Facial Expression Recognition Based on Radon and Discrete Wavelet Transform using Support Vector Machines

H. Ali[1], V. Sritharan[1], M. Hariharan[1], M. Elshaikh[2] and W.Khairunizam[1]
[1]*School of Mechatronic Engineering, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia.*
[2]*School of Computer and Communication Engineering, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia.*
hasimahali@unimap.edu.my

***Abstract*—Extracting facial features remains a difficult task because of unpredictable of facial features largely due to variations in pixel intensities and subtle changes of facial features. The Radon transform inherits rotational and translational properties that are capable of preserving pixel intensities variations and also is used to derive the directional features. Thus, this paper presents a new pattern framework for facial expression recognition based on Radon and wavelet transform using Support Vector Machines classifier to recognize the seven facial emotions. Firstly, the pre-processed facial images are projected into Radon space via Radon transform at a specified angle. Then, the obtained Radon space or sinogram that represent the facial emotions is subjected to wavelet transform. In this framework, the Radon space is decomposed into four sub-band at a different level of decomposition. The approximate coefficients sub-band are independently extracted and used as intrinsic features to recognize the facial emotion. To reduce the data dimensionality, principal component analysis (PCA) is applied to the extracted features. Then, the Support Vector Machines (SVM) classifier is adopted as a classifier to classify seven (anger, disgust, fear, happiness, neutral, sadness and surprise) facial emotions. To evaluate the effectiveness of the proposed method, the JAFFE database has been employed. Experimental results show that the proposed method has achieved 93.89% accuracy.**

***Index Terms*—Facial expression; Support Vector Machines; Radon projection; Wavelet transform.**

I. INTRODUCTION

Facial expression is a non-verbal communication type that can be considered as the most powerful communication tools in imparting a person's emotional state as well as conveying his/ her intentions or messages. Thus, facial emotion analysis has received a great deal of attention over the last two decades ranging from a psychological perspective to behavior science research and recently towards human-computer interaction (HCI). Facial emotion recognition (FER) has shown potential impacts on healthcare monitoring, robotics, customer services and etc. In spite of the rapid achievement in FER, the task in FER remains challenging largely because of complexity, subtlety, and variability of unpredictable facial features. In the literature, many techniques have been proposed in recognizing FER such as fuzzy inference system [1], expert systems [2], linear subspaces [3], pixel intensities [4], shape and texture [5], Gabor wavelet coefficients [6], and local binary pattern [7].

In 1998, for instance, Zhang et al. [6] have compared the performance of FER based on two different types of features using multi-layer perceptron. The first type of features is a set of fiducial points geometric-based features located on the face region while the second type of features is a set of multi-scales and multi-orientation of Gabor wavelet coefficients at the fiducial points. Their findings show that the Gabor wavelet coefficients are much powerful compared to the geometric fiducial points. Donato et al. [8] recognized FER based on action units of the face for six upper faces and six lower faces. They compared different types of features including Gabor wavelet coefficients, optical flow, PCA, ICA and local features for analyzing FER. They found that the Gabor wavelet coefficients and ICA achieved the best performances compared to others. In contrast, Gabor wavelets lead to the high computational cost of the FER. Meanwhile, Deng et al. [9] have suggested a new technique to resolve the issue in [8] by down-sampling the global Gabor wavelet filters having (m frequencies x n orientations) into a new local-Gabor wavelet filters by employing only a part of the m entire frequencies and part of n orientation parameters. They found the proposed new local-Gabor filters (m x n) not only shorten feature extraction time, but also reduces storage and the computational data of FER. In spite the significant information Gabor filter posses, the output of the Gabor filter banks are not mutually orthogonal which may result in a significant correlation [10] between texture features. Moreover, these transformations are usually not reversible which limits their application for texture synthesis.

The Radon transform is based on the parameterization of the straight line of the image domain and on the evaluation of the integrals of the image along these straight lines. The Radon transform helps the implementation of very effective detection algorithms, but it does not provide itself sufficient information for recognition purposes [11]. The human visual system analyzes the image at several spatial resolution scales. In image representation, high frequencies carry detail information about the image while low frequencies carry coarse information. An appropriate wavelet transform can result in robust image representation since it provides a precise and unifying framework for the analysis and characterization of an image at different scales [12].

This research is the extension of our previous work[13] to evaluate the analysis for a different level of DWT decomposition using SVM classifier. Besides, in [13] the analysis was focused on first level decomposition with k-NN classifier. In this research, we adopted the combination of the Radon transform and wavelet transform (WT) using multi-class SVM classifier in recognizing facial emotions. In this framework, we calculated the Radon transformation of the image to project a 2D spatial image into Radon space or sinogram. We then further applied DWT on projected Radon space by decomposing the space into four different sub bands and extracted the informative features. There are a large number of coefficients, and hence principal component analysis (PCA) was applied on the extracted features. . The reduced features were classified using SVM. Even though the study in [14] has a similar approach with the proposed method, however, their study has focused on face recognition application. In contrast, in this work, we focused on recognizing the seven emotions, namely anger, disgust, fear, happiness, neutral, sadness and surprises based on radon and wavelet transform characteristic using SVM classifier.

The outline of this paper is as follows: Section II explains the methods used in FER such as Radon transform, WT, PCA and SVM. Section III illustrates the experimental results and also discussion of the proposed method. Finally, Section IV concludes the findings of the proposed method.

II. Methodology

Figure 1 shows the framework of the proposed method. The principle of each working block is discussed as followed:

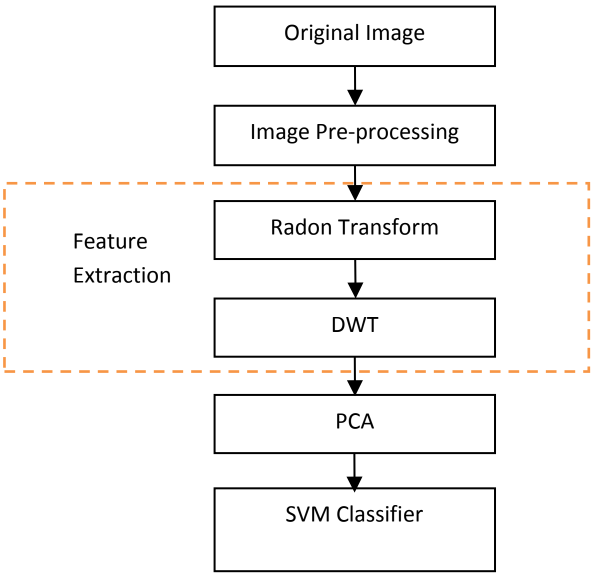


Figure 1: The framework of the proposed method

A. *Image Pre-Processing*

In this work, Japanese Female Facial Expression (JAFFE) database is used. It consists of 213 images which performing seven facial expressions. Firstly, the original facial image of size 256 by 256 pixels is pre-processed into 128 by 96 pixels as shown in Figure 2 in order to extract the local face region from clutter background. The pre-processed image then is subjected to histogram equalization for eliminating the illumination effects, before it is further analyzed using the Radon transform.

Figure 2: Example of six emotions and neutral

B. *Radon Transform*

Radon transform is one of the techniques used to detect features within the image. It is based on the parameterization of the straight line of the image domain and on the evaluation of the integrals of the image along these straight lines. Due to inherent properties of the Radon transform, it is a useful tool to capture the directional features of the images [15]. Furthermore, the Radon transform is a translation and rotation invariant [16], hence the variations in pixel intensities are preserved. The Radon transform of a 2D image function $f(x, y)$ in (s, θ) plane[17] is defined as:

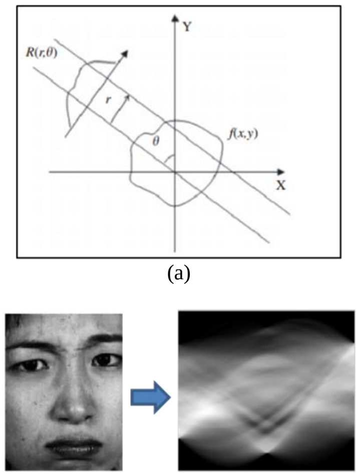
$$R(s,\theta)[f(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(s - x\cos\theta - y\sin\theta)\,dxdy \qquad (1)$$

where: $\delta(\cdot)$ is the Dirac delta function

$s \in [-\infty, \infty]$ = Perpendicular distance of a line from the origin

$\theta \in [0, \pi]$ = Angle formed by the distance vector as shown in Figure 3 (a)

The Radon transform has been widely applied for deriving the local features in edge detection, textural classification and retrieve the image in computer tomography. Recently, Radon transform has been successfully applied in face recognition [14,15] and also in medical images[16]. Thus, in this work Radon transform is used to project a 2D facial image into sinogram based on the rationale that Radon transform will preserve the directional features of the facial image. The sinogram of Radon transform of respective facial emotion is depicted in Figure 3(b).



(a)

(b)

Figure 3: (a) Basic concept of radon transform, (b) The emotion anger and it corresponding sinogram (Radon transform) for angle 0-179°

C. *2D Wavelet Transform*

The discrete wavelet transform (DWT) is a multiresolution technique used to decompose any non-stationary signals at a different time and frequency band which utilizes the basis function. Traditionally, DWT is widely used to analyze the 1-D signal in many areas. Due to advantages of DWT, then one-dimensional wavelet transform can be extended into 2D DWT which defined as [14].

$$W_{\varphi}(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\varphi_{j_0,m,n}(x,y) \qquad (2)$$

$$W_\psi^i(j,m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \psi^i_{j,m,n}(x,y) \quad (3)$$

$$i = \{H, V, D\}$$

where: index i = Directional wavelets in horizontal (H), vertical (V) and diagonal (D)
j_0 = Arbitrary starting scale
$W_\varphi\ (j_0, m, n)$ = Coefficient that define an approximation of $f\ (x, y)$ at scale j_0

The $W^i_\psi\ (j_0, m, n)$ coefficients add horizontal, vertical and diagonal details for scale $j \geq j_0$. The 2D-DWT can be implemented using a digital filter and down samplers. The advantages of the wavelet transform are the availability of both spatial and spectral information. The working principle of the wavelet transform is shown in Figure 4 which an image $x[m,n]$ is used as input at scale zero. Convolving the rows with h[n] (high pass) and g[n] (low pass) and down-sampling the column, resulting two sub-images whose horizontal resolutions are reduced by a factor of 2. The high pass or detail components characterize the image's high-frequency information with vertical orientation; the low pass, the approximate components contains its low-frequency, vertical information. Both sub-images are then filtered column wise and down-sampled to produce four quarter size output sub-images the approximation and horizontal, vertical and diagonal details. Figure 5 shows the concept of wavelet decomposition in (a) and (b) resulting in the first level decomposition of Radon space into four sub-band.

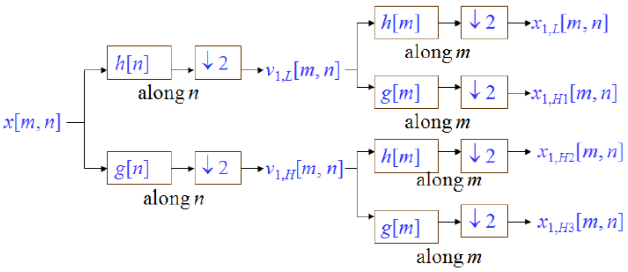
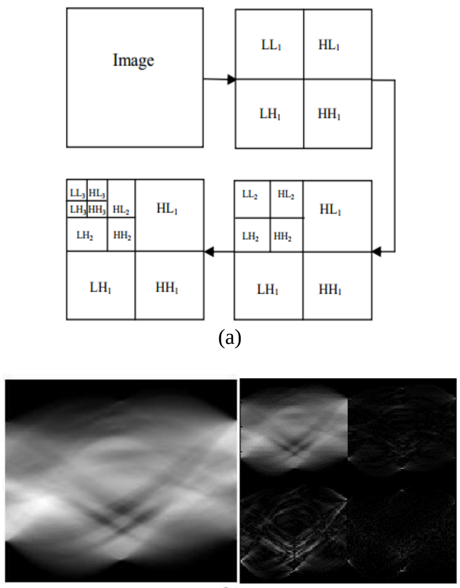


Figure 4: The analysis of filter bank [14]



(a)

(b)

Figure 5: The resulting decomposition of wavelet transform

D. Principle Component Analysis

The objective of PCA is to find an orthonormal set of axes in the direction of greatest variability in the data[18]. It calculates the principal components which are the basis vectors of directions in decreasing order of variability. The first principle component has the highest variance. The second principal component has the next highest variance and orthogonal to the first principal components and so on. PCA can be computed as follows:

Step 1. The covariance matrix is computed as:

$$S = (X - \bar{x})(X - \bar{x})^T \quad (4)$$

where: X = Data matrix of DWT coefficient in a sub band of N x 213 dimensions
N = Total number of features
$\bar{x}$ = Global mean of X

Step 2. Eigenvalues λ_i and eigenvectors v_i of S, are calculated as:

$$Sv_i = \lambda_i v_i, i = 1,2,3,\ldots,n \quad (5)$$

Step 3. The eigenvectors v_i that associated to the largest eigenvalues λ_i are sorted in descending order. Then, the first k principal components are chosen. The k PCs are the corresponding to the k largest eigenvalues. The k principal components of the observed vector x are given by,

$$y = W^T(X - \bar{x}) \quad (6)$$

In this work, the first 100 columns of projected data were considered as the 100 features based on containment of 99% of the total variability. Then, the 100 features were used for pattern recognition using classifiers.

E. Support Vector Machines

SVM is a popular machine learning technique introduced by Vapnik[19] due to attractive features and competitive performance. SVM has been successfully applied in many fields such as bioinformatics, text and image recognition. SVM can be categorized as linear and non-linear separable case. Figure 6 shows the SVM with a linear separable case or binary class classification. In this framework, SVM tries to maximize the margin between separating hyperplane of two class (i.e., class 1 and class 2). The optimal hyperplane is achieved when the margin of class 1 and class 2 is maximized. The margin is the distance between the closest data to the hyperplane. The patterns lying on the margin are called support vectors.

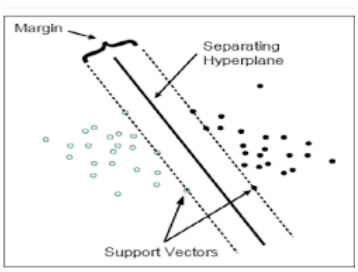


Figure 6: SVM with linear separable case

In the case of a nonlinear separable problem, a kernel function is used to map the feature space into higher dimensional feature space so as to make the problem become linearly separable. A common choice of kernel function is linear, polynomial, radial basis function and sigmoid.

III. Results and Discussion

To evaluate the effectiveness of the proposed method, we have used a public database of Japanese Female Facial Expression (JAFFE). The database consists 10 subjects performing seven facial expressions, 3 to 4 images per emotion with total 213 images (30 anger, 29 disgust, 33 fear, 30 happiness, 30 neutral, 31 sadness and 30 surprises). In extracting the facial region, the facial image of size 256 by 256 pixels is cropped into 128 by 96 pixels. Then, histogram equalization is applied on the image to enhance the image contrast before subjected to the Radon transform and followed by DWT.

In DWT framework, the image in Radon space or sinogram is decomposed into four sub-bands which are approximate coefficients (cA), horizontal coefficients (cH), vertical coefficients (cV) and diagonal coefficients (cD) with Daubechies wavelet of order 2 (db2) because of its symmetry, compact support and the use of overlapping windows to reflect all changes between pixel intensities (Jadhav et al., 2012) in deriving the multiresolution features.

In this work, we decomposed the radon space at the different level of decomposition using DWT (level one, level two and level three) for performance analysis. Then, the sub-band of approximate coefficients at each level (cA1, cA2, and cA3) are independently extracted and used as features as it contains important information that characterized the behavior of the emotion features. The extracted features are concatenated to form a feature vectors as [f_1, f_2, f_3, ...f_n] where n is the number of features. Since the original data contain a high dimensional data features, thus we have adopted PCA as it can reduce a high dimensional data into a lower dimensional data by a linear transformation. As far as we concern, the aim of PCA is to obtain linear transformation from high dimensional data into a lower dimensional data in the direction of variance is maximized. In this work, the first 100 principal components are used as significant features before fed to SVM classifier to recognize the seven facial emotions.

A. Analysis of Applied Radon on Facial Image

Radon transform has a useful tool in capturing the directional features of the images and preserving the variations in pixel intensities. Figure 4 shows the Radon transform (sinogram) of the seven facial emotions (anger, disgust, fear, happiness, sadness surprises + neutral) applied on facial emotion images for angle 0 to 179˚ with a step of 1˚ for better performances. The number of feature vectors is the number of Radon projections of an image in different directions. As observed in Figure 4 that the Radon operator maps the spatial domain (x, y) to the projection (s, θ), in which each point corresponds to a line in the spatial domain. Conversely, each point in the spatial domain becomes a sine curve in the projection domain.

One of the important concept of sinogram is that it shows the properties of the object or image (. For example, if sinogram is symmetric, it reveals that the object is symmetric, and if sinogram is symmetric about the image center is means the object or image is symmetric and parallel to x and y-axes. From observation, the sinogram (Radon transform) of seven facial emotions have shown the symmetrical property which implies that the face does pose the symmetrical configuration of geometrical features. So, it is clear that all the information regarding the shape of Radon transform curves is retained so that the facial emotions can still be distinguished.

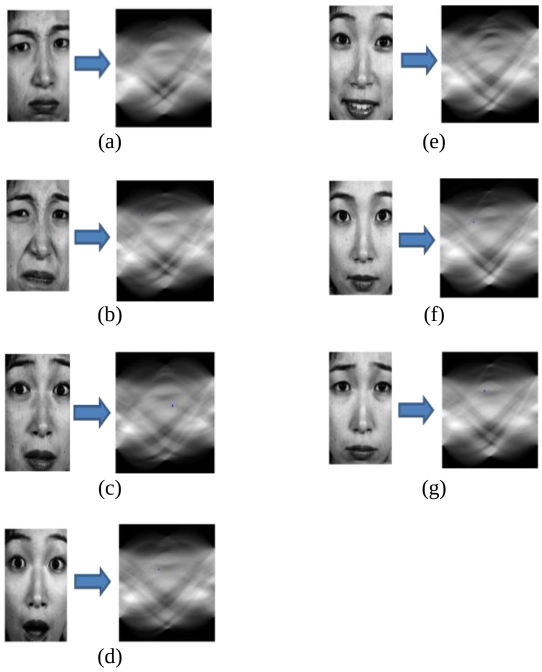


Figure 7: Seven facial expressions; ((a) anger, (b) disgust, (c) fear, (d) surprises, (e) happiness, (f) neutral and (g) sadness) with the respective Radon transforms for an angle of 0-170˚

B. Analysis of Applied Wavelet Transform on Radon on Space

Wavelet transform aims to derive multiresolution features based on the multiscale analysis. Figures 8, 9 and 10 depict the first-, second- and third-level DWT decomposition of Radon space, respectively. As observed in Figure 8, the Radon space of the seven facial emotions are decomposed into four sub-bands which results into approximate coefficients (cA1) horizontal coefficients (cH1), vertical coefficients (cV1) and diagonal coefficients (cD1).

The approximate coefficients (cA) at each level are extracted and used as features since it contains useful information that characterizes the distinctive features of the facial emotions. The low-frequency region in decompositions at different levels is the blurred version of the input image, while the high-frequency regions contain the finer detail or edge information contained in the input.

Table 1 shows the confusion matrix of Radon and DWT based features using SVM classifier first level decomposition. Based on Table 2 it can be seen that the average recognition rate of seven emotions has achieved 93.89%. It observed that emotion of *anger* and *neutral* had yielded the highest recognition rates which are 100%, whereas emotion *fear* gave the lowest recognition rate which is 84.84% or 5 out of 33 are misclassified. This mean that system is confuse to recognize between emotion fear and sad

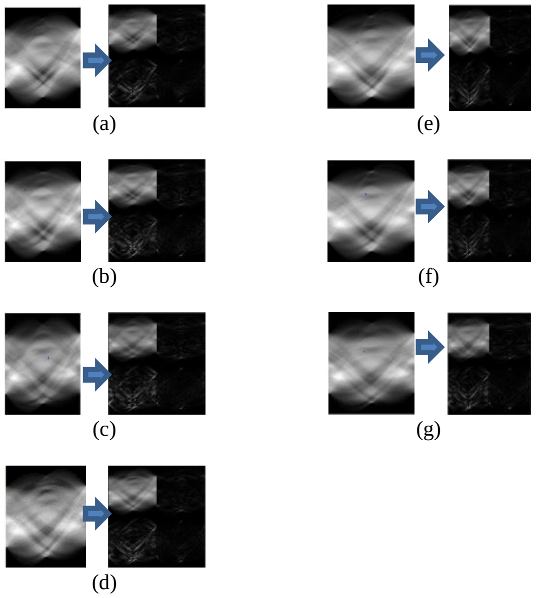


Figure 8: The 1st level of the approximate coefficient (cA1) DWT decomposition of Radon space (sinogram) for seven facial emotions; (a) anger, (b) disgust, (c) fear, (d) surprises, (e) happiness, (f) neutral and (g) sadness

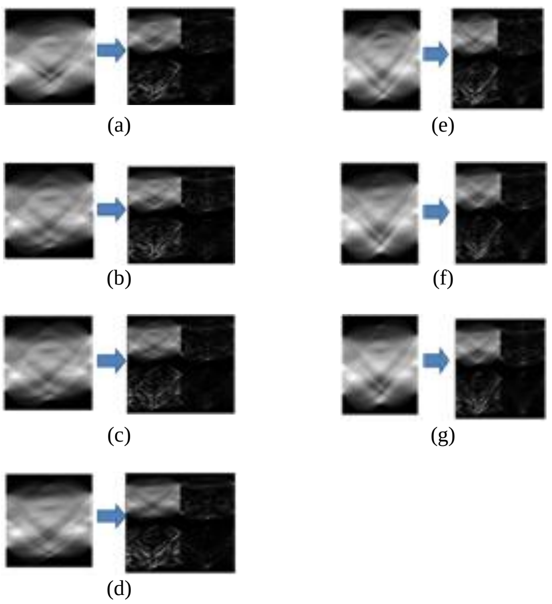


Figure 9: The 2nd level of the approximate coefficient (cA2) DWT decomposition of Radon space (sinogram) for seven facial emotions; (a) anger, (b) disgust, (c) fear, (d) surprises, (e) happiness, (f) neutral and (g) sadness

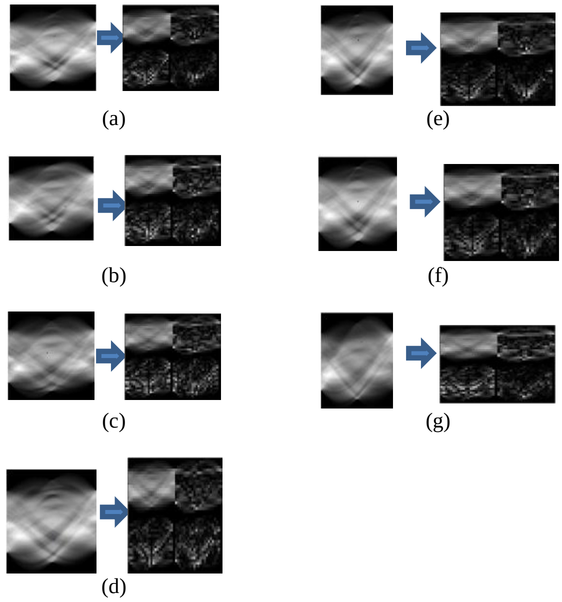


Figure 10: The 3rd level of the approximate coefficient (cA2) DWT decomposition of Radon space (sinogram) for seven facial emotions; (a) anger, (b) disgust, (c) fear, (d) surprises, (e) happiness, (f) neutral and (g) sadness

Table 1
Confusion Matrix of Radon and DWT (1st level DWT) Based Features using SVM Classifier

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	*Average*
Angry	30	1	0	0	0	0	0	
Disgust	0	27	0	0	0	0	0	
Fear	0	0	28	1	0	3	1	
Happy	0	0	0	29	0	0	1	
Neutral	0	0	1	0	30	1	0	
Sad	0	1	4	0	0	27	0	
Surprise	0	0	0	0	0	0	28	
Total	*30*	*29*	*33*	*30*	*30*	*31*	*30*	*93.89%*

Table 2 shows the recognition rates of using Radon and DWT based approach using SVM classifier at second level decomposition. As observed, the average recognition rate has achieved 93.43%, which slightly decreases from first level decomposition. In terms of individual emotion, the emotion of *anger* and *neutral* still achieved maximum rate which is 100%, whereas emotion *fear* contributed to the lowest recognition rate which is 84.85% or 5 out of 33 are misclassified.

Table 2
Confusion Matrix of Radon and DWT Based Features Using SVM Classifier (2nd Level Decomposition)

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	*Average*
Angry	30	2	0	0	0	0	0	
Disgust	0	26	1	0	0	1	0	
Fear	0	0	28	1	0	1	1	
Happy	0	0	0	28	0	0	1	
Neutral	0	0	0	1	30	0	0	
Sad	0	1	3	0	0	29	0	
Surprise	0	0	1	0	0	0	28	
Total	*30*	*29*	*33*	*30*	*30*	*31*	*30*	*93.43%*

Table 3 presents the confusion matrix at third level DWT decomposition on radon space using SVM classifier. As observed in Table 3 that the highest recognition rate was achieved by a *neural* expression which is 100%. On the other hand, the lowest recognition rate was obtained by emotion *sad* which is 80.64% which is 6 out of 33 are misclassified with fear, disgust, anger and neutral. As overall, the average recognition rate at third level decomposition DWT has achieved 91.08% slight decrease as compared second level DWT.

Table 3
Confusion matrix based on radon and DWT (3rd level DWT) features using SVM classifier.

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	*Average*
Angry	29	2	0	0	0	1	0	
Disgust	0	27	0	0	0	1	0	
Fear	0	0	28	1	0	3	1	
Happy	0	0	0	27	0	0	1	
Neutral	0	0	0	2	30	1	0	
Sad	1	0	3	0	0	25	0	
Surprise	0	0	2	0	0	0	28	
Total	*30*	*29*	*33*	*30*	*30*	*31*	*30*	*91.08%*

In short, the recognition rate of the first level decomposition DWT using SVM classifier has outperformed the second and third-level decomposition. This means that at the coarser levels only the strongest features of the sequence will be noticeable. Moreover, the intrinsic features at first level of sub-band frequency have significant information to characterize the behavior of the emotions thus improved the recognition rate.

IV. CONCLUSION

This paper has presented a new framework for facial expression recognition of using Radon and wavelet transform using SVM classifier. Radon transform preserves the pixels variations in deriving directional features of facial images. The wavelet transform is used to extract the Radon domain into multi resolution features. The approximation coefficients are extracted and used as features to classify the facial expression. Based on results obtained, the approximation coefficients at first level decomposition has achieved the highest recognition rate which is 93.89% outperformed the second and third decomposition. However, further study should be conducted to investigate the optimum features by utilizing optimization techniques such as differential evolutionary and Taguchi method in order to enhance the recognition results.

.

REFERENCES

[1] J. Clerk Maxwell, A M. Ilbeygi and H. Shah-Hosseini, “A novel fuzzy facial expression recognition system based on facial feature extraction from color face images.“ Engineering Applications of Artificial Intelligence, Vol. 25 (1), 2012, p. 130–146.

[2] M. Pantic, and L. J. M. Rothkrantz: Expert system for automatic analysis of facial expressions. Image and Vision Computing, 18, 2000, 881–905.

[3] N. Aifanti, and A. Delopoulos, “Linear subspaces for facial expression recognition. Signal Processing: Image Communication, 29(1), 2014, p. 177–188.

[4] T. Danisman, I. M. Bilasco, J. Martinet, and C. Djeraba, “Intelligent pixels of interest selection with application to facial expression recognition using multilayer perceptron. Signal Processing, 93(6), 2013, p. 1547–1556.

[5] Kotsia, S. Zafeiriou, and I. Pitas, “Texture and shape information fusion for facial expression and facial action unit recognition. Pattern Recognition, 41(3), 2008, p. 833–851.

[6] Z. Zhang, M. Lyons, M. Schuster, and S.Akamatsu: Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. In IEEE International Conference on Automatic Face and Gesture Recognition, 1998. (pp. 454–459.

[7] X. Feng, M. Pietikäinen, and A. Hadid, “Facial Expression Recognition with Local Binary Patterns and Linear Programming. Pattern Recognition and Image Analysis, 15(2), 2005, p. 546–548.

[8] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, & T. J. Sejnowski, “Classifying facial actions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21, 1999.

[9] H. Deng, L. Jin, L. Zhen, and J. Huang, “ A new facial expression recognition method based on local gabor filter bank and PCA plus LDA. International Journal of Information Technology, 11(303), 2005, p. 86–96.

[10] E. Owusu, Y. Zhan, and Q. R. Mao, “A neural-AdaBoost based facial expression recognition system. Expert Systems with Applications, 41(7), 2014, p. 3383–3390.

[11] E. Magli, G. Olma, and L. Lo Presti, “Pattern recognition by means of the Radon transform and the continuous wavelet transform, Signal Processing, Vol. 73, 1999, p. 277-289.

[12] S. Arivazhagan and L. Ganesan, “Texture classification using wavelet transform. Pattern Recognition Letters, 24, 2003, p.1513-1521.

[13] H. Ali, V. Sritharan, M. Hariharan, M. Shaikh, K. Wan. Feature Extraction Using Radon Transform and Discrete Wavelet Transform for Facial Emotion Recognition. 2016 2nd IEEE International Symposium on Robotics and Manufacturing Automation.

[14] D. V. Jadhav, and R. S. Holambe, “ Feature extraction using Radon and wavelet transforms with application to face recognition. Neurocomputing, 72(7-9), 2009, p. 1951–1959.

[15] J. A. Dargham, A. Chekima, E. Moung, and S.Omatu: Radon transform for face recognition. Artificial Life and Robotics, 15(3), 2010, p.359–362.

[16] U, R. Acharya, C. K. Chua, E. Y. K. Ng, W. Yu, and C.Chee: Application of higher order spectra for the identification of diabetes retinopathy stages. Journal of Medical Systems, 32(6), 2008, p. 481–488.

[17] R. C. Gonzalez, and R. E. Woods, “Digital Image Processing (3rd Edition). 3nd edition. Prentice Hall, New York, 2007.

[18] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 1997, p. 711–720.

[19] V. Vapnik, (1998). Statistical learning Theory. John Wiley and Sons.Inc., New York.