Facial Expression Recognition using 2D Stationary Wavelet Transform and Gray-Level Co-occurrence Matrix

Nikunja Bihari Kar
Department of Computer Science and Engineering
National Institute of Technology
Rourkela, Odisha - 769 008, India
nikunjakar@gmail.com

Korra Sathya Babu
Department of Computer Science and Engineering
National Institute of Technology
Rourkela, Odisha - 769 008, India
ksathyababu@nitrkl.ac.in

ABSTRACT

This paper presents an automated facial expression recognition (FER) system based on two dimensional stationary wavelet transform (2D-SWT) and gray-level co-occurrence matrix (GLCM). The proposed scheme employs 2D-SWT to decompose the image into a set of sub-bands. Then GLCM features are obtained from the 2D-SWT sub-bands. Subsequently, linear discriminant analysis (LDA) is harnessed to select the most relevant features. Finally, these features are used for classification of facial emotions using least squares variant of support vector machine (LS-SVM) with radial basis function (RBF) kernel. The performance of the pro-posed system is evaluated on two standard datasets namely, Extended Cohn-Kanade (CK+) and Japanese female facial expression (JAFFE). Experimental results based on 5-fold cross validation strategy indicate that the proposed scheme earns an accuracy of 96.72% and 99.79% over CK+ and JAFFE dataset respectively, which are superior to other competent schemes.

Keywords

Facial expression recognition; stationary wavelet transform; gray-level co-occurrence matrix.

1. INTRODUCTION

Emotion plays a significant role in real life commutation and interaction. The ability to recognize the emotional states of people surrounding us is an important part of natural communication. Affective computing is the study of systems and devices. It can translate, process, and perceive human affects. The primary issue with affective computing is to recognize emotion from facial expressions, or body gesture, or speech, or electroencephalogram (EEG). Affective computing is useful in sentiment analysis, intelligent systems, stress analysis, etc.

Facial expression analysis is the primary research area of psychology, and a large body of works have been published in the past years. However, the performance still requires potential improvement in order to meet real time scenario. In general, facial expression recognition (FER) system works in three stages: preprocessing, feature extraction and classification. Selection of proper features plays an important role in final decision making. A variety of feature extraction schemes have been used in the existing systems. Refer [1] to see a list of FER systems. The existing FER method has been used some feature extraction methods such as discrete wavelet transform (DWT) [2], local ternary pattern (LTP), local binary pattern (LBP) [3], stationary wavelet transform (SWT) [4], ripplet transform type II (ripplet-II) [5], Gabor wavelets [6], and histograms of oriented gradients (HOG) [7]. These feature extraction techniques extract features either from the whole face or some part of the face without any prior knowledge.

Happy and routray [8] developed a novel FER framework using appearance features of different facial patches. Siddiqi et al. [9] harnessed stepwise linear discriminant analysis (SWLDA) and hidden conditional random fields (HCRFs) for recognizing emotion category. In the work of Mlakar and potocinik [10] the HOG features are calculated for neutral and peak expression images. Then the SVM classifier uses the difference vector for recognition. Kar et al. [7] introduced a novel FER approach based on HOG and PCA+LDA for feature extraction and reduction. After that back propagation neural network (BPNN) is appended for expression classification.

Kazmi et al. [11] combined 3 level 2D-DWT with seven parallel SVMs to detect emotion. Siddiqi and Lee [12] employs Symlet wavelet for feature extraction. The use of LDA reduces the feature set. Then, HMM is harnessed for expression labeling. Zhang et al. [13] developed a FER sys-tem using biorthogonal wavelet entropy (BWE). A variant of SVM called Fuzzy SVM is trained for emotion classification. Wang et al. [4] harnessed an intelligent FER sys-tem based on stationary wavelet entropy (SWE) and Jaya learning algorithm. For emotion labeling, single hidden layer feed-forward neural network (SHLFNN) is used. Qayyum et al. [14] used SWT and discrete cosine transform (DCT) for features and BPNN for emotion classification. In previous existing works, it has been noticed that the wavelet trans-form (WT) technique is an ultimate and widespread feature extraction technique for systems [4][12][13][14]. But, due to some shortcoming of non-supportiveness to anisotropy and restricted directionality, the WT technique cannot extract the intrinsic and subtle features in the face expression images, which are required for detecting emotions.

From the literature mentioned above, a few strategies utilized DWT as a feature extraction tool, which is translation-variant. SWT can deal with the shift invariance problem. Hence, we conveyed 2D-SWT as feature extraction technique which is translation-invariant. Gray-level co-occurrence matrix (GLCM) is captured for all 2D-SWT high frequency coefficients, which decomposed upto 3-levels. GLCM is able to capture the texture features but is not directly usable. Hence, we further extracted 4 statistical features namely entropy, energy, autocorrelation, correlation from GLCM. Then, we applied LDA method to extract most discriminant features. Further, SVM is used in most cases for recognition which needs more computational overhead. Hence, a computational efficient variant of SVM called least-squares support vector machine (LS-SVM) is used for classification in this study.

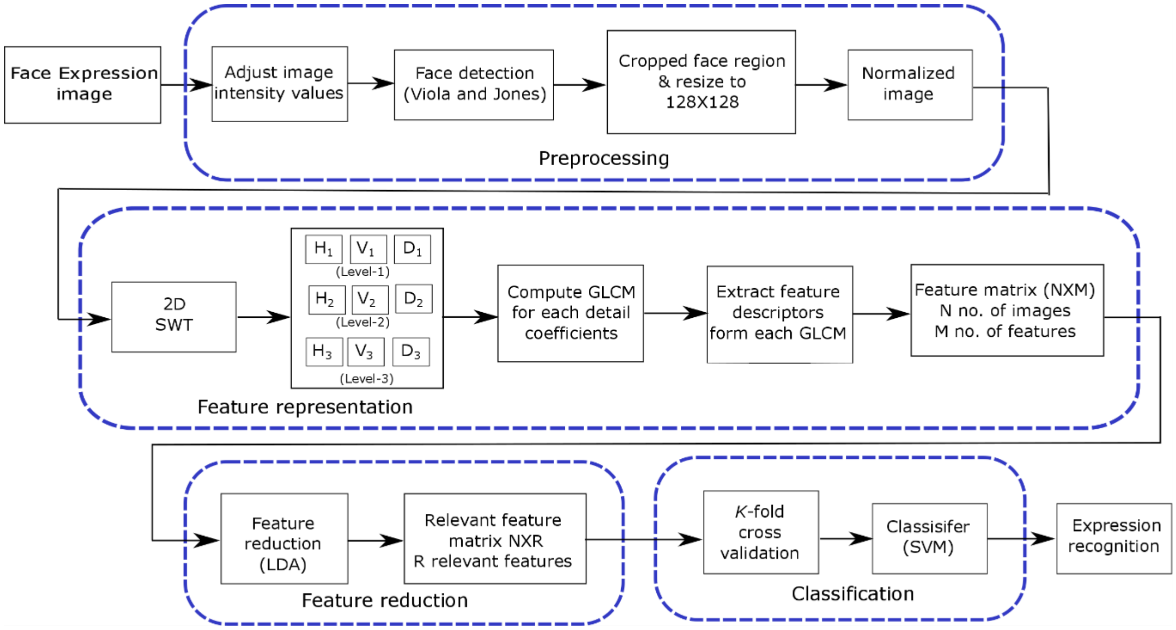


Figure 2: Architecture of the proposed scheme for recognizing emotion from expression images

The remaining structure of the paper is organized in following way. In Section 2, we defined the datasets. Section 3 introduced the working procedure of each step of the pro-posed methodology. Section 4 presents the results and discussions. Finally, Section 5 concluded the paper.

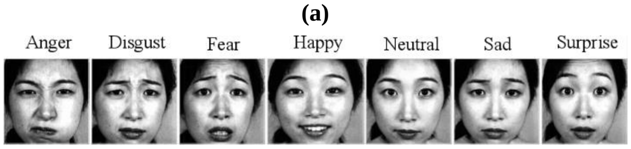
2. DATASET

This study employed, CK+ [15] and JAFFE [16] datasets to validate the proposed method. The number of images retained for our experiment is listed in Table 1. The CK+ dataset consists of 593 image sequences from 123 subjects. Among these 327 were labeled with seven basic emotion categories. JAFFE dataset comprises of 213 expression images from 10 individuals. Each of them has seven emotion expressions. In our experiment, we have used 213 images and 327 images from JAFFE and CK+ datasets respectively. The preprocessed sample images for the dataset of the experiment are shown in Figure 1.

Table 1: Images from each emotions for JAFFE and CK+ dataset of the experiment

	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral	Contempt	All
JAFFE	30	29	32	31	31	30	30	-	213
CK+	45	59	25	69	28	83	-	18	327

(a)



(b)

Figure 1: Some Normalized sample images from (a) CK+ and (b) JAFFE dataset of the experiment

3. PROPOSED METHODOLOGY

The proposed methodology follows in four (4) main phases such as (i) pre-processing of facial expression images, (ii) feature extraction, (iii) feature dimension and reduction, and (iv) classification phase. Figure 2 shows the block diagram of the proposed methodology and each block description of the proposed method is reported in next section.

3.1 Preprocessing

Initially, the expression images are converted to gray scale images. Then the contrast of the image increased by eliminating the 1% top and bottom of all pixel values. After that, the faces are detected by Viola and Jones face detection algorithm [17]. Afterward, the identified face region is cropped and normalized to the size of 128 × 128.

3.2 Feature extraction

In this study, we have extracted the SWT-based texture features from the high frequency sub-bands. A brief description of SWT and GLCM features are given below.

3.2.1 2D SWT

SWT is a variant of DWT designed to overcome the deficiency of translation invariant property of DWT. To achieve translation-invariant property the up-samplers and down-samplers are removed from DWT coefficients. Then, the filter coefficients are up-sampled by a factor of of $2(j-1)$ in the j^{th} level. Originally, the SWT is a redundant method. The dimension of the SWT coefficients at each level are same as the input sample. In our study, SWT with three level Haar wavelet is utilized to decompose the face image. The high-frequency SWT sub-bands are retained to analyze the edge and texture features of the image. The approximation sub-bands of SWT carries little energy for which it is not considered for texture analysis. The results of 2D-SWT decomposition up to three-level is shown in Figure 3, in which 9 SWT sub-bands are stored as D1, D2, D3, H1, H2, H3, V1, V2, and V3.

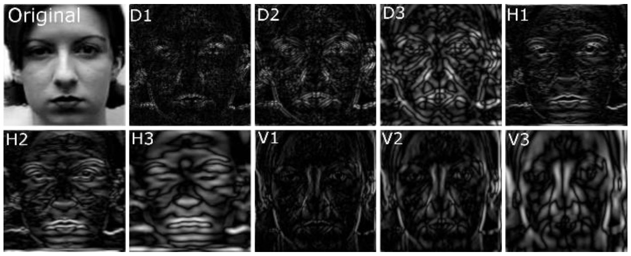


Figure 3: 3-level SWT decomposition of a neutral face image indicating all high frequency components

3.2.2 GLCM

GLCM is a matrix of frequencies, at which the two pixels separated by a certain distance occurred in an image. GLCM is computed by analyzing the spatial relationship between the two-pixel values, i.e., one is the reference pixel, and another one is a neighbor pixel.

Let $r(x, y)$ be the element of GLCM of the given image I of size $N * N$. The gray values G contains in the image I range from 0 to $G-1$. The texture-context information provided in the matrix r, is calculated by measuring the frequency of the two neighboring pixels at specific distance d in the image I, one with gray level x and other with gray level y. GLCM matrix [18] can be defined as follows,

$$r(x,y)=\sum_{p=1}^{N}\sum_{q=1}^{N}\begin{cases}1 & \text{if } I(p,q)=x \text{ and } I(p+\Delta p, q+\Delta q)=y\\ 0 & \text{Otherwise}\end{cases}$$

Where $r(x, y)$ is the reference pixel, and $r(x, y)$ is the neighbor pixel. The size of the GLCM is same as the number of gray levels in the input image. GLCM is computed based on two parameters d and θ. Where d is the relative distance between neighboring pixels whose value is limited to integral multiples of the pixel size and θ is the relative orientation.

Formally, the orientation values are 0^0, 45^0, 90^0, and 135^0. Here, the offset values are [0, 1], [1, 0], [1, 1], and [1, 1]. The offset values represent distance D=1 with four neighboring pixel orientation i.e. $\theta = 0^0, 45^0, 90^0$, and 135^0 with respect to the reference pixel. We have used four textural features in our work. Let $r(x, y)$ be the $(x, y)^{th}$ entry in the normalized GLCM. The description of these four features are described in Table 2.

Table 2: Textural features

Name	Definition
Energy	$f1=\sum_x\sum_y r(x,y)^2$
Entropy	$f2=-\sum_x\sum_y r(x,y)log(r(x,y))$
Correlation	$f3=\left(\sum_x\sum_y(x,y)r(x,y)-\mu_i\mu_j\right)/(\sigma_i\sigma_j)$
Autocorrelation	$f4=\sum_x\sum_y(xy)r(x,y)$

Where μ_i, μ_j and σ_i, σ_j are the mean and standard deviation of r_x and r_y.

For all extracted 9 SWT coefficients the GLCM is computed for four relative orientations i.e. $\theta = 0^0, 45^0, 90^0$, and 135^0. Then four textural features are calculated from the normalized GLCM. Hence, a total of 9 (SWT sub-bands) × 4 (Orientation) × 4 (Textural features) = 144 features are extracted.

3.3 Feature Reduction

In our study, linear discriminant analysis (LDA) approach [19] is adopted to enhance the class separability of all the expression images for recognition purpose. It finds the most discriminant projection vectors which map high dimensional feature space into low dimensional feature space. LDA projection vectors help all projected samples to form minimum within-class scatter and maximum between class scatter. In this work, the LDA is used to reduce the 144 number of features to just 6 for CK+ and JAFFE dataset. These prominent features selected by LDA improved accuracy and computation speed and also incurred less computational complexity. These are the initial principal components (PCs) with maximum variance.

3.4 Classification

After feature reduction, the feature matrix is fed to the classifier for prediction of emotion. The conventional SVM does not perform well on validating more massive datasets, and the computational overhead is also more. Hence, an efficient variant of SVM named as LS-SVM [20] used in our scheme to improve accuracy and computational complexity. LS-SVM is a supervised machine learning approach used to classify samples of two or more classes using linear or non-linear hyperplanes. Various experiments carried out using LS-SVM with kernels such as RBF, linear, and polynomial. Experimental results indicate that LS-SVM with RBF kernel outperforms other kernels.

Let there be N samples$\{p_i, q_i\}_{i=1}^{N}$. Where, $p_i \in R^n$ is the i^{th} input data and $q_i \in R$ is the i^{th} output label. The LS-SVM classifier decision function can be obtained as,

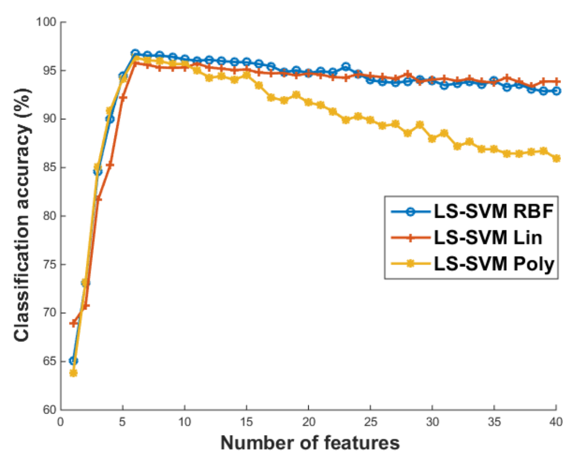
$$f(x) = sign\left[\sum_{i=1}^{N}\alpha_i y_i \kappa(x, x_i) + b\right]$$

Where, κ (.,.) is the kernel function, and α_i is the Lagrange multiplier.

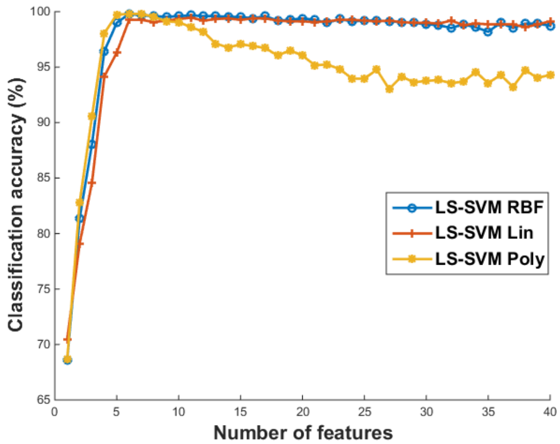
4. RESULTS AND DISCUSSION

The experiments were observed on a 3.4 GHz Core i7 processor and 4GB RAM PC running under Windows OS framework. Matlab tool is used to simulate the proposed system.

The results obtained from LDA on both the datasets are shown in Figure 4. It is observed that the highest accuracy has been achieved by using only six PCs because it preserves maximum variance. Hence for performance comparison, we have used only six features in our experiment.



(a)



(b)

Figure 4: Performance results of (a) CK+ and (b) JAFFE dataset in terms of number of features

Table 3: Classification results obtained by LS-SVM RBF on CK+ 7 dataset

Emotions	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	Accuracy (%)
Anger	44	263	19	1	98.00	93.00	93.88
Contempt	17	306	3	1	94.00	99.00	98.78
Disgust	56	254	14	3	95.00	95.00	94.80
Fear	23	300	2	2	92.00	99.00	98.78
Happy	63	246	12	6	91.00	95.00	94.50
Sad	27	292	7	1	96.00	98.00	97.55
Surprise	81	233	11	2	98.00	95.00	96.02
Average					96.00	97.00	96.72

Table 4: Classification results obtained by LS-SVM RBF on JAFFE 7 dataset

Emotions	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	Accuracy (%)
Anger	30	182	1	0	100.00	99.00	99.53
Disgust	29	184	0	0	100.00	100.00	100.00
Fear	32	181	0	0	100.00	100.00	100.00
Happy	31	181	1	0	100.00	99.00	99.53
Neutral	30	182	1	0	100.00	99.00	99.53
Sad	31	182	0	0	100.00	100.00	100.00
Surprise	30	183	0	0	100.00	100.00	100.00
Average					100.00	100.00	99.80

Table 5: Results of classification by different classifier (CK+ 7 dataset)

Classifiers	Sensitivity (%)	Specificity (%)	Accuracy (%)
KNN	86.90	96.90	86.85
RF	89.90	97.60	89.90
BPNN	92.40	98.30	92.35
LS-SVM Linear	94.60	95.95	95.76
LS-SVM Poly	94.92	96.41	96.33
LS-SVM RBF	96.07	96.78	96.72

Table 6: Results of classification by different classifier (JAFFE 7 dataset)

Classifiers	Sensitivity (%)	Specificity (%)	Accuracy (%)
KNN	99.10	99.80	99.06
RF	98.10	99.70	98.12
BPNN	99.50	99.90	99.53
LS-SVM Linear	99.54	99.22	99.26
LS-SVM Poly	100.00	99.77	99.79
LS-SVM RBF	100.00	99.77	99.79

The classification results of LS-SVM RBF on CK+ and JAFFE datasets are presented in Tables 3 and 4. To demonstrate the performance of the LS-SVM classifier along other classifiers like KNN, RF, and BPNN have also been implemented. All the classifiers are trained and tested with the same set of data obtained from LDA. The classifiers use 5-fold stratified cross-validation (SCV) to make them more generalize and stable to independent datasets. Parameters for different classifiers have been tuned experimentally. Finally, the parameters with least SCV error are chosen.

Experimental results showed in Tables 5 and 6, indicate that the performance of LS-SVM with RBF kernel obtained higher accuracy than other classifiers on both the datasets. Therefore we select LS-SVM RBF as the classifier for this study. Finally, we compare our proposed approach with other seven state-of-the-art methods on both CK+ and JAFFE datasets. It is observed from Table 7 that our method achieves highest classification accuracy on JAFFE dataset and achieves second best result on CK+ dataset.

5. CONCLUSION

In this study, we have developed an improved FER system which combines SWT and GLCM to derive significant features. SWT is used to decompose the preprocessed image into number of sub-bands. Then, the features are derived by calculating the GCLM features over the SWT sub-bands. In order to derive discriminant features, we have applied LDA. Finally, the resultant features are subjected to LS-SVM classifier for detecting emotions. The experimental results on benchmark datasets show that the results are promising compared to the state-of-the-art methods.

The performance of the proposed system can be further increased by applying other powerful feature extraction and machine learning schemes. In future, we plan to investigate the application of deep learning algorithms for emotion recognition.

6. REFERENCES

[1] Maja Pantic and Leon JM Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

[2] Shi Dongcheng and Jiang Jieqing. The method of acial expression recognition based on DWT-PCA/LDA. In *International Congress on Image and Signal Processing (CISP)*, volume 4, pages 1970–1974. IEEE, 2010.

[3] Tommaso Gritti, Caifeng Shan, Vincent Jeanne, and Ralph Braspenning. Local features based facial expression recognition with face registration errors. In *8th IEEE International Conference on Automatic Face & Gesture Recognition, FG'08*., pages 1–8. IEEE, 2008.

[4] Shui-Hua Wang, Preetha Phillips, Zheng-Chao Dong, and Yu-Dong Zhang. Intelligent facial emotion recognition based on stationary wavelet entropy and jaya algorithm. *Neurocomputing*, 272:668–676, 2018.

[5] Nikunja Bihari Kar, Korra Sathya Babu, Arun Kumar Sangaiah, and Sambit Bakshi. Face expression recognition system based on ripplet transform type ii and least square svm. *Multimedia Tools and Applications*, pages 1–24, 2017.

[6] Marian Stewart Bartlett, Gwen Littlewort, Mark Frank, Claudia Lainscsek, Ian Fasel, and Javier Movellan. Fully

Table 7: Comparison of proposed system with existing schemes on CK+ and JAFFE database

References	Facial feature	Classifier	Accuracy (%) CK+	Accuracy (%) JAFFE
Zhang *et al.* [21] ′11	Patch based Gabor	SVM	94.48	99.23
Wang *et al.* [22] ′13	HOG+WLD	KNN	93.97	95.86
Mlakar and potocinik [10] ′15	HOG difference vector	SVM	95.64	87.82
Happy *et al.* [8] ′15	Salient facial patches	RBF SVM	94.09	91.79
Siddiqi*et al.* [9] ′15	SWLDA	HCRF	96.83	96.33
Uccar *et al.* [23] ′16	Local Curvlet transform	OSELM-SC	95.17	94.65
Qayyum *et al.* [14] ′17	SWT+DCT	BPNN	96.61	98.83
Proposed method	**SWT+GLCM+LDA**	**LS-SVM RBF**	**96.72**	**99.79**

automatic facial action recognition in spontaneous behavior. In *7th International Conference on Automatic Face and Gesture Recognition*, pages 223–230. IEEE, 2006.

[7] Nikunja Bihari Kar, Korra Sathya Babu, and Sanjay Kumar Jena. Face expression recognition using histograms of oriented gradients with reduced features. In *Proceedings of International Conference on Computer Vision and Image Processing*, pages 209–219. Springer, 2017.

[8] SL Happy and Aurobinda Routray. Automatic facial expression recognition using features of salient facial patches. *IEEE transactions on Affective Computing*, 6(1):1–12, 2015.

[9] Muhammad Hameed Siddiqi, Rahman Ali, Adil Mehmood Khan, Young-Tack Park, and Sungyoung Lee. Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. *IEEE Transactions on Image Processing*, 24(4):1386–1398, 2015.

[10] Urosˇ Mlakar and Boˇzidar Potoˇcnik. Automated facial expression recognition based on histograms of oriented gradient feature vector differences. *Signal, Image and Video Processing*, pages 1–9, 2015.

[11] Sidra Batool Kazmi, M Arfan Jaffar, et al. Wavelets-based facial expression recognition using a bank of support vector machines. *Soft Computing*, 16(3):369–379, 2012.

[12] Muhammad Hameed Siddiqi and Sungyoung Lee. Human facial expression recognition using wavelet transform and hidden markov model. In *International Workshop on Ambient Assisted Living*, pages 112–119. Springer, 2013.

[13] Yu-Dong Zhang, Zhang-Jing Yang, Hui-Min Lu, Xing-Xing Zhou, Preetha Phillips, Qing-Ming Liu, and Shui-Hua Wang. Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access*, 4:8375–8385, 2016.

[14] Huma Qayyum, Muhammad Majid, Syed Muhammad Anwar, and Bilal Khan. Facial expression recognition using stationary wavelet transform features. *Mathematical Problems in Engineering*, 2017.

[15] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The Extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 94–101. IEEE, 2010.

[16] Michael Lyons, Shota Akamatsu, Miyuki Kamachi, and Jiro Gyoba. Coding facial expressions with Gabor wavelets. In *Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205. IEEE, 1998.

[17] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, volume 1, pages I–511. IEEE, 2001.

[18] L-K Soh and Costas Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on geoscience and remote sensing*, 37(2):780–795, 1999.

[19] Hua Yu and Jie Yang. A direct LDA algorithm for high-dimensional datawith application to face recognition. *Pattern recognition*, 34(10):2067–2070, 2001.

[20] J.A.K. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999.

[21] Ligang Zhang and Dian Tjondronegoro. Facial expression recognition using facial movement features. *IEEE Transactions on Affective Computing*, 2(4):219–229, 2011.

[22] Xiaohua Wang, Chao Jin, Wei Liu, Min Hu, Liangfeng Xu, and Fuji Ren. Feature fusion of HOG and WLD for facial expression recognition. In *International Symposium on System Integration (SII)*, pages 227–232. IEEE, 2013.

[23] Ay¸segu¨l U¸car, Yakup Demir, and C¨uneyt Gu¨zeli¸s. A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering. *Neural Computing and Applications*, 27(1):131–142, 2016.