I.J. Image, Graphics and Signal Processing, **2021, 1, 50-57**
Published Online February 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2021.01.05

Emotion Recognition from Faces Using Effective Features Extraction Method

Htwe Pa Pa Win
University of Computer Studies, Hpa-an
Email: hppwucsy@gmail.com

Phyo Thu Thu Khine
University of Computer Studies, Hpa-an
Email: phyothuthukhine@gmail.com

Zon Nyein Nway
University of Computer Studies, Yangon
Email: zonnyeinnway@ucsy.edu.mm

Received: 19 July 2020; Accepted: 28 October 2020; Published: 08 February 2021

Abstract: With the rapid development and requirement of application with Artificial Intelligent (AI) technologies, the researches related to human-computer interaction are always active and the emotional status of the users is very essential for most of the environment. Facial Emotion Recognition, FER is one of the important visual information providers for the AI systems. This paper proposes a FER system using an effective feature extraction methodology and classification technologies. Local features of the face are more effective features for recognition and Scale Invariant Feature Transform, SIFT can give a better representation of the face. The bag of the visual word (BOVW) is the good encoding method and the advancement of that model Vector of Locally Aggregate Descriptor, VLAD provides the better encoder for SIFT features and used these benefits for feature extraction environments. The power of SVM includes unknown class recognition problems and this advantage is used for classification. This system used the standard basement JAFEE dataset to measure the success of the proposed methods and prepared to compare with other systems. The proposed system achieves the better result when it compared with some of the other previous systems because of the combination of effective feature extraction and encoding method.

Index Terms: Artificial Intelligent (AI), BOVW, Facial Emotion Recognition (FER), JAFEE, Local Features, SIFT, SVM, VLAD

1. Introduction

Some of the ability of intelligent systems needs to detect and recognize people's emotion as that skill is the essential requirements for communications and interactions on a daily basis of human. The importance of emotion in the systems can be seen in many domains; the emotional status of patients for the health domain, the customer satisfaction for marketing environments, the emotional decision making of the pilot, and the teachers' and students' relation in the E-learning system. AI technologies are trying to participate in the described environments in many ways and some include robots. The status of human emotion can find in sound and visual effects and facial expression is the most important one. Facial expression recognition, FER systems are the basic part to find the human feeling in an evolutionary approach [1-3].

The Facial Emotion Recognition needs to classify the basic facial expressions of six groups, namely, Anger, Disgust, Fear, Sadness, Happiness, and Surprise and the Neutral can be added as the seventh emotion. The main parts of the emotion recognition system from face include feature extraction and classification [4].

Facial Emotion Recognition (FER) system is one of the reorganization processes of the emotional state of the personality of the individual people based on the facial image captured with different devices and with different resolutions. Since the advancement of the technologies, the high-quality devices capture the face images with colors and that images can produce better accuracy rates. However, various techniques have been proposed by different researchers to enhance the recognition rate by finding the appropriate feature descriptor. Moreover, various issues such as the face length, eyebrows' angles, lips' angles, forehead curves, and different resolutions face images remain as the challenging problems for FER systems. Especially facial expression recognition in the low-quality image requires to be emphasized

as these mechanisms can be applied for various environments of the facial emotion recognition system. Therefore, this paper emphasizes the low-quality image of a standard face dataset called JAFEE [5,6].

Local invariant features of the image have commonly been used in face recognition and other computer vision systems. The features need to be unchangeable to image rotation, scaling, illumination, and viewpoint changes. The descriptors to construct features need to be extremely distinctive and resistant for any transformations and moreover to be convenient for matching or classification. The SIFT descriptor may be the best appealing technique for practical application in today's world [7]. The VLAD, Vector of Locally Aggregate Descriptor encoding technique has achieved great success in several computer vision tasks. It can be combined with SIFT features and suitable for scene classification and recognition process [8]. The achievement of SVM can be seen in many recognition process and image processing systems. Therefore, this paper proposes the emotion recognition from the face using SVM and the effective combination of SIFT and VLAD encoding techniques. The rest of the paper is presented with five sections. Section 2 discusses the previous works, and Section 3 proposes the emotion recognition system and experiment evolution is done in Section 4. The recognition system is finished in Section 5 as the conclusion section.

2. Related Works

Although there has been a lot of works for facial emotion recognition system, some of the related works that used JAFFE for testing and producing the results have been analyzed.

The authors in [1] presented a facial expression recognition system for emotion recognition applying Deep Learning Neural Network. The face images are pre-processed by using the HOG algorithm for the purpose of finding the face boundary and determine with SVM to classify these HOG features. The face area is cropped based on these coordinates and blurred to remove noise before producing local binary pattern (LPB) images due to improving the Action Unit values (AUs). They extract holistic features from LBP images and landmarks information. Emotions are classified using appearance feature-based CNN. They used CK+, JAFFE dataset for their evolution, perform standard cross-validation techniques, and achieve 91.27% in JAFFE, and more results are achieved in CK+ as it is the color image sets.

The groups of [3] performed the prediction system for customer satisfaction from facial emotional expression. They used the JAFFE dataset to experiment the effectiveness of their system and used the SVM classifier for the categorization of the emotion. They used the landmark points with the geometric features of the face as the representation vectors. They obtained these features by converting the face image data into geometric primitive's data such as mouth, eyes, noses and chin, the relative position of them, width and distance between them. They classified just only 3 groups for satisfaction, non-satisfaction, and neutral and achieve 0.92 for the ROC curve.

The researchers in [9] worked for a facial emotion recognition system using deep neural networks. Their approach based on a convolutional network where attention is focused on the rich feature of the face parts to reduce the network layers, to be less than 10 layers, instead of using deeper networks. They apply a visualization technique to highlight the most salient regions of the face image to improve the classifier's outcome. They used SIFT, HOG, and LBP as the features for the system and FER-2013, CK+, JAFEE, and FERG are used as the datasets. They achieved 92.8% for the JAFEE dataset by testing 70 images.

The analyzer in [10] presented a facial expression recognition model using Gabor filter and LGC-HD, local gradient code horizontal diagonal features to generate descriptors. The histogram equalization technique, Viola-Jones algorithm, Haar filtering method, and Adboost classification mechanisms are applied at the preprocessing steps to detect face and normalize the face to the standard size of 112x96 pixels. The chi-square distance is used to classify the descriptors. They described eyebrow, eye and mouth have higher weights than the other facial regions. They used JAFEE and CK datasets to test the system and achieve a 93.33% recognition rate. They reduced computational time and improved recognition rate by using the weighted mask technique in Gabor's image.

The research groups of [11] presented a human emotion recognition system from the face for IoT devices using CNN. They emphasize on sub-network of deep learning and name their system "HERO". They combine the advantages of ensemble learning methods based on CNNs, by integrating the weak classifiers to get a new strong classifier instead of using normal features extraction procedures. They created 8 images using translation and flip operation from the single input test image to enhance the accuracy rate. They used local features of eye and mouth and test their system with JAFFE, AffectNet, and FER2013 datasets. They achieve 96.44% for JAFEE dataset for 42 test images. They achieve high accuracy results on happy, surprise, neutral, and disgust and achieve relatively low results on sad, fear, and angry as these expressions confuse with others.

3. Proposed System Design

The proposed framework consists of three parts: preprocessing, feature extraction, and classification. The detailed design of proposed system is described in Fig. 1.

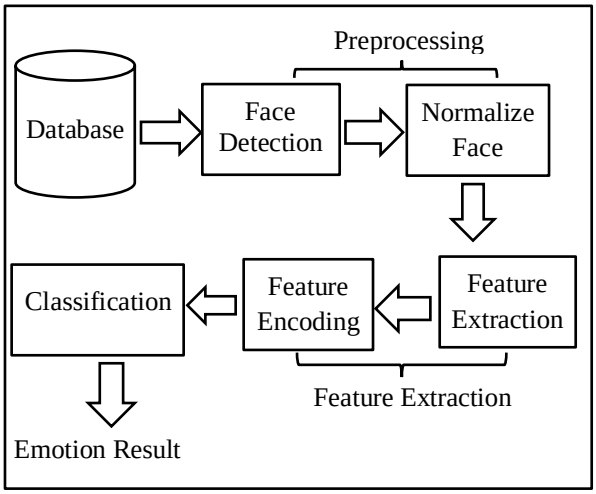


Fig. 1. Proposed Face Emotion Recogntion System

A. Preprocessing

This process includes face detection and normalization. The images in original database are sized by 256x256 pixels and consist of unnecessary parts. The main parts of face portion are eyes, eyebrows, noses and mouth. Therefore, the input images are automatically detected for face portion using the classic Viola-Jones face detector [12] and normalize into 128x128 pixels size. The filtering mechanism such as blurring is tried for the main face porting but it is not suitable for SIFT feature nature because original image is low resolution. Therefore, the original cropped face porting is delivered to the feature extraction process.

B. Feature Extraction

When the normalized images are gained, the features can then be extracted to represent the information as the emotion. The represented information is extracted using the SIFT feature. And the encoding procedure used the VLAD method.

- **Scale Invariant Feature Transform**

Scale Invariant Feature Transform, SIFT, can be seen as one of the most robust methods in computer vision. It is a method of detecting and extracting the local feature descriptors from the images that can never change even there may be changes in rotation, illumination, noise, and scaling. Moreover, it can detect the points that can't alter in the little changes of viewpoint. The SIFT algorithm consists of the following steps.

Step1: Scale-space Extrema Detection

The SIFT operator finds the most stable IPs of an input image using a Difference of Gaussians, "DoG" in scale space. DOG is calculated from the two adjacent images in the same octave using the following equation:

$$D(x,y,\sigma) = \big(G(x,y,k\sigma) - G(x,y,\sigma)\big) \otimes I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma) \quad (1)$$

SIFT algorithms remove the low contrast Interest-point, IPs, and bad localization on edges to improve the IPs detection capability.

Step2: Orientation Assignment

The different orientations are allocated on IPs to get the invariance of rotation. The gradient magnitude and the orientation of each IP is calculated as

$$m(x,y) = \sqrt{\big(L(x+1,y) - L(x-1,y)\big)^2} \quad (2)$$

$$\theta(x,y) = \arctan\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \quad (3)$$

These orientations build a 36 bins histogram.

Step3: Keypoint Description

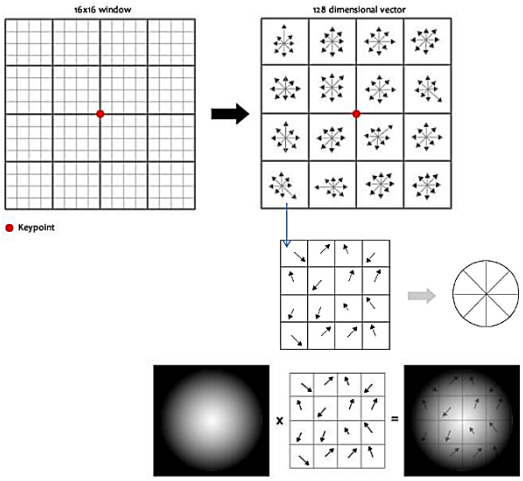
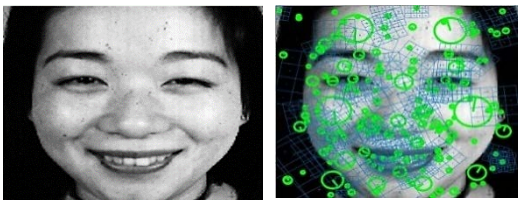


Fig. 2. Process of SIFT feature extraction

Finally, a 128 bins descriptor defines a selected IP. The key point descriptors generating is displayed in Fig. 2 [13-15]. The sample output of the SIFT features of the used dataset is depicted in Fig. 3.



(a) (b)

Fig. 3. Normalized image and SIFT descriptor image

- **Vector of Locally Aggregated Descriptors**

Vector of Locally Aggregated Descriptors, VLAD encoding is the representation of a super-efficient vector encoding method created by [16]. VLAD is the extension of the bag of words, BOW model and it is more computationally efficient than the BOW model. It encodes the features into the high dimensional vector using the clustering centroids of the k-means algorithm. To find the centroids, the VLAD feature is obtained as o residuals by using the following equation.

$$v_k = \sum_{i=1}^{N} \alpha_{ik}(x_i - u_k) \tag{4}$$

Where $\{x_i\}$ is set of image features, α_{ik} is the association of data x_i to u_k , $\alpha_{ik} \geq 0$ and $\sum_k^K \alpha_{ik} = 1$.
For hard association, the nearest neighbor is found by using the following equation.

$$\alpha_{ik} = \begin{cases} 1 \; if \; || \; x_i - \mu_k || \; 2 \leq || \; x_i - \mu_l || \; 2 \; \forall l \neq k \\ 0 \; otherwise. \end{cases} \tag{5}$$

Then, the VLAD feature of the image I is stacked as

$$\Phi(I) = [\dots v_k^T \dots]^T \tag{6}$$

It is used as a featurization technique to aggregate the transformation of SIFT features into a fixed-sized representation of a vector. The idea is how to involve the assignments of the feature descriptors into the nearest group of the dictionary. The workflow of the VLAD is shown in Fig. 4 [17, 18]. This system used this encoding mechanism as the supplement to improve classification performance and reduce the complexity.

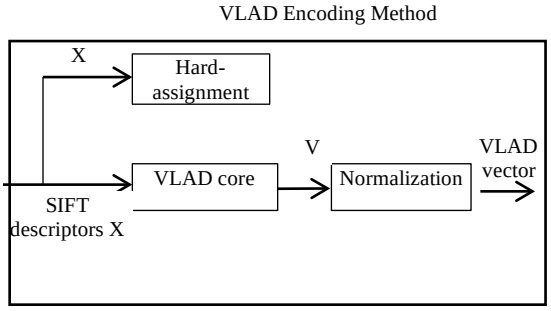


Fig. 4. The process of VLAD

C. Classification

The classification process is done using the Support Vector Machine, SVM. SVM is used to detect emotions from VLADs by using "one against one" implementation to classify multiclass to emotion stages. It constructs a hyperplane between the spaces to separate different emotion classes. The optimal hyperplane is generated iteratively to minimize the error and produce the maximum hyperplane to divide the VLAD vectors into equivalent classes.

4. Experiments

A. Dataset preparation

The Japanese Female Facial Expression, JAFFE dataset is used to test the emotion recognition. It was published from Kyushu University [19], it is the standard facial expression dataset. The most previous works have experimented on this dataset. It contains total number of 213 facial grayscaled images with seven emotions: angry, fear, disgust, happy, neutral, sad, surprise. The images are taken from the posting of 10 Japanese females and fully labeled. The images are 256x256 pixels in each. The test set is randomly extracted 20% from the original and the rest 80% is for the training set. Therefore, the training set consists of 171 images and the different 42 images are in the test set.

B. Results

The experiment is started with the extraction of SIFT features after the detection of faces process from the database and then encoding with the VLAD method. As the technique of encoding method is based on the number of clusters and experiments are carried with different cluster numbers. The different confusion matrix for each emotion from different cluster numbers 64, 96, and 128 are described in fig 5, fig. 6, and fig. 7 respectively.

Emotion	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Disgust	0.00	0.67	0.33	0.00	0.00	0.00	0.00
Fear	0.00	0.00	0.83	0.00	0.00	0.00	0.17
Happy	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Neutral	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Sad	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Surprise	0.00	0.00	0.17	0.00	0.00	0.00	0.83

>>

Fig. 5. Confusion Matrix for each emotion with cluster number 64

Emotion	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	0.83	0.00	0.00	0.00	0.00	0.17	0.00
Disgust	0.00	0.67	0.17	0.00	0.00	0.17	0.00
Fear	0.00	0.00	0.83	0.00	0.00	0.00	0.17
Happy	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Neutral	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Sad	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Surprise	0.00	0.00	0.00	0.00	0.00	0.00	1.00

>>

Fig. 6. Confusion Matrix for each emotion with cluster number 96

Emotion	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Disgust	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Fear	0.00	0.17	0.83	0.00	0.00	0.00	0.00
Happy	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Neutral	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Sad	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Surprise	0.00	0.00	0.00	0.00	0.00	0.00	1.00

>>

Fig. 7. Confusion Matrix for each emotion with cluster number 128

As can be seen from the confusion matrix the proposed feature extraction method is least efficient for 'Fear' emotion, however, it is the stable method for various cluster number of the encoding method, as the percentage of 'Disgust' and 'Surprise' emotion are changed and improved by the number of clusters. However, the result on the 'Angry' feature is strange as when the cluster is increased from 64 to 96, the accuracy is decreased by 0.17%. But when the cluster size is increased to 128, all the accuracy for emotion is increased. The SIFT feature is needed to enhance to recognize the "Fear" emotion correctly as its rate is stable even the number of clusters are changed. This may also be due to the nature of expression of the people in database for that expression and the use of face detector. The Viola-Jones algorithm detect some fear face not including the chin even this is not the main part of the face, it may reduce the classification rate and the better face detector techniques may be required. The average accuracy is calculated and described in Fig 8 to be seen clearly how n the numbers of clusters affect the performance.

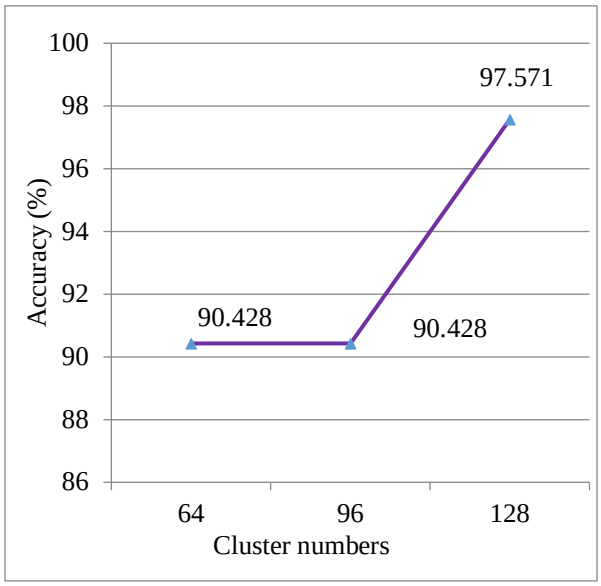


Fig. 8. Comparison of average accuracy on number of clusters

Then the performance is needed to compare with the other previous works with the same dataset and environments. Although there are many research works, some of them are not clearly described the testing environments and some others are not the same analysis for the proposed system. Therefore, the same testing experiments of the results is compared in Table 1. The same analysis mean they used 80% of the database and the remaining is used for testing and choose randomly from the original dataset. The detail of the features and classifiers they used are also described with the accuracy results in the table.

Table 1. Comparison for the proposed system with other previous system

Works	Classifier	Features	Accuracy (%)
Gabor [8], 2016	Template Matching	Gabor+LGC-HD	93.30
HERO [9], 2019	CNNs	Local features	96.44
Proposed, 2020	SVM	SIFT+VLAD (128 cluster)	97.57

5. Conclusion

This paper proposed a combination of effective feature extraction methods for the Facial Emotion Recognition system. The local features are the most effective features and most used features for facial expression and the advancement of the feature extraction methodology based on these methods are always emerged. The simple base method of SIFT features descriptors are used and the effective encoding method of the VLAD technique improves the SIFT feature descriptors. The power of SVM supporting for recognition of unknown datasets improves the facial emotion system performance. Therefore, the system outperforms some of the other previous systems described in the experiment section. The recommendation can be stated that the use of effective filtering mechanism and face detection technique may improve the accuracy rate. Although the extraction of separate SIFT features from the main parts of the face such as eyes, eyebrows, nose and mouth cannot improve the accuracy rate, other feature extraction techniques from separate parts may improve the accuracy. However, the improvement is always necessary to be better that can be deployed in a real-world environment for all faces that are not in the dataset and for the cross-matching environment. These works may be also the future works for all active learners.

References

[1] Kim, J., Kim, B, Roy, P., P., and Jeong, D., "Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure", IEEE Access. DOI:10.1109/ACCESS.2019.2907327

[2] Balasubramanian, B., Diwan, P., Nadar, R., and Bhatia, A., "Analysis of Facial Emotion Recognition", Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019), IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8

[3] Bouzakraoui, M. S., Sadiq, A. and Ala, A., Y., "Appreciation of Customer Satisfaction through analysis Facial Expressions and Emotions Recognition",

[4] Taha, B., and Hatzinakos, D., "Emotion Recognition from 2D Facial Expressions", IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 2019. DOI: 10.1109/CCECE.2019.8861751

[5] Verma, G., Verma, H. Hybrid-Deep Learning Model for Emotion Recognition Using Facial Expressions. Rev Socionetwork Strat (2020). https://doi.org/10.1007/s12626-020-00061-6

[6] Kumari N., Bhatia R. (2020) Comparative Study and Analysis of Various Facial Emotion Recognition Techniques. In: Kapur P., Singh G., Klochkov Y., Kumar U. (eds) Decision Analytics Applications in Industry. Asset Analytics (Performance and Safety Management). Springer, Singapore. https://doi.org/10.1007/978-981-15-3643-4_11

[7] Yang, H., Cheng, G., and Chen, H., "High Efficient Local Feature Matching", 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC 2018).DOI**:** 10.1109/IMCEC.2018.8469593

[8] Li, X., Li, X., Chen, X., Jian, Z. and Zheng, L., "Towards Optimal VLAD for Visual Recognition", IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS), 2019. DOI: 10.1109/CCIS48116.2019.9073752

[9] Minaee, S., and Abdolrashidi, A., "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network", Computer Vision and Pattern Recognition (cs.CV), Mon, 4 Feb 2019. DOI: https://arxiv.org/abs/1902.01019v1

[10] S. A. M. Al-Sumaidaee, S. S. Dlay, W. L. Woo and J. A. Chambers, "Facial Expression Recognition using Local Gabor Gradient Code-Horizontal Diagonal Descriptor", 2nd IET International Conference on Intelligent Signal Processing 2015 (ISP), 1-2 Dec. 2015. DOI**:** 10.1049/cp.2015.1766

[11] Hua,W.; Dai, F.; Huang, L.; Xiong, J.; Gui, G. HERO: Human emotions recognition for realizing intelligent Internet of Things. IEEE Access 2019, 7, 24321–24332. DOI: 10.1109/ACCESS.2019.2900231

[12] Viola, P.; Jones, M.J. Robust real-time face detection. Int. J. Comput. Vis. 2004, 57, 137–154.:VISI.0000013087.49260.fb.

[13] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60(2), pp. 91-110, 2004. DOI: https://doi.org/10.1023/B:VISI.0000029664.99615.94

[14] Karami, E., Prasad, S., and Shehata, M., "Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images", In Proceedings of the 2015 Newfoundland Electrical and Computer Engineering Conference,St. johns, Canada, November, 2015, Computer Vision and Pattern Recognition (cs.CV).DOI: https://arxiv.org/abs/1710.02726v1

[15] Wei, M., and Xiwei, P., "WLIB-SIFT: A Distinctive Local Image Feature Descriptor", 2nd IEEE International Conference on Information Communication and Signal Processing, 28-30 Sept. 2019**.** DOI: 10.1109/ICICSP48821.2019.8958587

[16] H. Jegou, F. Perronnin, M. Douze, J. S´anchez, P. Perez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 9, pp. 1704–1716, 2012. DOI: 10.1109/TPAMI.2011.235

[17] Li, Q., Peng, Q., and Yan, C., "Multiple VLAD encoding of CNNs for image classification", Computing in Science & Engineering (Volume: 20, Issue: 2 , Mar./ Apr. 2018). DOI: 10.1109/MCSE.2018.108164530

[18] Zhigang Tu, Z., et.al, "Action-Stage Emphasized Spatio-Temporal VLAD for Video Action Recognition", IEEE Transactions on Image Processing (Volume: 28 , Issue: 6 , June 2019), DOI: 10.1109/TIP.2018.2890749

[19] Lyons, M.J.; Akamatsu, S.; Kamachi, M.; Gyoba, J.; Budynek, J. "The Japanese female facial expression (JAFFE) database". In Proceedings of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 14–16 April 1998; pp. 14–16.

Authors' Profiles

Htwe Pa Pa Win received her Ph.D (IT) from University of Computer Studies, Yangon, Myanmar in 2012. She is currently working as a Lecturer at the University of Computer Studies, Hpa-an, Myanmar. Her research interests include Image Processing, Speech processing and Digital Signal processing.

Phyo Thu Thu Khine received her Ph.D (IT) from University of Computer Studies, Yangon, Myanmar in 2012. She is currently working as a Lecturer at the University of Computer Studies, Hpa-an, Myanmar. Her research interests include Image Processing, Speech processing, Digital Signal processing, Database Management System and Big Data.

Zon Nyein Nway received her Ph.D (IT) from University of Computer Studies, Yangon, Myanmar in 2014. She is currently working as a Lecturer at the University of Computer Studies, Yangon, Myanmar. Her research interests include Cyber Security, Image Processing, Database Management System and Big Data.

How to cite this paper: Htwe Pa Pa Win, Phyo Thu Thu Khine, Zon Nyein Nway, " Emotion Recognition from Faces Using Effective Features Extraction Method", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.13, No.1, pp. 50-57, 2021.DOI: 10.5815/ijigsp.2021.01.05

