video sequencescollected from 15 subjects. Five different expressions exposed by the subjects are considered for experimental purpose and from each subject a video of 5 minutes duration is recorded. The images are captured by using Microsoft HD web camera at a rate of 25fps. A few example expressions are shown in Figure 5.

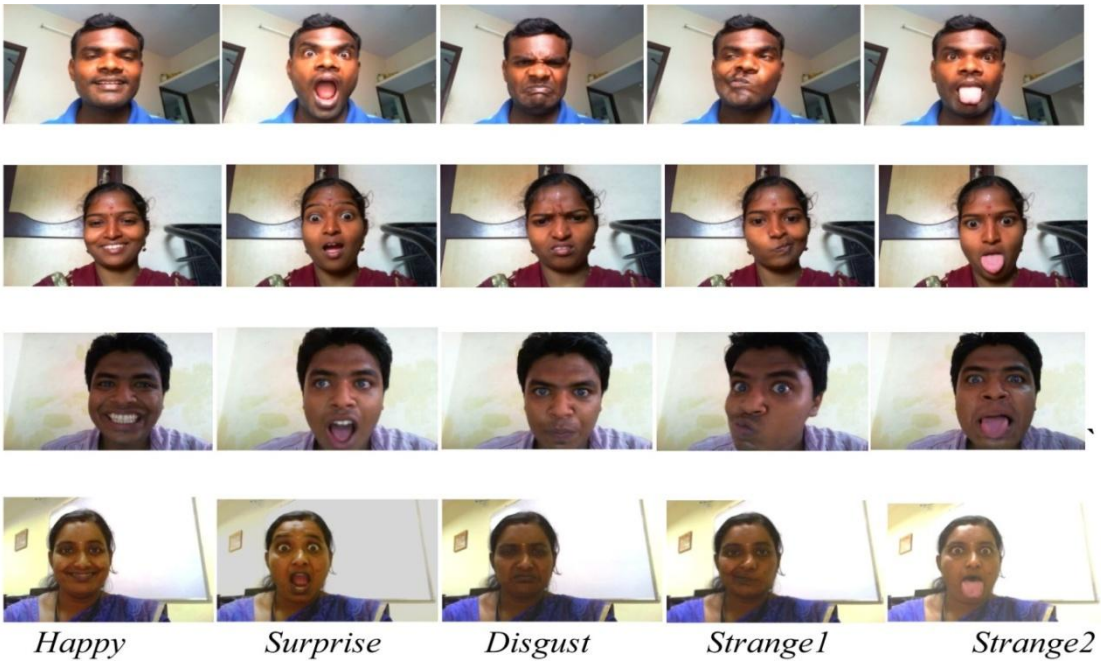

Fig. 5 Sample Expressions exposed by the subjects

4.1 Results and Discussions

A hybrid combination of the DWT and the DCT is used to extract features from MHI face images, which is further fed to classification using SVM and Bayesian network. The feature is obtained for all the training and testing frames. For training, expressions given by 9 subjects are used and expressions given by 6 subjects are used for testing. The training samples are labeled as 0 for Happy, 1 for Disgust, 2 for Surprise, 3 for Strange1 and 4 for Strange2 expressions.

In the proposed work, Level 2 Haar wavelet decomposition is employed on the face MHI image . Level 1 decomposition of original image in Figure 6 produces one approximation and three orientation detail images as shown in Figures 6(a), 6(b), 6(c) and 6(d) i.e., LL, LH, HL and HH sub-bands respectively. Figure 6(a) holds the highest energy of the image concentrated on low frequency components while the other three sub-bands hold much lesser energies. Similarly, for Level 2 Haar wavelet decomposition the LL sub-band image Figure 6(a) is decomposed, which produces four images as shown in Figures 6(e), 6(f), 6(g) and 6(h). In the proposed work, 2D-DWT is used to extract the coefficients of lowest frequency range in sub-bands. Therefore, the LL sub-band part having the highest image energy is selected for feature extraction.

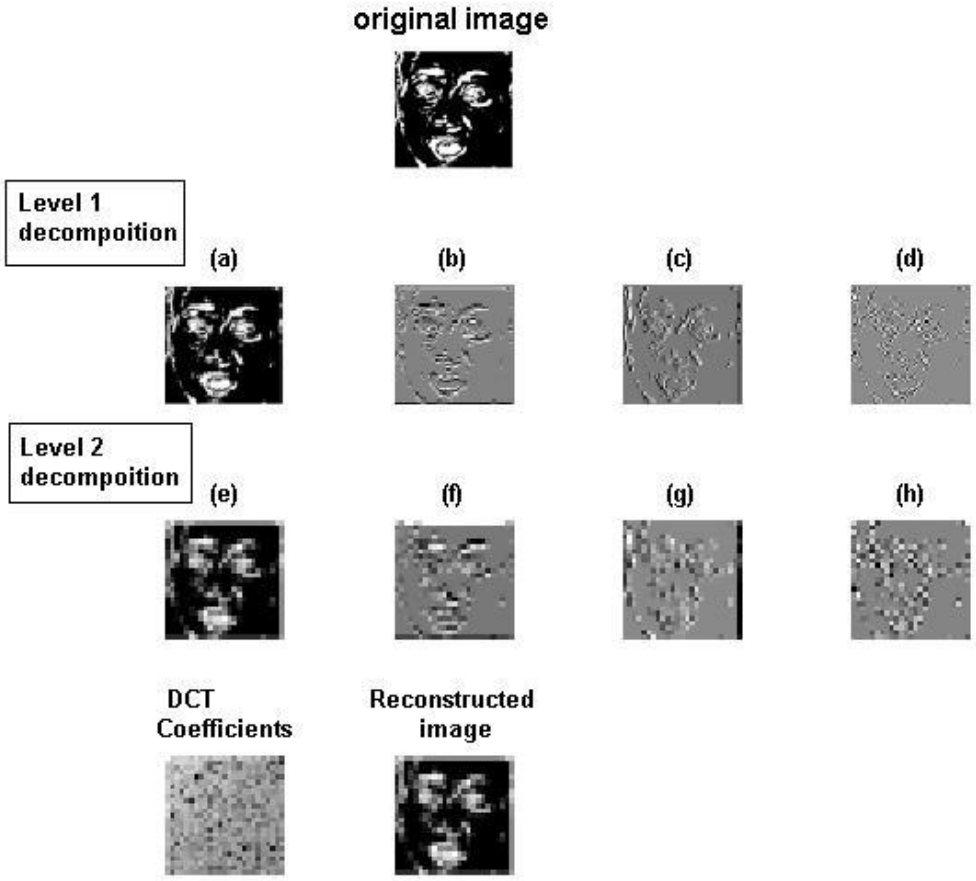

Fig 6. Result of Level 2 Haar wavelet Decomposition

Secondly, the 2D-DCT of level 2 LL sub-band component is computed. The reconstructed image by computing the 2D-IDCT is depicted in Figure 6.

The zig-zag approach is used to select the DCT co-efficient and 30 features are selected using this approach and given as input to SVM and Bayesian Network. The experimental results shows that SVM polynomial kernel outperforms Bayesian network and is able to recognize the expression of the test samples with 89.90% accuracy using SVM polynomial kernel and 87.32% with Bayesian Network respectively as shown in Figure 7. Table 1 shows the F1-measure obtained for the individual expressions considered in this work and it is observed that the system is well able to recognize strange2 and happy expressions when compared to that of other facial expressions.

4.2 Evaluation Metrics

A systematic study of performance measure for classification tasks is presented in [17].Precision (P) and Recall (R) are the generally used assessment metrics and these measures are used to evaluate the performance of the proposed system. These measures provide the best perspective on a classifier's performance for classification. The measures are defined as follows:

$$\text{Precision(P)} = \frac{\text{No. of True Positives}}{\text{No. of True Positives} + \text{False Positives}}$$

$$\text{Recall(R)} = \frac{\text{No. of True Positives}}{\text{No. of True Positives} + \text{False Negatives}}$$

The work used F1-measure(F1) as the collective measure of Precision (P) and recall (R) for calculating accuracy which is defined as follows:

$$F1 - measure = \frac{2.Precision.Recall}{Precision + Recall}$$

Table 1 F1-measure for various Expressions

Classifier	Happy (%)	Surprise (%)	Disgust (%)	Strange1 (%)	Strange2 (%)
SVMPolynomial kernel(degree 3)	93.2	83.9	92.6	85.4	94.3
SVMRBF kernel	61.4	57.3	60.5	55.2	62.3
Bayesian Network	90.2	82.8	89.3	84.2	91.4

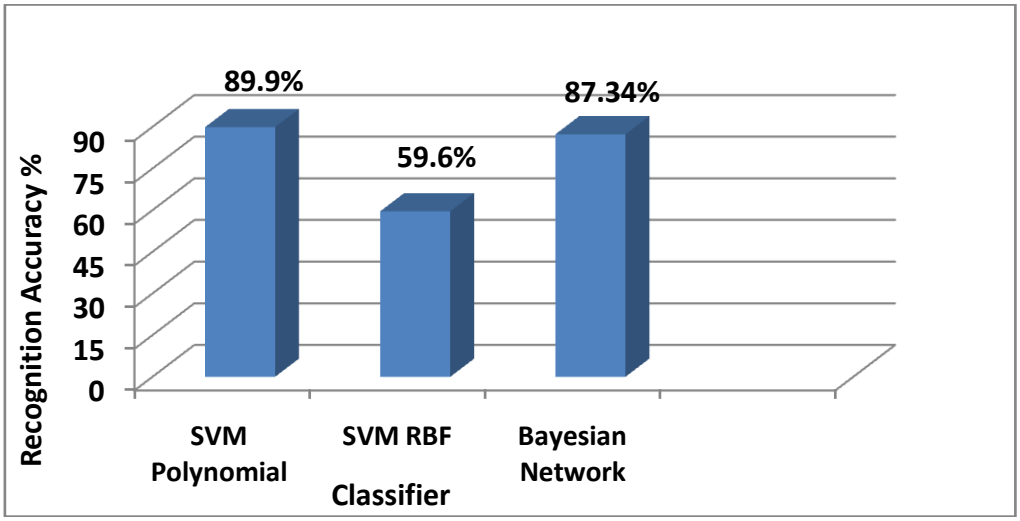

Fig.7 Overall Recognition Accuracy

5. Conclusion

In this paper, a hybrid transform based approach to recognize facial expressions from a sequence of image is presented. The work exploits temporal templates to capture the dynamics of face actions and then uses hybrid transform features to characterize the facial expressions.The approach achieves overall recognition accuracy of 89.90% with SVM Polynomial kernel and 87.32% Bayesian Network respectively. The future work aims to apply the proposed method on occluded images to recognize facial expressions .

References

[1] Kotsia, I.; Buciu, I. Pitas, I. , 2008. "An analysis of facial expression recognition under partial facial image occlusion", Image Vision Computing, 26: 1052–1067 .

[2] Pantic, M., and Rothkrantz, L, 2000. "Automatic analysis of facial expressions: the state of art", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (12) : 1424-1445.

[3] Ting Wu, Siyao Fu, Guosheng Yang , 2012. "Survey of the Facial Expression Recognition, Advances in Brain Inspired Cognitive Systems", Lecture Notes in Computer Science , 7366: 392-402 .

[4] Lienhart, R., Fasel, B., and Luettin, J, 2003. "Automatic facial expression analysis: a survey"**,** Pattern Recognition ,259-275.

[5] Caifeng Shan,Shaogang Gong,Peter W. McOwan, 2009. "Facial expression recognition based on Local Binary Patterns: A comprehensive study",Image and Vision Computing,.27(6): 803-816.

[6] Y. Zhu , L.C. De Silva, C.C. Ko, 2002. "Using moment invariants and HMM in facial expression recognition", Pattern Recognition Letters,.23 (1-3): 83-91.

[7] Philipp Michel, Rana L Kaliouby, 2003. Real time facial expression recognition in video using support vector machines, Proceedings of the 5th international Conference on Multimodal interfaces,ACM,NewYork,pp 258-264.

[8] Xue-wen Chen, Thomas Huang, 2003. "Facial expression recognition: A clustering based approach", Pattern recognition Letters, 24(9-10):1295-1302.

[9] Sidra BatoolKazmi, Qurat-ul-Ain ,M. ArfanJaffar, 2012."Wavelets-based facial expression recognition using a bank of support vector machines ",Soft Computing, 16:369–379.

[10] Fu-Song Hsu ,Wei-Yang Lin , Tzu-Wei Tsai, 2014. "Facial expression recognition using bag of distances", Multimed Tools Applications,73:309–326.

[11] Jae-Khun Chang ,Seung-TaekRyoo, 2014. "Real-time emotion retrieval scheme in video with image sequence features", Journal of Real-Time Image Processing , 9: 541–547 .

[12] P., & Jones, M. J., 2004. "Robust real-time face detection",International Journal of Computer Vision, 57: 137–154.

[13] Atiqur Rahman Ahad ,J. K. Tan ,H. Kim , S. Ishikawa , 2013. "Motion history image: its variants and applications" , Machine Vision and Applications ,23:255-281.

[14] Vapnik, V, 1998. Statistical Learning Theory. Wiley, NY.

[15] Lewis, J. P 2004 . Tutorial on SVM. CGIT Lab, USC.

[16] Friedman, N., Geiger, D. and Goldszmidt, M., 1997. "Bayesian Network Classifiers.,Machine Learning", Kluwer Academic Publishers, Boston, 29: 131-163.

[17] Marina Sokolova , Guy Lapalme , 2009. "A systematic analysis of performance measures for classification tasks", Information Processing and Management 45, 427–437.