

Received December 27, 2016; accepted January 13, 2017, date of publication March 1, 2017, date of current version April 24, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2676238

# Facial Expression Recognition Utilizing Local Direction-Based Robust Features and Deep Belief Network

MD. ZIA UDDIN<sup>1</sup>, (Member, IEEE), MOHAMMAD MEHEDI HASSAN<sup>2</sup>, (Member, IEEE), AHMAD ALMOGREN<sup>2</sup>, ATIF ALAMRI<sup>2</sup>, (Member, IEEE), MAJED ALRUBAIAN<sup>2</sup> AND GIANCARLO FORTINO<sup>3</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Informatics, University of Oslo, Oslo, Norway

<sup>2</sup>College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

<sup>3</sup>Department of Informatics, Modeling, Electronics, and Systems, University of Calabria, 87036 Rende, Italy

Corresponding author: M. M. Hassan (mmhassan@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research at King Saud University through Research Group RGP-281.

**ABSTRACT** Emotional health plays very vital role to improve people's quality of lives, especially for the elderly. Negative emotional states can lead to social or mental health problems. To cope with emotional health problems caused by negative emotions in daily life, we propose efficient facial expression recognition system to contribute in emotional healthcare system. Thus, facial expressions play a key role in our daily communications, and recent years have witnessed a great amount of research works for reliable facial expressions recognition (FER) systems. Therefore, facial expression evaluation or analysis from video information is very challenging and its accuracy depends on the extraction of robust features. In this paper, a unique feature extraction method is presented to extract distinguished features from the human face. For person independent expression recognition, depth video data is used as input to the system where in each frame, pixel intensities are distributed based on the distances to the camera. A novel robust feature extraction process is applied in this work which is named as local directional position pattern (LDPP). In LDPP, after extracting local directional strengths for each pixel such as applied in typical local directional pattern (LDP), top directional strength positions are considered in binary along with their strength sign bits. Considering top directional strength positions with strength signs in LDPP can differentiate edge pixels with bright as well as dark regions on their opposite sides by generating different patterns whereas typical LDP only considers directions representing the top strengths irrespective of their signs as well as position orders (i.e., directions with top strengths represent 1 and rest of them 0), which can generate the same patterns in this regard sometimes. Hence, LDP fails to distinguish edge pixels with opposite bright and dark regions in some cases which can be overcome by LDPP. Moreover, the LDPP capabilities are extended through principal component analysis (PCA) and generalized discriminant analysis (GDA) for better face characteristic illustration in expression. The proposed features are finally applied with deep belief network (DBN) for expression training and recognition.

**INDEX TERMS** Facial expressions recognition (FER), deep belief network (DBN), depth image, generalized discriminant analysis (GDA), local directional pattern (LDP), principal component analysis (PCA).

## I. INTRODUCTION

Recently, ubiquitous healthcare systems have attracted a lot of researchers due to their prominent application the field of human computer interactions (HCI) [1]. In a ubiquitous healthcare system, HCI systems could considerably be improved if computers could be able to recognize the emotions of the people from their facial expressions and react in a friendly manner according to the users' necessities.

When humans experience any situation in their daily lives, they express their mental states through emotions that can effect to their behaviors, thoughts and feelings. Positive emotions can represent healthy mental states by delivering position expression such as happiness and pleasure. On the contrary, negative emotions can represent negative emotions such as sadness and anger. Thus, both positive and negative emotions can affect emotional health in our daily lives

quite substantially. Emotional health refers to the ability to manage feeling to deal with problems.

People with nice emotional health can positively control themselves as well as address negative emotions whereas people with bad emotional health usually experience difficulties with controlling their behaviors and feelings i.e., they often cannot manage themselves to cope with negative emotions. In the worst case of bad emotional health, they might become psychological patients as a result. Therefore, bad emotional health can lead people to social and mental health problems. To improve emotional health, an efficient facial expression recognition system can play a vital role to understand mental states over time can generate mental health-log for the analysis of mental behavior patterns.

With respect to emotional state representation, video-based facial appearance or expression recognition (FER) is getting huge considerations by numerous scientists these days as it is thought to be a standout amongst the most appealing research topics in robot vision and image processing [2]. The vast majority of the FER works followed Principal Component Analysis (PCA) [3]–[10]. In [3], it was utilized for detecting various facial action units for FER. In [5], it was utilized for analysis of facial activity coding system. Later, some works utilized Independent Component Analysis (ICA) approach in FER works [11]–[22]. In [14], ICA was utilized to attain local statistically unbiased features for classification of various facial expressions. ICA was also used to recognize the facial activity unites in [15].

Except ICA, local Binary patterns (LBP) has additionally been focused currently for facial features analysis [23]–[25]. LBP features are tolerant in opposition to illumination modifications and computationally they are simple. In a while, Local Directional Pattern (LDP) that represents local face features was adopted by means of that specialize in face pixel's gradient records [26]. Whilst extracting LDP capabilities, after acquiring a pixel's directional strengths, the binary values are assigned primarily based at the top  $t$  variety of strengths where the price of  $t$  is decided empirically which varies in different experiment settings.

In this paper, a classic LDP is reformed to achieve greater robust features than LDP. After determining the directional strength of a depth pixel, top directional strengths are acquired in descending order and corresponding power sign bits are blended with the top directional power positions in binary to represent strong functions. This approach is named as local directional position pattern (LDPP). In typical LDP for a pixel, the directions with top edge strength are considered by assigning the bit 1 for them and 0 for rest of the directions. It never considers the strength signs and the order of the directions with top strengths which may result into same pattern for two opposite kind of edge pixels having opposite dark and bright regions. This problem can be overcome by the proposed LDPP. Basically, for an edge pixel, dark region mostly represents negative strength whereas bright region shows positive strength. As LDP does not consider the signs of the directions strengths, two edge

pixels with opposite dark and bright regions may exchange the strength signs keeping their strength order same which should generate the same LDP code for these two edge pixels where they should be very different patterns. Besides, LDP represents flat bit 1 to directions with top strengths and 0 to others without considering the orders of the strengths. In such cases, LDP becomes a weak approach to generate features whereas considering strength sign bits along with the top directional strength positions in binary may resolve this issue very strongly to represent robust features such as LDPP. Thus, top directional strength positions' are considered in binary with sign bits in LDPP and then, LDPP histogram is generated to represent robust features for whole face. To make the LDPP features more robust, General Discriminant Analysis (GDA) is utilized after applying PCA for dimension reduction. GDA is considered to be an efficient tool to discriminate the features from different classes [27]–[30].

### A. RELATED FER WORKS

Regarding utilization of cameras for facial expression analysis, RGB cameras have been most popular as face images are very easily obtained from those cameras and besides, they are cheap and commonly used for daily applications such as video chatting through different kind of internet-based software. Though RGB cameras are very famous but images captured through them can change the face pixel intensities very rapidly due to illuminations changes in the scene. Hence, distance-based image can be a better option for facial expression recognition but RGB cameras cannot generate such distance-based face images to describe person-independent facial expressions in the images.

However, depth based cameras can overcome such limitations by providing the depth information of the face parts based on the distance to the camera that would allow one to come up with more efficient expression recognition systems than RGB-based ones. Besides, depth cameras would make it possible to solve some privacy issues such as hiding person's identification in the depth images whereas RGB cameras cannot hide person's identification in RGB images and hence, depth cameras can be utilized regardless of person's identity. As a result, depth images have been getting very much attentions by many researchers in a wide range of computer vision and image processing applications such as body motion recognition [31]–[52], hand motion recognition [53]–[62], and face recognition [63]–[76]. In [31], the authors analyzed depth videos for distinguished human activity recognition. In [33], the authors analyzed surface histograms on depth images for human activity recognition. In [34], the authors did moving body parts analysis from depth data for robust human activity recognition. Yang et al. used Depth Motion Maps (DMM) for obtaining temporal motion energies in [36]. In [40], Koppula et al. applied depth videos for human-object interactions [40]. In [41], Yang et al. obtained Eigen joints obtained using depth video to be used with Naive-Bayes-Nearest-Neighbor (NBNN) for human action analysis. In [42], Sung et al used Maximum

Entropy Markov Model (MEMM) for action recognition in depth videos. Besides human activities, depth information-based hand movement researches were also done for human computer interaction [44]–[52]. In [45], particle swarm optimization was done for analysis of interacting hands in depth videos.

In [46], the authors applied depth images for object dependent hand pose tracking. In [47], the authors focused on depth information-based hand movements and applied random forests for hand part segmentation. American Sign Language (ASL) was also focused in various hand gesture researches [49]–[52]. Similar kind of research information for depth data analysis for gesture recognition can be found in [53]–[62]. Depth information was also used in head pose/face position estimation in some works [63]–[72]. For instance, in [66], the authors applied neural networks for head estimation from depth image. In [68], the authors focused on nose position in a depth face. In [71], the authors did face recognition from low quality depth images where the depth images were obtained from stereo cameras. In [73]–[76], the authors focused on depth image-based upper face researches.

For training and recognition of expression features, Hidden Markov Model (HMM) has been considered in some FER works such as [77], [78]. Nowadays, Deep Neural Network (DNN) picked up a ton of consideration since DNN can create a few elements from the crude info, in spite of alternate classifiers. Likewise, DNN could overcome a few constraints of a perceptron that is not ready to perform general pattern recognition. However, DNN had required a lot of preparing time [79]. In 2004, Hilton et al. proposed an enhanced form of DNN, called Deep Belief Network (DBN), which uses Restricted Boltzmann Machine (RBM) for proficient preparing [80]. There are some different works proposed in [81]–[88].

## B. PROPOSED WORK

A novel FER approach is proposed in this work using LDPP, PCA, GDA, and DBN based on a depth sensor-based video camera images. The LDPP features are extracted first from the facial expression depth images which are then PCA is applied for dimension reduction. Furthermore, the face features are classified by GDA to make them more robust. Finally, the features are applied to train a DBN to be applied later for recognition on cloud. Fig. 1 depicts the basic architecture of the proposed FER system.

## II. FEATURE EXTRACTION

A depth camera is first used to capture various depth images from depth videos of the objects that generate RGB and depth information at the same time. The depth sensor video information or data demonstrate the scope of every pixel in the science as a gray level power or intensity. Figs. 2(a) and (b) represent a happy RGB and depth image respectively. The depth images indicate the bright pixel values for near and dark ones for far distant face parts. Fig. 3 shows a sequence of gray and depth faces from surprise and disgust expressions respectively.

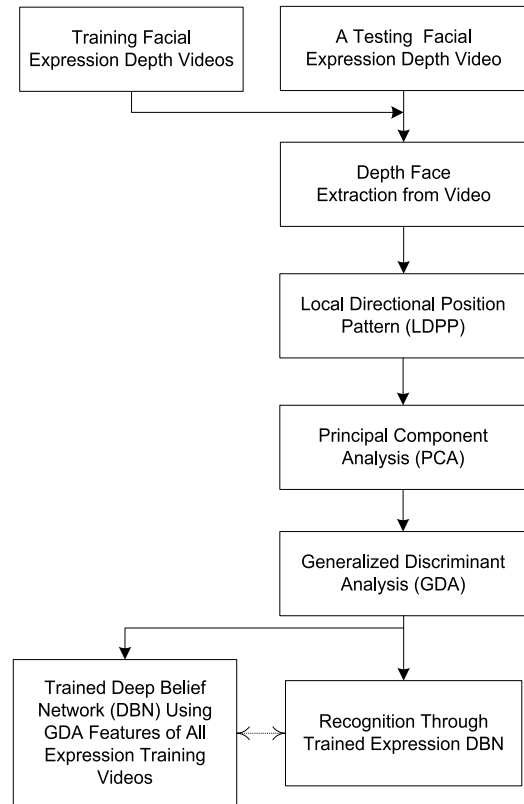
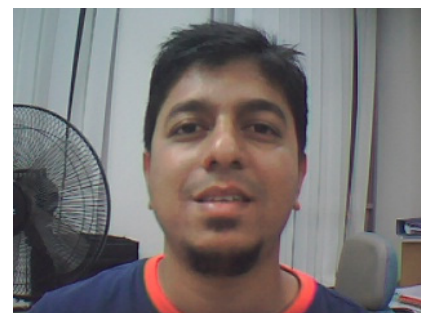


FIGURE 1. Basic architecture of proposed FER system.



(a)



(b)

FIGURE 2. (a) An RGB image and (b) corresponding depth image of a happy expression.

## A. LOCAL DIRECTIONAL POSITION PATTERN (LDPP)

For every pixel of an input depth face, the LDPP appoints an eight-bit binary code. This sample pattern is figured by

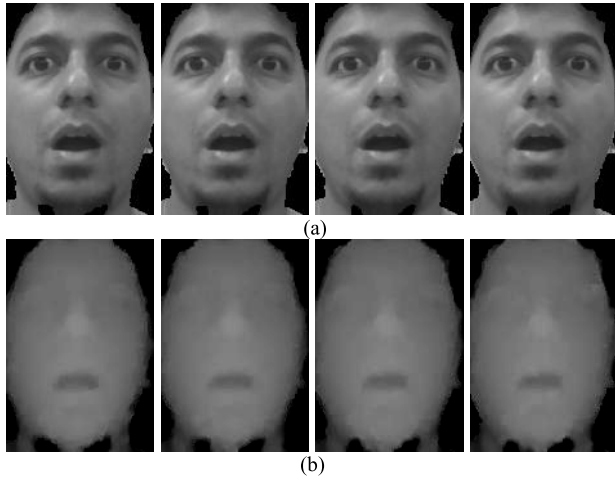


FIGURE 3. (a) Sample gray faces converted from RGB and (b) depth faces from a surprised facial expression.

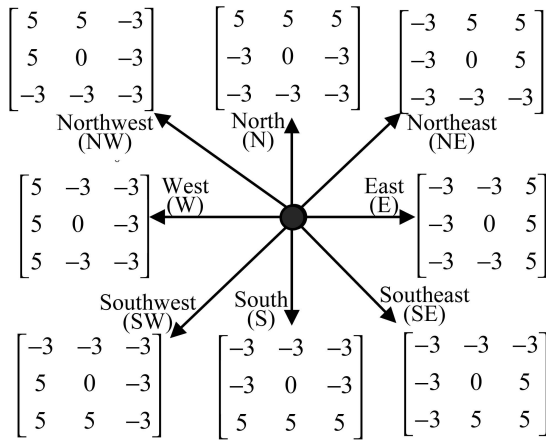


FIGURE 4. Kirsch edge masks in eight directions.

considering top two edge strength position with sign from eight distinct directions. For pixel in the picture, the eight directional edge reaction values  $\{D_i\}$  where  $i = 0, 1, \dots, 7$  are calculated by Kirsch masks. Fig. 4 shows the Kirsch masks. After applying the mask, the directional positions are determined as

$$D_{0,1,\dots,7} \in E, SE, S, SW, W, NW, N, NE. \quad (1)$$

Thus, LDPP code for a pixel  $x$  is derived as

$$LDPP(x) = \sum_{i=0}^7 L_i \times 2^i, \quad (2)$$

$$L = A || K, \quad (3)$$

$$A = B(g) || \text{binary}(\text{Arg}(D(g))), \quad (4)$$

$$K = B(e) || \text{binary}(\text{Arg}(D(e))), \quad (5)$$

$$g \in (R_0, R_1, R_2, \dots, R_7), \quad (6)$$

$$e \in (R_0, R_1, R_2, \dots, R_7), \quad (7)$$

where  $g$  represents the highest edge response direction,  $e$  second highest edge response direction,  $R$  rankings of

$D_5$	$D_6$	$D_7$
$D_4$	x	$D_0$
$D_3$	$D_2$	$D_1$

(a)

$B_5$	$B_6$	$B_7$
$B_4$	x	$B_0$
$B_3$	$B_2$	$B_1$

(b)

$R_5$	$R_6$	$R_7$
$R_4$	x	$R_0$
$R_3$	$R_2$	$R_1$

(c)

FIGURE 5. (a) Edge response to eight directions, (b) sign bit of the edge responses in corresponding direction, (c) ranking of the edge response.

the edge responses to the corresponding directions. Fig. 5 depicts the edge responses, sign bit of the edge responses, and edge response ranking to eight directions. The highest edge response is set to eighth rank. Then, the second highest response is set to seventh, and so on. Fig. 6 shows two examples of LDPP codes where typical LDP makes same patterns for different edges but LDPP can generate separate pattern. In the upper part of figure, the highest edge response is 2422 and hence the first bit of LDPP code for the pixel is the sign bit of 2422 which is 0 and the following three bits are the binary of the direction where the highest strength position lies i.e., 001 which is binary of 1 from  $D_1$ . The second highest edge response is  $-1578$  and hence the fifth bit of LDPP code for the pixel is the sign bit of  $-1578$  which is 1 followed by three bits that are the binary of the direction where the second highest strength position lies i.e., 100 which is binary of 4 from  $D_4$ . Hence, the LDPP code for upper pixel is 00011100 and for lower pixel 10010100. On the other hand, LDP codes for both of them are same, which is 01110011 as their directional rankings are same in both the pixels. Hence, LDPP code represents better features than LDP.

Consequently, the LDPP code is used to transform an image into the LDPP map. The textual feature of the image is exhibited by the histogram of the LDPP map of which the  $s^{th}$  bin can be defined as

$$Z_s = \sum_{x,y} I \{LDPP(x, y) = s\}, s = 0, 1, \dots, n - 1 \quad (8)$$

where  $n$  represents the no. of the LDPP histogram bins for an image  $I$ . Hence, the histogram of the LDPP map is expressed as

$$H = (Z_0, Z_1, \dots, Z_{n-1}). \quad (9)$$

Now, to portray the LDPP highlights or features, a depth silhouette picture is separated into non-overlapping rectangle region and the histogram is processed for every region as appeared in Fig. 7. Moreover, the entire LDPP highlight or feature  $A$  is presented as a concatenated sequence of histograms.

$$A = (H^1, H^2, \dots, H^g) \quad (10)$$

where  $g$  denotes the no. non-overlapped areas or regions in the image.



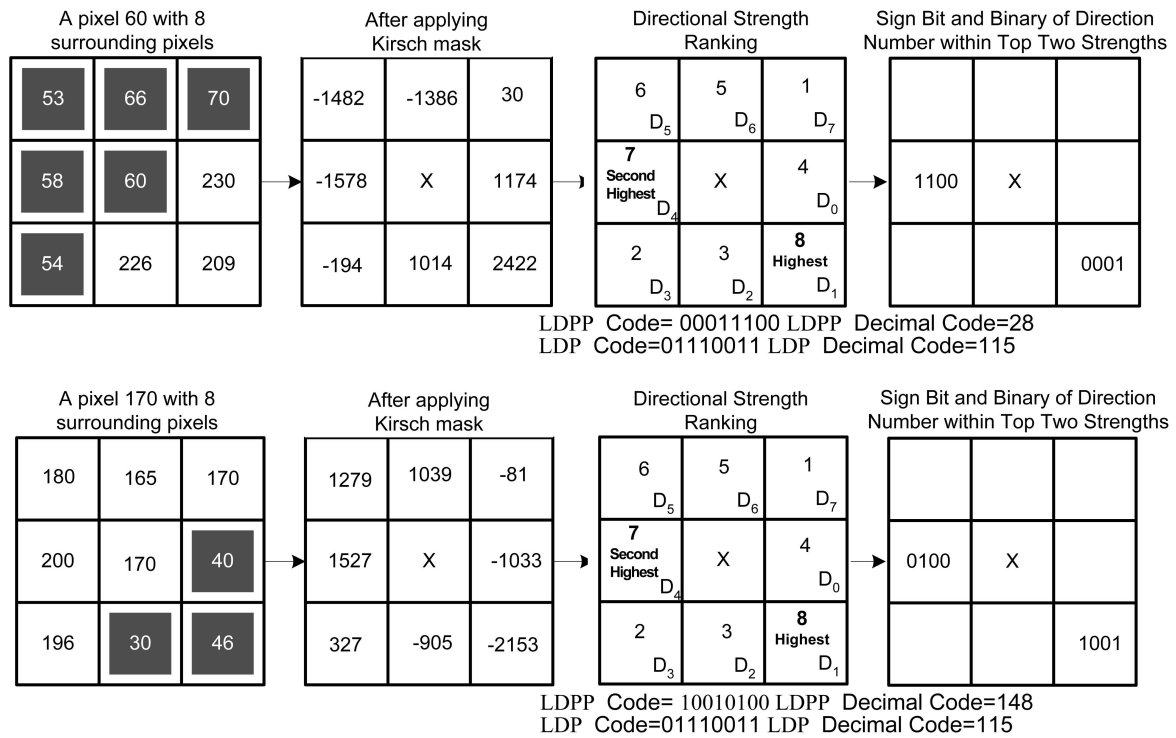


FIGURE 6. Two examples for opposite edge pixels where LDP fails to separate them but LDPP does successfully.

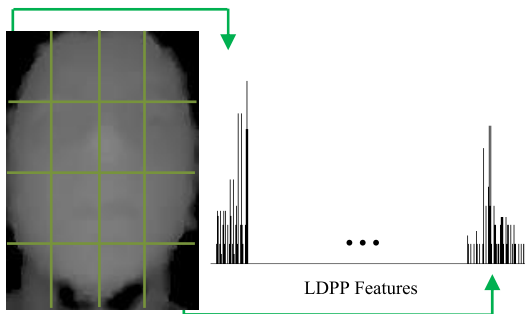


FIGURE 7. A depth expression image is divided into small regions and the regions' LDPP histograms are concatenated to represent features for a face.

**B. PRINCIPAL COMPONENT ANALYSIS (PCA)**

Once locally salient LDPP features are obtained for all the trained facial expression depth images, the feature dimension becomes high and hence, PCA is adopted in this work for dimension reduction. PCA is used to look for the directions of maximum variation in data. Considering  $J$  as a covariance matrix of LDPP feature vectors, PCA on  $J$  should find out the principal components with high variances. Thus, PCA on  $J$  can be described as

$$Y = E^T J E \tag{11}$$

where  $E$  indicates the eigenvector matrix representing principal components (PCs). In this work, we considered 150 PCs after PCA over  $J$ . Fig. 8 depicts the top 150 eigenvalues

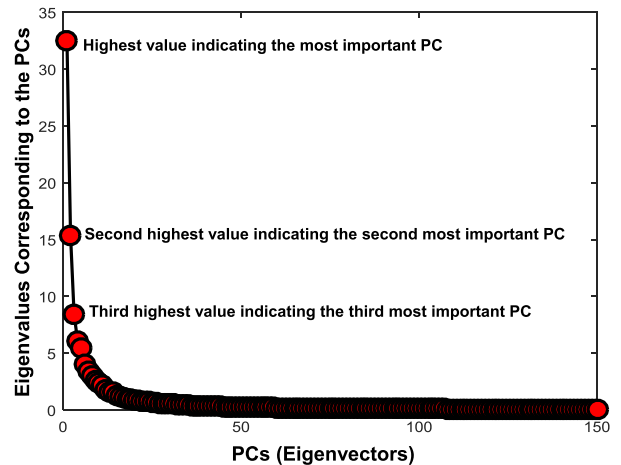


FIGURE 8. Top 150 eigenvalues after applying PCA on LDPP features.

corresponding to the first 150 PCs once PCA is applied on LDPP features. The eigenvalues basically indicates the importance of the corresponding PCs. It can be noticed in the figure that after first few positions, the eigenvalues are descending to zero, indicating the considered number of dimensions should reduce the LDPP feature dimension well with negligible loss of original features. Thus, the reduced dimensional LDPP features after PCA can be shown as

$$C = A E. \tag{12}$$

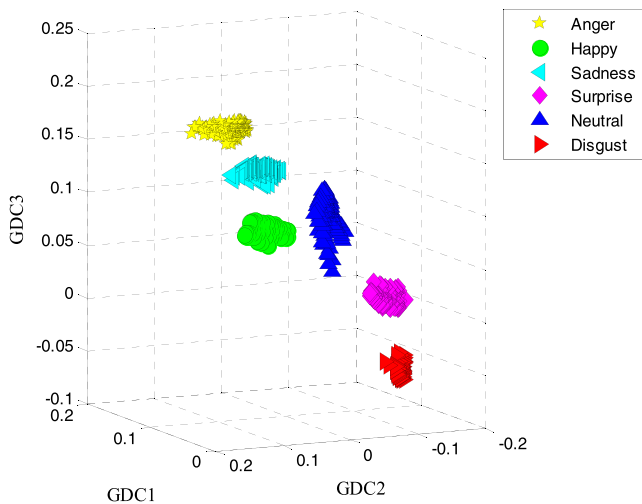
**C. GENERALIZED DISCRIMINANT ANALYSIS (GDA)**

The final step of the feature extraction from a depth image of facial expression is to apply generalized discriminant analysis (GDA) to make the features more robust. GDA, a generalized method of linear discriminant analysis (LDA) which is basically based on an eigenvalue resolution problem to make between inner-class scatterings minimum and inter-class scatterings maximum. GDA first represents the inputs into a high dimensional feature space where it tries to solve the problem by applying LDA method on the feature space. Thus, the fundamental idea of GDA is to map the training data into a high dimensional feature space  $M$  by a nonlinear Gaussian kernel function to apply LDA on  $M$ . Hence, the main goal of GDA is to maximize the following equations.

$$\Lambda_{GDA} = \frac{|\Lambda^T G_B \Lambda|}{|\Lambda^T G_T \Lambda|} \tag{13}$$

where  $G_B^\theta$  and  $G_T^\theta$  are the between-class and total scatter matrices of the features. Finally, the PCA features  $C$  is projected on GDA feature space  $\Lambda_{GDA}$  as

$$O = \Lambda_{GDA}^T C. \tag{14}$$



**FIGURE 9.** 3-D plot of GDA features of depth faces from six expressions.

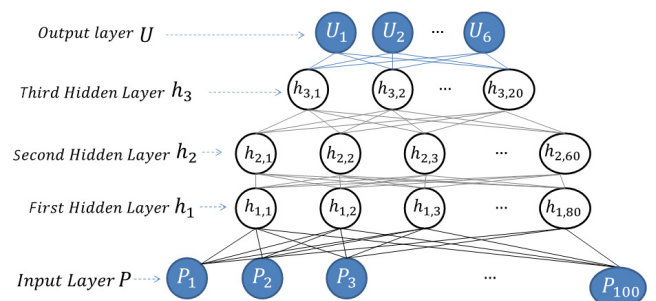
Fig. 9 shows a 3-D plot of the GDA features of training expression images which shows good separation among the samples of different classes which indicates the robustness of GDA in this regard. The LDPP-PCA-GDA features for each image in a video of length  $r$  are augmented further to feed into DBN as

$$T = [O^1 || O^2 || O^3 || \dots || O^r]. \tag{15}$$

**III. DBN FOR EXPRESSION MODELING**

Training a DBN consists of two main parts that are pre-training and fine-tuning. The pre-training phase consists of Bolt Restricted Boltzmann Machine (RBM). When the network is pre-trained then the weights of the networks are

adjusted later on by fine-tuning algorithm. RBM is basically very useful for unsupervised learning that contributes for local optimum error avoidance. One of the key benefits of using DBN is the ability of DBN to extract and select prominent features from the input data. Each layer of RBM is updated depending on the previous layer. Once the first layer is done with computing the weight matrix, the weights are then considered as an input for the second layer and so 'on. This process continues to train RBMs one after another. Besides, the input during this process is reduced layer by layer and hence, the selected features at the hidden nodes of the last layer can be considered as a vector of features for current layer. The algorithm of Contrastive Divergence-1 (CD-1) can be utilized to update the matrix of weights layer by layer [87].



**FIGURE 10.** Structure of a DBN used in this work with 100 input neurons, 80 neurons in hidden layer1, 60 neurons in hidden layer2, 20 neurons in hidden layer3, and 6 output neurons.

Fig. 10 shows a sample DBN where three hidden layers consisting of different number of neurons in different layers such as 100 for input layer, 80 for hidden layer1, 60 for hidden layer2, 20 for hidden layer3, and 6 for output layers indicating to train and recognize 6 classes i.e., expressions in this regard. For initialization of the network, a greedy layer-wise training methodology is applied. Once the weights of the first RBM are trained,  $h_1$  becomes fixed. Then, the weights of the second RBM are adjusted for training using the fixed  $h_1$ . Then, the third RBM is trained with the help of previous RBM. The process of training a typical RBM involves some crucial steps. First of all, initialization is done where a bias vector for the visible layer  $P$ , a bias vector for the hidden layer  $H$ , a weight matrix  $T$  are set to zero.

$$h_1 = \begin{cases} 1, & f(H + p_1 T^T) > r \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

$$precon = \begin{cases} 1, & f(P + h_1 T) > r \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

$$h_{recon} = f(H + precon T^T) \tag{18}$$

Next, utilizing the eq. (16), the binary state of the hidden layer  $h_1$  is calculated and later on it is used to reconstruct the binary state of the visible layer  $precon$  using (17). Then, the hidden layer  $h_{recon}$  is re-computed given  $precon$  where

$$f(t) = 1/(1 + \exp(-t)) \tag{19}$$

The threshold value  $r$  is learnt with the weights to determine the output of the sigmoid function in the network and the weight difference is computed as

$$\Delta T = \left( \frac{h_1 p_1}{L} \right) - (h_{recon} \cdot p_{recon}) / L \quad (20)$$

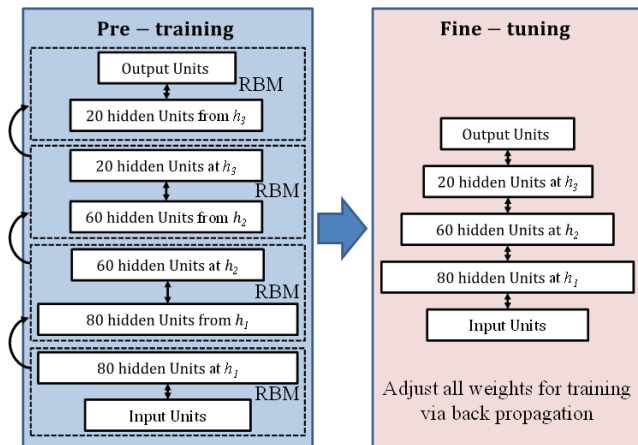


FIGURE 11. The pre-training and fine-tuning processes of DBN used in this work.

where  $L$  is considered as batch size. Finally, the current weight becomes a summation of the previous weights. These steps are repeated for all the batches. When RBM process is done, a typical back propagation algorithm is applied for adjustments of all parameters for fine-tuning. The pre-training and fine-tuning steps are shown in Fig. 11.

TABLE 1. Expression recognition using RGB faces using PCA with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>37.50</b>	2.50	2.50	10	7.50	40
Happy	7.50	<b>47.50</b>	10	10	20	10
Sadness	0	12.50	<b>70</b>	10	2.50	5
Surprise	10	5	2.50	<b>75</b>	2.50	5
Neutral	7.50	2.50	20	5	<b>60</b>	5
Disgust	7.50	2.50	25	0	0	<b>65</b>
Mean	<b>58</b>					

#### IV. EXPERIMENTS AND RESULTS

A depth database was built based on [85], [86] for this work containing six facial expressions: namely Anger, Happy, Sad, Surprise, Disgust, and Neutral. There were 40 videos for each expression where each video consisted of 10 sequential frames. For the experiments, four-fold cross validation was applied to generate four groups of datasets where for each fold, 30 videos were used for training and other 10 for testing without overlapping the videos in training and testing. Hence, a total of 120 videos were applied for training and 40 for testing respectively.

TABLE 2. Expression recognition using RGB faces using PCA-LDA with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>50</b>	10	10	10	5	15
Happy	5	<b>55</b>	10	10	15	0
Sadness	0	10	<b>75</b>	2.50	7.50	5
Surprise	10	5	0	<b>72.50</b>	7.50	20
Neutral	0	7.50	30	7.50	<b>55</b>	0
Disgust	0	10	20	0	0	<b>70</b>
Mean	<b>61.50</b>					

TABLE 3. Expression recognition using RGB faces using ICA with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>75</b>	0	10	0	5	10
Happy	7.50	<b>82.50</b>	10	5	5	
Sadness	0	7.5	<b>82.50</b>	10	0	0
Surprise	0	0	0	<b>80</b>	12.50	7.50
Neutral	0	0	7.50	10	<b>80</b>	0
Disgust	2.50	5	7.50	0	0	<b>85</b>
Mean	<b>80.50</b>					

TABLE 4. Expression recognition using RGB faces using LBP with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>82.50</b>	0	7.50	0	0	10
Happy	5	<b>80</b>	10	10	0	
Sadness	0	10	<b>80</b>	10	0	0
Surprise	0	0	0	<b>80</b>	12.50	7.50
Neutral	0	5	7.50	5	<b>82.50</b>	0
Disgust	0	2.50	10	10	0	<b>77.50</b>
Mean	<b>81.25</b>					

TABLE 5. Expression recognition using RGB faces using LDP with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>82.50</b>	5	2.50	5	2.50	2.50
Happy	5	<b>85</b>	7.50	2.50	0	
Sadness	0	7.5	<b>82.50</b>	10	0	0
Surprise	0	0	0	<b>85</b>	10	5
Neutral	0	7.50	0	7.50	<b>82.50</b>	2.50
Disgust	5	2.50	7.50	5	0	<b>80</b>
Mean	<b>82.91</b>					

##### A. RGB CAMERA-BASED EXPERIMENTS

The FER experiments were started based on the videos obtained from RGB camera. The RGB video-based FER experimental results are shown in confusion matrix from Table I-Table VII. PCA with HMM could achieve 58% mean recognition rate only, which is very poor and the lowest recognition rate amongst our all experiments. Then, PCA-LDA was tried for FER which achieved mean recognition

**TABLE 6.** Expression recognition using RGB faces using LDPP-PCA-GDA with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>90</b>	0	5	0	0	5
Happy	2.50	<b>90</b>	0	5	5	
Sadness	0	2.50	<b>87.50</b>	10	0	0
Surprise	0	0	0	<b>92.50</b>	7.50	0
Neutral	0	5	10	5	<b>90</b>	0
Disgust	0	2.50	12.20	0	0	<b>87.50</b>
Mean	<b>89.58</b>					

**TABLE 7.** Expression recognition using RGB faces with LDPP-PCA-GDA with DBN.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>92.50</b>	2.50	0	2.50	0	5
Happy	2.50	<b>90</b>	2.50	0	5	
Sadness	0	2.50	<b>92.50</b>	5	0	5
Surprise	2.50	0	2.50	<b>95</b>	0	0
Neutral	0	5	7.50	0	<b>95</b>	0
Disgust	5	0	2.50	0	2.50	<b>90</b>
Mean	<b>92.50</b>					

**TABLE 8.** Expression recognition using depth faces using PCA with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>50</b>	7.50	12.50	10	17.50	2.50
Happy	7.50	<b>52.50</b>	10	10	20	10
Sadness	0	17.5	<b>70</b>	7.50	2.50	2.50
Surprise	2.50	2.50	7.50	<b>80</b>	5	2.50
Neutral	10	2.50	15	5	<b>60</b>	7.50
Disgust	10	2.50	12.50	7.50	5	<b>62.50</b>
Mean	<b>62</b>					

rate of 61.50%. We proceed to apply ICA and HMM on the RGB facial expression images, obtained 80.50% mean recognition rate. Furthermore, LBP was applied with HMM for FER that achieved the mean recognition rate of 81.25%. Then, LDP was then tried which achieved 82.91%, better than others so far. Later on, LDPP features were combined with PCA and GDA features to be tried with HMMs that achieved 89.58% mean recognition rate. Then, LDPP-PCA-GDA features were tried with DBN for better FER on the RGB faces which could achieve mean recognition rate of 92.50%, the highest in RGB camera-based experiments.

**B. DEPTH CAMERA-BASED EXPERIMENTS**

After the RGB video-based FER experiments, the experiments were continued to the depth camera-based ones. The depth video-based experimental results are shown in confusion matrix from Table VIII-Table XIV. We can see that when PCA was combined with HMM, it could bring us a mean recognition rate of 62%, which is the lowest performance

**TABLE 9.** Expression recognition using depth faces with PCA-LDA with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>55</b>	0	25	0	10	10
Happy	10	<b>60</b>	10	5	15	0
Sadness	0	17.5	<b>75</b>	7.50	0	0
Surprise	0	0	0	<b>75</b>	10	15
Neutral	0	10	20	0	<b>62.50</b>	7.50
Disgust	5	5	10	12.50	0	<b>67.50</b>
Mean	<b>65</b>					

**TABLE 10.** Expression recognition using depth faces using ICA with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>82.50</b>	2.50	5	0	5	5
Happy	2.50	<b>82.50</b>	5	2.50	2.50	5
Sadness	5	2.50	<b>82.50</b>	2.50	5	2.50
Surprise	0	5	2.50	<b>85</b>	2.50	5
Neutral	5	2.50	5	5	<b>82.50</b>	0
Disgust	2.50	5	5	2.50	0	<b>85</b>
Mean	<b>83.50</b>					

**TABLE 11.** Expression recognition using depth faces using LBP with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>87.50</b>	0	12.50	0	0	0
Happy	10	<b>87.50</b>	0	0	2.50	0
Sadness	0	0	<b>85</b>	0	15	0
Surprise	5	2.50	0	<b>92.50</b>	0	0
Neutral	2.50	7.50	2.50	0	<b>87.50</b>	0
Disgust	0	0	12.50	0	0	<b>87.50</b>
Mean	<b>87.91</b>					

**TABLE 12.** Expression recognition using depth faces using LDP with HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>87.50</b>	0	7.50	0	0	5
Happy	2.50	<b>90</b>	2.50	0	5	0
Sadness	0	0	<b>87.50</b>	0	10	2.50
Surprise	5	0	2.50	<b>92.50</b>	0	0
Neutral	0	2.5	7.50	0	<b>90</b>	0
Disgust	0	0	12.5	0	0	<b>87.50</b>
Mean	<b>89.16</b>					

amongst the depth face-based experiments. On the other hand, using ICA with HMM, the average recognition rate is 83.50%. This result indicates a better performance of ICA as compared with PCA and PCA-LDA (i.e., 65%) on depth faces. Next, we tried to use LBP-HMM and LDP-HMM on depth FER database and found that those two approaches could bring us the mean recognition rate of 87.91% and



**TABLE 13. Expression recognition using depth faces using LDPP-PCA-GDA with HMM.**

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>92.50</b>	2.50	2.50	0	0	2.50
Happy	2.50	<b>90</b>	2.50	0	5	0
Sadness	2.50	5	<b>87.50</b>	2.50	0	2.50
Surprise	0	0	0	<b>97.50</b>	0	2.50
Neutral	0	0	10	0	<b>90</b>	0
Disgust	2.50	5	2.50	0	0	<b>90</b>
Mean	<b>91.25</b>					

**TABLE 14. Expression recognition using depth faces using LDPP-PCA-GDA with DBN.**

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	<b>95</b>	2.50	0	0	0	2.50
Happy	0	<b>97.50</b>	0	2.50	0	0
Sadness	0	0	<b>97.50</b>	0	2.50	0
Surprise	2.50	0	0	<b>97.50</b>	0	0
Neutral	0	2.50	0	0	<b>97.50</b>	0
Disgust	5	0	0	0	0	<b>95</b>
Mean	<b>96.67</b>					

89.16% respectively. Then, proposed LDPP-PCA-GDA features were applied with HMMs on depth faces that achieved 91.67% mean recognition rate which is best recognition rate in HMM-based experiments. At last, the highest average recognition rate of 96.67% was achieved when the proposed LDPP-PCA-GDA features were applied with DBN.

## V. CONCLUSION

Facial expression recognition (FER) is the most natural way of human emotion expression. For last couple of decades, it has been a very prominent research area for enormous computer vision and image processing applications. An FER system basically consists of three fundamental parts: face image preprocessing that captures the image and improves the quality of the image by eliminating unnecessary details from the background, feature extraction that obtains distinguishable robust features for each expression, and a recognition part that recognizes facial expression in video by applying features on a robust model. Facial image features are very sensitive to noise and illumination and often generates complexity by merging to with each other in the feature space. Hence, performance of an FER system is very much dependent on the extraction of good features. In this study, we have proposed a new approach for emotion recognition from facial expression depth videos, where a novel feature extraction method consisting of LDPP, PCA, and GDA has been investigated. The proposed method consists of tolerance against illumination variation and extracts salient features by utilizing prominent directional strengths of the pixels by considering the highest strength directional position and the signs of the strengths. Besides, the proposed features can

overcome critical problems which could not be resolved by conventional LDP feature extraction sometimes such as generating different patterns for edge pixels of reverse directions of bright and dark parts. Regarding depth faces, one major advantage of depth face over RGB is that FER with depth map can be implemented without revealing the identity of the subject since depth faces are represented with respect to the distance to the camera. Hence, the original identity of a person is hidden, which seems to resolve privacy issues regarding permission of the subjects in database. The robust LDPP-PCA-GDA features have been further combined with a state-of-the-art machine learning technique, Deep Belief Network (DBN) for modeling the expressions as well as recognition. Furthermore, the proposed FER method was compared with traditional approaches and its recognition performance was superior over them.

## REFERENCES

- [1] P. Baxter, S. Lemaignan, and J. G. Trafton, "Cognitive architectures for social human-robot interaction," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, Bielefeld, Germany, Mar. 2016, pp. 504–505.
- [2] S. Tadeusz, "Application of vision information to planning trajectories of adept six-300 robot," in *Proc. 21st Int. Conf. Methods Models Autom. Robot. (MMAR)*, Międzyzdroje, Poland, Aug./Sep. 2016, pp. 1069–1075.
- [3] D.-S. Kim, I.-J. Jeon, S.-Y. Lee, P.-K. Rhee, and D.-J. Chung, "Embedded face recognition based on fast genetic algorithm for intelligent digital photography," *IEEE Trans. Consum. Electron.*, vol. 52, no. 3, pp. 726–734, Aug. 2006.
- [4] C. Padgett and G. Cottrell, "Representing face images for emotion classification," in *Advances in Neural Information Processing Systems*, vol. 9. Cambridge, MA, USA: MIT Press, 1997.
- [5] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 3, pp. 311–324, May 2007.
- [6] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 974–989, Oct. 1999.
- [7] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.
- [8] M. Meulders, P. D. Boeck, I. Van Mechelen, and A. Gelman, "Probabilistic feature analysis of facial perception of emotions," *Appl. Statist.*, vol. 54, no. 4, pp. 781–793, 2005.
- [9] A. J. Calder, A. M. Burton, P. Miller, A. W. Young, and S. Akamatsu, "A principal component analysis of facial expressions," *Vis. Res.*, vol. 41, no. 9, pp. 1179–1208, Apr. 2001.
- [10] S. Dubuisson, F. Davoine, and M. Masson, "A solution for facial expression representation and recognition," *Signal Process., Image Commun.*, vol. 17, no. 9, pp. 657–673, Oct. 2002.
- [11] I. Buciu, C. Kotropoulos, and I. Pitas, "ICA and Gabor representation for facial expression recognition," in *Proc. Int. Conf. Image Process. (ICIP)*, Sep. 2003, pp. 855–858.
- [12] F. Chen and K. Kotani, "Facial expression recognition by supervised independent component analysis using MAP estimation," *IEICE Trans. Inf. Syst.*, vol. E91-D, no. 2, pp. 341–350, 2008.
- [13] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY, USA: Wiley, 2001.
- [14] Y. Karklin and M. S. Lewicki, "Learning higher-order structures in natural images," *Netw., Comput. Neural Syst.*, vol. 14, no. 3, pp. 483–499, 2003.
- [15] M. S. Bartlett, G. Donato, J. R. Movellan, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Face image analysis for expression measurement and detection of deceit," in *Proc. 6th Joint Symp. Neural Comput.*, 1999, pp. 8–15.
- [16] C.-F. Chuang and F. Y. Shih, "Recognizing facial action units using independent component analysis and support vector machine," *Pattern Recognit.*, vol. 39, no. 9, pp. 1795–1798, Sep. 2006.
- [17] A. J. Calder, A. W. Young, and J. Keane, "Configural information in facial expression perception," *J. Experim. Psychol., Human Perception Perform.*, vol. 26, no. 2, pp. 527–551, 2000.

- [18] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 200–205.
- [19] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–1464, Nov. 2002.
- [20] C. Liu, "Enhanced independent component analysis and its application to content based face image retrieval," *IEEE Trans. Syst., Man, B, Cybern.*, vol. 34, no. 2, pp. 1117–1127, Apr. 2004.
- [21] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.
- [22] M. Z. Uddin, J. J. Lee, and T.-S. Kim, "An enhanced independent component-based human facial expression recognition from video," *IEEE Trans. Consum. Electron.*, vol. 55, no. 4, pp. 2216–2224, Nov. 2009.
- [23] T. Ojala, M. Pietikäinen, and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [24] C. Shan, S. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2005, pp. 370–373.
- [25] 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.
- [26] T. Jabid, M. H. Kabir, and O. Chae, "Local directional pattern (LDP)—A robust image descriptor for object recognition," in *Proc. IEEE Adv. Video Signal Based Surveill. (AVSS)*, Aug./Sep. 2010, pp. 482–487.
- [27] P. Yu, D. Xu, and P. Yu, "Comparison of PCA, LDA and GDA for palmprint verification," in *Proc. Int. Conf. Inf., Netw. Autom.*, Oct. 2010, pp. 148–152.
- [28] L. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," *Comput. Vis. Image Understand.*, vol. 91, pp. 160–187, Jul. 2003.
- [29] B. M. Asl, S. K. Setarehdan, and M. Mohebbi, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artif. Intell. Med.*, vol. 44, no. 1, pp. 51–64, Sep. 2008.
- [30] M. Z. Uddin, "A depth video-based facial expression recognition system utilizing generalized local directional deviation-based binary pattern feature discriminant analysis," *Multimedia Tools Appl.*, vol. 75, no. 12, pp. 6871–6886, Jun. 2016.
- [31] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *Proc. Workshop Human Activity Understand. 3D Data*, 2010, pp. 9–14.
- [32] W. Li, Z. Zhang, and Z. Liu, "Expandable data-driven graphical modeling of human actions based on salient postures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1499–1510, Nov. 2008.
- [33] O. Oreifej and Z. Liu, "HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 716–723.
- [34] A. W. Vieira, E. R. Nascimento, G. L. Oliveira, Z. Liu, and M. F. M. Campos, "STOP: Space-time occupancy patterns for 3D action recognition from depth map sequences," in *Proc. Prog. Pattern Recognit., Image Anal., Comput. Vis., Appl.*, 2012, pp. 252–259.
- [35] J. Wang, Z. Liu, J. Chorowski, Z. Chen, and Y. Wu, "Robust 3D action recognition with random occupancy patterns," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 872–885.
- [36] X. Yang, C. Zhang, and Y. Tian, "Recognizing actions using depth motion maps-based histograms of oriented gradients," in *Proc. ACM Int. Conf. Multimedia*, 2012, pp. 1057–1060.
- [37] J. Lei, X. Ren, and D. Fox, "Fine-grained kitchen activity recognition using RGB-D," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 208–211.
- [38] A. Jalal, M. Z. Uddin, J. T. Kim, and T.-S. Kim, "Recognition of human home activities via depth silhouettes and R transformation for smart homes," *Indoor Built Environ.*, vol. 21, no. 1, pp. 184–190, 2011.
- [39] Y. Wang, K. Huang, and T. Tan, "Human activity recognition based on R transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [40] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from RGB-D videos," *Int. J. Robot. Res.*, vol. 32, no. 8, pp. 951–970, 2013.
- [41] X. Yang and Y. L. Tian, "EigenJoints-based action recognition using Naïve-Bayes-nearest-neighbor," in *Proc. Workshop Human Activity Understanding 3D Data*, Jun. 2012, pp. 14–19.
- [42] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from RGBD images," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 842–849.
- [43] A. McCallum, D. Freitag, and F. Pereira, "Maximum entropy Markov models for information extraction and segmentation," in *Proc. Int. Conf. Mach. Learn.*, 2000, pp. 591–598.
- [44] H. Hamer, K. Schindler, E. Koller-Meier, and L. Van Gool, "Tracking a hand manipulating an object," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1475–1482.
- [45] I. Oikonomidis, N. Kyriazis, and A. A. Argyros, "Tracking the articulated motion of two strongly interacting hands," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1862–1869.
- [46] H. Hamer, J. Gall, T. Weise, and L. Van Gool, "An object-dependent hand pose prior from sparse training data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 671–678.
- [47] D. D. Luong, S. Lee, and T.-S. Kim, "Human computer interface using the recognized finger parts of hand depth silhouette via random forests," in *Proc. 13th Int. Conf. Control, Autom. Syst.*, Oct. 2013, pp. 905–909.
- [48] G. J. Iddan and G. Yahav, "Three-dimensional imaging in the studio and elsewhere," *Proc. SPIE*, vol. 4298, pp. 48–55, Apr. 2001.
- [49] S. C. W. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 873–891, Jun. 2005.
- [50] P. Yin, T. Starner, H. Hamilton, I. Essa, and J. M. Rehg, "Learning the basic units in American sign language using discriminative segmental feature selection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 4757–4760.
- [51] H. D. Yang, S. Sclaroff, and S. W. Lee, "Sign language spotting with a threshold model based on conditional random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 7, pp. 1264–1277, Jul. 2009.
- [52] P. Dreuwe et al., "The SignSpeak project—Bridging the gap between signers and speakers," in *Proc. Int. Conf. Lang. Resour. Eval.*, May 2010, pp. 476–481.
- [53] X. Liu and K. Fujimura, "Hand gesture recognition using depth data," in *Proc. Int. Conf. Autom. Face Gesture Recognit.*, May 2004, pp. 529–534.
- [54] Z. Mo and U. Neumann, "Real-time hand pose recognition using low-resolution depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 1499–1505.
- [55] P. Breuer, C. Eckes, and S. Müller, "Hand gesture recognition with a novel IR time-of-flight range camera—A pilot study," in *Proc. 3rd Int. Conf. Comput. Vis./Comput. Graph. Collaboration Techn.*, 2007, pp. 247–260.
- [56] S. Soutschek, J. Penne, J. Hornegger, and J. Kornhuber, "3-D gesture-based scene navigation in medical imaging applications using time-of-flight cameras," in *Proc. Workshop Time Flight Camera Based Comput. Vis.*, Jun. 2008, pp. 1–6.
- [57] K. Kollorz, J. Penne, J. Hornegger, and A. Barke, "Gesture recognition with a time-of-flight camera," *Int. J. Intell. Syst. Technol. Appl.*, vol. 5, nos. 3–4, pp. 334–343, Nov. 2008.
- [58] J. Penne, S. Soutschek, L. Fedorowicz, and J. Hornegger, "Robust real-time 3D time-of-flight based gesture navigation," in *Proc. Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–2.
- [59] Z. Li and R. Jarvis, "Real time hand gesture recognition using a range camera," in *Proc. Austral. Conf. Robot. Autom.*, 2009, pp. 1–7.
- [60] H. Takimoto, S. Yoshimori, Y. Mitsukura, and M. Fukumi, "Classification of hand postures based on 3D vision model for human-robot interaction," in *Proc. Int. Symp. Robot Human Interact. Commun.*, Sep. 2010, pp. 292–297.
- [61] M. Van den Bergh and L. Van Gool, "Combining RGB and ToF cameras for real-time 3D hand gesture interaction," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Jan. 2011, pp. 66–72.
- [62] J. Marnik, "The polish finger alphabet hand postures recognition using elastic graph matching," in *Computer Recognition Systems 2*, Berlin, Germany: Springer, vol. 45, 2007, pp. 454–461.
- [63] M. D. Breitenstein, D. Kuettel, T. Weise, L. Van Gool, and H. Pfister, "Real-time face pose estimation from single range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
- [64] Q. Cai, D. Gallup, C. Zhang, and Z. Zhang, "3D deformable face tracking with a commodity depth camera," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 229–242.

- [65] L. P. Morency, P. Sundberg, and T. Darrell, "Pose estimation using 3D view-based eigenspaces," in *Proc. IEEE Int. Workshop Anal. Modeling Faces Gestures*, Oct. 2003, pp. 45–52.
- [66] E. Seemann, K. Nickel, and R. Stiefelhagen, "Head pose estimation using stereo vision for human-robot interaction," in *Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2004, pp. 626–631.
- [67] A. Mian, M. Bennamoun, and R. Owens, "Automatic 3D face detection, normalization and recognition," in *Proc. 3rd Int. Symp. 3D Data Process., Vis., Transmiss.*, Jun. 2006, pp. 735–742.
- [68] X. Lu and A. K. Jain, "Automatic feature extraction for multiview 3D face recognition," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, Apr. 2006, pp. 585–590.
- [69] T. Weise, B. Leibe, and L. Van Gool, "Fast 3D scanning with automatic motion compensation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [70] T. Weise, S. Bouaziz, H. Li, and M. Pauly, "Realtime performance-based facial animation," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–10, Jul. 2011.
- [71] M. D. Breitenstein, J. Jensen, C. Høilund, T. B. Moeslund, and L. Van Gool, "Head pose estimation from passive stereo images," in *Proc. 16th Scand. Conf. Image Anal.*, 2009, pp. 219–228.
- [72] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 617–624.
- [73] Y. Sun and L. Yin, "Automatic pose estimation of 3D facial models," in *Proc. Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.
- [74] M. P. Segundo, L. Silva, O. R. P. Bellon, and C. C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 40, no. 5, pp. 1319–1330, Oct. 2010.
- [75] K. I. Chang, W. Bowyer, and P. J. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1695–1700, Oct. 2006.
- [76] P. Nair and A. Cavallaro, "3-D face detection, landmark localization, and registration using a point distribution model," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 611–623, Jun. 2009.
- [77] P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 1, pp. 3–11, Mar. 2006.
- [78] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [79] M. Minsky and S. A. Papert, *Perceptrons: An Introduction to Computational Geometry*, vol. 165. Cambridge, MA, USA: MIT Press, 1969, pp. 780–782.
- [80] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [81] M. M. Hassan, K. Lin, X. Yue, and J. Wan, "A multimedia healthcare data sharing approach through cloud-based body area network," *Future Generat. Comput. Syst.*, vol. 66, pp. 48–58, Jan. 2017.
- [82] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, "iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization," *Future Generat. Comput. Syst.*, vol. 66, pp. 30–35, Jan. 2017.
- [83] Y. Zhang, "GroRec: A group-centric intelligent recommender system integrating social, mobile and big data technologies," *IEEE Trans. Serv. Comput.*, vol. 9, no. 5, pp. 786–795, Sep./Oct. 2016.
- [84] 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.
- [85] M. Z. Uddin and A. M. J. Sarkar, "A facial expression recognition system from depth video," in *Proc. WorldComp*, 2014, pp. 1–6.
- [86] M. Z. Uddin, "Facial expression recognition using depth information and spatiotemporal features," in *Proc. ICACT*, Jan./Feb. 2016, pp. 726–731.
- [87] Y. Zhang, D. Zhang, M. M. Hassan, A. Alamri, and L. Peng, "CADRE: Cloud-assisted drug recommendation service for online pharmacies," *Mobile Netw. Appl.*, vol. 20, no. 3, pp. 348–355, Jun. 2014.
- [88] A. Fischer and C. Igel, "Training restricted Boltzmann machines: An introduction," *Pattern Recognit.*, vol. 47, no. 1, pp. 25–39, Jan. 2014.



**MD. ZIA UDDIN** (M'12) received the Ph.D. degree in biomedical engineering in 2011. He was with the Electronics Engineering Department, Inha University, South Korea, as an Assistant Professor (non-tenure track), from 2011 to 2013. He was with the Computer Education Department, Sungkyunkwan University, South Korea, as an Assistant Professor (non-tenure track), from 2013 to 2016. Since 2017, he has been a Post-Doctoral Research Fellow with the Robotics and Intelligent Systems Research Group, Department of Informatics, University of Oslo, Norway. His research interests include computer vision, image processing, and pattern recognition. More specifically, he has been involved in color and 3-D depth video-based human activity, gait, and facial expression analysis. In this regard, he has developed several novel methodologies using color and depth videos.



**MOHAMMAD MEHEDI HASSAN** (M'12) received the Ph.D. degree in computer engineering from Kyung Hee University, South Korea, in 2011. He is currently an Associate Professor with the Information Systems Department, College of Computer and Information Sciences (CCIS), King Saud University, Riyadh (KSU), Saudi Arabia. He has authored over 100 research papers in journals and conferences of international repute. His research interests include cloud federation, multimedia cloud, sensor-cloud, Internet of Things, big data, mobile cloud, cloud security, IPTV, sensor networks, 5G network, social networks, publish/subscribe systems, and recommender systems. He received a Best Paper Award from the CloudComp conference, China, in 2014. He also received the Excellence in Research Award from CCIS, KSU, in 2015 and 2016. He has served as the Chair and a Technical Program Committee Member in numerous international conferences/workshops, such as the IEEE HPCC, the ACM BodyNets, the IEEE ICME, the IEEE ScalCom, the ACM Multimedia, the ICA3PP, the IEEE ICC, the TPMC, and the IDCS. He has also been a Guest Editor of several international ISI-indexed journals.



**AHMAD ALMOGREN** received the Ph.D. degree in computer sciences from Southern Methodist University, Dallas, TX, USA, in 2002. He was an Assistant Professor of Computer Science and a member of the scientific council at the Riyadh College of Technology. He served as the Dean of the College of Computer and Information Sciences and the Head of the Council of Academic Accreditation, Al Yamamah University. He is currently an Associate Professor and the Vice Dean for the development and quality at the College of Computer and Information Sciences, King Saud University, Saudi Arabia. His research interests include mobile and pervasive computing, computer security, sensor and cognitive networks, and data consistency. He has served as a Guest Editor for several computer journals.



**ATIF ALAMRI** (M'12) received the B.Sc. and M.Sc. degrees in information systems from the College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia, in 2000 and 2004, respectively, and the Ph.D. degree in computer science from the School of Information Technology and Engineering, University of Ottawa, Canada, in 2010. He is currently an Associate Professor with CCIS, KSU. He is one of the founding members of the Chair of Pervasive and Mobile Computing at KSU and successfully manages its research program, which transformed the chair as one of the best chairs of research excellence in the college. His research interests include multimedia assisted health systems, ambient intelligence, and service-oriented architecture, multimedia cloud, sensor-cloud, Internet of Things, big data, mobile cloud, social networks, and recommender systems.



**MAJED ALRUBAIAN** received the M.S. degree in information systems and the Ph.D. degree in computer and information sciences from the College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, in 2004 and 2016, respectively. He is currently involved in the academic field of one of the major colleges with security responsibilities in Saudi Arabia. His research interests include social media analysis and mining, social computing, information credibility, human–computer interaction, social-IoT, and cyber security.



**GIANCARLO FORTINO** (SM'12) was a Visiting Researcher with the International Computer Science Institute, Berkeley, CA, USA, in 1997 and 1999, and a Visiting Professor with the Queensland University of Technology, Brisbane, Australia, in 2009. He was nominated as a Guest Professor in Computer Engineering with the Wuhan University of Technology, in 2012. He has been an Associate Professor of Computer Engineering with the Department of Informatics, Modeling, Electronics and Systems, University of Calabria (Unical), Rende, Italy, since 2006. He is currently a Co-Founder and the CEO of SenSysCal S.r.l., a spinoff of Unical, focused on innovative sensor-based systems for e-health and demotics. He holds the Italian National Habilitation for Full Professorship. He has authored over 230 publications in journals, conferences and books. His research interests include distributed computing, wireless sensor networks, software agents, cloud computing, and Internet of Things systems. He is the Founding Editor of the Springer Book Series *Internet of Things: Technology, Communications and Computing* and serves on the Editorial Board of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, the *Journal of Networks and Computer Applications*, the *Engineering Applications of Artificial Intelligence*, the *Information Fusion*, and the *Multi Agent and GRID Systems*.

...