VOL. 11, NO. 1, JANUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

AUTOMATED RECOGNITION SYSTEM FOR FACIAL EXPRESSION BASED ON THE FUSION OF SPATIAL AND FREQUENCY DOMAIN FEATURES

R. Suresh and S. Audithan
Department of Computer Science and Engineering, PRIST University, Tanjore, Tamilnadu, India
E-Mail: mkrsuresh.phd@gmail.com

ABSTRACT

The development of facial expression recognition system proliferates now-a-days due to its various fields of application such as Human Computer Interaction (HCI), behavioral studies, facial nerve grading in medicine, automated tutoring system, synthetic face animation, and robotics. Among the various behavioral traits such as voice and gaits, facial expression is the most and effectual communicative way of humans. It is also a natural, non-verbal and non-intrusive communicative source. In this study, multidirectional approach based robust and automated facial expression recognition system is proposed. Contourlet transform is adopted as multi resolution and multi directional approach for feature computation along with Discriminative Robust Local Ternary Pattern (DRLTP) and Gray Level Co-occurrence Matrix (GLCM). On account of classification K-Nearest Neighbor (KNN) classifier is used based on city block distance measure. The standard Japanese Female Facial Expression (JAFFE) database is utilized to evaluate the performance of the proposed algorithm. Experimental result shows that the proposed system achieves satisfactory performance of over 91%.

Keywords: facial expression recognition, human computer interaction, contourlet transform, discriminative robust local ternary pattern, gray level co-occurrence matrix.

1. INTRODUCTION

Facial expression is a visible manifestation of the affective state, cognitive activity, intention, personality, and psychopathology of a person. It not only expresses our emotions but also provides important communicative cues during social interaction [1]. Automated facial expression by a computer is considered to be more objective than those labeled by people and it can be used in clinical psychology, psychiatry, and neurology [2]. Various research works has been done in the area of facial expression analysis. Some of the accomplished works are discussed in this section.

A method to perform facial expression recognition on images in the encrypted domain is presented in [3] based on local fisher discriminant analysis. This system solves the problem of needing to trust servers since the test image for facial expression recognition can remain in encrypted form at all times without needing any decryption, even during the expression recognition process. The aging effect on computational facial expression recognition based on two data bases descried in [4]. The feature dimensionality problem is investigated by using manifold learning techniques. The spatiotemporal monogenic binary patterns are used to describe both appearance and motion information of the dynamic sequences in [5]. Two-layer structure is utilized to represent the facial image by monogenic signal analysis, phase-quadrant encoding method, local XOR and spatiotemporal local binary pattern.

An automated facial expression recognition technique based on Gauss-Laguerre (GL) filter using infrared images is described in [6]. GL filter of circular harmonic wavelets are used to extract the features from infrared images. A set of redundant wavelets are generated, which enable an accurate extraction of complex texture features by using GL filters with properly tuned parameters. KNN classifier is used for classification. Partial Active Appearance Model (AAM) fitting is applied on mouth and eyes to achieve better alignment for facial features in [7]. Multi level optical flow is used to determine the initial positions of facial feature models and stable partial AAM. Dynamic face recognition system is used to recognize different users and select the trained fitting model in recognizing the facial expressions.

A method for head pose invariant facial expression recognition that is based on 2D geometric features is implemented in [8]. To achieve head pose invariance, the coupled scaled Gaussian process regression model for head pose normalization is used. Three automatic emotion recognition systems based on interval type-2 fuzzy set (IT2FS), interval approach-IT2FS, and General type-2 fuzzy sets is presented in [9]. These systems use the background knowledge about a large face database with known emotion classes to classify an unknown facial expression. All the schemes first construct a fuzzy face space, and then infer the emotion class of the unknown facial expression by determining the maximum support of the individual emotion classes using the pre-constructed fuzzy face space. The class with the highest

support is assigned as the emotion of the unknown facial expression.

A new approach for integrated face and facial expression recognition system for robotic applications is explained in [10]. Initially, facial images are acquired from web camera. AAM is applied in facial images to generate texture model. The modified Lucas-Kanade image alignment algorithm is used to find the possible facial features. The acquired parameters are used to train back propagation neural network for facial expression recognition. A meta-analysis of the first such challenge in automatic recognition of facial expressions is presented in [11]. It details the challenge data, evaluation protocol, and the results attained in two sub challenges, which are action unit detection and classification of facial expression imagery in terms of a number of discrete emotion categories.

Multimodal spontaneous facial expression database of natural visible and infrared facial expressions (NVIE) analyzed in [12]. NVIE is analyzed by four methods, which are the effectiveness of emotion elicitation, the interrater reliability for expression annotation, the relationship between spontaneous expressions and affective states, and the differences between posed and spontaneous expressions. A novel facial expression recognition system in video sequences based on Hough forest algorithm is described in [13]. The non rigid morphing facial expressions are analyzed and eliminate the person specific effects through patch features extracted from facial motion due to different facial expressions. Finally, classification and localization of the center of the facial expression in the video sequences are performed by using a Hough forest. Automated facial expression recognition from image sequences is discussed in [14]. Two approaches, color normalization, and local binary pattern are used to extract facial features.

In this study, an automated facial expression recognition system is proposed using facial images based on Contourlet Transform, DRLTP, GLCM and KNN classifier. The rest of this paper is organized as follows: The mathematical background of Contourlet Transform, GLCM and DRLTP is described in section 2. The proposed facial expression recognition approach is presented in section 3. The experimental results of the proposed system are discussed in section 4 and conclusion is made in section 5.

2. MATHEMATICAL PRELIMINARIES

The proposed facial expression recognition system is built based on Contourlet Transform, DRLTP, GLCM and KNN classifier. The mathematical preliminaries of the aforementioned techniques are discussed in this section.

2.1 Contourlet transform

The Contourlet Transform consists of a double iterated filter bank in [15]. First the Laplacian Pyramid (LP) is used to detect the point discontinuities of the image and then a Directional Filter Bank (DFB) to link point discontinuities into linear structures. The general idea behind this image analysis scheme is the use of wavelet like transform to detect the edges of an image and then the utilization of a local directional transform for contour segment detection. This scheme provides an image expansion that uses basic elements like contour segments, and thus is named as Contourlet. An advantageous characteristic of Contourlet is that they have elongated support at various scales, directions and aspect ratios, allowing the Contourlet transform to efficiently approximate a smooth contour at multiple resolutions. It is ideal for images with smooth curves as it requires far less descriptors to represent such shapes, compared to other transforms such as the discrete wavelet transforms. Additionally in the frequency domain it provides multi scale and directional decomposition.

The separation of multi scale and directional decomposition stages provides a fast and flexible transform, at the expense of some redundancy (up to 33%) due to the Laplacian Pyramid. This problem has been addressed and a critically sampled Contourlet transform is proposed [16], called CRISP Contourlet, utilizing a combined iterated non separable filter bank for both multi scale and directional decomposition. A variety of filters can be used for both the LP and the DFB. In this work, the debauches 9-7 filters have been utilized for the LP. For the DFB, these filters are mapped into their corresponding 2-D filters using the McClellan Transform [17] as proposed in [15].

2.1.1 Laplacian pyramid

At each level, the LP decomposition creates a down sampled low pass version of the original image and the difference between the original and the prediction, resulting in a band pass image. An overview of the LP decomposition process utilized for the Contourlet transform is shown in Figure-1. H and G are the low pass analysis and synthesis filters, while M is the sampling matrix, $a[n]$ is the coarse image, while $b[n]$ is the difference between the signal and the prediction, containing the supplementary high frequencies.

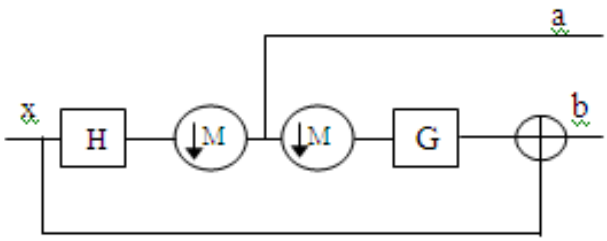

Figure-1. Laplacian pyramid decomposition process for 1-level of decomposition.

2.1.2 Directional filter bank

The DFB proposed in [18] is a 2D dimensional filter bank that can achieve perfect reconstruction. The DFB implementation utilizes *l*-level binary tree decomposition and leads to 2^l directional sub bands with wedge shaped frequency partitioning. An example of wedge shaped frequency partitioning is shown in Figure-2. The DFB, involves the modulation of the input image and the use of quincunx filter banks with diamond shaped filters has been constructed. The use of complicated tree expanding rule in order to obtain the desired frequency partition for finer directional sub band is the disadvantage of DFB. The simplified DFB proposed in [15] consists of two stages. The first is the two channel quincunx filter bank with fan filters that divides a 2-D spectrum in to vertical and horizontal directions. A quincunx filter bank consists of low pass and high pass analysis and synthesis filters and M- fold up sampler and down samplers. At filter bank shown on Figure-3, Q is a matrix used to decimate the sub band signal. In case of quincunx matrix, the filter bank is termed quincunx filter bank. Reordering of samples by Shearing operator is the second stage. Modulating the input signal is avoided by using the new construction method and for the decomposition tree's expansion it follows a simpler rule.

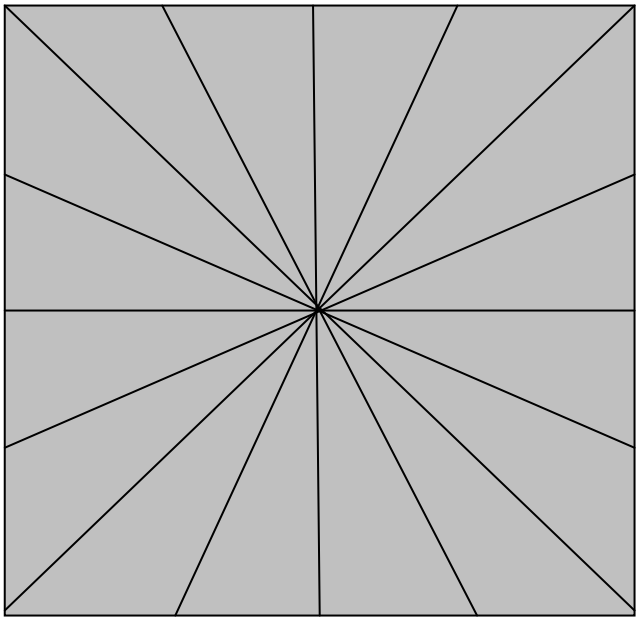

Figure-2. Wedge shaped frequency partitioning with 2^l directional sub bands (l=3).

2.1.3 Pyramid directional filter bank

The DFB is designed to capture the high frequency content of an image, which represents its directionality. DFB alone does not provide a sparse representation for images. The removal of the low frequencies from the input image before the application of the DFB is the solution to this problem. This can be achieved by combining the DFB with multi scale decomposition like the LP. By combining the LP and the DFB, a double filter bank named Pyramidal DFB (PDFB) is obtained.

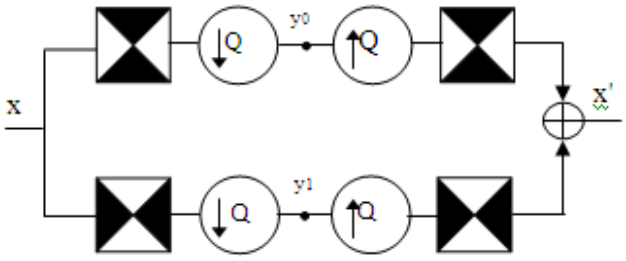

Figure-3. 2D dimensional spectrum partition using quincunx filter banks with fan filter.

Band pass images decomposed using the LP is fed into the DFB in order to capture the directional information. This scheme can be iterated on the coarse image and the iteration number is restricted only by the size of the original image due to the down sampling in each level. A double iterated filter bank which decomposes into directional sub bands at multiple scales is the combined result which named as "Contourlet filter

bank". Figure-3 shows an overview of 2D dimensional spectrum partition using quincunx filter banks with fan filter. Q is a quincunx sampling matrix and the black areas represent the ideal frequency support of each filter.

Considering $a_0[n]$ as the input image, the output of J level LP decomposition is a low pass image $a_1[n]$ and J band pass images $b_j[n], j = 1,2,..J$ from finer to coarse scale. At each level j, the image $a_{j-1}[n]$ is decomposed in to coarser image $a_j[n]$ and a detailed image $b_j[n]$.Considering l_j as the DFB decomposition level at the j^{th} level of the Laplacian pyramid's decomposition, each band pass image $b_j[n]$ is decomposed by an l_j-level DFB into 2 power l_j band pass directional images $c_{j,k}^{(l_j)}[n]$. The computational complexity of the discrete contourlet transform is $O[N]$ for N- pixel images when finite impulse response filters is used. In contourlet transform, the LP provides a down sampled low pass and a band pass version of the image in each level. The band pass image is fed into the DFB. This scheme is iterated in the low pass image.

2.2 Gray level co-occurrence matrix

The use of co-occurrence probabilities using GLCM for extracting various texture features is described [19]. GLCM is also called gray level dependency matrix. It is defined as "A two dimensional histogram of gray levels for a pair of pixels, which are separated by a fixed spatial relationship."

2.2.1 Contrast

Contrast is a measure of intensity or gray level variations between the reference pixel and its neighbor. In the visual perception of the real world, contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view. It is defined by (1)

$$Contrast = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \right\}, \quad where\ n = |i-j| \qquad (1)$$

where $p(i,j)$ is the $(i,j)^{th}$ entry in a normalized GLCM matrix and N_g is the number of distinct gray levels. When i and j are equal, the cell is on the diagonal and $i-j=0$. These values represent pixels entirely similar to their neighbor, so they are given a weight of 0. If i and j differ by 1, there is a small contrast, and the weight is 1. If i and j differ by 2, the contrast is increasing and the weight is 4. The weights continue to increase exponentially as $(i-j)$ increases.

2.2.2 Energy

Energy is also called angular second moment where it measures the textural uniformity. If an image is completely homogeneous then the energy will be of maximum. It is given by (2).

$$Energy = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j)^2 \qquad (2)$$

2.2.3 Homogeneity

Homogeneity is also named as Inverse Difference Moment (IDM), which measures the local homogeneity of an image. IDM feature obtains the measures of the closeness of the distribution of the GLCM elements to the GLCM diagonal. It is given by (3).

$$Homogenity = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{1}{1+(i-j)^2} p(i,j) \qquad (3)$$

2.2.4 Correlation

The correlation measures the linear dependency of grey levels on those of neighboring pixels. It is defined by (4).

$$Correlation = \frac{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \qquad (4)$$

where μ_x, μ_y, σ_x and σ_y are the means and standard deviations.

2.3 Discriminative robust local ternary pattern

Local Ternary Pattern (LTP) is an extended version of Local Binary Pattern (LBP). A well known problem with LBP is that it is sensitive to noise in the near uniform image regions. Thus, three-valued coding scheme named LTP is implemented by using a thresholding function around zero to evaluate the local gray scale difference to makes LPB more discriminant and less sensitive to noise [20]. The difference *d* between the gray value of the pixel *x* from the gray values of one in its neighborhood *u* is encoded by the three values according to the following threshold rule:

$$d=\begin{cases} 1 & u \geq x+\tau \\ 0 & x-\tau \leq u < x+\tau \\ -1 & otherwise \end{cases} \quad (5)$$

where u is neighbouring pixel, x is center pixel and τ is threshold value. From a computational point of view, a ternary pattern is split into two binary patterns by considering its positive and negative components. The histograms computed from these two descriptors are then concatenated. However, discrimination between a bright object against a dark background inherent in LTP. Thus to resolve the issue of brightness reversal of object and background, DRLTP is designed for texture information capture.

The k^{th} weighted LTP bin value of a MxN image block is as follows:

$$h_{ltp}(k) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \omega_{x,y} \delta(LTP_{x,y}, k) \quad (6)$$

The RLTP histogram is created from (13) as follows

$$h_{rltp}(k) = \begin{cases} h_{ltp}(k), & k = 0 \\ h_{ltp}(k) + h_{ltp}(-k), & 0 < k < \dfrac{3^B+1}{2} \end{cases} \quad (7)$$

where $h_{rltp}(k)$ is the k^{th} RLTP bin value.

The absolute difference between the bins of a LTP code and its inverted representation is taken to form Difference of LTP (DLTP) histogram as follows:

$$h_{dltp}(k) = | h_{ltp}(k) - h_{ltp}(-k) |, \qquad 0 < k < \frac{3^B+1}{2} \quad (8)$$

where $h_{dltp}(k)$ is the k^{th} DLTP bin value. RLTP and DLTP are concatenated to form DRLTP as follows:

$$h_{drltp}(l) = \begin{cases} h_{rltp}(l) & 0 \leq l < \dfrac{3^B+1}{2} \\ h_{dltp}(l - \dfrac{3^B-1}{2}), & \dfrac{3^B+1}{2} \leq l < 3^B \end{cases} \quad (9)$$

The computed DRLTP contains *both* edge and texture information.

3. PROPOSED SYSTEM

In this study, an automated facial expression recognition system is proposed using pattern recognition and machine learning approaches. The proposed automated classification of facial expressions system consists of two sequential modules: feature extraction and recognition. The process of feature extraction is of key importance to the entire recognition process. Figure-4 shows the schematic model of the proposed facial expression recognition system.

3.1 Feature extraction

Generally, feature extraction reduces the dimensionality of the input space, by retaining essential information with high discrimination power and high stability. In this study features are extracted from facial images by applying Contourlet transformation. In this module, multi-resolution and directional Contourlet transform is applied to facial images up to N-scale that produces directional sub-bands which represents the input image in different directions. From the directional sub-bands, discriminative features such as mean, entropy and DRLTP are computed as feature vectors. Along with the above statistical features, texture features such as contrast, correlation, energy, and homogeneity are also extracted by applying GLCM on facial images directly. Consequently, the extracted features from frequency and spatial domains are fused together and the feature extraction process is repeated for all the training facial images. Finally, extracted features are stored in feature database for further recognition process.

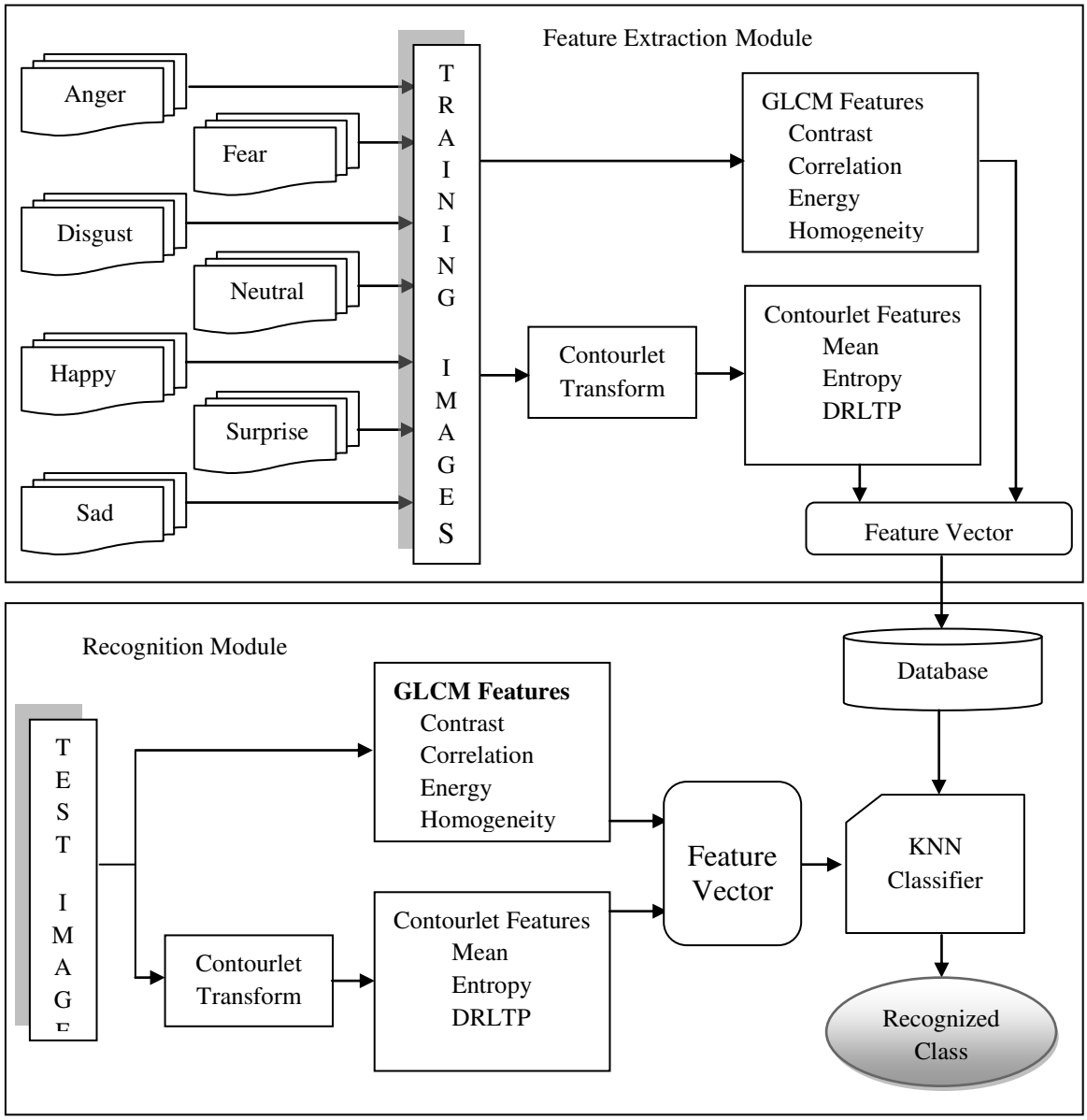

Figure-4. Schematic model of the proposed facial expression recognition system.

3.2 Recognition

The second module of the proposed approach is recognition or classification. In order to recognize the expression of a test image, the same frequency and spatial domain features are extracted and fused. Then, the extracted test features and stored database are fed into the KNN classifier, whereas minimum distance between test and stored feature database are measured using city block measure. Seven facial expressions such as anger, disgust, fear, happy, neutral, sad and surprise are taken into the analysis. Eventually, test image is recognized into one of the seven predefined expression states. The distance between test and feature database is computed using the following eqn.

$$\text{Distance}(u,v) = |x_1 - x_2| + |y_1 - y_2| \quad (10)$$

where x_1 and y_1 indicates test feature's point x_2 and y_2 indicates database feature's points. If the points have *n*-dimensions such as $u = (x_1, x_2, x_3, \ldots x_n)$ and $v = (y_1, y_2, y_3, \ldots y_n)$ then the generalized city block distance formula between these points is

$$\text{Distance}(u,v) = |x_1 - y_1| + |x_2 - y_2| + \cdots\ldots |x_n - y_n|$$

$$= \sum_{i=1}^{n} |x_i - y_i| \quad (11)$$

4. RESULTS AND DISCUSSIONS

In order to assess the performance of the proposed facial expression recognition system, experiments are carried out using benchmark facial image dataset named JAFFE database [21]. Totally 213 images are available in JAFFE database with seven categories of expression (6 vital + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. Table-1 shows the classification accuracy obtained by the proposed system.

Table-1. Recognition accuracy of the proposed emotion recognition system using fusion approach.

Emotional state	Level of decomposition								
	1	2	3	4	5	6	7	8	9
Anger	93.33	93.33	90.00	96.67	93.33	93.33	96.67	90.00	96.67
Disgust	90.00	90.00	93.33	93.33	90.00	90.00	90.00	96.67	90.00
Fear	81.25	90.63	93.75	87.50	93.75	90.63	90.63	93.75	93.75
Happy	78.13	78.13	78.13	81.25	84.38	81.25	84.38	81.25	84.38
Neutral	93.33	96.67	96.67	96.67	100.00	100.00	100.00	100.00	100.00
Sad	83.87	80.65	87.10	87.10	83.87	87.10	87.10	87.10	87.10
surprise	83.33	83.33	90.00	80.00	86.67	90.00	90.00	90.00	90.00
Average	86.18	87.53	89.85	88.93	90.29	90.33	91.25	91.25	91.70

It is observed from the Table-1 that the maximum classification accuracy of 91.70% is achieved at 9th level of Contourlet decomposition. It is due to the fact that, high level decomposition produces more informative features of facial images which tends to improve the classification accuracy. Among the seven facial expressions, only the neutral emotional state is classified without misclassification. And except Happy and Sad state, the classification accuracy of all other emotional states are over 90%. Figure-5 shows the graphical representation of average recognition accuracy corresponding with each level of decomposition. It shows the average recognition accuracy while using the features mean, entropy, fusion of mean and DRLTP features, entropy and DRLTP features and fusion of all features.

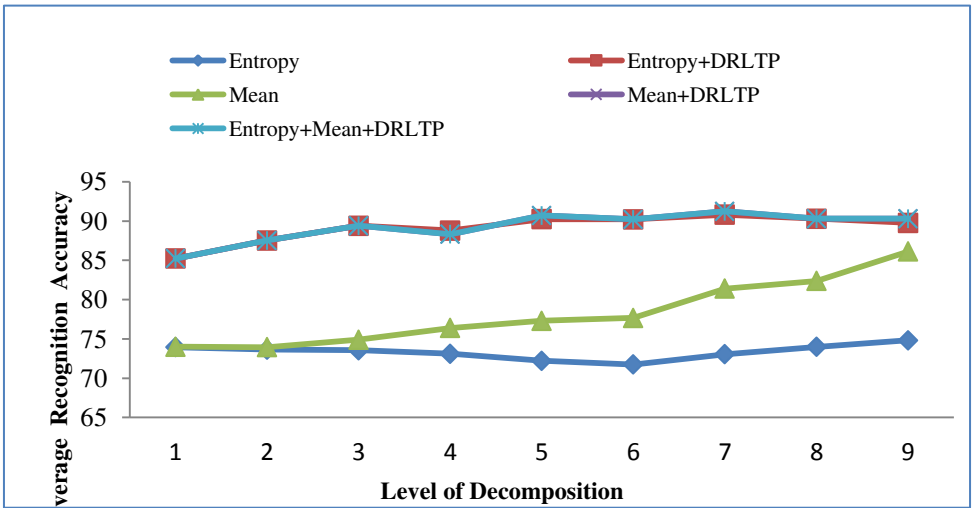

Figure-5. Average recognition accuracy Vs decomposition level.

It is evident from the Figure-5 that the proposed system produces higher classification accuracy while fusing features such as mean, entropy, and DRLTP together rather than their individual performance. Also it is

noted that the performance of the proposed system increases with respect to Contourlet decomposition level.

5. CONCLUSIONS

In this study a robust and reliable automated facial expression recognition system is proposed using spatial and frequency domain features. It is implemented by designing two sequential modules. At first, the appropriate features are extracted from facial images in the feature extraction module, whereas Contourlet features are introduced along with GLCM and DRLTP features. They are effectively discriminates the various facial expressions. KNN classifier is adopted for facial recognition stage, whereas minimum distance measure is exploited. Hence, it directly classifies the test image into one of the seven standard expressions such as anger, disgust, fear, happy, sad, surprise, and neutral with great accuracy. Experimental results stats that 91.70% classification accuracy is achieved by the proposed methodology for facial expression recognition.

REFERENCES

[1] Lajevardi, S. 2011. Automatic recognition of facial expressions, Doctor of Philosophy (PhD), Electrical and Computer Engineering, RMIT University.

[2] Vadivel A, Shanthi P and Shaila S.G. 2015. Estimating Emotions Using Geometric Features from Facial Expressions. Encyclopedia of Information Science and Technology. pp. 3754-3761.

[3] Rahulamathavan Yogachandran, R. Phan, J. Chambers and D. Parish. 2012. Facial Expression Recognition in the Encrypted Domain based on Local Fisher Discriminant Analysis. IEEE Transactions on Affective Computing. 4(1): 83-92.

[4] Guo Guo Dong, Rui Guo and Xin Li. 2013. Facial Expression Recognition Influenced by Human Aging. IEEE Transactions on Affective Computing. pp. 291-298.

[5] Huang X., Zhao G., Zheng W. and Pietikainen M. 2012. Spatiotemporal local monogenic binary patterns for facial expression recognition. IEEE signal Processing Letters. 19(5): 243-246.

[6] Poursaberi A., Yanushkevich S. and Gavrilova M. 2013. An Efficient Facial Expression Recognition System in Infrared Images. Fourth International Conference on Emerging Security Technologies. 25-28.

[7] Luo R.C., Lin P. H., Wu Y.C. and Huang C.Y. 2012. Dynamic face recognition system in recognizing facial expressions for service robotics. ASME International Conference on Advanced Intelligent Mechatronics. pp. 879-884.

[8] Rudovic O., Pantic M. and Patras I. 2013. Coupled Gaussian Processes for Pose-Invariant Facial Expression Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 35(6): 1357-1369.

[9] Halder A., Konar A., Mandal R., Chakraborty A., Bhowmik P., Pal N. R. and Nagar A.K. 2013. General and interval type-2 fuzzy face-space approach to emotion recognition. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 43(3): 587-605.

[10] Song K. T. and Chen Y. W. 2011. Design for integrated face and facial expression recognition. IEEE 37th Annual Conference on Industrial Electronics Society. pp. 4306-4311.

[11] Valstar M. F., Mehu M., Jiang B., Pantic M. and Scherer K. 2012. Meta-analysis of the first facial expression recognition challenge. IEEE Transactions on Systems, Man, and Cybernetics. 42(4): 966-979.

[12] Wang S., Liu Z., Wang Z., Wu G., Shen P., He S. and Wang X. 2012. Analyses of a Multi-modal Spontaneous Facial Expression Database. IEEE Transactions on Affective Computing. 4(1): 34-46.

[13] Hsu C. T., Hsu S. C. and Huang C.L. 2013. Facial expression recognition using Hough forest. IEEE Association Annual Summit conference on Signal and Information Processing. pp. 1-9.

[14] Sarawagi V and Arya K.V. 2013. Automatic facial expression recognition for image sequences, Sixth International Conference on Contemporary Computing. 278-282.

[15] Do M.N and Vetterli. M. 2005. The contourlet transform: an efficient directional multi resolution image representation. IEEE Transactions on Image Processing. 14(12): 2091-2106.

[16] Lu. Y and Do M.N. 2003. CRISP contourlet: a critically sampled directional multi resolution image representation. Proceedings of SPIE Conference on

Wavelet applications in signal and image processing. pp. 655-665.

[17] Mersereau R., Mecklenbrauker W, Quatieri T. 1976. McClellan transformations for two dimensional digital filtering, part-1: Design. IEEE Transactions on Circuits and Systems. 23(7): 405-414.

[18] Bamberger R. H. and Smith M. J. 1992. A filter bank for the directional decomposition of images: Theory and Design. IEEE Transactions on signal Processing. 44(4): 882-893.

[19] Haralick R. M., Shanmugam K. and Dinstein I. H. 1973. Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics. 610-621.

[20] Tan X. and Triggs B. 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Transactions on Image Processing. 19(6): 1635-1650.

[21] JAFFE database: http://www.kasrl.org/jaffe.html.