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

Md. Zia Uddin, Member, IEEE, Mohammed Mehedi Hassan, Member, IEEE Ahmad Almogren, Atif Alamri, Member, IEEE, Majed Alrubaian and Giancarlo Fortino, Senior Member, IEEE

***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. Hence, facial expression analysis from video data is considered to be a very challenging task in the research areas of computer vision, image processing, and pattern recognition. The accuracy of a FER system is pretty much reliant on the extraction of robust features. In this work, a novel feature extraction method is proposed to extract prominent 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. Furthermore, the LDPP features are extended by Principal Component Analysis (PCA) and Generalized Discriminant Analysis (GDA) to make them more robust for better face feature representation in expression. The proposed features are finally applied with Deep Belief Network (DBN) for expression training and recognition.**

***Index Terms*—FER, DBN, Depth Image, GDA, LDP, 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.

In regard to emotional state representation, video-based facial expression recognition (FER) is getting significant attentions by many researchers nowadays as it is considered to be one of the most attractive research topics in robot vision and image processing [2]. Most of the FER works adopted Principal Component Analysis (PCA) [3]-[10]. In [3], PCA was applied to focus different facial action units for FER. In [5], it was used for facial action coding system analysis for FER. Independent Component Analysis (ICA), a higher level statistical approach than PCA has been

This work was full financially supported by the King Saud University, through Vice Deanship of Research Chairs.

Md. Zia Uddin is with the department of Computer Education, Sungkyunkwan University, Seoul, and Republic of Korea (e-mail: ziauddin@skku.edu).

M. M. Hassan, A. Almogren, A. Alamri and M. Rubaian are with the Research Chair of Pervasive and Mobile Computing, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia (e-mail: {mmhassan, ahalmogren, atif, malrubai}@ksu.edu.sa). M. M. Hassan is the corresponding author.

G. Fortino is with the Department of Informatics, Modeling, Electronics, and Systems, University of Calabria, Italy (e-mail: g.fortino@unical.it).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2676238, IEEE Access

adopted later in FER works for it local face image feature extraction [11]-[22]. In [14], ICA was adopted to obtain local statistically independent features for classification of different facial expressions. ICA was also used to identify the facial action unites in [15]. Besides ICA, Local Binary Patterns (LBP) has also been focused recently for facial expression analysis [23]-[25]. LBP features are tolerant against illumination changes and computational they are very simple. Later on, Local Directional Pattern (LDP) to represent local face features was adopted by focusing on face pixel's gradient information [26]. While extracting LDP features, after obtaining a pixel's directional strengths, the binary values are assigned based on the top t number of strengths where the value of t is determined empirically which varies from experiment to experiment.

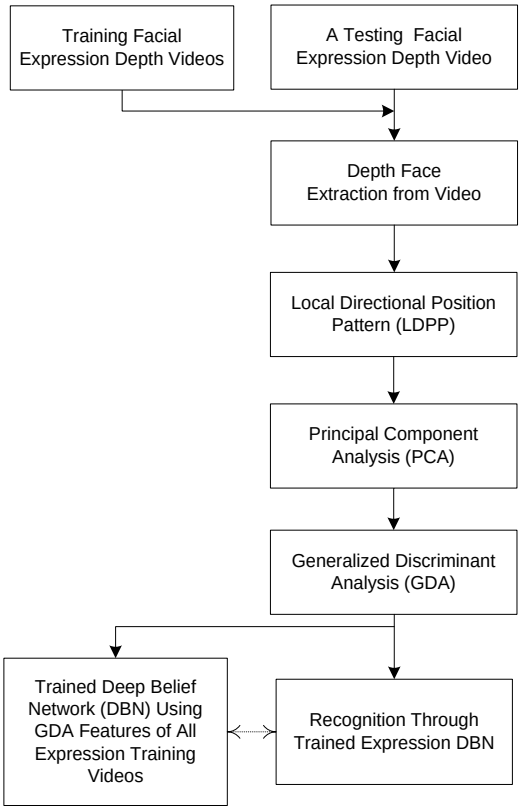


Fig. 1. Basic architecture of proposed FER system.

In this work, a typical LDP is modified to obtain more robust features than LDP. After determining the directional strength of a depth pixel, top directional strengths are obtained in descending order and corresponding strength sign bits are combined with the top directional strength positions in binary to represent robust features. 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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2676238, IEEE Access

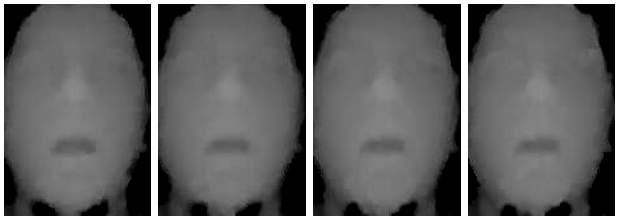
Maximum Entropy Markov Model (MEMM) for action recognition in depth videos. Besides human activities, depth information-based hand movement researches were also done for human computer interaction [44]-[52]. In [45], particle swarm optimization was done for analysis of interacting hands in depth videos.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

(a)

(b)

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

(a)



(b)

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

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) gained a lot of attention since DNN can generate some features from the raw input, contrary to the other classifiers. Also DNN could overcome some limitations of a perceptron that is not able to perform in general pattern recognition However DNN had required too much training time [79]. In 2004, Hilton et al. proposed an improved version of DNN, called Deep Belief Network (DBN), which utilizes Restricted Boltzmann Machine (RBM) for efficient training [80]. There are some other works proposed in [81-85].

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

The depth images are acquired first by a depth camera [2] where the depth videos generate RGB and depth information simultaneously for the objects captured by the camera. The depth sensor video data shows the range of every pixel in the scene as a gray level 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2676238, IEEE Access

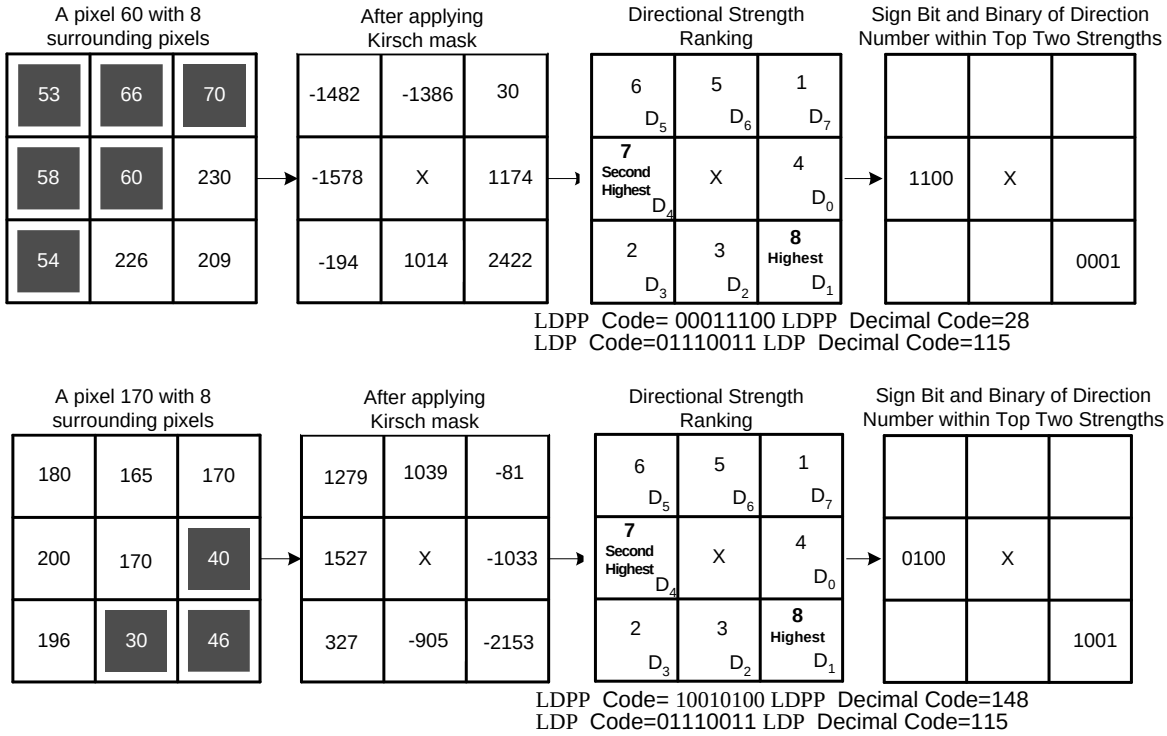


Fig. 6. Two examples for opposite edge pixels where LDP fails to separate them but LDPP successfully.

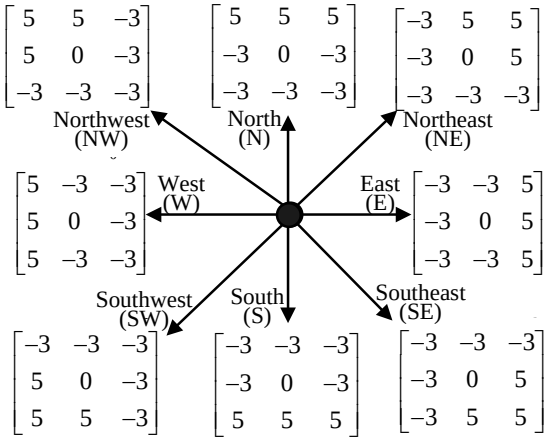


Fig. 4. Kirsch edge masks in eight directions.

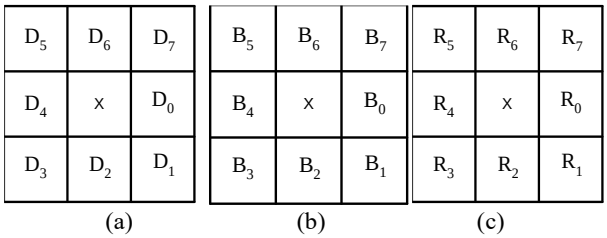


Fig. 5. (a) Edge response to eight directions, (b) sign bit of the edge responses in corresponding direction, (c) ranking of the edge response.

A. Local Directional Position Pattern (LDPP)

The Local Directional Position Pattern (LDPP) assigns an eight-bit binary code to each pixel of an input depth face. This pattern is calculated by considering top two edge strength position with sign from eight different directions. For pixel in the image, the eight directional edge response values $\{D_i\}$ where i=0,1,..,7 are calculated by Kirsch masks. Fig. 4 shows the Kirsch masks. After applying the mask, the directional positions are determined as

$$D_{0,1,..,7} \in E, SE, S, SW, W, NW, N, NE. \tag{1}$$

Thus, LDPP code for a pixel x is derived as

$$LDPP(x) = \sum_{i=0}^{8} L_i \times 2^i, \tag{2}$$

$$L = A \,||\, K, \tag{3}$$

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

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

$$g \in (R_0, R_1, R_2, ..., R_7), \tag{6}$$

$$e \in (R_0, R_1, R_2, ..., R_7), \tag{7}$$

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2676238, IEEE Access

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.

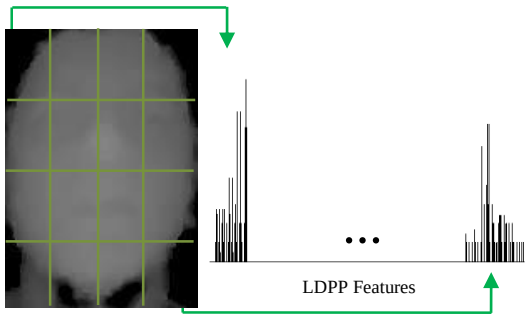


Fig. 7. A depth expression image is divided into small regions and the regions' LDPP histograms are concatenated to represent features for a face.

Thus, an image is transformed in to the LDPP map using LDPP code. The image textual feature is presented by the histogram of the LDPP map of which the s^{th} bin can be defined as

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

where n is the number of the LDPP histogram bins for an image I. Then, the histogram of the LDPP map is presented as

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

To describe the LDPP features, a depth silhouette image is divided into non-overlapping rectangle regions and the histogram is computed for each region as shown in Fig. 7. Furthermore, the whole LDPP feature A is expressed as a concatenated sequence of histograms

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

where g represents the number of non-overlapped regions in the image.

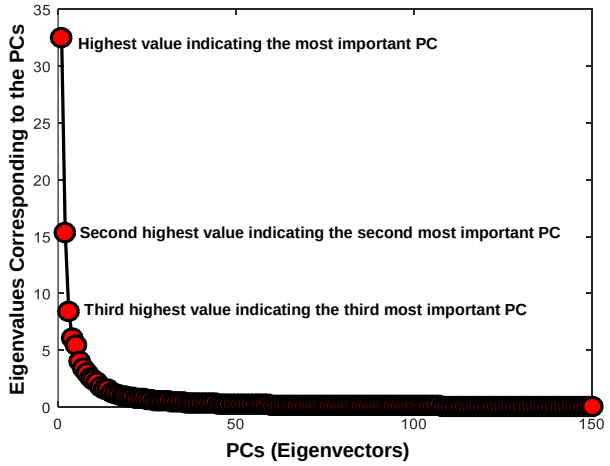


Fig. 8. Top 150 eigenvalues after applying PCA on LDPP features.

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

$$Y = E^T JE \quad (11)$$

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 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

$$C = AE \quad (12)$$

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.

$$\Lambda_{GDA} = \frac{\left|\Lambda^T G_B \Lambda\right|}{\left|\Lambda^T G_T \Lambda\right|} \quad (13)$$

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

$$O = \Lambda_{GDA}^T C. \quad (14)$$

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2676238, IEEE Access

image in a video of length r are augmented further to feed into DBN as

$$T = [O^1 \| O^2 \| O^3 \| ... \| O^r]. \quad (15)$$

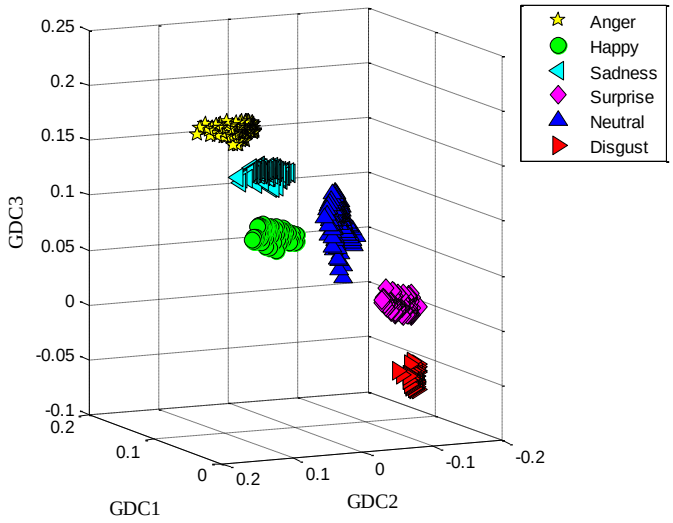


Fig. 9. 3-D plot of GDA features of depth faces from six expressions.

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 [86].

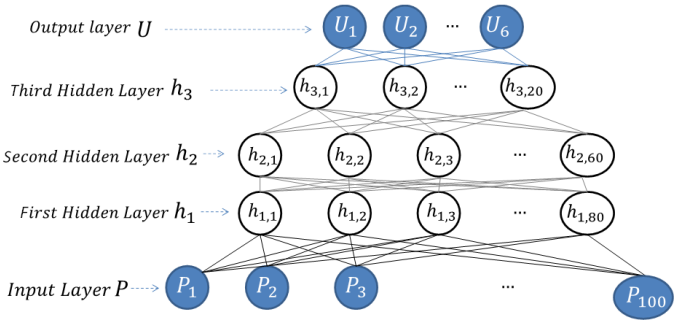


Fig. 10. Structure of a DBN used in this work with 100 input neurons, 80 neurons in hidden layer1, 60 neurons in hidden layer2, 20 neurons in hidden layer3, and 6 output neurons.

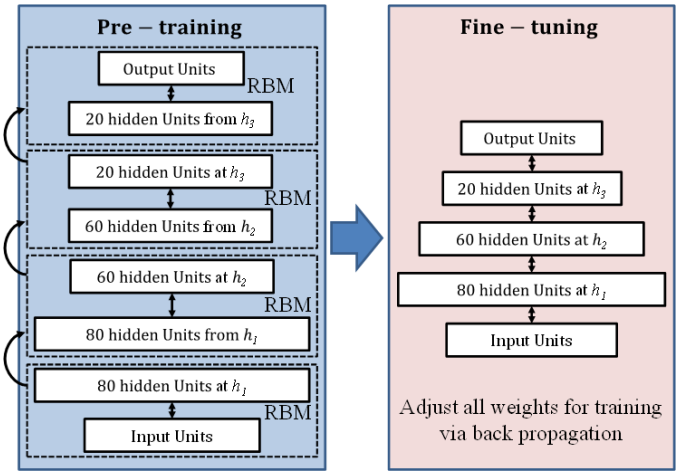


Fig. 11. The pre-training and fine-tuning processes of DBN used in this work.

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.

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

$$p_{recon} = \begin{cases} 1, & f(P + h_1 T) > r \\ 0, & otherwise \end{cases} \quad (17)$$

$$h_{recon} = f(H + p_{recon} T^T) \quad (18)$$

Then, the binary state of the hidden layer h_1 is computed using (16). Later on, the binary state of the visible layer p_{recon} is reconstructed from the binary state of the hidden layer using (17). Then, the hidden layer h_{recon} is re-computed given p_{recon} where

$$f(t) = 1/(1 + exp(-t)) \quad (19)$$

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

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

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2676238, IEEE Access

adjustments of all parameters for fine-tuning. The pre-training and fine-tuning steps are shown in Fig. 11.

IV. Experiments and Results

A depth database was built 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.

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. The mean recognition rate using PCA with HMM on depth faces is 58%. Then, PCA-LDA was tried for FER which achieved mean recognition 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 achieved the recognition rate of 92.50%, the highest in RGB camera-based experiments.

TABLE I
EXPRESSION RECOGNITION USING RGB FACES USING PCA WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**37.50**	0	0	12.50	10	40
Happy	7.50	**47.50**	10	10	20	10
Sadness	0	17.50	**70**	12.50	0	0
Surprise	0	0	0	**75**	15	10
Neutral	0	5	25	10	**60**	0
Disgust	0	7.50	27.50	0	0	**65**
Mean	**58**					

TABLE II.
EXPRESSION RECOGNITION USING RGB FACES USING PCA-LDA WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**50**	0	20	0	0	30
Happy	5	**55**	10	10	15	0
Sadness	0	15	**75**	15	0	0
Surprise	0	0	0	**72.50**	7.50	20
Neutral	0	7.50	30	7.50	**55**	0
Disgust	0	10	20	0	0	**70**
Mean	**61.50**					

TABLE III
EXPRESSION RECOGNITION USING RGB FACES USING ICA WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**75**	0	10	0	5	10
Happy	7.50	**82.50**	10	5	5	
Sadness	0	7.5	**82.50**	10	0	0
Surprise	0	0	0	**80**	12.50	7.50
Neutral	0	0	7.50	10	**82.50**	0
Disgust	0	0	17.50	0	0	**82.50**
Mean	**80.50**					

TABLE IV
EXPRESSION RECOGNITION USING RGB FACES USING LBP WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**82.50**	0	7.50	0	0	10
Happy	5	**80**	10	10	0	
Sadness	0	10	**80**	10	0	0
Surprise	0	0	0	**80**	12.50	7.50
Neutral	0	5	7.50	5	**82.50**	0
Disgust	0	2.50	10	10	0	**77.50**
Mean	**81.25**					

TABLE V
EXPRESSION RECOGNITION USING RGB FACES USING LDP WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**85**	0	5	0	0	10
Happy	0	**82.50**	7.50	10	0	
Sadness	0	7.5	**82.50**	10	0	0
Surprise	0	0	0	**82.50**	10	7.50
Neutral	0	5	7.50	0	**85**	0
Disgust	0	0	10	10	0	**80**
Mean	**82.91**					

TABLE VI
EXPRESSION RECOGNITION USING RGB FACES USING LDPP-PCA-GDA WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**90**	0	5	0	0	5
Happy	2.50	**90**	0	5	5	
Sadness	0	2.50	**87.50**	10	0	0
Surprise	0	0	0	**92.50**	7.50	0
Neutral	0	5	10	5	**90**	0
Disgust	0	2.50	12.20	0	0	**87.50**
Mean	**89.58**					

TABLE VII
EXPRESSION RECOGNITION USING RGB FACES WITH LDPP-PCA-GDA WITH DBN.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**90**	0	0	0	0	10
Happy	2.50	**92.50**	0	0	5	
Sadness	0	2.50	**92.50**	7.50	0	0
Surprise	0	0	0	**95**	5	0
Neutral	0	5	7.50	0	**95**	0
Disgust	0	0	10	0	0	**90**
Mean	**92.50**					

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2676238, IEEE Access

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. The mean recognition rate using PCA with HMM on depth faces is 62%. The mean recognition rate utilizing ICA representation with HMM on the depth facial expression images is 83.50%, indicating better performance than applying PCA as well as PCA-LDA (i.e., 65%) on depth faces. Then, LBP was tried with HMM on the same database that achieved the mean recognition rate of 87.91%. LDP with HMM was then employed and achieved the better recognition rate than LBP i.e., 89.16%. 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. Finally, the proposed LDPP-PCA-GDA features were applied with DBN which showed its superiority over the other methods by achieving the highest mean recognition rate (i.e., 96.67%).

TABLE VIII
EXPRESSION RECOGNITION USING DEPTH FACES WITH PCA WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**50**	0	20	0	15	15
Happy	7.50	**52.50**	10	10	20	10
Sadness	0	17.5	**70**	12.50	0	0
Surprise	0	0	0	**80**	15	5
Neutral	0	5	25	10	**60**	0
Disgust	0	5	27.50	0	0	**62.50**
Mean	**62**					

TABLE IX
EXPRESSION RECOGNITION USING DEPTH FACES WITH PCA-LDA WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**55**	0	25	0	10	10
Happy	10	**60**	10	5	15	0
Sadness	0	17.5	**75**	7.50	0	0
Surprise	0	0	0	**75**	10	15
Neutral	0	10	27.50	0	**62.50**	0
Disgust	0	0	20	12.50	0	**67.50**
Mean	**65**					

TABLE X
EXPRESSION RECOGNITION USING DEPTH FACES USING ICA WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**80**	0	10	0	0	10
Happy	0	**85**	10	5	0	0
Sadness	0	0	**85**	15	0	0
Surprise	0	0	0	**82.50**	17.50	0
Neutral	0	5	15	0	**85**	0
Disgust	0	0	15	2.50	0	**82.50**
Mean	**83.50**					

TABLE XI
EXPRESSION RECOGNITION USING DEPTH FACES USING LBP WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**87.50**	0	12.50	0	0	0
Happy	10	**87.50**	0	0	2.50	0
Sadness	0	0	**85**	0	15	0
Surprise	7.50	0	0	**92.50**	0	0
Neutral	0	2.50	10	0	**87.50**	0
Disgust	0	0	12.50	0	0	**87.50**
Mean	**87.91**					

TABLE XII
EXPRESSION RECOGNITION USING DEPTH FACES USING LDP WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**87.50**	0	10	0	0	2.50
Happy	10	**90**	0	0	0	0
Sadness	0	0	**87.50**	0	12.50	0
Surprise	7.50	0	0	**92.50**	0	0
Neutral	0	2.5	7.50	0	**90**	0
Disgust	0	0	12.5	0	0	**87.50**
Mean	**89.16**					

TABLE XIII
EXPRESSION RECOGNITION USING DEPTH FACES USING LDPP-PCA-GDA WITH HMM.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**90**	0	10	0	0	0
Happy	7.50	**92.50**	0	0	0	0
Sadness	0	0	**90**	10	0	0
Surprise	0	0	0	**95**	0	5
Neutral	0	0	10	0	**90**	0
Disgust	0	0	10	0	0	**90**
Mean	**91.25**					

TABLE XIV
EXPRESSION RECOGNITION USING DEPTH FACES USING LDPP-PCA-GDA WITH DBN.

	Anger	Happy	Sadness	Surprise	Neutral	Disgust
Anger	**95**	0	0	0	0	5
Happy	0	**97.50**	0	2.50	0	0
Sadness	0	0	**97.50**	0	2.50	0
Surprise	0	2.50	0	**97.50**	0	0
Neutral	0	0	2.50	0	**97.50**	0
Disgust	5	0	0	0	0	**95**
Mean	**96.67**					

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 and J. G. Trafton, "Cognitive Architectures for Human-Robot Interaction," *in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14,* Bielefeld, Germany: ACM Press, pp. 504–505, 2014.

[2] S. Tadeusz, "Application of vision information to planning trajectories of Adept Six-300 robot," *in Proceedings of* 21st *International Conference on Methods and Models in Automation and Robotics (MMAR)*, Miedzyzdroje, Poland, pp. 1069-1075, 2016.

[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 Transactions on Consumer Electronics*, vol. 52, no. 3, pp. 726–734, 2006.

[4] C. Padgett and G. Cottrell, "Representation face images for emotion classification," *Advances in Neural Information Processing Systems*, vol. 9, Cambridge, MA, MIT Press, 1997.

[5] S. Mitra and T. Acharya, "Gesture Recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311-324, 2007.

[6] G. Donato, M. S. Bartlett, J. C. Hagar, P. Ekman, and T. J. Sejnowski, "Classifying Facial Actions," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, 1999.

[7] P. Ekman and W. V. Priesen, "Facial Action Coding System: A technique for the Measurement of Facial Movement," *Consulting Psychologists Press*, Palo Alto, CA, 1978.

[8] M. Meulders, P. D. Boeck, I. V. Mechelen, and A. Gelman, "Probabilistic feature analysis of facial perception of emotions," *Applied Statistics*, 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," *Vision Research*, vol. 41, pp. 1179-1208, 2001.

[10] S. Dubuisson, F. Davoine, and M. Masson, "A solution for facial expression representation and recognition," *Signal Processing: Image Communication*, vol. 17, pp. 657-673, 2002.

[11] I. Buciu, C. Kotropoulos, and I. Pitas, "ICA and Gabor Representation for Facial Expression Recognition," *in Proceedings of the IEEE*, pp. 855-858, 2003.

[12] F. chen and K. Kotani, "Facial Expression Recognition by Supervised Independent Component Analysis Using MAP Estimation," *IEICE Transactions on Information and Systems*, vol. E91-D, no. 2, pp. 341-350, 2008.

[13] A. Hyvarinen, J. Karhunen, and E. Oja, "Independent Component Analysis," *John Wiley & Sons*, 2001.

[14] Y. Karklin and M.S Lewicki, "Learning higher-order structures in natural images," *Network: Computation in Neural Systems*, vol. 14, 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 Proceedings of the Sixth Joint Symposium on Neural Computation*, pp. 8-15, 1999.

[16] C. Chao-Fa and F. Y. Shin, "Recognizing Facial Action Units Using Independent Component Analysis and Support Vector Machine," *Pattern Recognition*, vol. 39, pp. 1795-1798, 2006.

[17] A. J. Calder, A. W. Young, and J. Keane, "Configural information in facial expression perception," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 2, pp. 527-551, 2000.

[18] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," *in Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp.200-205, 1998.

[19] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face Recognition by Independent Component Analysis," *IEEE Transaction on Neural Networks*, vol. 13, no. 6, pp. 1450-1464, 2002.

[20] C. Liu, "Enhanced independent component analysis and its application to content based face image retrieval," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 34, no. 2, pp. 1117-1127, 2004.

[21] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," Image and Vision Computing, vol. 16, 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 Transactions on Consumer Electronics, vol. 55, no. 4, pp. 2216–2224, 2009.

[23] T. Ojala, M. Pietikäinen, T. Mäenpää, "Multiresolution gray scale and rotation invariant texture analysis with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 24, pp. 971-987, 2002.

[24] C. Shan, S. Gong, P. McOwan, "Robust facial expression recognition using local binary patterns," *in Proceedings of IEEE International Conference on Image Processing (ICIP)*, pp. 370-373, 2005.

[25] C. Shan, S. Gong, and P. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis. Comput.*, vol. 27, pp. 803-816, 2009.

[26] T. Jabid, M. H. Kabir, O. Chae, "Local Directional Pattern (LDP) A Robust Image Descriptor for Object Recognition", in Proceedings of the IEEE Advanced Video and Signal Based Surveillance (AVSS), pp. 482–487, 2010.

[27] P. Yu, D. Xu, and P. Yu, "Comparison of PCA, LDA and GDA for Palm print Verification," *in Proceedings of the International Conference on Information, Networking and Automation*, pp.148-152, 2010.

[28] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: temporal and static modeling," *Computer Vision and Image Understanding*, vol. 91, pp. 160-187, 2003.

[29] B.M. Asl, S.K. Setarehdan, and M. Mohebbi, "Support vector mashine-based arrhythmia classification using reduced features of heart rate variability signal," *Artif. Intell. Med.*, vol. 44, pp. 51–64, 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 and Applications*, vol. 75, no. 12, pp. 6871-6886, 2016.

[31] W Li., Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3d points," *in Proceedings of Workshop on Human Activity Understanding from 3D Data*, pp. 9-14, 2010

[32] W Li., Z. Zhang, and Z. Liu, "Expandable data-driven graphical modeling of human actions based on salient postures," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1499-1510, 2008.

[33] O. Oreifej and Z. Liu, "Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences," *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 716-723,* 2013.

[34] A. Vieira, E. Nascimento, G. Oliveira, Z. Liu, and M. Campos,"Stop: Space-time occupancy patterns for 3d action recognition from depth map sequences," *in Proceedings of Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications,* pp. 252-259, 2012.

[35] J. Wang., Z. Liu, J. Chorowski, Z. Chen, and Y. Wu, " Robust 3d action recognition with random occupancy patterns," *in Proceedings of European Conference on Computer Vision, pp.* 872-885, 2012.

[36] X. Yang, C. Zhang, and Y Tian, " Recognizing actions using depth motion mapsbased histograms of oriented gradients," *in Proceedings of ACM International Conference on Multimedia,* pp. 1057-1060, 2012

[37] J. Lei, X. Ren, qnd D. Fox, "Fine-grained kitchen activity recognition using rgb-d," *in Proceedings of ACM Conference on Ubiquitous Computing,* pp.208-211, 2012.

[38] A. Jalal ., M.Z. Uddin, J.T. Kim, and T.S. Kim, "Recognition of human home activities via depth silhouettes and transformation for smart homes," Indoor and Built Environment, vol. 21, no 1, pp. 184-190, 2011.

[39] Y. Wang, K. Huang, and T. Tan, "Human activity recognition based on r transform," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[40] H.S. Koppula, R. Gupta, and A. Saxena, "Human activity learning using object affordances from rgb-d videos," *International Journal of Robotics Research*, vol. 32, no. 8, pp. 951-970, 2013

[41] X. Yang and Y. Tian, "Eigenjoints-based action recognition using naive-bayesnearest-neighbor," *in Proceedings of Workshop on Human Activity Understanding from 3D* Data, pp. 14-19, 2012.

[42] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from rgbd images," *in Proceedings of IEEE International Conference on Robotics and Automation*, pp. 842-849, 2012.

[43] A. McCallum, D. Freitag, and F.C.N. Pereira, "Maximum entropy markov models for information extraction and segmentation," *in Proceedings of International Conference on Machine Learning,* pp. 591-598,2000.

[44] H. Hamer, K. Schindler, E. Koller-Meier, and L. Van Gool, "Tracking a hand manipulating an object" *in Proceedings of IEEE International Conference on Computer Vision*, pp. 1475-1482, 2009.

[45] I. Oikonomidis, N. Kyriazis, and A.A. Argyros, "Tracking the articulated motion of two strongly interacting hands ," *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1862-1869, 2012.

[46] H. Hamer, J. Gall, T. Weise, and L. Van Gool, "An object-dependent hand pose prior from sparse training data," *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. pp. 671-678, 2010.

[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 Proceedings of 13th International Conference on Control, Automation and Systems,* pp. 905-909, 2013.

[48] G. J. Iddan and G. Yahav, "3D imaging in the studio (and elsewhere…)," *in Proceedings of SPIE,* vol. 4298, pp. 48-55, 2001.

[49] S. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 6, pp. 873-891, 2005.

[50] T. Pei, T. Starner, H. Hamilton, I. Essa, and J. Rehg, "Learnung the basic units in american sign language using discriminative segmental feature selection," *in Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing,* pp. 4757-4760, 2009.

[51] H.D. Yang, S. Sclaroff, and S.W. Lee, "Sign language spotting with a threshold model based on conditional random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, no.7, pp.1264-1277, 2009.

[52] P. Dreuw, H. Ney, G. Martinez, O. Crasborn, J. Piater, J.M. Moya, and M. Wheatley, "The signspeak project - bridging the gap between signers and speakers," *in Proceedings of International Conference on Language Resources and Evaluation,* pp. 476-481, 2010.

[53] X. Liu and K. Fujimura, "Hand gesture recognition using depth data," *in Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp. 529-534, 2004.

[54] Z. Mo and U. Neumann, "Real-time hand pose recognition using low-resolution depth images," *in Proceedigns of IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1499-1505, 2006

[55] P. Breuer, C. Eckes, and S. Muller, "Hand gesture recognition with a novel IR time-of-flight range camera: a pilot study," in Proceedings of the 3rd international conference on Computer vision/computer graphics collaboration techniques, pp. 247-260, 2007.

[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 Proceedings of Workshop On Time of Flight Camera based Computer Vision,* pp. 1-6, 2008.

[57] E. Kollorz, J. Penne, J. Hornegger, and A. Barke, "Gesture recognition with a time-of-flight camera," *International Journal of Intelligent Systems Technologies and Applications,* vol. 5, pp. 334-343, 2008.

[58] J. Penne, S. Soutschek, L. Fedorowicz, and J. Hornegger, "Robust real-time 3d time-of-flight based gesture navigation," *in Proceedings of International Conference on Automatic Face and Gesture Recognition,* pp. 1-2,2008.

[59] Z. Li and R. Jarvis, "Real time hand gesture recognition using a range camera," *in Proceedings of Australasian Conference on Robotics and Automation,* 2009.

[60] H. Takimoto, S. Yoshimori, Y. Mitsukura, and M. Fukumi, "Classification of hand postures based on 3d vision model for human-robot interaction," *in Proceedings of International Symposium on Robot and Human Interactive Communication,* pp. 292-297, 2010.

[61] M. Van den Bergh, and L. Van Gool, "Combining rgb and tof cameras for real-time 3d hand gesture interaction," *in Proceedings of IEEE Workshop on Applications of Computer Vision,* pp. 66-72,. 2011.

[62] J. Marnik, "The polish finger alphabet hand postures recognition using elastic graph matching," *Computer Recognition Systems 2,* vol. 45, pp.454-461, 2007

[63] M.D. Breitenstein, D. Kuettel, T. Weise, L. Van Gool, and H. Pfister, "Real-time face pose estimation from single range images," *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1-8, 2008.

[64] Q. Cai, D. Gallup, C. Zhang, and Z. Zhang, "3d deformable face tracking with a commodity depth camera," *in Proceeding of European Conference on Computer Vision,* pp.229-242, 2010

[65] L.P. Morency, P. Sundberg, and T. Darrell, "Pose estimation using 3d view-based eigenspaces," in Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 45-52, 2003.

[66] E. Seemann, K. Nickel, and R. Stiefelhagen, "Head pose estimation using stereo vision for human-robot interaction," *in Proceedings of Sixth IEEE International Conference on on Automatic Face and Gesture Recognition*, pp. 626-631, 2004.

[67] A. Mian, M. Bennamoun, and R. Owens, "Automatic 3d face detection, normalization and recognition," *in Proceedings of Third International Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 735-742, 2006.

[68] X. Lu and A.K. Jain, "Automatic feature extraction for multiview 3d face recognition," *in Proceedings of 7th International Conference on Automatic Face and Gesture Recognition,* pp. 585-590, 2006.

[69] T. Weise, B. Leibe, and L. Van Gool,"Fast 3d scanning with automatic motion compensation," *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1-8, 2007.

[70] T. Weise, S. Bouaziz, H. Li, and M. Pauly, "Realtime performance-based facial animation," *ACM Transactions on Graphics,* vol. 30, no. 4, article 77, pp. 1-10, 2011.

[71] M.D. Breitenstein, J. Jensen, C. Hoilund, T.B. Moeslund, and L. Van Gool, "Head pose estimation from passive stereo images," *in proceedings of 16th Scandinavian Conference on Image Analysis*, pp. 219-228, 2009.

[72] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,* pp. 617-624, 2011.

[73] Y. Sun and L. Yin, "Automatic pose estimation of 3d facial models," *in Proceedings of International Conference on Pattern Recognition,* pp. 1-4, 2008.

[74] M. Segundo, L. Silva, O. Bellon, and C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 40, no. 5, pp. 1319-1330, 2010.

[75] K.I. Chang, K.W. Bowyer, and P.J. Flynn, "Multiple nose region matching for 3d face recognition under varying facial expression," *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1695-1700, 2006.

[76] P. Nair and A. Cavallaro, "3-d face detection, landmark localization, and registration using a point distribution model," *IEEE Transactions* on Multimedia, vol. 11, no. 4, pp. 611-623, 2009.

[77] P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," *IEEE Transaction on Information and Security*, vol. 1, pp. 3-11, 2006.

[78] L. R. Rabiner, "A Tutorial on Hidden Markov Modes and selected application in speech recognition," *in Proceedings of IEEE*, vol. 77, pp. 257-286, 1989.

[79] M. Minsky and S. Papert, "Perceptrons. An Introduction to Computational Geometry," M.I.T Press, Cambridge, vol. 165, no. 3895, pp. 780-782, 1969.

[80] G. E. Hinton, S. Osindero, Y. The, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527-1554, 2006.

[81] M. M. Hassan, K. Lin, K., X. Yue, J. Wan, "A multimedia healthcare data sharing approach through cloud-based body area network". Future Generation Computer Systems, Vol. 66, pp. 48–58, 2017

[82] Y. Zhang, M. Chen, D. Huang, D. Wu, Y. Li, "iDoctor: Personalized and Professionalized Medical Recommendations Based on Hybrid Matrix Factorization", Future Generation Computer Systems, vol. 66, pp. 30-35, 2017.

[83] Y. Zhang, "Grorec: A group-centric intelligent recommender system integrating social, mobile and big data technologies", IEEE Transactions on Services Computing, vol. 9, no. 5, pp. 786-795, 2016.

[84] Y. Zhang, M. Chen, S. Mao, L. Hu, V. Leung, "CAP: Crowd Activity Prediction Based on Big Data Analysis", IEEE Network, Vol. 28, No. 4, pp. 52-57, July 2014.

[85] Y. Zhang, D. Zhang, M. M. Hassan, A. Alamri, L. Peng, "CADRE: Cloud-Assisted Drug REcommendation Service for Online Pharmacies", ACM/Springer Mobile Networks and Applications, Vol. 20, No. 3, pp. 348-355, 2015.

[86] A. Fischer and C., "Training restricted Boltzmann machines: An introduction," *Pattern Recognition*, vol. 47, no. 1, pp. 25-39, 2014.