IEEE Access

Multidisciplinary : Rapid Review : Open Access Journal

An Emotion Recognition System for Mobile Applications

M. SHAMIM HOSSAIN[1], (Senior Member, IEEE), and GHULAM MUHAMMAD[2], (Member, IEEE)
[1]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: M. S. Hossain (mshossain@ksu.edu.sa)

ABSTRACT Emotion-aware mobile applications have been increasing due to their smart features and user acceptability. To realize such an application, an emotion recognition system should be in real time and highly accurate. As a mobile device has limited processing power, the algorithm in the emotion recognition system should be implemented using less computation. In this paper, we propose an emotion recognition with high performance for mobile applications. In the proposed system, facial video is captured by an embedded camera of a smart phone. Some representative frames are extracted from the video, and a face detection module is applied to extract the face regions in the frames. The Bandlet transform is realized on the face regions, and the resultant subband is divided into non-overlapping blocks. Local binary patterns' histograms are calculated for each block, and then are concatenated over all the blocks. The Kruskal–Wallis feature selection is applied to select the most dominant bins of the concatenated histograms. The dominant bins are then fed into a Gaussian mixture model-based classifier to classify the emotion. Experimental results show that the proposed system achieves high recognition accuracy in a reasonable time.

INDEX TERMS Emotion recognition, mobile applications, feature extraction, local binary pattern.

I. INTRODUCTION

Due to the ongoing growth along with the extensive use of smart phones, services and applications, emotion recognition is becoming an essential part of providing emotional care to people. Provisioning emotional care can greatly enhance users' experience to improve the quality of life. The conventional method of emotion recognition may not cater to the need of mobile application users for their value-added emergent services. Moreover, because of the dynamicity and heterogeneity of mobile applications and services, it is a challenge to provide an emotion recognition system that can collect, analyze, and process emotional communications in real time and highly accurate manner with a minimal computation time.

There exist a number of emotion recognition systems in the literature. The emotion can be recognized from speech, image, video, or text. There are many applications of the emotion recognition in mobile platforms. In mobile applications, for example, the text of the SMS can be analyzed to detect the mood or the emotion of the users. Once the emotion is detected, the system can automatically put a corresponding 'emoji' in the SMS. By analyzing a video in the context of

emotion, a smart phone can automatically change the wallpaper, or play some favorite songs to coop with the emotion of the user. The same can be applied using oral conversation through a smart phone; an emotion can be detected from the conversational speech, and an appropriate filtering can be applied to the speech.

To realize an emotion-aware mobile application, the emotion recognition engine must be in real-time, should be computationally less expensive, and give high recognition accuracy [1]. Most of the available emotion recognition systems do not address all of these issues, as they are designed mainly to work offline and for desktop applications. These applications are not bounded by storage, processing power, or time. For example, many online game applications gather the video or the sound of the users, process them in a Cloud, and analyze them for a later improvement (next version) of the game [6]. In this case, the Cloud can provide unlimited storage and processing power, and the game developer can have enough time to analyze. On the contrary, emotion-aware mobile applications cannot have this luxury. Mehmood and Lee proposed an emotion recognition system from brain signal pattern using late positive potential features [2].
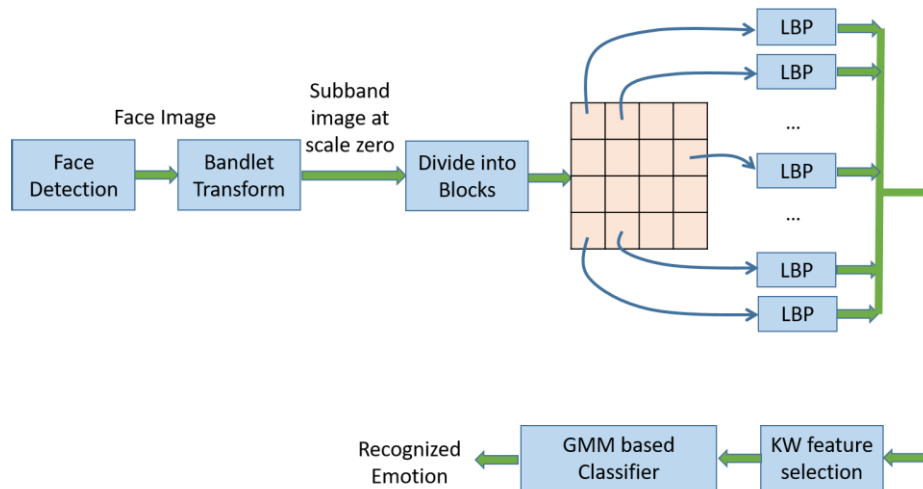
FIGURE 1. Block diagram of the proposed emotion recognition system.

This system needs EEG sensors to be associated to the mobile device, which is an extra burden to the device. Chen *et al.* proposed CP-Robot for emotion sensing and interaction [3]. This proposed robot uses a Cloud for image and video processing. A distant learning system LIVES through interactive video and emotion detection was proposed in [4]. The LIVES can recognize the emotion of the students from the video, and adjust the content of the lecture.

Several emotion-aware applications were proposed in [5] and [6]. Hossain *et al.* [5] used audio and visual modalities to recognize emotion in 5G. They used speech as an audio modality, and image as a visual modality. They found that these two modalities complement each other in terms of emotion information. Hossain *et al.* [6] proposed a cloud-gaming framework by embedding emotion recognition module in it. By recognizing emotion from audio and visual cues, the game screen or content may be adjusted. Y. Zhang *et al.* [29], [30] proposed techniques that integrate mobile and big data technologies along with emotional interaction [31].

Emotion can be recognized by either speech or images. For example, Zhang *et al.* [7] recognized emotion from Chinese speech using deep learning. Their method was quite robust; however, it needed a huge amount of training data. A good survey of emotion recognition from faces can be found in [8]. Another recent emotion recognition system from both audio and visual modalities involves multi-directional regression and ridgelet transform [9]. This system achieved a good accuracy in two databases. Kahou *et al.* [26] proposed a deep learning based approach to recognize emotion in the wild from video clips and achieved 47.67% accuracy. Multiscale temporal modeling was used for emotion recognition from video in [27]. An extreme learning machined based classifier was used together with an active learning for emotion recognition in [28].

Some current healthcare systems use emotion recognition to improve the patient care. Ragsdale *et al.* [10] proposed

a teaching method to interpret emotions of the patients to healthcare providers. Emotion recognition from faces for healthcare was proposed in [11] and [12]. The same from EEG signals was proposed in [13]. To investigate social psychology, mobile phones were equipped with emotion recognition software, called 'EmotionSense' [14]. In a smart home, the emotions of aging adults were recognized in [15]. All these emotion recognition systems either used huge training data, or are computationally expensive.

In this paper, we propose a high-performance emotion recognition system for mobile applications. The embedded camera of a smart phone captures video of the user. The Bandlet transform is applied to some selective frames, which are extracted from the video, to give some subband images. Local binary patterns (LBP) histogram is calculated from the subband images. This histogram describes the features of the frames. A Gaussian mixture model (GMM) based classifier is used as a classifier. The proposed emotion recognition system is evaluated using several databases. The contribution of this paper is as follows: (i) the use of the Bandlet transform in emotion recognition, (ii) the use of the Kruskal-Wallis (KW) feature selection to reduce the time requirement during a test phase, and (iii) an achievement of higher accuracies in two publicly available databases compared with those using other contemporary systems.

The rest of this paper is organized as follows. In Section II, the proposed emotion recognition system is described. Experimental results are reported and analyzed in Section III. Finally, we conclude the paper in Section IV.

II. PROPOSED EMOTION RECOGNITION SYSTEM
Fig. 1 shows a block diagram of the proposed emotion recognition system for mobile applications. In the following subsections, we describe the components of the proposed system.
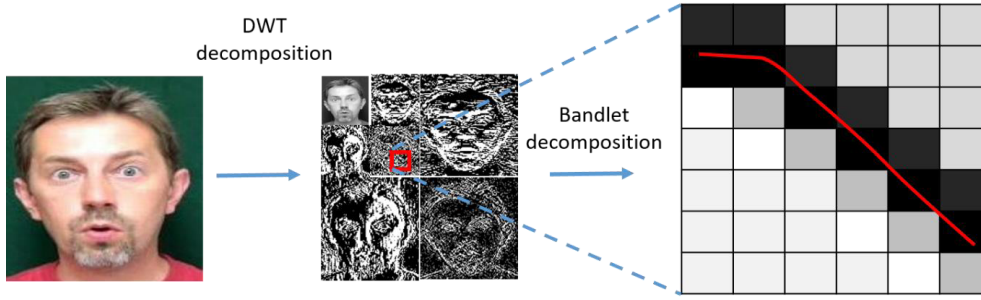
FIGURE 2. Illustration of the Bandlet decomposition.

A. CAPTURING VIDEO

An embedded camera of the smart phone captures the video of the user. As most of the time, the user faces towards the screen of the phone, the video mainly captures the face, the head, and some body parts of the user.

B. SELECTING REPRESENTATIVE FRAMES

As there are many frames in the video sequence, we need to select some representative frames from the sequence to reduce the burden of the processing. To select frames, first, all the frames are converted to the gray scale. Then, histograms are obtained from each frame. A chi-square distance is calculated between the histograms of two successive frames. We select a frame, when the distances between the histogram of this frame and that of the previous frame, and the next frame are minimal. In this way, we select a frame, which is stable in nature.

C. FACE DETECTION

Once we select the frames, the face areas in the frames are detected by the Viola-Jones algorithm [16]. This algorithm works fast, and is suitable for a real-time implementation. Now a day, many smart phones have the face detection functionality embedded into the mobile system.

D. BANDLET TRANSFORM

The Bandlet transform is applied to the detected face area. A face image has many geometric structures that carry valuable information about the identity, the gender, the age, and the emotion of the face. A traditional wavelet transform does not take care much about the geometric structure of an image, especially in sharp transitions; however, representing sharp transitions using geometrical structures can improve the representation of image. One of the major obstacles of using geometrical structure is a high computation complexity. The Bandlet transform overcomes the obstacle to represent geaometric structure of an image by calculating the geometric flow in the form of Bandlet bases [18]. This transform works on the gray scale images; therefore, in the proposed method, the input color images are converted into gray scale images.

To form orthogonal Bandlet bases, the image needs to be divided into regions consisting of geometric flows. In fact, the image is divided into small square blocks, where each block contains at most one contour. If a block does not contain a contour, the geometric flow in that block is not defined.

The Bandlet transform approximates the regions (Ω) by using the wavelet basis in the L2 (Ω) domain as follows.

$$\begin{cases} \varphi_{i,n}(x) = \varphi_{i,n1}(x1)\varphi_{i,n2}(x2) \\ \psi_{i,n}^{H}(x) = \varphi_{i,n1}(x1)\psi_{i,n2}(x2) \\ \psi_{i,n}^{V}(x) = \psi_{i,n1}(x1)\varphi_{i,n2}(x2) \\ \psi_{i,n}^{D}(x) = \psi_{i,n1}(x1)\psi_{i,n2}(x2) \end{cases}$$

where, $\Psi(.)$ and $\varphi(.)$ are the wavelet and the scaling functions, i is the dilation, $n1 \times n2$ is the dimension of the input image, x is the pixel location. φ, Ψ^{H}, Ψ^{V}, Ψ^{D} are the coarse level (low-frequency) approximation, high-frequency representations along horizontal, vertical, and diagonal directions of the image. Fig. 2 illustrates the Bandlet decomposition.

Once the wavelet bases are computed, the geometric flow is calculated in the region by replacing the wavelet bases by the Bandlet orthonormal bases as follows.

$$\begin{cases} \varphi_{i,n}(x) = \varphi_{i,n1}(x1)\varphi_{i,n2}(x2 - c(x1)) \\ \psi_{i,n}^{H}(x) = \varphi_{i,n1}(x1)\psi_{i,n2}(x2 - c(x1)) \\ \psi_{i,n}^{V}(x) = \psi_{i,n1}(x1)\varphi_{i,n2}(x2 - c(x1)) \\ \psi_{i,n}^{D}(x) = \psi_{i,n1}(x1)\psi_{i,n2}(x2 - c(x1)) \end{cases}$$

In the above equation, $c(x)$ is the geometric flow line of the fix translation parameter $x1$ as follows.

$$c(x) = \sum_{u=x_{\min}}^{x} c'(u)$$

From the above equation, we understand that the block sizes affect the geometric flow direction. In practice, the smaller block size gives better representation of geometric flow than the larger block size.

In our proposed system, we investigated the effect of different block sizes, and scale decompositions on the emotion recognition accuracy.

E. LBP

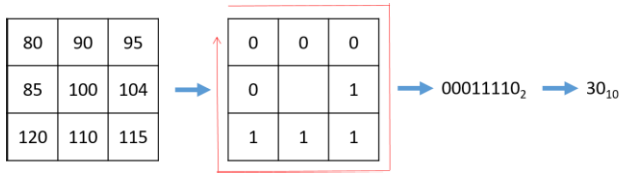The next step of the proposed system is to apply the LBP on the subband images of the Bandlet transform. The LBP

80	90	95
85	100	104
120	110	115

0	0	0
0		1
1	1	1

$\rightarrow$ 00011110_2 $\rightarrow$ 30_{10}

FIGURE 3. Illustration of the LBP calculation.

is a powerful gray-level image descriptor, which operates in real-time [17]. It has been used in many image-processing applications including face recognition, gender recognition, and ethnicity recognition. The basic LBP operates in a 3×3 neighborhood, where the neighboring pixels' intensities are threshold by the center pixel intensity. If a pixel intensity is higher than the center pixel's intensity, then a '1' is assigned to that pixel position, otherwise a '0' is assigned. All the eight neighboring pixels' assigned values are concatenated to produce an 8-bit binary number, which is next converted to a decimal value. This decimal value corresponds to the LBP value of the center pixel. A histogram is created from these LBP values to describe the image. For an interpolated LBP, a circular interpolation of radius R and P number of pixels is used. A uniform LBP is a pattern, where there are at most two bit-wise transitions from zero to one or vice versa. Fig. 3 illustrates the LBP calculation.

In the proposed system, the subband obtained after the Bandlet decomposition is divided into non-overlapping blocks. The LBP histograms are calculated for all the blocks and then concatenated to describe the feature set for the image.

F. KW FEATURE SELECTION

The number of features for the image is very huge; many features slow down the process, and they also bring the 'curse of dimensionality'. Therefore, a feature selection technique can be applied to reduce the dimension of the feature vector. There are many feature selection techniques, each of which has its own advantage and disadvantage. In our proposed system, we adopt the KW technique for its simplicity and low computational complexity, keeping in mind that the system is to deploy in mobile applications. The KW is a non-parametric one-way analysis of variance test, where the null hypothesis is that the samples from different groups have the same median. It returns a value p; if the value of p is close to zero, the null hypothesis is rejected, and the corresponding feature is selected. The feature, which results in a big value of p, is discarded because it is considered to be non-discriminative.

G. GMM-BASED CLASSIFIER

The selected features are fed into a GMM based classifier. During training, models of different emotions are created from the feature set. During testing, log-likelihood scores are obtained for each emotion using the feature set and the models of the emotion. The emotion corresponding to the maximum score is the output of the system. In the

experiments, different numbers of Gaussian mixtures are investigated. We choose the GMM based classifier because it can operate in real-time, it is more stable than the neural network based classifiers.

III. EXPERIMENTAL RESULTS AND ANALYSIS

To validate the proposed system, we used a number of experiments using two publicly available databases, namely Canade-Kohn (CK) [19] and the Japanese female facial expression (JAFFE) [20] databases. In the following sub-sections, we briefly describe the databases, and present the experimental results and discussion.

A. DATABASES

In our experiments, we used two databases. The JAFFE database consists of emotional face images of Japanese actresses. There are total 213 face images of 10 female Japanese. All the images are gray, and have a resolution of 256×256. The original images were printed, scanned, and digitized. The faces are frontal. There are seven emotion classes, which are anger, happiness, sadness, disgust, afraid, and surprise, in addition to neural.

The CK database was created by the faces of 100 university-level students. After careful observations, the faces of four students were discarded because they were not properly showing the emotions. The students were ethnically diverse. Video sequences were captured in a controlled environment, and the participants posed with six emotions as mentioned previously. There were 408 video sequences, and three most representative image frames were selected from each sequence. The first frame of each sequence was treated as a neutral expression. The total number of images is 1632.

B. EXPERIMENTAL RESULTS AND DISCUSSION

First, we used the JAFFE database in our experiments to set up different parameters of the proposed system. We chose this because it a smaller size database than the CK database. Once we fix the parameters using the JAFFE database, we used the CK database.

During classification, we adopted a 5-fold approach, where the database was divided into five equal groups. In each iteration, four groups were trained and the rest was tested. After five iterations, all the five groups were tested. With this approach, the biasness of the system towards any particular images could be reduced. For feature selection using the KW method, we set the value of p to be 0.2, which gave the optimal results for almost all the cases.

In the Bandlet decomposition, we tried different sizes of blocks, and different scales of subband images. Fig. 4 shows the average accuracy (%) of the system using various block sizes and scales. The average accuracy was obtained over seven emotion categories. From the figure, we find that the 2×2 block size and scale 0 subband produced the highest accuracy, which is 92.4%. If we increased the block size, the accuracy decreased. The accuracy also decreased with the increase of scale.
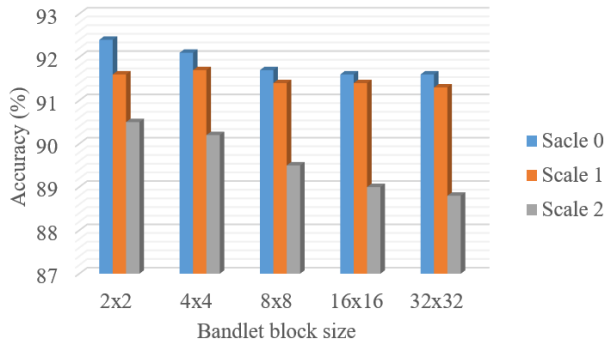
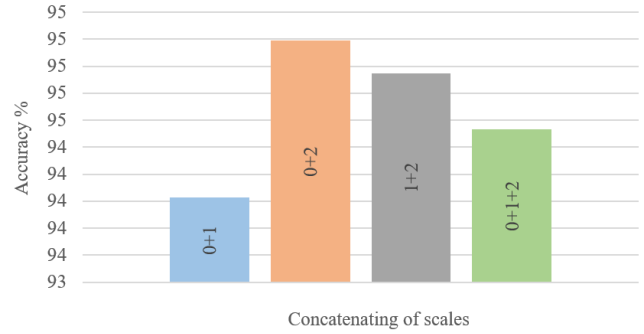**FIGURE 4.** Accuracies in different Bandlet scales and block sizes.

In the next experiment, we normalized the Bandlet coefficients, and investigated its effect. In the normalization, the coefficients were divided by the length of the vector [21]. Fig. 5 shows the average accuracy (%) of the system using the normalized Bandlet coefficients. These accuracies are better than those without normalization. For example, the normalized Bandlet with block size 2×2 and scale 0 achieved accuracy of 93.5%, which is better than 92.4% obtained without normalization. The similar fact is true for all the block sizes and scales. In the following experiments, we used only normalized Bandlet coefficients.
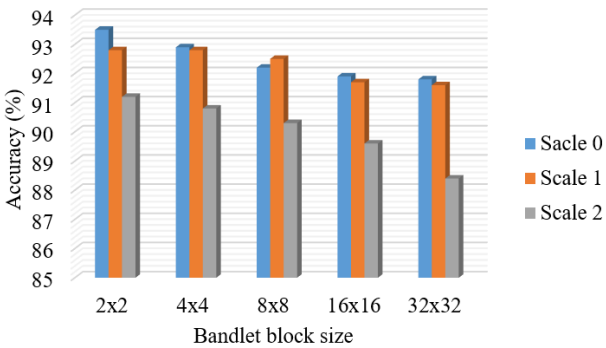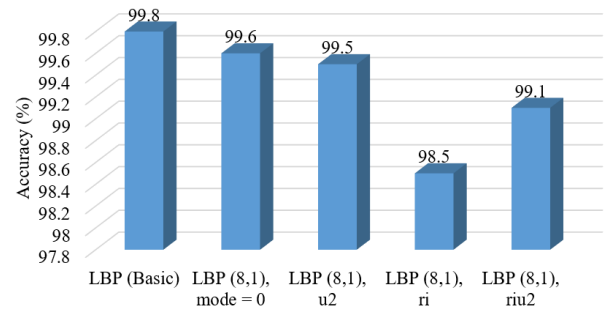

FIGURE 5. Accuracies in different Bandlet scales and block sizes using normalized Bandlet coefficients.

We investigated the effect of concatenating the coefficients of different scales. In this case, we used only 2×2 block size. Fig. 6 shows the accuracy of the system in various concatenating scales. The concatenation of scales 0 and 2 produced the highest accuracy, followed by scale 1 and 2. It can be noted that concatenating scales increased the processing time too much, which may hinder the goal of the system being used in mobile applications. Therefore, we did not proceed with the concatenating scales, rather we used the LBP on scale 0 only to increase the performance.

In the proposed system, various invariants of the LBP were investigated. We applied the LBP on the normalized Bandlet with the block size 2×2 and scale 0. The variants of the LBP used are the basic LBP, the interpolated LBP with mode 0, uniform (u2), rotation invariant (ri), and rotation invariant uniform (riu2). The block size of the LBP was 16×16, and

FIGURE 6. Accuracies in different concatenating Bandlet scales using block size of 2×2 using normalized Bandlet coefficients.

FIGURE 7. Accuracies using different LBP variants in Bandlet scale 0 and block size of 2×2.
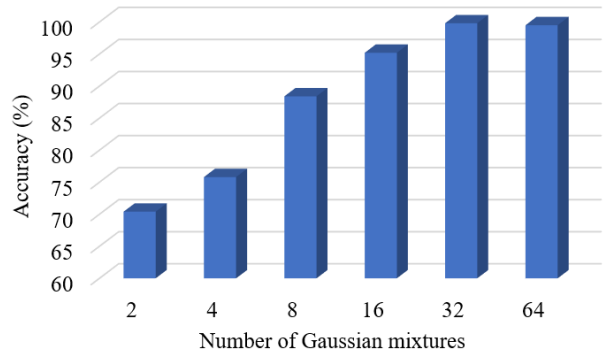
FIGURE 8. Accuracies in different number of Gaussian mixtures in the GMM.

the number of Gaussian mixtures in the classification stage was 32. The average accuracies are shown in Fig. 7. The basic LBP obtained the best accuracy of 99.8%. In fact, all the accuracies using the LBP were above 98%, which indicates the feasibility of the proposed system.

To examine the effect of Gaussian mixtures in the classification stage, we fixed the parameters of the Bandlet transform as the block size of 2×2, scale 0, and normalized coefficients, and the basic LBP with block size of 16×16. The average accuracy of the system using various numbers of mixtures is shown in Fig. 8. As we see from the figure, 32 mixtures gave the highest accuracy of 99.8%, followed by 99.5% obtained with 64 mixtures. It can be mentioned that even we used

TABLE 1. Confusion matrix of the proposed system using the JAFFE database.

Output→ Input ↓	Anger	Happy	Sad	Disgust	Afraid	Surprise	Neutral
Anger	99.7	0	0.1	0.1	0.1	0	0
Happy	0	99.9	0	0.1	0	0	0
Sad	0.1	0	99.8	0	0	0.1	0
Disgust	0.1	0	0.1	99.7	0	0.1	0
Afraid	0	0.1	0	0	99.8	0	0.1
Surprise	0.1	0	0	0.1	0	99.8	0
Neutral	0	0	0	0	0.1	0	99.9

16 mixtures (lesser computation than 32 mixtures), the accuracy would be above 94%.

Table 1 shows the confusion matrix of the proposed system using the JAFFE database. The proposed system has the following parameters: Normalized Bandlet coefficients, Bandlet block size of 2×2, scale 0, the basic LBP with block size 16×16, and 32 Gaussian mixtures. From the confusion matrix, we find that the happy and the neutral emotions had the highest accuracy of 99%, and the angry emotion had the lowest accuracy of 99.7%. On the average, the system took 1.2 seconds to recognize one emotion.

TABLE 2. Confusion matrix of the proposed system using the CK database.

Output→ Input ↓	Anger	Happy	Sad	Disgust	Afraid	Surprise	Neutral
Anger	97.8	0.1	0	0.8	0.6	0.3	0.4
Happy	0.1	99.7	0.1	0	0.1	0	0
Sad	0	0	99.8	0	0.1	0.1	0
Disgust	0.1	0.1	0	99.8	0	0	0
Afraid	0.1	0	0	0	99.9	0	0
Surprise	0.2	0.1	0	0	0	99.7	0
Neutral	0.1	0	0	0	0	0	99.9

Table 2 shows the confusion matrix of the proposed system using the CK database. The parameters are the same as described in the previous paragraph. From the table, we find that the anger emotion was mostly confused with the disgust and the afraid emotions. Other emotions were recognized almost correctly. The system took 1.32 seconds on the average to recognize one emotion.

TABLE 3. Confusion matrix of the proposed system using both the JAFFE and CK database.

Systems	Accuracy (%)	
	JAFFE	CK
[23]	85.4	96.4
[24]	81.59	94.88
[25]	94.37	-
Proposed	99.8	99.7

Table 3 gives a performance comparison of different emotion recognition systems. The proposed system outperforms all other related systems in both the databases.

IV. CONCLUSION

An emotion recognition system for mobile applications has been proposed. The emotions are recognized by face images. The Bandlet transform and the LBP are used as features, which are then selected by the KW feature selection method. The GMM based classifier is applied to recognize the emotions. Two publicly available databases are used to validate the system. The proposed system achieved 99.8% accuracy using the JAFFE database, and 99.7% accuracy using the CK database. It takes less than 1.4 seconds to recognize one instance of emotion. The high performance and the less time requirement of the system make it suitable to any emotion-aware mobile applications. In a future study, we want to extend this work to incorporate different input modalities of emotion.

REFERENCES

[1] M. Chen, Y. Zhang, Y. Li, S. Mao, and V. C. M. Leung, "EMC: Emotion-aware mobile cloud computing in 5G," *IEEE Netw.*, vol. 29, no. 2, pp. 32–38, Mar./Apr. 2015.

[2] R. M. Mehmood and H. J. Lee, "A novel feature extraction method based on late positive potential for emotion recognition in human brain signal patterns," *Comput. Elect. Eng.*, vol. 53, pp. 444–457, Jul. 2016.

[3] M. Chen *et al.*, "CP-Robot: Cloud-assisted pillow robot for emotion sensing and interaction," in *Proc. Industrial IoT*, Guangzhou, China, Mar. 2016, pp. 81–93.

[4] M. Chen, Y. Hao, Y. Li, D. Wu, and D. Huang, "Demo: LIVES: Learning through interactive video and emotion-aware system," in *Proc. ACM Mobihoc*, Hangzhou, China, Jun. 2015, pp. 399–400.

[5] M. S. Hossain, G. Muhammad, M. F. Alhamid, B. Song, and K. Al-Mutib, "Audio-visual emotion recognition using big data towards 5G," *Mobile Netw. Appl.*, vol. 21, no. 5, pp. 753–763, Oct. 2016.

[6] M. S. Hossain, G. Muhammad, B. Song, M. M. Hassan, A. Alelaiwi, and A. Alamri, "Audio–visual emotion-aware cloud gaming framework," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 12, pp. 2105–2118, Dec. 2015.

[7] W. Zhang, D. Zhao, X. Chen, and Y. Zhang, "Deep learning based emotion recognition from Chinese speech," in *Proc. 14th Int. Conf. Inclusive Smart Cities Digit. Health (ICOST)*, vol. 9677. New York, NY, USA, 2016, pp. 49–58.

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

[9] M. S. Hossain and G. Muhammad, "Audio-visual emotion recognition using multi-directional regression and Ridgelet transform," *J. Multimodal User Interfaces*, vol. 10, no. 4, pp. 325–333, Dec. 2016.

[10] J. W. Ragsdale, R. Van Deusen, D. Rubio, and C. Spagnoletti, "Recognizing patients emotions: Teaching health care providers to interpret facial expressions," *Acad. Med.*, vol. 91, no. 9, pp. 1270–1275, Sep. 2016.

[11] S. Tivatansakul, M. Ohkura, S. Puangpontip, and T. Achalakul, "Emotional healthcare system: Emotion detection by facial expressions using Japanese database," in *Proc. 6th Comput. Sci. Electron. Eng. Conf. (CEEC)*, Colchester, U.K., 2014, pp. 41–46.

[12] Y. Zhang, M. Chen, D. Huang, D. Wu, Y. Li, "iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization," *Future Generat. Comput. Syst.*, vol. 66, pp. 30–35, Jan. 2017.

[13] M. Ali, A. H. Mosa, F. A. Machot, and K. Kyamakya, "EEG-based emotion recognition approach for e-healthcare applications," in *Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Vienna, Austria, 2016, pp. 946–950.

[14] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas, "EmotionSense: A mobile phones based adaptive platform for experimental social psychology research," in *Proc. 12th ACM Int. Conf. Ubiquitous Comput. (UbiComp)*, Copenhagen, Denmark, Sep. 2010, pp. 281–290.

[15] J. C. Castillo *et al.*, "Software architecture for smart emotion recognition and regulation of the ageing adult," *Cognit. Comput.*, vol. 8, no. 2, pp. 357–367, Apr. 2016.

[16] R. Nielek and A. Wierzbicki, "Emotion aware mobile application," in *Computational Collective Intelligence. Technologies and Applications* (Lecture Notes in Computer Science), Berlin, Heidelberg, Springer: vol. 6422. 2010, pp. 122–131.

[17] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[18] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[19] S. Mallat and G. Peyré, "A review of bandlet methods for geometrical image representation," *Numer. Algorithms*, vol. 44, no. 3, pp. 205–234, 2007.

[20] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Mar. 2000, pp. 46–53.

[21] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 200–205.

[22] F. S. Hill, Jr., and S. M. Kelley, *Computer Graphics Using OpenGL*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2006, p. 920.

[23] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI J.*, vol. 32, no. 5, pp. 784–794, 2010.

[24] X. Zhao and S. Zhang, "Facial expression recognition based on local binary patterns and kernel discriminant isomap," *Sensors*, vol. 11, no. 10, pp. 9573–9588, 2011.

[25] Y. Rahulamathavan, R. C.-W. Phan, J. A. Chambers, and D. J. Parish, "Facial expression recognition in the encrypted domain based on local Fisher discriminant analysis," *IEEE Trans. Affect. Comput.*, vol. 4, no. 1, pp. 83–92, Jan./Mar. 2013.

[26] S. E. Kahou *et al.*, "EmoNets: Multimodal deep learning approaches for emotion recognition in video," *J. Multimodal User Interfaces*, vol. 10, no. 2, pp. 99–111, Jun. 2016.

[27] L. Chao, J. Tao, M. Yang, Y. Li, and Z. Wen, "Multi-scale temporal modeling for dimensional emotion recognition in video," in *Proc. 4th Int. Workshop Audio/Vis. Emotion Challenge (AVEC)*, Orlando, FL, USA, Nov. 2014, pp. 11–18.

[28] G. Muhammad and M. F. Alhamid, "User emotion recognition from a larger pool of social network data using active learning," *Multimedia Tools Appl.*, accessed on Sep. 19, 2016, doi: 10.1007/s11042-016-3912-2.

[29] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. Leung, "CAP: Community activity prediction based on big data analysis," *IEEE Netw.*, vol. 28, no. 4, pp. 52–57, Jul./Aug. 2014.

[30] Y. Zhang, "GroRec: A group-centric intelligent recommender system integrating social, mobile and big data technologies," *IEEE Trans. Services Comput.*, vol. 9, no. 5, pp. 786–795, Sep./Oct. 2016.

[31] M. Chen, Y. Zhang, Y. Li, M. M. Hassan, and A. Alamri, "AIWAC: Affective interaction through wearable computing and cloud technology," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 20–27, Feb. 2015.

M. SHAMIM HOSSAIN (SM'09) received the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Canada. He is currently an Associate Professor with King Saud University, Riyadh, Saudi Arabia. He has authored or co-authored around 120 publications, including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. His current research interests include serious games, social media, Internet of Things, cloud and multimedia for healthcare, smart health, and resource provisioning for big data processing on media clouds. He was a recipient of a number of awards, including the best conference paper award, the 2016 *ACM Transactions on Multimedia Computing, Communications and Applications* Nicolas D. Georganas Best Paper Award and the Research in Excellence Award from King Saud University. He has served as the Co-Chair, the General Chair, the Workshop Chair, the Publication Chair, and a TPC Member for over 12 IEEE and ACM conferences and workshops. He currently serves as a Co-Chair of the 7th IEEE ICME workshop on Multimedia Services and Tools for E-health MUST-EH 2017. He is on the Editorial Board of the IEEE Access, *Computers and Electrical Engineering* (Elsevier), *Games for Health Journal*, and the *International Journal of Multimedia Tools and Applications* (Springer). He currently serves as a Lead Guest Editor of the *IEEE Communication Magazine*, the IEEE Transactions on Cloud Computing, the IEEE Access, and *Sensors* (MDPI). He served as a Guest Editor of the IEEE Transactions on Information Technology in Biomedicine (currently JBHI), *International Journal of Multimedia Tools and Applications* (Springer), *Cluster Computing* (Springer), *Future Generation Computer Systems* (Elsevier), *Computers and Electrical Engineering* (Elsevier), and *International Journal of Distributed Sensor Networks*. He is a member of ACM and ACM SIGMM.

GHULAM MUHAMMAD (M'07) received the Ph.D. degree in electrical and computer engineering with specialization in speech recognition from the Toyohashi University of Technology, Japan, in 2006. He is currently a Professor with the Computer Engineering Department, King Saud University, Riyadh, Saudi Arabia. He has authored or co-authored around 150 publications, including refereed IEEE/ACM/Springer/Elsevier journals, and conference papers. He holds a U.S. patent on environment. His research interest includes speech processing, image processing, smart health, and multimedia signal processing. He is a member of the ACM. He was a recipient of the Japan Society for Promotion of Science Post-Doctoral Fellowship.

· · ·