

## Research Article

### Object Analysis of Human Emotions by Contourlets and GLCM Features

R. Suresh and S. Audithan

Department of Computer Science and Engineering, PRIST University, Tanjore, Tamilnadu, India

**Abstract:** Facial expression is one of the most significant ways to express the intention, emotion and other nonverbal messages of human beings. A computerized human emotion recognition system based on Contourlet transformation is proposed. In order to analyze the presented study, seven kind of human emotions such as anger, fear, happiness, surprise, sadness, disgust and neutral of facial images are taken into account. The considered emotional images of human are represented by Contourlet transformation that decomposes the images into directional sub-bands at multiple levels. The features are extracted from the obtained sub-bands and stored for further analysis. Also, texture features from Gray Level Co-occurrence Matrix (GLCM) are extracted and fused together with contourlet features to obtain higher recognition accuracy. To recognize the facial expressions, K Nearest Neighbor (KNN) classifier is used to recognize the input facial image into one of the seven analyzed expressions and over 90% accuracy is achieved.

**Keywords:** Contourlet transform, emotion recognition, facial expression, nearest neighbor classifier

## INTRODUCTION

The emotional state of a human being describes their interaction with other people. Therefore, the recognition of facial expression is appropriate for the development of systems that require human-machine interaction. Human emotion recognition from facial expressions based on fuzzy relational approach is discussed in Chakraborty *et al.* (2009). Facial expressions are analyzed by segmenting and localizing the individual frames into regions of interest such as eye opening, mouth opening and the length of eyebrow constriction. An automated face recognition system based on feature based approach is introduced in Balasubramani *et al.* (2008). Seven facial expressions such as joy, smile, surprise, disgust, anger, sadness and fear are used for the feature based approach. Facial features points are detected and segmented. The distance between the features points is evaluated and used through data mining technique to generate a set of relevant prediction for facial expression recognition.

Novel facial expression recognition system based on key frame selection is explained in Guo *et al.* (2006). The four basic expressions; happiness, anger, sadness and surprise are analyzed. Initially, key frame selection of the facial motion from an image sequence is designed. An optical flow is used to determine changes in a facial expression. A recognition system for facial expression in low resolution images is described in Chang *et al.* (2009). Based on the transition matrices, six kinds of facial expressions are identified. The

boosted tree classifiers and template matching techniques are used to locate and crop the effective face region to characterize the facial expressions.

Multistage facial expression recognition system is implemented in Anderson and Peter (2006). A spatial ratio template tracker algorithm is used to locate the faces. Optical flow of the face is evaluated and the motion signatures are produced to classify the facial expression by using Support Vector Machine (SVM) classifier. An automatic recognition system for facial expression by using neural network classifier is discussed in Jyh-Yeong and Jia-Lin (1999). To extract the contours of the eyebrows, eyes and mouth of a face image three methods are used. They are rough contour estimation routine, mathematical morphology and point contour detection. Around 30 facial characteristic points are used to describe the position and shape of the above mentioned organs. Then the facial expressions can be described by combining different action units that are used for describing the basic muscle movement of a human face.

A new approach for facial expression recognition is presented in Jyh-Yeong and Jia-Lin (2001). Two inner canthi are detected as reference points from the input face images for searching the expression features such as contour and displacement of eyebrows, eyes and mouth. The extracted expression features are given to the Elman neural network of classifiers as input to recognize expressions. An automated facial expression recognition technique based on Gauss-Laguerre (GL) filter using infrared images is described in Poursaberi

**Corresponding Author:** R. Suresh, Department of Computer Science and Engineering, PRIST University, Tanjore, Tamilnadu, India

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

*et al.* (2013). 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.

Active Appearance Models (AAM) based facial expression recognition is implemented in Martin *et al.* (2008). To yield more robustness against varying lighting conditions AAM approach is applied in edge images. Three different classifiers such as AAM classifier, SVM and multi-layer perceptron are used and compared with each other. An automatic recognition system for facial expression is developed in Franco and Alessandro (2001) by using local unsupervised processing. Initially, local unsupervised processing stage is inserted within a neural network constructed to recognize facial expressions order. It also reduces the dimensionality of the input data.

Partial AAM fitting is applied on mouth and eyes to achieve better alignment for facial features in Luo *et al.* (2012) 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 new technique of facial expression recognition system is illustrated in Ching and Li (2008) based on neural network with fuzzified characteristic distances weights. The characteristic distances are calculated from different feature area such as mouth, eye or eyebrow. The obtained characteristic distances are multiplied with the fuzzified weights and sent to a neural network system for recognition of the facial expressions. It is composed of self organizing map neural network and Back Propagation Neural Network (BPNN).

A new approach for integrated face and facial expression recognition system for robotic applications is explained in Song and Yi-Wen (2011). 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 BPNN for face and facial expression recognition. An automated recognition system for facial expressions technique for image sequences is discussed in Sarawagi and Arya (2013). Two approaches, color normalization and local binary pattern are used to extract facial features. A novel facial expression recognition system in video sequences based on Hough forest algorithm is described in Chi-Ting *et al.* (2013). 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.

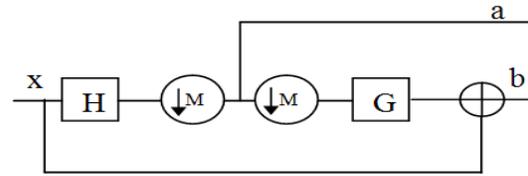


Fig. 1: Laplacian pyramid decomposition process for 1-level of decomposition

## METHODOLOGY

**Contourlet transform:** The Contourlet Transform consists of a double iterated filter bank in Do and Vetterli (2005). 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 contourlets. An advantageous characteristic of contourlets 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 in Lu and Do (2003) who proposed a critically sampled contourlet transform, called CRISP contourlets, 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 study, 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 in Mersereau *et al.* (1976) as proposed in Do and Vetterli (2005).

**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 Fig. 1.  $H$  and  $G$  are the low pass analysis and synthesis filters, while  $M$  is the sampling matrix is the coarse image, while is the difference between the signal and the prediction, containing the supplementary high frequencies.

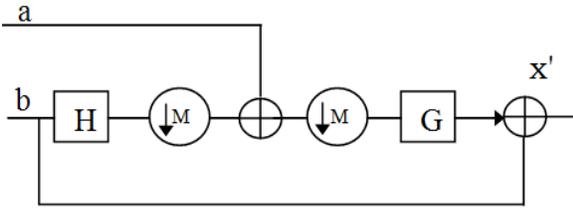


Fig. 2: Reconstruction method of contourlet transform

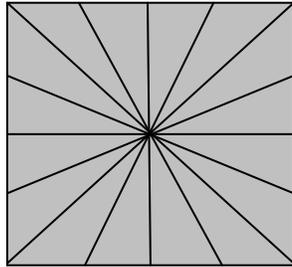


Fig. 3: Wedge shaped frequency partitioning with  $2^l$  directional sub bands ( $l = 3$ )

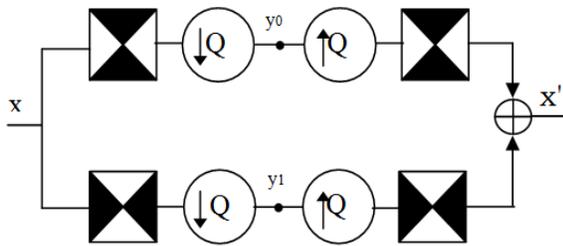


Fig. 4: 2D dimensional spectrum partition using quincunx filter banks with fan filter

A coarse image with a lower frequencies and a more detailed image with the supplementary high frequencies are obtained. The detailed image contains the point discontinuities of the original image. This scheme can be repeated continuously in the low pass image and is restricted only the size of the original image due to the down sampling. A disadvantage of the LP is the implicit over sampling. However, at each pyramidal level, it generates only one band pass image which does not have scrambled frequencies. Frequency scrambling can occur in the wavelet filter bank when the spectrum of a high pass channel is folded back in to the low frequency band after down sampling and is reflected.

In the original LP reconstruction method, the signal is obtained by adding back the difference to the prediction from the coarse signal. However, for the Contourlet Transform a new method, shown to be more efficient in the presence of noise compared to the original reconstruction method is utilized. Orthogonal filters along with the optimal linear reconstruction method using the dual frame operator are used for the reconstruction of the image as shown in Fig. 2.

**Directional filter bank:** The DFB proposed in Bamberger and Smith (1992) 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 Fig. 3. 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 Do and Vetterli (2005) 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 Fig. 4, 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 4 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.

**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. 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".

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  $b_j[n]$ ,  $J = 1, 2, \dots, J$  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.

### GRAY LEVEL COOCCURRENCE MATRIX

Haralick *et al.* (1973) described the use of co-occurrence probabilities using GLCM for extracting various texture features. GLCM is also called as 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".

**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:

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

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.

**Energy:** Energy is also called angular second moment where it measures textural uniformity. If an image is completely homogeneous then the energy will be maximum:

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

**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:

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

**Correlation:** The correlation measures the linear dependency of grey levels on those of neighboring pixels. It is defined by:

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

where,  $\mu_x, \mu_y, \sigma_x$  and  $\sigma_y$  are the means and standard deviations.

**Proposed method:** The proposed automated emotion recognition system composed of two important steps, which are feature extraction and classification. Feature extraction is one of the key techniques in pattern recognition. A successful feature extraction technique will suppress superfluous image content which in turn reduces the processing time, transmission time and bandwidth, storage requirements. Also, the classifier performance depends on the discriminating power of the extracted features. In the proposed approach, contourlet transformation is used to represent the facial expression images and features are extracted from that representation.

The first module of the proposed emotion recognition system is feature extraction. In this module features are extracted from the input training images. Initially, an efficient multi resolution contourlet transformation is applied to the input image at N level. It produces directional sub-bands that represent the input image in different directions. The number of obtained directional sub-bands depends on the level of decomposition. From the directional sub-bands, the informative features are extracted from the input images. The features such as mean and entropy are calculated by taking average and entropy of the attained coefficients of each directional sub-bands.

Also, the textural feature such as contrast, correlation, energy and homogeneity are extracted from the computed GLCM of the input image without applying contourlet transformation. Then the extracted features from frequency and spatial domains are fused together and the feature extraction process is repeated for all the training images. Finally, the obtained features of the training images are stored in feature database corresponding with facial expression details for classification. The proposed approach for human emotion recognition system is shows in Fig. 5.

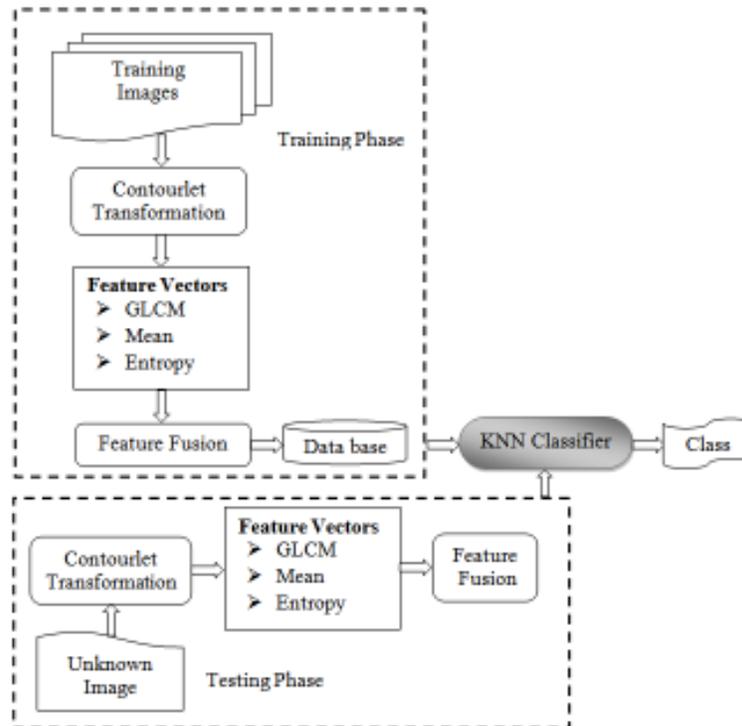


Fig. 5: Block diagram of the proposed contourlet transform based human emotion recognition system

The second module of the proposed approach is classification. An efficient KNN classifier is used as classifier. In order to recognize the facial expression of a test image, the same feature extraction method as what have done for training images is applied to it. In the classification stage, the emotion of the given image is recognized by KNN classifier which calculates the distance between the unknown features metrics to recognize into one of the seven analyzed expression. In order to achieve better classification accuracy city block is taken into account as suitable distance measure.

## RESULTS AND DISCUSSION

The proposed human emotion recognition system is carried out by using JAFFE database (JAFFE database: <http://www.kasrl.org/jaffe.html>). It composed of 213 images with seven kind of expression (6 vital + 1 neutral) pose by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. The photos were taken at the Psychology Department in Kyushu University. The facial expressions of a Japanese female are shown in Fig. 6.

The contourlet transform based feature such as mean, entropy and as well as GLCM features are considered. The contourlet based features are extracted from the decomposed images and GLCM features are extracted from the input facial image. As the proposed

system is based on contourlet transform, a multi direction approach, the performance is evaluated by varying the level of contourlet decomposition. Table 1 shows the classification accuracy of the proposed system using fusion approach. In the fusion approach, all the extracted features from contourlet and GLCM are fused serially and recognized by KNN classifier.

It is clearly shown from the Table 1 that the highest average recognition accuracy of over 90% is obtained at 9th level of decomposition. Among the various facial expressions, the happy emotional state achieves lower recognition accuracy. However, the accuracy of the proposed system is improved by increasing the level of decomposition. Figure 7 shows the graphical representation of average recognition accuracy corresponding with level of decomposition. It shows the average of all facial expression while using the features mean, entropy, fusion of mean and GLCM features, fusion of entropy and GLCM features and fusion of all features.

It is clearly inferred from the figure that the proposed fusion of mean and entropy of contourlet and GLCM based features produces better results than their individual performance. Also it is observed that irrespective of the level of decomposition, the higher average accuracy is obtained by the fusion of all extracted features. As the level of decomposition increases, the recognition accuracy also increases. This is due to that higher level of decomposition produces more information about the facial expression in different directional sub-bands.



Fig. 6: Facial expressions of a Japanese female from left to right are anger, disgust, fear, happy, neutral, sad and surprise

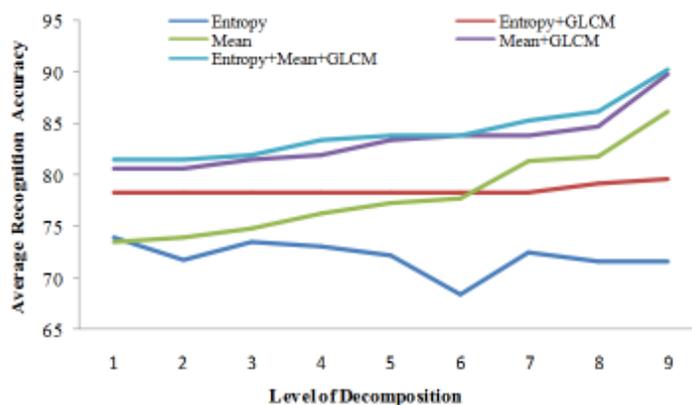


Fig. 7: Average recognition accuracy vs. decomposition level

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	80.00	80.00	80.00	80.00	80.00	83.33	93.33	90.00	90.00
Disgust	86.67	86.67	86.67	90.00	90.00	90.00	90.00	93.33	93.33
Fear	84.38	84.38	84.38	84.38	84.38	84.38	78.13	84.38	87.50
Happy	68.75	68.75	68.75	68.75	68.75	68.75	75.00	78.13	84.38
Neutral	90.00	90.00	93.33	93.33	93.33	93.33	93.33	93.33	90.00
Sad	74.19	74.19	74.19	80.65	80.65	80.65	80.65	83.87	96.77
Surprise	86.67	86.67	86.67	86.67	90.00	86.67	86.67	80.00	90.00
Average	81.52	81.52	82.00	83.40	83.87	83.87	85.30	86.15	90.28

### CONCLUSION

In this study, an efficient automated human emotion recognition system is designed to recognize seven kind of facial expressions based on contourlet transform. Contourlet transform is a multi resolution technique that represents the input data in directional sub bands at various scales. The features that distinguish different facial expressions are extracted from the contourlet representation and fed to KNN classifier. The proposed system composed of decomposition, feature extraction and classification stages. KNN is utilized as classifier with opted distance metric to achieve maximum recognition accuracy for various emotional states. The results show that the recognition accuracy achieved by the proposed human emotion recognition system is over 90%.

### REFERENCES

Anderson, K. and W.M. Peter, 2006. A real-time automated system for the recognition of human facial expressions. *IEEE T. Syst. Man Cy. B*, 36(1): 96-105.

Balasubramani, A., K. Kalaivanan, R.C. Karpagalakshmi and R. Monikandan, 2008. Automatic facial expression recognition system. *Proceeding of the International Conference on Computing, Communication and Networking*, pp: 1-5.

Bamberger, R.H. and M.J.T. Smith, 1992. A filter bank for the directional decomposition of images: Theory and design. *IEEE T. Signal Proces.*, 44(4): 882-893.

Chakraborty, A., K. Amit, K.C. Uday and C. Amita, 2009. Emotion recognition from facial expressions and its control using fuzzy logic. *IEEE T. Syst. Man Cy. A*, 39(4): 726-743.

Chang, Y., C. Lien and L. Lin, 2009. A new appearance-based facial expression recognition system with expression transition matrices. *Proceedings of the 3rd International Conference on Innovative Computing Information and Control*, pp: 538.

Chi-Ting, H., H. Shih-Chung and H. Chung-Lin, 2013. Facial expression recognition using Hough forest. *Proceeding of the Signal and Information Processing Association Annual Summit and Conference (APSIPA, 2013)*, pp: 1-9.

- Ching, Y.L. and C.L. Li, 2008. Recognition of facial expression by using neural-network system with fuzzified characteristic distances weights. Proceedings of the IEEE International Conference on Fuzzy Systems, pp: 1694-1699.
- Do, M.N. and M. Vetterli, 2005. The contourlet transform: An efficient directional multi resolution image representation. IEEE T. Image Process., 14(12): 2091-2106.
- Franco, L. and T. Alessandro, 2001. A neural network facial expression recognition system using unsupervised local processing. Proceeding of the 2nd International Symposium on Image and Signal Processing and Analysis, pp: 628-632.
- Guo, S.M., Y.A. Pan, L. Yueh-Ching, C.Y. Hsu, S. H.T. Jason and C.I. Chang, 2006. A key frame selection-based facial expression recognition system. Proceeding of the 1st International Conference on Innovative Computing, Information and Control, 3: 341-344.
- Haralick, R.M., S. Karthikeyan and D. Its'Hak, 1973. Textural features for image classification. IEEE T. Syst. Man Cyb., SMC-3(6): 610-621.
- Jyh-Yeong, C. and C. Jia-Lin, 1999. A facial expression recognition system using neural networks. Proceedings of the International Joint Conference on Neural Networks, 5: 3511-3516.
- Jyh-Yeong, C. and C. Jia-Lin, 2001. Automated facial expression recognition system using neural networks. J. Chinese Inst. Eng., 24(3): 345-356.
- Lu, Y. and M.N. Do, 2003. Crisp-contourlets: A critically sampled directional multiresolution image representation. Proceeding of SPIE Conference on Wavelet Applications in Signal and Image Processing, pp: 655-665.
- Luo, R.C., H.L. Pei, C.W. Yen and C.Y. Huang, 2012. Dynamic face recognition system in recognizing facial expressions for service robotics. Proceedings of the ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp: 879-884.
- Martin, C., W. Uwe and H.M. Gross, 2008. A real-time facial expression recognition system based on active appearance models using gray images and edge images. Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition, pp: 1-6.
- Mersereau, R.M., W. Mecklenbrauker and T. Quatieri, 1976. McClellan transformations for two-dimensional digital filtering-Part I: Design. IEEE T. Circuits Syst., 23(7): 405-414.
- Poursaberi, A., Y. Svetlana and G. Marina, 2013. An efficient facial expression recognition system in infrared images. Proceeding of the 4th International Conference on Emerging Security Technologies (EST, 2013), pp: 25-28.
- Sarawagi, V. and K.V. Arya, 2013. Automatic facial expression recognition for image sequences. Proceeding of the 6th International Conference on Contemporary Computing (IC3), pp: 278-282.
- Song, K.T. and C. Yi-Wen, 2011. A design for integrated face and facial expression recognition. Proceeding of the 37th Annual Conference on IEEE Industrial Electronics Society, pp: 4306-4311.