Contourlet Tranform Based Human Emotion Recognition System

R. Suresh and S. Audithan
Department of computer science and Engineering, PRIST University, Tanjore, India
Email: {mkrsuresh.phd, saudithan}@gmail.com

P. Elakkiya
Department of communication systems, PRIST University, Tanjore, India
Email: palanieswari04@gmail.com.

***Abstract*—In this study, an automated human emotion recognition system based on contourlet transform is presented. The various human emotional stats such as surprise, sadness, disgust, anger, fear, happiness and neutral are analyzed. The emotional states are identified by the motion of the muscles in the skin of the human face. In order to classify the emotional state, the patterns generated by the muscles motions should be represented efficiently. In the proposed approach, these patterns are characterized by contourlet transform. The high dimensional contourlet transformed face image is translated into low dimensional space in the feature extraction stage. This can be achieved by taking the average of contourlet coefficients in each sub-band and fed into the classifier. The number of features used for classification depends on the level of decomposition used while transforming the face image by contourlet transform. The emotional states are classified by nearest neighbor classifier and the classifier performance is analyzed by various distance measures; euclidean, city block, cosine and correlation efficiently. The results show that the proposed approach based on contourlet transform produces over 80% recognition accuracy on JApanese Female Facial Expression (JAFFE) database.**

***Index Terms*—contourlet transform, nearest neighbour classifier, facial expressions, emotion recogniton**

I. Introduction

Human emotion recognition is one of the active research topics in computer vision and psychology study. It has many useful applications in human computer interaction, human behavior understanding and perceptual user interfaces. The aging effect on computational facial expression recognition based on two data bases descried in [1]. The feature dimensionality problem is investigated by using manifold learning techniques. Three automatic emotion recognition systems based on interval type-2 fuzzy set (IT2FS), interval approach-IT2FS, and General type-2 fuzzy sets (GT2FS) is presented in [2]. 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 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. A method for head pose invariant facial expression recognition that is based on 2D geometric features is implemented in [4]. To achieve head pose invariance, the coupled scaled gaussian process regression model for head pose normalization is proposed.

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. A meta-analysis of the first such challenge in automatic recognition of facial expressions is presented in [6]. 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 [7]. 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 lattice computing extension of the fuzzy ARTMP neural classifier is designed for facial expression recognition in [8]. Statistical local features based facial expression recognition system is evaluated in [9]. The face images are represented by using Local Binary Patterns (LBP).

Manuscript received January 28, 2014; revised April 28, 2014.

©2014 Engineering and Technology Publishing
doi: 10.12720/ijsps.2.1.7-11

An approach for facial expression by discovering associations between visual feature and LBP is proposed in [10]. It automatically tracks the facial area and segments face into meaningful areas based on description of LBP and then it accumulates the probabilities throughout the frames from video data to capture the temporal characteristics of facial expressions by analyzing facial expressions. An automatic solution to identify human expressions as well as overcoming facial expressions variation and intensity problems is described in [11]. Facial features component are automatically detected and segmented. Then, facial expression deformations are done by detected facial feature points.

Based on support vector machine, the capability of polynomial kernel function and radial basis function kernel in the facial expression recognition using the JAFFE expressions library is analyzed in [12]. A simple algorithm based on hidden Markov model for automatic recognition of facial expressions is proposed in [13]. The variations in feature point distances are used to characterize the transition from one emotion to another for recognize the facial expressions.

The organization of the paper is as follows. The background of contourlet transform is introduced in section 2. The proposed human emotion recognition system based on contourlet features is presented in Section 3. The experimental results are presented in Section 4 and from the experimental results; conclusion is made in Section 5.

II. Contourlet Transform

The Contourlet Transform consists of a double iterated filter bank [14]. 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 [14], at the expense of some redundancy (up to 33%) due to the Laplacian Pyramid. This problem has been addressed in [15] 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 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 [16] as proposed in [14].

A. 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. $a[n]$ is the coarse image, while $b[n]$ is the difference between the signal and the prediction, containing the supplementary high frequencies.

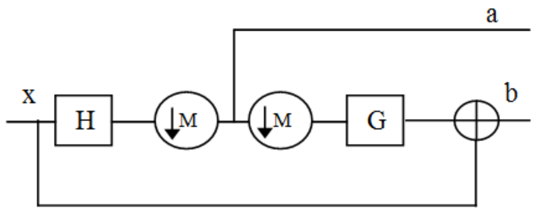

Figure. 1. Laplaican Pyramid decomposition process for 1-level of decomposition

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.

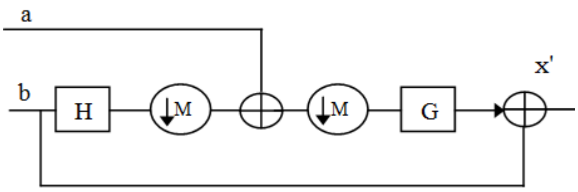

Figure. 2. Reconstruction method of contourlet transform

B. Directional Filter Bank

The DFB proposed in [17] 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 [14] 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.

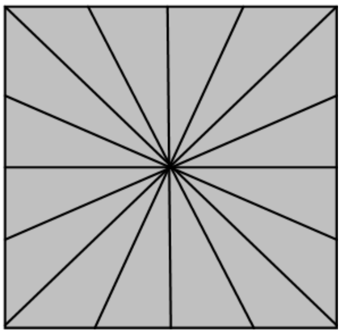

Figure. 3. wedge shaped frequency partitioning with 2^l directional sub bands (l =3)

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. Fig. 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.

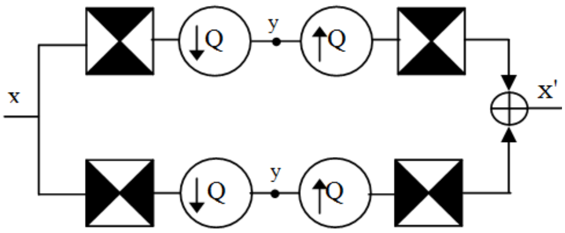

Figure. 4. 2D dimensional spectrum partition using quincunx filter banks with fan filter

C. 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 $b_j[n], j=1,2,..J$ 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.

III. Proposed System

The two main important modules in the proposed human emotion system are feature extraction and recognition modules. In the feature extraction module, contourlet transform is used and nearest neighbor classifier is used in the classification module. The block diagram of the proposed approach is shown in Fig. 5.

The human emotion recognition system is considered as a pattern recognition technique. In such a system feature extraction is the most important step. In the proposed approach contourlet transform based features are extracted. In order to classify human emotions, the features that represent human emotions must be extracted first and fed into the classifier for training. First, the emotions from the training images are decomposed by using contourlet transform at predefined level. The decomposition produces number of sub-bands that depends on the level of decomposition.

From the sub-bands, the energies are calculated. It can be calculated by taking the magnitude of coefficients in sub-bands. This high dimensional space is reduced by averaging energies in each sub-band. The reduced average energy is used as features for the corresponding emotional image. This course of action is repeated for the training images and the features are stored in the database with corresponding emotional state. The same features are extracted for the test image and compared with the database using nearest neighbor classifier in order to classify the emotional state of the test image based on euclidean distance measure.

©2014 Engineering and Technology Publishing

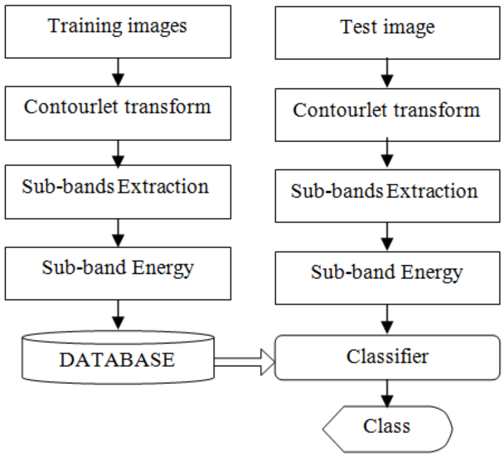

Figure. 5. Block diagram of proposed human emotion recognition system

IV. Results and Discussions

The performance of the proposed human emotion recognition system is evaluated by using JAFFE database [18]. The database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed 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.

Figure. 6. Facial expression from top to bottom and left to right are anger, disgust, fear, happy, neutral, sad and surprise

In the proposed approach, the average energies extracted from the sub-bands of contourlet transformed facial image is used as features. As the contourlet transform is a multi scale transformation, the performance is evaluated by varying the level of decomposition. Table I shows the recognition accuracy of the proposed system obtained by using contourlet features from 1-level to 5-level.

TABLE I. Recognition Accuracy of the Proposed System upto 5^{TH} Level of Contourlet Transform

Emotional State	Level of decomposition				
	1	2	3	4	5
Anger	80	80	76.67	76.67	80
Disgust	73.33	73.33	70	73.33	80
Fear	71.88	71.88	78.13	71.88	68.75
Happy	65.63	68.75	71.88	71.88	71.88
Neutral	70	73.33	83.33	83.33	83.33
Sad	67.74	70.97	74.19	70.97	77.42
surprise	73.33	73.33	70	73.33	73.33

It is observed from Table I, the maximum recognition accuracy for neutral, anger and disgust is over 80% and the remaining emotions such as fear, happy, sad and surprise is less than 80%. In order to achieve more recognition accuracy, the level of decomposition is further increased up to 9^{th} level and the obtained accuracy is tabulated and shown in Table II.

TABLE II. Recognition Accuracy of the Proposed System from 6^{TH} Level of Contourlet Transform

Emotional State	Level of decomposition			
	6	7	8	9
Anger	83.33	80	86.67	90
Disgust	76.67	80	86.67	96.67
Fear	68.75	81.25	81.25	84.38
Happy	75	71.88	71.88	84.38
Neutral	86.67	86.67	93.33	86.67
Sad	74.19	77.42	70.97	77.42
surprise	73.33	76.67	83.33	86.67

As the level of decomposition increases, the recognition accuracy also increases due to the capture of texture patterns in contourlet sub-band images at higher resolution level. However, the maximum recognition accuracy achieved for the emotion; sad is 77.42% which is very less in comparison with other emotional states. The graphical representation of average recognition accuracy with respect to level of decomposition is shown in Fig. 7.

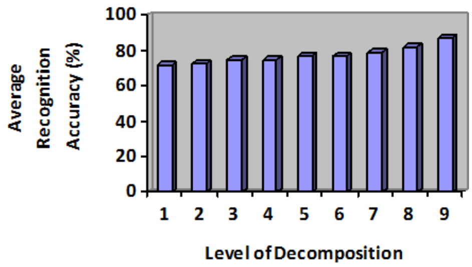

Figure. 7. Average classfication accuracy vs level of decomposition

Table III shows the comparative analysis of proposed method with existing methods in terms of recognition accuracy.

TABLE III. Comparision of Recognition Accuracy of the Proposed System with Other Method

Reference	Method	Accuracy (%)
Shinohara [19]	HLAC+Fisher weight maps	69.4
Mingwei Huang [20]	SNE+SVM	73
Proposed Method	Contourlet	86.6

V. Conclusion

Contourlet transform based human emotion recognition system is described in this paper. The basic 6 emotional states and of a human are identified by the proposed approach. They are surprise, sadness, disgust, anger, fear, happiness and also neutral. The results show the capability of contourlet transform to represent the human emotion states and the maximum average recognition accuracy achieved is 86.6%. In future, the recognition accuracy will be increased by adopting other transformation techniques and also computing the statistical properties of contourlet transform.

©2014 Engineering and Technology Publishing

REFERENCES

[1] G. D. Guo, R. Guo, and X. Li, "Facial expression recognition influenced by human aging," *IEEE Transactions on Affective Computing*, pp. 291-298, 2013.

[2] A. Halder, A. Konar, *et al.*, "General and interval type-2 fuzzy face-space approach to emotion recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 3, pp. 587-605, 2013.

[3] Y. Rahulamathavan, R. Phan, J. Chambers, and D. Parish, "Facial expression recognition in the encrypted domain based on local fisher discriminant analysis," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 83-92, 2012.

[4] O. Rudovic, M. Pantic, and I. Patras, "Coupled gaussian processes for pose-invariant facial expression recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1357-1369, 2013.

[5] X. H. Huang, G. Y. Zhao, W. M. Zheng, and M. Pietikainen, "Spatiotemporal local monogenic binary patterns for facial expression recognition," *IEEE Signal Processing Letters*, vol. 19, no. 5, pp. 243-246, 2012.

[6] M. F. Valstar, M. Mehu, B. H. Jiang, M. Pantic, and K. Scherer, "Meta-analysis of the first facial expression recognition challenge," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 4. pp. 966-979, 2012.

[7] S. F. Wang, Z. L. Liu, *et al.*, "Analyses of a multi-modal spontaneous facial expression database," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 34-46, 2012.

[8] V. G. Kaburlasos, S. E. Papadakis, and G. A. Papakostas, "Lattice computing extension of the FAM neural classifier for human facial expression recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 10, pp. 1526-1538, 2013.

[9] R. Verma and M.-Y. Dabbagh, "Fast facial expression recognition based on local binary patterns," in *Proc. 26th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2013, pp. 1-4.

[10] H. Q. Zhang, S. Luo, and O. Yoshie, "Facial expression recognition by analyzing features of conceptual regions," in *Proc. IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS)*, 2013, pp. 529-534.

[11] H. Mliki, N. Fourati, S. Smaoui, and M. Hammami, "Automatic facial expression recognition system," in *Proc. ACS International Conference on Computer Systems and Applications (AICCSA)*, 2013, pp. 1-4.

[12] F. G. Wang, K. T. He, Y. Liu, L. Li, and X. G. Hu, "Research on the selection of kernel function in SVM based facial expression recognition," in *Proc. 8th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2013, pp. 1404-1408.

[13] M. Lahbiri, A. Fnaiech, M. Bouchouicha, M. Sayadi, and P. Gorce, "Facial emotion recognition with the hidden Markov model," in *Proc. International Conference on Electrical Engineering and Software Applications (ICEESA)*, 2013, pp. 1-6.

[14] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multi resolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091-2106, 2005.

[15] Y. Lu and M. N. Do, "CRISP contourlet: A critically sampled directional multi resolution image representation," in *Proc. SPIE Conference on Wavelet Applications in Signal and Image Processing*, 2003, pp. 655-665.

[16] R. M. Mersereau, W. Mecklenbrauker, and Q. T. Jr., "McClellan transformations for two dimensional digital filtering, part-1: Design," *IEEE Transactions on Circuits and Systems*, vol. 23, no. 7, pp. 405-414, 1976.

[17] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp 882-893, 1992.

[18] JAFFE database. [Online]. Available: http://www.kasrl.org/jaffe.html

[19] Y. Shinohara and N. Otsu, "Facial Expression Recognition Using Fisher Weight Maps," in *Proc. IEEE Conf. on Automatic Face and Guesture Recognition*, 2004, pp. 499-504.

[20] M. W. Huang, Z. Wang, Z. L. Ying, "Facial expression recognition using Stochastic Neighbor Embedding and SVMs," in *Proc. ICSSE*, 2011, pp. 671-674.

R. Suresh is currently working as a vice principal in modern polytechnic college in the year 2012- till date Tamilnadu, India. He received his B.Tech degree from Arasu Engineering College (ANNA UNIVERSITY) in the year 2006 and M.Tech degrees from SASTRA UNIVERSITY - Thanjavur in the year 2008 respectively. He has worked as a Lecturer in Annai polytechnic college, in the year 2008 to 2011. He has presented more than 10 papers in conferences.

P. Elakkiya is currently working as a HOD in ECE department modern polytechnic college in the year 2012- till date Tamilnadu, India. He received his B.E degree from Periyar Maniammai College of technology for women (Anna University) in the year 2005 and M.E degrees from Saranathan College of engineering in the year 2007 respectively. He has worked as a Lecturer in Annai polytechnic college, in the year 2008 to 2011. He has presented more than 10 papers in conferences.

Audithan Sivaraman is currently working as professor in PRIST University Tamilnadu, India. He received his BE degree from Bharathi Dhasan University in 2000 and ME and P.HD degrees from Annamalai University in 2006 and 2011 respectively. He has worked as a network engineer in RBCOMTEC at Hydrabad from 2000 to 2003 and professor at SASTRA University from 2006 to 2007. He was the research assistant in annmalai university under university research fellowship scheme from 2007 to 2011. Ha has presented and published more than 15 papers in referred conferences and high impact factored journals. He had visited many foreign countries like Dubai, Singapore; Thailand etc. His research interests include Image processing, Network Security, and Artificial Intelligence.

©2014 Engineering and Technology Publishing