Hindawi
Mathematical Problems in Engineering
Volume 2017, Article ID 9854050, 9 pages
https://doi.org/10.1155/2017/9854050

Research Article

Facial Expression Recognition Using Stationary Wavelet Transform Features

Huma Qayyum,[1] Muhammad Majid,[1] Syed Muhammad Anwar,[2] and Bilal Khan[3]

[1]*Department of Computer Engineering, University of Engineering and Technology, Taxila, Taxila 47050, Pakistan*
[2]*Department of Software Engineering, University of Engineering and Technology, Taxila, Taxila 47050, Pakistan*
[3]*Department of Electrical Engineering, COMSATS Institute of Information Technology, Abbottabad 22010, Pakistan*

Correspondence should be addressed to Muhammad Majid; m.majid@uettaxila.edu.pk

Received 9 October 2016; Revised 11 December 2016; Accepted 18 December 2016; Published 11 January 2017

Academic Editor: Simone Bianco

Humans use facial expressions to convey personal feelings. Facial expressions need to be automatically recognized to design control and interactive applications. Feature extraction in an accurate manner is one of the key steps in automatic facial expression recognition system. Current frequency domain facial expression recognition systems have not fully utilized the facial elements and muscle movements for recognition. In this paper, stationary wavelet transform is used to extract features for facial expression recognition due to its good localization characteristics, in both spectral and spatial domains. More specifically a combination of horizontal and vertical subbands of stationary wavelet transform is used as these subbands contain muscle movement information for majority of the facial expressions. Feature dimensionality is further reduced by applying discrete cosine transform on these subbands. The selected features are then passed into feed forward neural network that is trained through back propagation algorithm. An average recognition rate of 98.83% and 96.61% is achieved for JAFFE and CK+ dataset, respectively. An accuracy of 94.28% is achieved for MS-Kinect dataset that is locally recorded. It has been observed that the proposed technique is very promising for facial expression recognition when compared to other state-of-the-art techniques.

1. Introduction

Emotions are a natural and powerful way of communication among living beings. Humans express their emotions by voice, face, body gestures, and behavioral changes. A reliable emotion perception scheme is required in order to translate human expression and behavioral changes into useful commands to control systems. Emotion recognition is a challenging task because humans do not always express themselves by words and gestures. Automatic human emotion recognition is a multidisciplinary area including psychology, speech analysis, computer vision, and machine learning. Facial expression is considered as a powerful mean of one to one communication after speech signals and plays a pivotal role in human computer interaction (HCI). Human emotional state is provoked by external stimuli resulting in changes in facial dimensions. The most commonly used system for face expression is developed by Ekman and Frieses and is famously known as Facial Action Coding System (FACS) [1, 2]. In [3] Ekman has defined six basic classes of facial expression, that is, anger, disgust, fear, happiness, sadness, and surprise, which are commonly used by researchers working in this area. Automatic facial expression recognition and analysis play a vital role in different application areas, such as human machine interaction, surveillance, information security, robotics, and video summarization.

Development of an affective facial recognition system still remains a challenging task. Facial images and videos are affected by illumination conditions, human age, and variations in how the emotion is expressed. A first attempt was made in 1978 by Suwa et al. [4] to develop automatic facial expression system by analyzing twenty known areas of image structure. Since then many attempts have been made to develop automatic emotion recognition system that can narrate affective human feelings. Automatic facial expression recognition system is divided into three main stages, that is, (1) face detection and tracking, (2) feature extraction, and (3) emotion classification [1, 2]. In the first stage, human

face is cropped by face detection, head tracking, and head pose estimation. In the second stage relevant image features are extracted that describes the changes in facial expressions. These features can be extracted in holistic or anatomical fashion. In holistic approach features are extracted from the whole face and are based on image texture or its transformation [5, 6]. On the other hand in anatomical approach feature are extracted from subportions of the face and are based on distance measures and geometric transformations [7, 8]. For facial expression recognition systems, extracted facial features have been generally based on geometry [9, 10] or appearance [11, 12]. Appearance based features used skin texture variants like furrows and wrinkles for studying facial expression and these features have been applied to the complete face as well as selected face regions. Donato et al. [6] have worked in plane image transform by applying Gabor Wavelets as texture descriptors, on posed data of 24 subjects and nearest neighbor as classifier. Wu et al. [13] used Gabor motion energy as texture descriptor in plane image transform with support vector machine (SVM) as emotion classifier and Cohn-Kanade (CK) [14] as database. However, texture deformation dynamics can also be incorporated for feature extraction. On the other hand geometric features analyze the variation of human face components such as nose, eyebrows, lips, eyes in terms of location, distance, and shape. Yacoob and Davis [15] have applied rule-based classifier and region based optical geometric methods on posed data of 32 subjects for emotion recognition. Wang et al. [16] recognized emotions by using geometrical B spline curve, with Euclidean classifier and applied algorithm on posed data of 8 different subjects. Valstar and Pantic [17] applied affine transformation registration technique on predefined MMI [18] and CK [14] databases by defining dynamics of 20 facial points with probabilistic actively learned support vector machine as emotion classifier. Hybrid approaches make use of both the geometrical and appearance based features for emotion recognition. A comprehensive survey of existing methods that have been adopted in facial expression recognition systems is presented in [19–21].

Most important aspect of facial emotion recognition system is to extract relevant features from either spatial or frequency domain. A number of algorithms based on frequency domain features have been proposed in literature that have used discrete cosine transform (DCT), discrete wavelet transform (DWT), Discrete Fourier Transform (DFT), Gabor Wavelet Transform, and curvelet transform [22–28]. In [22], three different types of features based on DCT, FFT, and signal value decomposition are extracted and then fed to SVM for facial expression recognition. A comparative study of DCT and two-dimensional principal component analysis (2D-PCA) for feature dimension reduction in facial expression recognition is presented in [23] and it is found that DCT gives better recognition rate with the same feature reduction as compared to 2D-PCA. In [24], wavelet based features are provided to bank of seven parallel SVMs. A particular facial expression is recognized by each SVM, which are then combined by maximum function. DWT, DFT, and DCT are used as a unique combination in [25] to extract features for face recognition system that are invariant in terms of pose, translation, and illumination. In [26], 18 filtered images are obtained by convolving input image with 18 Gabor filters. The amplitude value of each filtered image from selected points is used as features, which are then classified by using Bayes classifier, SVM, and AdaBoost. In [27], facial element and muscle movement are used as features for facial expression recognition, which are obtained from patch based 3D Gabor filters. In [28], entropy, standard deviation, and mean of curvelet transform coefficients of each region are used as features for facial expression recognition. The extracted features are passed to online sequential extreme learning machine with radial basis function.

Current frequency domain approaches have not fully utilized the facial elements and muscle movements for recognition. Therefore, hybrid frequency domain features are required to fully cover these aspects, which are necessary for facial expression recognition. In this paper, we aim to improve the performance of facial expression recognition system by extracting feature in frequency domain using stationary wavelet transform (SWT). The main reason for utilizing this transformational technique is due to its good localization characteristics, in both spectral and spatial domains [29]. Once SWT is applied on the detected face, feature vector length is further reduced by applying DCT and selecting a few number of DCT coefficients. For classification of facial expression, an artificial neural network is trained on extracted features from the images. The major contributions of this study are as follows:

(1) Stationary wavelet transform for facial expression recognition with a special emphasis on horizontal and vertical subbands is used.

(2) A new dataset of videos is generated using MS-Kinect with seven general emotions.

(3) A significant increase in accuracy of classification is achieved when compared to the other frequency based emotion classifiers.

The remainder of the paper is organized as follows: In Section 2, an overview of stationary wavelet transform is presented. Section 3 describes the proposed methodology in detail. Section 4 presents the facial expression recognition results of the proposed method and demonstrates a comparison with state-of-the-art methods followed by conclusion in Section 5.

2. Stationary Wavelet Transform

Conventionally, Fourier Transform (FT) is used as a signal analysis tool that converts the signal into constituent sinusoids of different frequencies. The major drawback with Fourier Transform is the loss of time information. Short Time Fourier Transform (STFT) is considered as a compromise between the time and frequency information. In STFT, a window is applied to the signal and then Fourier Transform is computed. The preciseness of STFT depends on window shape and size. Wavelet transform preserves both the time and frequency information by decomposing the signal in

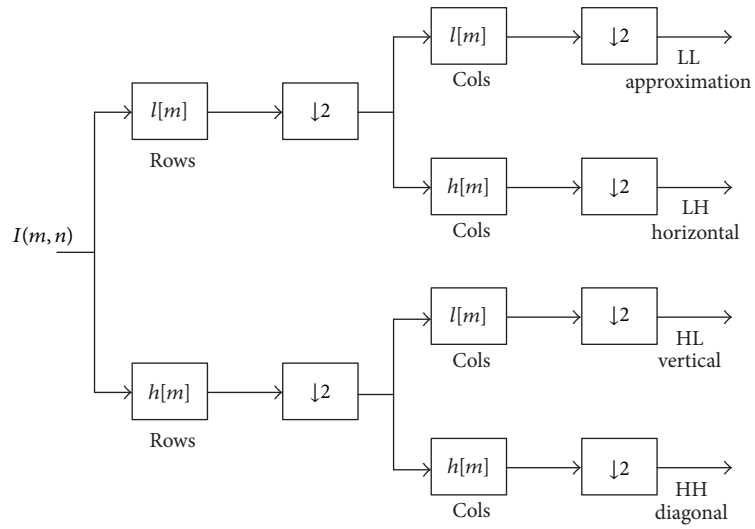

Figure 1: Single level discrete wavelet transform decomposition of image into four subbands.

a hierarchy of increasing resolution. Wavelet transform of signal $x(t)$ is represented as

$$W(a,b) = \int_{-\infty}^{\infty} x(t)\,\psi_{a,b}(t)\,dt, \tag{1}$$

where $\psi_{a,b}(t)$ is the dilated and translated version of the mother wavelet ψ and is calculated as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-a}{b}\right), \tag{2}$$

where a and b are real and positive number representing dilation and translation. Similarly, discrete wavelet transform of signal $x[n]$ is represented as

$$W(k,l) = \sum_{m=-\infty}^{\infty} x[m]\,\psi_{k,l}[m], \tag{3}$$

where $\psi_{k,l}[m]$ is the dilated and translated version of the mother wavelet ψ and is calculated as

$$\psi_{k,l}[m] = 2^{-k/l}\psi\left[2^{-k}m - l\right]. \tag{4}$$

Discrete wavelet transform (DWT) can be implemented using filter bank approach and lifting scheme. In filter bank approach, input signal is passed through low $l[m]$ and high pass $h[m]$ filters and then decimated by a factor of two to get approximation and detailed coefficients. In case of image, DWT is applied in each dimension separately. Figure 1 shows the single level wavelet decomposition of image, which results in four subbands, that is, LL, LH, HL, and HH representing low resolution approximation, horizontal, vertical, and diagonal information of the input image. DWT is a spatial variant transform that means the DWT of shifted version of signal is not equivalent to the shift in DWT of signal. The spatial variant nature of DWT occurs because of the decimation operation that can be carried out by either choosing even indexed or odd indexed elements.

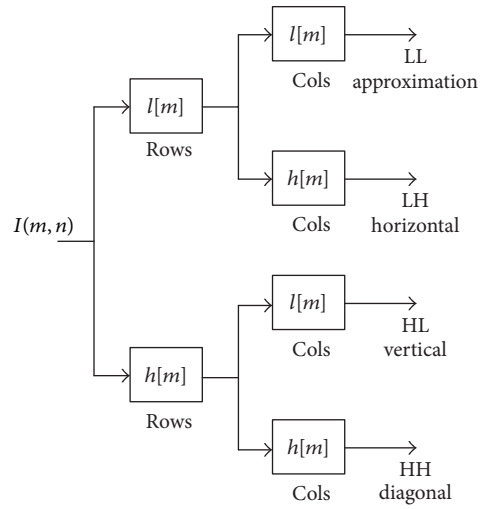

Figure 2: Single level stationary wavelet transform decomposition of image into four subbands.

Stationary wavelet transform (SWT) solves this problem of shift invariance. SWT differs from conventional DWT in terms of decimation and shift invariance, which makes it feasible for change detection, pattern recognition, and feature extraction. In conventional DWT, at each level of transform input signal is firstly convolved with low $l[m]$ and high $h[m]$ pass filter and then decimated by a factor of two to obtain wavelet transform coefficients. The resolution after DWT remains the same as the input signal. In SWT, the input signal is convolved with low $l[m]$ and high $h[m]$ pass filter in a similar manner as in DWT but no decimation is performed to obtain wavelet coefficients of different subbands. As there is no decimation involved in SWT, therefore the number of coefficients is twice that of the samples in the input signal. Figure 2 shows single level decomposition of the input image using SWT. It is clear from Figure 2 that M × N image is

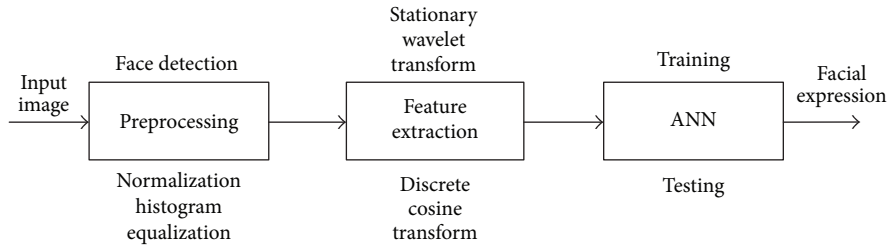

Figure 3: General framework of the proposed methodology.

decomposed into four subbands, that is, LL, LH, HL, and HH representing low resolution approximation, horizontal, vertical, and diagonal information of the input image all having the same M × N resolution.

3. Proposed Methodology

Figure 3 illustrates the proposed methodology for facial expression recognition, which is composed of three stages, namely, preprocessing, feature extraction, and classification. A detailed discussion is presented in following subsections.

3.1. Preprocessing. Preprocessing stage is divided into two steps, that is, face detection and normalization. In the first step, human face is detected from input image with the help of Viola and Jones algorithm [30]. The main advantage of this algorithm is its capability to quickly detect faces with a high detection rate. The algorithm is composed of three phases.

First, the input image I is represented in the form of an integral image I_{int}, which is calculated as

$$I_{\text{int}}(x, y) = \sum_{x_i \le x, y_i \le y} I(x_i, y_i). \tag{5}$$

The computation is performed on the entire image in a single pass using the following pair of equations:

$$\begin{aligned} s(x, y) &= s(x, y-1) + I(x, y), \\ I_{\text{int}}(x, y) &= I_{\text{int}}(x-1, y) + s(x, y), \end{aligned} \tag{6}$$

where $s(x, y)$ represents the row sum, $s(x, -1)$, and $I_{\text{int}}(-1, y)$ equals zero. The use of integral image allows rapid feature computation in image subregions and is independent of the size of neighborhood selected. Secondly, AdaBoost learning algorithm is used to select few critical features from the complete set of features. Thirdly, the classifiers are combined in a cascade manner resulting in rapid discarding of background regions and spending more time on the computation of face like regions. The features used in these classifiers are based on the area of rectangular neighborhood of pixels. For a rectangular area in the integral image with four corner pixel values of P, Q, R, and S, the area is calculated as

$$A = P + S - Q - R. \tag{7}$$

In the second step, image normalization and histogram equalization are performed on detected face to remove unrelated and unwanted parts, which are present in the background of dataset. The normalized image I_{norm} is obtained as

$$I_{\text{norm}}(x, y) = \frac{I_d(x, y) - \min(I_d(x, y))}{\max(I_d(x, y)) - \min(I_d(x, y))}, \tag{8}$$

where $I_d(x, y)$ is the subimage detected as face region and $\min(\cdot)$ and $\max(\cdot)$ are functions used to find the minimum and maximum pixel values, respectively. Image normalization changes the intensity of images to the new intensity range [0-1]. Equalization is used to enhance the global contrast of the image giving better dynamic range.

3.2. Feature Extraction. The process of feature extraction is shown in Figure 4. The detected and preprocessed face from input image is firstly decomposed into different subbands using stationary wavelet transform. SWT differs from conventional DWT in terms of decimation and shift invariance at the cost of redundant information. The mathematical proof of the shift invariance of SWT is discussed in detail in [31] when subbands are not decimated.

In SWT, input image is convolved with low pass and high pass filter to obtain approximated and detailed coefficients without decimation. For the detected face image I_d of size M × N, the SWT at jth level is given as

$$\begin{aligned} \text{LL}_{j+1}(a, b) &= \sum_x \sum_y l_x^j l_y^j \text{LL}_j(a + x, b + y), \\ \text{LH}_{j+1}(a, b) &= \sum_x \sum_y h_x^j l_y^j \text{LL}_j(a + x, b + y), \\ \text{HL}_{j+1}(a, b) &= \sum_x \sum_y l_x^j h_y^j \text{LL}_j(a + x, b + y), \\ \text{HH}_{j+1}(a, b) &= \sum_x \sum_y h_x^j h_y^j \text{LL}_j(a + x, b + y), \end{aligned} \tag{9}$$

where $a = 1, 2, \ldots, M$, $b = 1, 2, \ldots, N$, and h and l represent the low pass and high pass filters. LL, LH, HL, and HH represent the approximate, horizontal, vertical, and diagonal subbands, respectively. In each SWT subband different information of the image is retained. The LL subband is the overall image approximation, LH, HL, and HH have the horizontal, vertical, and diagonal information, respectively, as shown in Figure 4. This piece of information helps in recognizing facial expressions that are dependent on the changes that occur in these orientations. For example, for

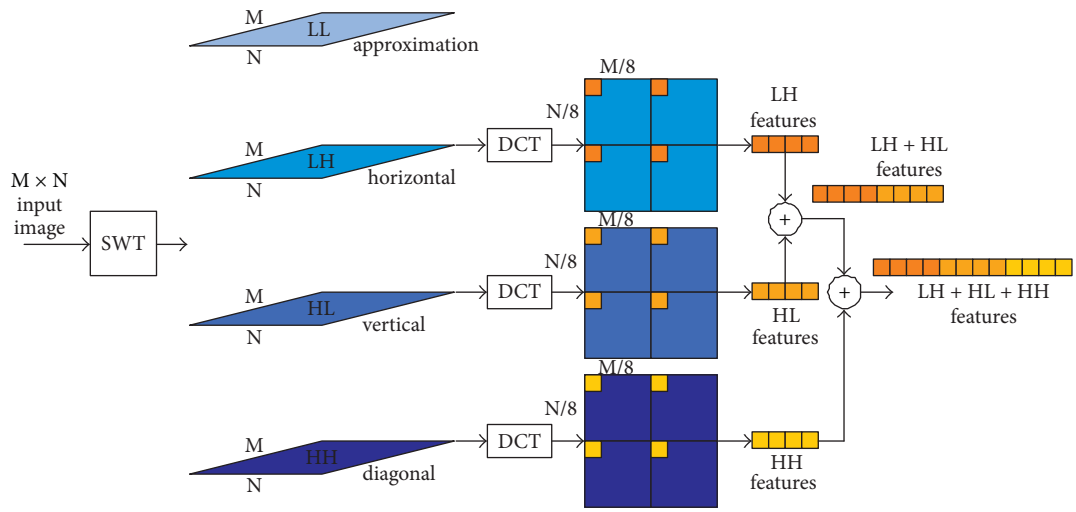

Figure 4: Proposed feature extraction process using SWT and DCT.

smile the major changes in face occur in horizontal direction due to movement of lips. The resulting SWT decomposition has the same size as the original image, which results in four times the number of coefficients as compared to the input image, since we have two-dimensional data. Therefore, some mechanism is required to reduce the features from the nondecimated wavelet coefficients.

To reduce feature vector length, 8×8 block DCT is applied to the LH, HL, and HH subbands of SWT. The DCT applied to each block is calculated as

$$X(u,v) = \frac{C(u)\,C(v)}{4} \sum_{m=0}^{7} \sum_{n=0}^{7} x[m,n] \cdot \cos\left(\frac{(2m+1)\,u\pi}{16}\right) \cos\left(\frac{(2n+1)\,v\pi}{16}\right), \tag{10}$$

where

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}}, & u = 0, \\ 1, & 1 \le u \le 7, \end{cases} \qquad C(v) = \begin{cases} \frac{1}{\sqrt{2}}, & v = 0, \\ 1, & 1 \le v \le 7. \end{cases} \tag{11}$$

DC coefficient from each block is selected as features for each subband because it represents majority of the energy of that subband. We have combined features from different subbands, that is, LH, HL, LH + HL, and LH + HL + HH, to obtain more descriptive features in terms of horizontal, vertical, and diagonal directions, which are then fed to artificial neural network for facial expression recognition. This combination of SWT and DCT resulted in improved classification results by utilizing redundant information from different SWT subbands, but also having a reduced feature vector length by utilizing only the DC coefficient of DCT.

3.3. Artificial Neural Network. The selected features are fed into a neural network that is trained to classify the seven common facial expressions. The neural network design consists of three fully connected layers as shown in Figure 5. It has k inputs (f_1–f_k), which is equivalent to the length of the feature vector and seven outputs (O_1–O_7) that correspond to the emotions being recognized. The training data is organized in pairs (F_i, Y_i), where F is the input feature vector and Y is the corresponding target output. The network uses feed forward connections for training and back propagation for optimization. The actual output of the network during training is given as O that differs slightly from the target output Y. The output layer is defined as a softmax function and is given as

$$O_i^L = \frac{e^{z_{L_i}}}{\sum_{z=1}^{Z} e^{z_{L_z}}}, \tag{12}$$

where, L represent the output layer and $Z = 7$. This gives a probability distribution of the classified emotions where the highest value is picked as the emotion classified. All other neurons use sigmoid function for activation and are given as

$$f(z) = \frac{1}{1+e^{-z}}, \qquad z^{l+1} = w^l a^l + b^l, \qquad a^{l+1} = f\left(z^{l+1}\right), \tag{13}$$

where z and a represent each neuron's input and output, respectively. The network is parametrized using the connection weights w and biases b. These network parameters are initialized using Gaussian distribution and are trained using backpropagation to minimize negative log probability that is used as the cost function.

The optimization is performed using stochastic gradient descent (SGD), which has the following cost function:

$$C = -\frac{1}{n}\sum_{n} \ln\left(a^L\right) + \lambda \|W\|^2, \tag{14}$$

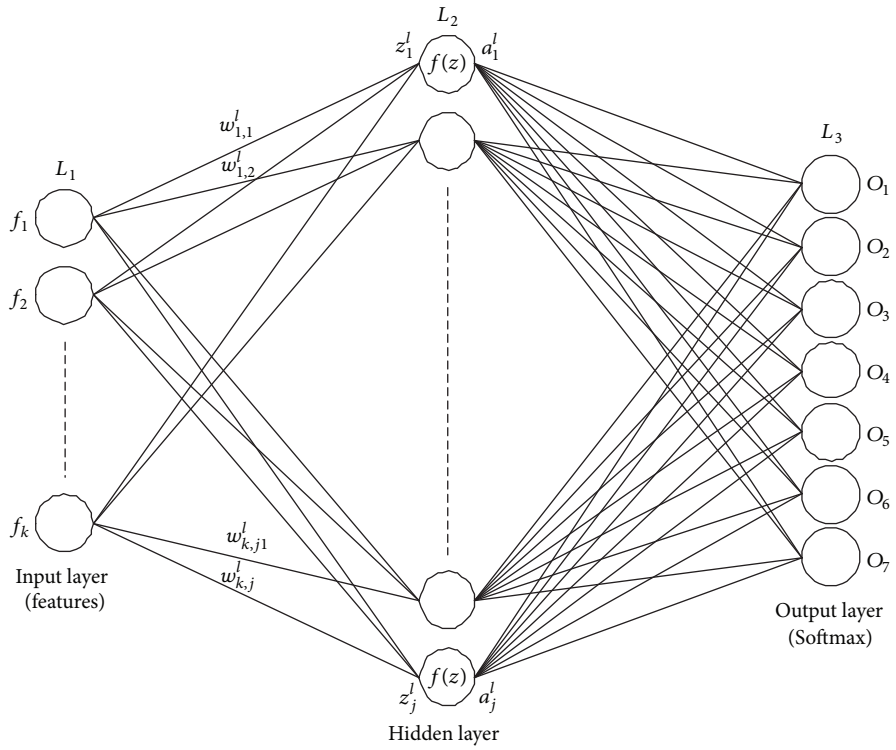

Figure 5: Architecture of the neural network used in the proposed framework.

where n is the total number of inputs and a^L is output of the final layer and λ is the regularization parameter. The error is back-propagated, and using SGD the algorithm converges to its optimal state. For backpropagation, gradient values are used to update the network parameters. The weights are updated using

$$\Delta w_{i,j}^{l+1} = -\gamma a_i^l \delta_j^{l+1}, \tag{15}$$

where γ is the learning rate and δ represents the error gradient given as

$$\delta_j^l = \frac{\partial E}{\partial a_i w_{i,j}^l}. \tag{16}$$

L2 regularization is also used to update the cost function to avoid overfitting. The system is designed to solve the following minimization problem to find the best weights as

$$\min_W \sum_n C\left((W, b), F\right) + \lambda \|W\|^2. \tag{17}$$

This design was particularly authenticated for the classification of seven ideal facial expressions of JAFFE and CK+ database. The same experiment was also conducted on our own collected database, where seven common facial expressions are classified. L2 penalty is used with the weight terms, so that the neural network generalizes better and does not overfit the sampling errors.

4. Simulation Results

In this section we present the recognition performance of the proposed approach and its comparison with the state-of-the-art frequency domain facial expression recognition methods. The recognition performance is computed in terms of correct recognition rate (CRR). Three different datasets were used to evaluate the facial expression recognition results. The datasets used are Japanese Female Facial Expression (JAFFE), Cohn-Kanade (CK+), and our own dataset acquired from MS-Kinect device. The JAFFE dataset consists of 213 gray scale images of seven different expressions by 10 different females. The seven different emotions present are anger (AN), happiness (HA), neutral (NE), sadness (SA), disgust (DI), fear (FE), and surprise (SU). The spatial resolution of each image is 256×256 and is rated for different facial expression by 60 subjects. The CK+ dataset consists of both posed and nonposed facial expression of 210 adults. In this work supervised learning is used; hence the posed data from CK+ dataset is used. The dataset consists of seven emotions. The spatial resolution of the frames used is 640×480. The MS-Kinect dataset is created for the same seven emotions and the purpose is to use facial expression recognition in MS-Kinect based applications. The MS-Kinect dataset consists of 210 images of seven expressions by 5 males and 5 females. The spatial resolution of each image is 640×480 and is rated for different emotions by 10 subjects.

In this work features were extracted using stationary wavelet transform. Applying DCT on each subband of SWT reduces the overall feature vector length. Four different

Table 1: CRRs of seven emotions (anger (AN), happiness (HA), sadness (SA), disgust (DI), fear (FE), surprise (SU), and neutral (NE)) on JAFFE, CK+, and MS-Kinect dataset.

Dataset	SWT subbands used in feature selection			
	LH	HL	LH + HL + HH	LH + HL
JAFFE	90.07%	90.12%	88.52%	98.83%
CK+	91.03%	90.89%	88.23%	96.61%
MS-Kinect	89.52%	89.52%	87.61%	94.28%

Table 2: Confusion matrix of seven emotions on the JAFFE dataset using LH + HL (horizontal and vertical) subband features.

	AN	HA	NE	SA	DI	FE	SU
AN	**96.66**	0	0	0	0	3.34	0
HA	0	**100**	0	0	0	0	0
NE	0	3.34	**96.66**	0	0	0	0
SA	0	0	0	**100**	0	0	0
DI	3.57	0	0	0	**96.43**	0	0
FE	0	0	0	0	0	**100**	0
SU	0	0	0	0	0	0	**100**

Table 3: Confusion matrix of seven emotions on the CK+ dataset using LH + HL (horizontal and vertical) subband features.

	AN	HA	NE	SA	DI	FE	SU
AN	**100**	0	0	0	0	0	0
HA	0	**93.33**	0	0	6.67	0	0
NE	0	0	**90**	0	6.67	3.33	0
SA	6.7	0	0	**86.67**	0	6.67	0
DI	0	0	0	0	**100**	0	0
FE	0	3.33	0	0	0	**90**	6.67
SU	0	0	0	0	0	0	**100**

features are created by combining DCT coefficients from different subbands, that is, LH, HL, LH + HL + HH, and LH + HL considering horizontal, vertical, horizontal + vertical + diagonal, and horizontal + vertical information. Twenty images for each expression are used for training purpose and the rest of the images are used for testing. Table 1 shows the CRRs of seven emotions on JAFFE and MS-Kinect dataset by considering different features vectors based on the selection of SWT subbands. It is evident from Table 1 that horizontal and vertical SWT subband contribute more in accurately recognizing the facial expression. The combination of features from LH and HL subband representing horizontal and vertical features of the face gives the best CRR of 98.83%, 96.61%, and 94.28% for JAFEE, CK+, and MS-Kinect dataset, respectively.

The confusion matrix of seven emotions on JAFFE dataset is presented in Table 2. Anger and disgust emotions are difficult to recognize with respect to other expressions having CRR of 96.6 and 96.4 percent, respectively. On the other hand the rest of the facial expressions are perfectly recognized. In terms of misrecognition rate, anger contributes the most in the reduction of overall performance. Table 3 represents the confusion matrix of CK+ dataset. Anger and surprise emotions are recognized with higher recognition rate while disgust, fear, and happy emotions are recognized with lesser recognition rate. The confusion matrix of seven emotions on MS-Kinect dataset is presented in Table 4. It is evident that sad, fear, and happy facial expressions are difficult to be recognized with the CRR of 86.6, 90, and 93 percent, respectively. On the other hand the rest of the facial expressions are perfectly recognized. In terms of misrecognition rate, sadness contributes the most to the reduction of overall performance.

Table 4: Confusion matrix of seven emotions on the MS-Kinect dataset using LH + HL (horizontal and vertical) subband features.

	AN	HA	NE	SA	DI	FE	SU
AN	**98.65**	0	1.35	0	0	3.34	0
HA	1.35	**96.85**	0	0	1.80	0	0
NE	0	2.28	**94.62**	1.46	1.64	0	0
SA	0	0	1.20	**97**	1.80	0	0
DI	5.1	0	0	0	**93.70**	1.20	0
FE	0	2.0	0	2.6	0	**95.40**	0
SU	0	0	0	0	0	0	**100**

Table 5: Comparison of the proposed facial expression recognition system with state-of-the-art methods for JAFEE dataset.

Method, year	Features	Overall CRR (emotions)
Proposed	SWT + DCT	98.83% (7)
[28], 2016	Curvelet transform	95.17% (7)
[27], 2011	Patch based Gabor	92.93% (6)
[26], 2005	Gabor + FSLP	91.00% (7)
[23], 2008	DCT	79.30% (7)

Table 5 represents the comparison of proposed facial expression recognition scheme with the state-of-the-art methods. The state-of-the-art schemes are selected because they used frequency domain features, similar testing strategy, and the same dataset. It is evident from the table that the proposed scheme outperforms state-of-the-art approaches when JAFFE dataset is used. The CRR of the proposed scheme is 19.53, 7.83, 5.9, and 3.66 percent higher than those in [23], [26], [27], and [28], respectively.

5. Conclusion

This study investigates the facial expression recognition system from images using stationary wavelet transform features. These features have also been compared with other state-of-the-art frequency domain features in terms of correct recognition rate and classification accuracy. The simulation results also reveal meaningful performance improvements due to the use of SWT features. Different emotions generate varying muscle movements. However, majority of the emotions bring horizontal and vertical muscle movement on the face. Therefore, features are combined from LH and HL subband of SWT representing horizontal and vertical direction, respectively. Decimation operation is not involved in SWT, which results in a large number of coefficients. Hence, DCT is performed to reduce the features dimension. The results indicate that SWT

based features show 19.53, 7.83, 5.9, and 3.66 percent better recognition results than DCT based, Gabor based, patched based, and curvelet based features when applied on JAFFE dataset. Moreover, the highest accuracy of proposed scheme is 98.8%, 96.61%, and 94.28% in case of JAFFE, CK+, and MS-Kinect, respectively, when LH + HL information is utilized. The proposed facial expression recognition scheme can play a vital role in HCI and Kinect based applications. In the future, we intend to use the proposed facial expression recognition scheme for the generation of personalized video summaries.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] P. Ekman and W. V. Friesen, *Facial Action Coding System*, Consulting Psychologists Press, Stanford University, Palo Alto, Calif, USA, 1977.

[2] J. C. Hagerand, P. Ekman, and W. V. Friesen, *Facial Action Coding System*, A Human Face, Salt Lake City, Utah, USA, 2002.

[3] P. Ekman, *The Argument and Evidence About Universals in Facial Expressions of Emotion*, John Wiley & Sons, Hoboken, NJ, USA, 1989.

[4] M. Suwa, N. Sugie, and K. Fujimora, "A preliminary note on pattern recognition of human emotional expression," in *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pp. 408–410, Kyoto, Japan, November 1978.

[5] A. J. Calder, A. M. Burton, P. Miller, A. W. Young, and S. Akamatsu, "A principal component analysis of facial expressions," *Vision Research*, vol. 41, no. 9, pp. 1179–1208, 2001.

[6] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[7] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 699–714, 2005.

[8] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "Dynamics of facial expression extracted automatically from video," *Image and Vision Computing*, vol. 24, no. 6, pp. 615–625, 2006.

[9] M. F. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 1, pp. 28–43, 2012.

[10] M. H. Siddiqi, R. Ali, A. M. Khan, E. S. Kim, G. J. Kim, and S. Lee, "Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection," *Multimedia Systems*, vol. 21, no. 6, pp. 541–555, 2015.

[11] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[12] S. M. Lajevardi and Z. M. Hussain, "Automatic facial expression recognition: feature extraction and selection," *Signal, Image and Video Processing*, vol. 6, no. 1, pp. 159–169, 2012.

[13] T. Wu, M. S. Bartlett, and J. R. Movellan, "Facial expression recognition using Gabor motion energy filters," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPRW '10)*, pp. 42–47, San Francisco, Calif, USA, June 2010.

[14] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG '00)*, pp. 46–53, IEEE, Grenoble, France, March 2000.

[15] Y. Yacoob and L. S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 636–642, 1996.

[16] M. Wang, Y. Iwai, and M. Yachida, "Expression recognition from time-sequential facial images by use of expression change model," in *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition (FG '98)*, pp. 324–329, IEEE, Nara, Japan, April 1998.

[17] M. F. Valstar and M. Pantic, "Combined support vector machines and hidden Markov models for modeling facial action temporal dynamics," in *Proceedings of the International Workshop on Human-Computer Interaction*, pp. 118–127, 2007.

[18] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[19] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[20] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: a comprehensive survey," *Image and Vision Computing*, vol. 30, no. 10, pp. 683–697, 2012.

[21] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: a survey of registration, representation, and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1113–1133, 2015.

[22] G. U. Kharat and S. V. Dudul, "Human emotion recognition system using optimally designed SVM with different facial feature extraction techniques," *WSEAS Transactions on Computers*, vol. 7, no. 6, pp. 650–659, 2008.

[23] B. Jiang, G.-S. Yang, and H.-L. Zhang, "Comparative study of dimension reduction and recognition algorithms of dct and 2DPCA," in *Proceedings of the 7th International Conference on Machine Learning and Cybernetics (ICMLC '08)*, pp. 407–410, Kunming, China, July 2008.

[24] S. B. Kazmi, Qurat-ul-Ain, and M. A. Jaffar, "Wavelets-based facial expression recognition using a bank of support vector machines," *Soft Computing*, vol. 16, no. 3, pp. 369–379, 2012.

[25] N. L. A. Krisshna, V. K. Deepak, K. Manikantan, and S. Ramachandran, "Face recognition using transform domain feature extraction and PSO-based feature selection," *Applied Soft Computing*, vol. 22, pp. 141–161, 2014.

[26] G. Guo and C. R. Dyer, "Learning from examples in the small sample case: face expression recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 3, pp. 477–488, 2005.

[27] L. Zhang and D. Tjondronegoro, "Facial expression recognition using facial movement features," *IEEE Transactions on Affective Computing*, vol. 2, no. 4, pp. 219–229, 2011.

[28] A. Uçar, Y. Demir, and C. Güzeliş, "A new facial expression recognition based on curvelet transform and online sequential

extreme learning machine initialized with spherical clustering," *Neural Computing and Applications*, vol. 27, no. 1, pp. 131–142, 2016.

[29] S. Chaplot, L. M. Patnaik, and N. R. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network," *Biomedical Signal Processing and Control*, vol. 1, no. 1, pp. 86–92, 2006.

[30] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[31] J.-C. Pesquet, H. Krim, and H. Carfantan, "Time-invariant orthonormal wavelet representations," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1964–1970, 1996.