Computers and Electrical Engineering 63 (2017) 114–125

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

A facial expression recognition system using robust face features from depth videos and deep learning☆

Md. Zia Uddin[a], Mohammed Mehedi Hassan[b,*], Ahmad Almogren[b], Mansour Zuair[b], Giancarlo Fortino[c], Jim Torresen[a]

[a] Department of Informatics, University of Oslo, Norway
[b] College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
[c] Department of Informatics, Modeling, Electronics, and Systems, University of Calabria, Italy

ARTICLE INFO

Article history:
Received 13 October 2016
Revised 21 April 2017
Accepted 21 April 2017
Available online 29 April 2017

Keywords:
Depth image
Facial expression recognition
Modified local directional patterns
Generalized discriminant analysis
Deep belief network

ABSTRACT

This work proposes a depth camera-based robust facial expression recognition (FER) system that can be adopted for better human machine interaction. Although video-based facial expression analysis has been focused on by many researchers, there are still various problems to be solved in this regard such as noise due to illumination variations over time. Depth video data in the helps to make an FER system person-independent as pixel values in depth images are distributed based on distances from a depth camera. Besides, depth images should resolve some privacy issues as real identity of a user can be hidden. The accuracy of an FER system is much dependent on the extraction of robust features. Here, we propose a novel method to extract salient features from depth faces that are further combined with deep learning for efficient training and recognition. Eight directional strengths are obtained for each pixel in a depth image where signs of some top strengths are arranged to represent unique as well as robust face features, which can be denoted as Modified Local Directional Patterns (MLDP). The MLDP features are further processed by Generalized Discriminant Analysis (GDA) for better face feature extraction. GDA is an efficient tool that helps distinguishing MLDP features of different facial expressions by clustering the features from the same expression as close as possible and separating the features from different expressions as much as possible in a non-linear space. Then, MLDP-GDA features are applied with Deep Belief Network (DBN) for training different facial expressions. Finally, the trained DBN is used to recognize facial expressions in a depth video for testing. The proposed approach was compared with other traditional approaches in a standalone system where the proposed one showed its superiority by achieving mean recognition rate of 96.25% where the other approaches could make 91.67% at the best. The deep learning-based training and recognition of the facial expression features can also be undertaken with cloud computing to support many users and make the system faster than a standalone system.

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. J. Wan.

* Corresponding author.
E-mail addresses: mdzu@ifi.uio.no (Md.Z. Uddin), mmhassan@ksu.edu.sa (M.M. Hassan), ahalmogren@ksu.edu.sa (A. Almogren), zuair@ksu.edu.sa (M. Zuair), g.fortino@unical.it (G. Fortino), jimtoer@ifi.uio.no (J. Torresen).

http://dx.doi.org/10.1016/j.compeleceng.2017.04.019
0045-7906/

1. Introduction

Human robot interactions (HRI) have attracted many researchers recently to contribute in developing smartly controlled healthcare systems [1,2]. Humans very often use nonverbal cues in their daily lives where the cues include hand gestures, facial expressions, and tone of the voice to express feelings. Hence, HRI systems should include these to make full benefit of natural interaction with the users. In a ubiquitous robotic healthcare system, HRI systems could be considerably improved if robots could understand peoples' emotions based on analyzing facial expressions and react as friendly as possible according to their current emotional states. Providing emotional healthcare support using robots could also be important to improve the quality of life. Humans' mental states are revealed through emotions in their daily situations where positive emotions represent healthy mental states by carrying positive facial expressions (e.g., pleasure and happiness). On the other hand, negative emotions can represent unhealthy mental states by carrying negative facial expressions such as anger and sadness. Thus, both positive and negative emotions can closely affect peoples' emotional health in their daily lives. To improve emotional health, a robust facial expression recognition (FER) system plays a key role in understanding mental states over time by the analysis of emotional behavior patterns.

Basically, vision-based FER systems can be categorized into two main types: pose-based and spontaneous. Pose-based FER systems usually recognize artificial facial expressions where expressions are produced by people by asking them to express a selection of expressions in sequence. On the contrary, spontaneous FER systems recognize the facial expressions that people does spontaneously in daily life such as during conservations and while watching movies. This work focuses on pose-based FER due to unavailability of pure spontaneous facial expression depth database. A video-based FER system consists of two types of classifications: frame-based and sequence-based. In the former one, only one frame is utilized to recognize different facial expressions. On the other hand, sequence-based methods apply temporal information in the frame sequences to recognize different facial expressions in videos. As image sequence-based FER systems contain more information than single frame-based one, this work utilizes face image sequences i.e., videos. A typical video-based FER system consists of three main parts: preprocessing, feature extraction, and recognition. In preprocessing, face area is detected in an image of a video. Feature extraction handles extracting salient features from each face to distinguish facial expressions. At last, face features are trained and used to recognize different facial expressions.

1.1. Organization of the paper

The rest of the paper is organized as follows. Section 2 discusses some significant research works related to the proposed system. Sections 3 and 4 illustrate the feature extraction process from depth images followed by expression modelling using deep learning, respectively. Furthermore, Section 5 describes experimental results using different approaches including the proposed one. Finally, Section 6 draws the conclusion of the work.

2. Related works

A huge amount of research works has been observed in developing reliable FER systems as it contributes to a large range of application domains in computer visions, image processing, and pattern classification [3–15]. A very challenging problem to solve in these is the ability of the computing systems to detect human faces to recognize underlying expressions in them such as angry, happy, neutral, sad, and disgust. Hence, accurate recognition of facial expressions is still considered to be a major challenge due to some parameters such as presence of noise from the environments due to illumination variation over time in the scene. Hence, robust FER still demands much attention to support various applications. The most important aspect for any FER approach is to find efficient feature representation in face images. Features in FER are considered as efficient when they can minimize within-class variations and maximize between-class variations. Hence, the main goal of feature extraction is to find a robust representation of face features which can provide robustness during the recognition process. Based on the features used in FER systems so far, feature extraction methods can be divided into two main categories i.e., geometric and appearance-based. In geometric feature-based FER, geometric feature vectors are formed considering geometric relationships such as angles and positions between different face parts such as eyes, ears, and nose. This has been a popular method but the effectiveness is highly dependent on accurate detection of facial components in the images, which can be a very challenging task in unconstrained environments and dynamic scenarios. Hence, researchers in this field moved their research directions over to appearance-based face analysis to obtain a more robust FER system. Appearance-based FER methods focus on facial appearance and try to do different analysis such as by applying filters on the whole face image. Among all the vision-based FER research, most of the work has been done using appearance-based methods. This work also describes an appearance-based FER system.

To represent facial expression features in a video, Principal Component Analysis (PCA) has been mostly applied in FER systems [3–6]. In [3], PCA was tried to recognize action units to represent and recognize different facial expressions. In [5], the authors applied PCA for providing a facial action coding system where different facial expressions were modelled and recognized. Independent Component Analysis (ICA), a higher level statistical approach than PCA was adopted later in FER works for statistically independent local face feature extraction [7–15]. In [10], ICA was adopted to obtain statistically independent features focusing on local face components (e.g., nose, lips, and eyebrows) in different facial expressions. In [11], ICA was adopted in FER to analyze facial action units in different facial expression images. For local feature extraction, some

researchers also focused on Local Binary Patterns (LBP) for various applications including FER [16]. LBP features have better tolerance than ICA against illumination changes in a scene. Besides, LBP features are computationally simple. In a facial expression image for feature extraction, edge pixels are considered to be the most important pixels amongst all pixels in the face. Hence, pixel features considering gradient information or edge information should be focused more for robust FER. As LBP features could not represent image pixel's gradient information, it was improved later and named as Local Directional Pattern (LDP). In FER, LDP represents local face features based on gradient information in eight prominent surrounding directions of a pixel [17]. While extracting typical LDP features, the binary values are assigned considering top number of edge strengths where the number is determined empirically and varies from experiment to experiment. One big limitation in LDP is that LDP does not consider directional strength signs for pixels whereas the signs of the direction strengths can contribute well in distinguishing edge pixels with same strength magnitudes but opposite signs. This limitation of typical LDP is resolved in this work, which makes the proposed FER system robust.

In our approach, after obtaining directional edge strengths for a depth pixel as for the regular LDP, top edge strengths are organized in descending order and then their signs are considered respectively to represent robust features. In the regular LDP, top directional strengths are considered irrespective of signs. Basically, dark pixels in an edge mostly represent negative directional strengths and bright pixels show positive strengths. Hence, considering only strengths excluding signs may result in the same LDP code for two opposite kinds of edge pixels. Considering signs should resolve this issue to represent robust features. This modification of typical LDP is denoted as Modified LDP or MLDP in this work. To make the MLDP features more robust, Generalized Discriminant Analysis (GDA) is adopted here since GDA can discriminate features from different facial expression classes in nonlinear spaces [18]. GDA is basically a generalization of typical Linear Discriminant Analysis (LDA) where inputs go through a kernel to help them clustering in nonlinear space.

This work focuses on depth camera-based FER system since RGB cameras have some limitations for person-independent facial expression analysis. For instance, RGB face images can reveal a person's identity very easily which can lead to problems related to privacy. On the other hand, depth images can resolve this issue as pixel intensities in depth images are distributed according to the camera distance and hence, person's identity can be hidden. Besides, different face parts can be distinguished in a depth face image based on the depth values. As depth images can represent more robust features than RGB images, depth images have been in various image processing and pattern recognition applications such as body motion recognition [19,20], hand motion recognition [21,22], and face analysis in biometric authentication [23,24].

To model salient face features in videos with facial expressions, Hidden Markov Models (HMMs) have been adopted in some works [10,15]. Recently, deep learning has been getting a lot of attentions in the research community since features are extracted from input data in addition to normal training process. Deep Neural Network (DNN) was the first deep learning approach proposed to train and recognize patterns from high dimensional data [25]. One major disadvantage of DNNs is that they are time-consuming to train. DNN sometimes causes overfitting problems as well. However, to solve different problems in DNN, it was improved and named as Deep Belief Network (DBN). DBN applies Restricted Boltzmann Machine (RBM) to train features in different hidden layers. Basically, it is RBM that makes DBN very robust and more practically implementable than a typical DNN [26]. Considering the robustness of DBN, we adopted it in this work to model and recognize different facial expressions in depth videos.

2.1. Proposed approach

The proposed system uses a unique approach called MLDP with GDA, and DBN for robust FER. Typical LDP features are modified first to create MLDP features from depth images. The MLDP features are then further enhanced by a supervised classification called GDA. Then, a DBN is trained using robust MLDP-GDA features obtained from different training facial expression depth videos. Finally, the trained expression DBN is used for testing a depth video of facial expression. Fig. 1 shows the basic flows of our FER system. Although the system can work well as a standalone system, cloud computing can be adopted to make the system faster and allow many concurrent users [27,28].

3. Feature extraction

A commercial depth camera is used in this work to acquire depth and RGB videos [2]. Fig. 2(a) and (b) represent a gray and corresponding depth image of a surprise expression, respectively. In a depth image of a facial expression, near face parts (e.g., nose) contain brighter pixel intensities than parts further away (e.g., eyes) which contain comparatively darker intensities. Fig. 3 shows a sequence of depth images of faces for surprise and disgust expressions.

First of all, Modified Local Directional Pattern (MLDP) features are obtained for each expression image. MLDP assigns a six-bit binary code for a pixel in a depth face image. This value is calculated to contain the relative edge strengths of a pixel in eight directions. Each pixel's edge strengths are calculated by typical Kirsch masks [17]. Fig. 4 shows the masks used in this work. In an image, the presence of an edge in any location indicates high response in that direction. Therefore, knowing prominent directional edge strengths for each pixel should contribute in representing robust features for each pixel, especially features based on which expressions are distinguished. For MLDP feature of a pixel, six top edge directional strength values are organized in descending order. Then, the corresponding signs of them are considered. Thus, MLDP code

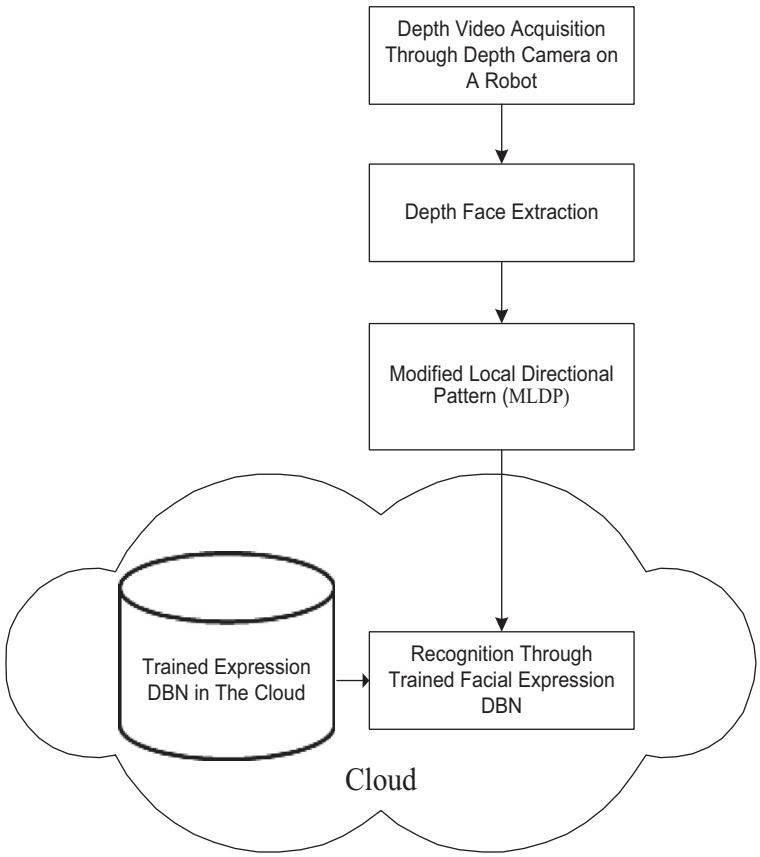

Fig. 1. Proposed steps for FER system using images from a camera on a robot.

Fig. 2. (a) A gray pixel image and (b) corresponding depth image of a surprise expression.

for a pixel is derived as

$$MLDP = \sum_{i=0}^{d} P(D_i) \times 2^i, \tag{1}$$

$$P(a) = \begin{cases} 1 & a < 0 \\ 0 & a \geq 0 \end{cases} \tag{2}$$

where d refers to the number of directions representing top strengths, and D_i the strength value of ith top direction. Function $P(a)$ determines the sign bit of a where a represents the edge strength of a pixel to a specific direction. In our experiments,

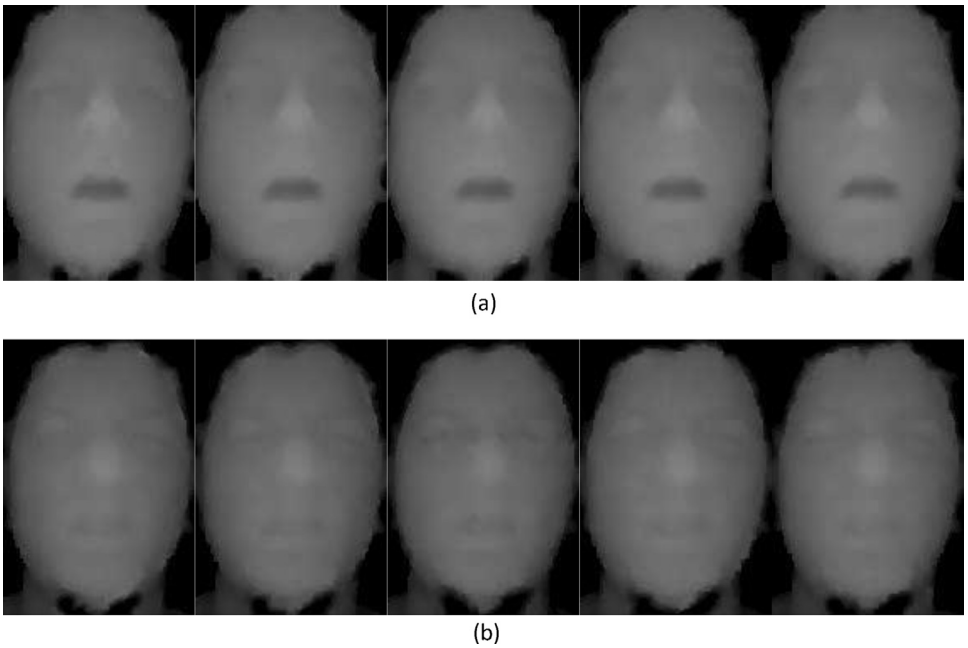

Fig. 3. A sequence of depth images of faces of (a) surprise and (c) disgust expressions.

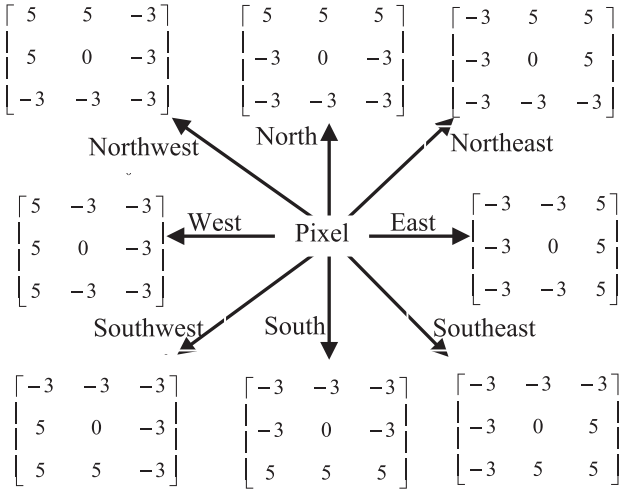

Fig. 4. Kirsch edge masks for eight directions.

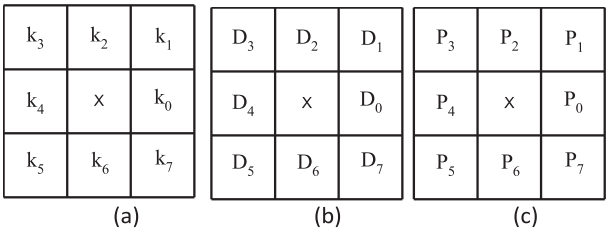

Fig. 5. (a) Edge mask response to eight directions, (b) directional strength ranking position, and (c) MLDP binary bit positions.

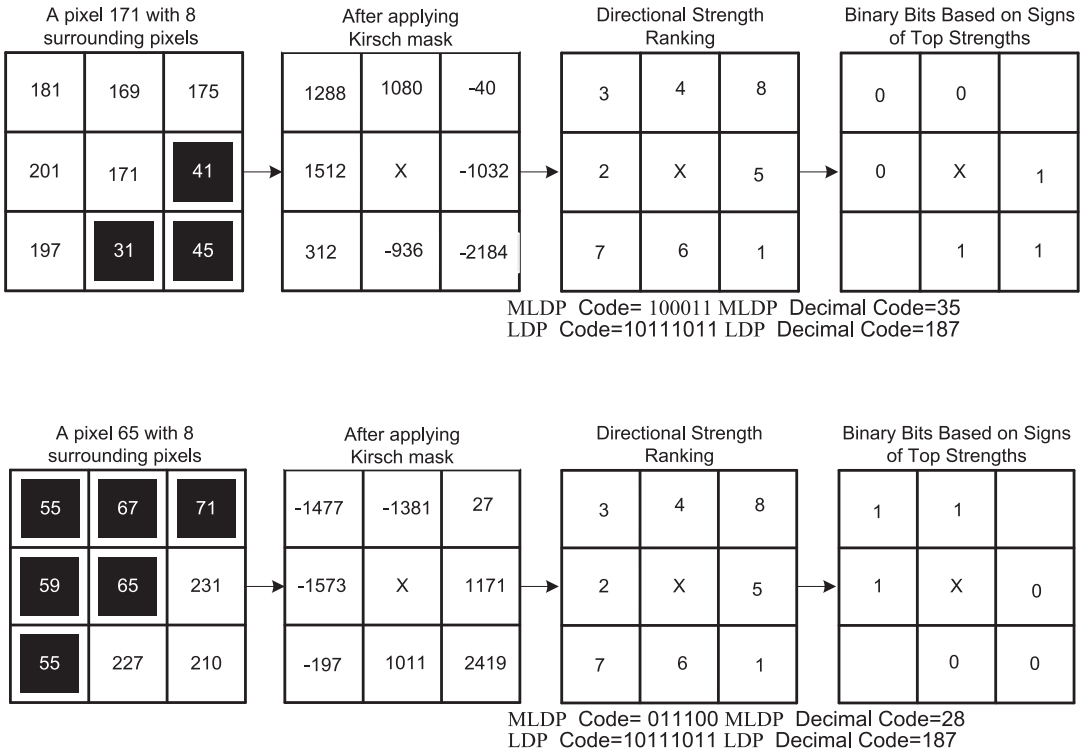

Fig. 6. Examples to generate feature patterns for two kind of pixels where LDP completely fails but MLDP succeeds to distinguish them well.

we considered $d=6$ as considering higher value of d could not bring us better results. Fig. 5 depicts the mask responses, directional strength rankings, and MLDP bit positions. Fig. 6 shows two examples of MLDP code computation for two different center pixels with value of 171 and 65, respectively. It is to be noticed in Fig. 6 that typical LDP generates the same code for edge pixels whereas our proposed MLDP generates different codes for both, indicating robustness of MLDP over LDP. For the upper pixel (i.e., 171) in the figure, the first or top ranked edge strength is −2184. Hence, the sign bit of −2184 (i.e., 1) is the first bit of the MLDP code for the pixel. The second top ranked edge response is 1512. Hence, the sign bit of 1512 (i.e., 0) should be the second bit of MLDP code for the pixel. Similarly, the third, fourth, fifth, and sixth bits are represented with their sign bits of the edge responses i.e., 0, 0, 1, and 1, respectively. Therefore, the MLDP code for the upper pixel is 100,011 and for the lower pixel is 01,1100. LDP fails in this situation by generating completely same code for both pixels (i.e., 10,011,101) as their directional rankings are the same although the pixels belong to completely opposite edges. Thus, MLDP indicates more robustness than typical LDP, especially to create patterns for pixels as mentioned above. Besides, the LDP code is fixed to eight bit but MLDP code can be represented with less than eight bits, resulting in reduced dimensional features being potentially more robustness. MLDP features represent detail information of a face image (e.g., edges and corner features). MLDP histogram calculated over the whole facial expression image considers only feature patterns for whole face irrespective of different locations in the face image. Hence, the whole image is divided in to some sub-regions and MLDP histograms are calculated for them. Then, the non-overlapped sub-regions' histograms are augmented to obtain histogram features for a single face image. For a sub-region in a face image, the image textual feature Z_s for sth bin in a histogram of the MLDP map is represented as

$$Z_s = \sum_{x,y} I\{MLDP(x,y) = s\}, s = 0, 1, \ldots n-1 \tag{3}$$

where n is the number of MLDP histogram bins, which is basically 256. Then, MLDP histogram for a local sub-region is represented as

$$H = (Z_0, Z_1, \ldots, Z_{n-1}). \tag{4}$$

Furthermore, histograms of all g sub-regions are concatenated to represent MLDP features A for a facial expression image as

$$A = (H^1, H^2, \ldots, H^g). \tag{5}$$

After MLDP, the features are enhanced more by applying a supervised classification method called Generalized Discriminant Analysis (GDA). GDA is a generalization of typical Linear Discriminant Analysis (LDA). Basically, LDA tries to distinguish

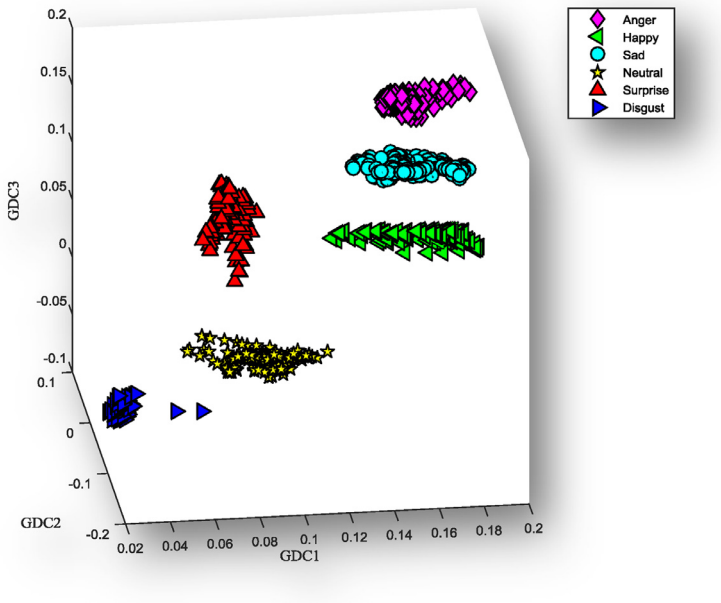

Fig. 7. 3-D plot of GDA on MLDP features of depth faces from six expressions.

the samples of different expressions in a linear space. Hence, it faces difficulties when the samples are not separable linearly. To overcome this limitation, LDA is modified and named as GDA to separate samples non-linearly where the features of each face from different expressions go through a kernel (e.g., Gaussian) before regular LDA is applied. The key finding of GDA is maximizing the scatterings of samples between classes while minimizing the scatterings of samples within classes. GDA criterion function can be defined as

$$G_{GDA} = \frac{\left|G^T S_B G\right|}{\left|G^T S_T G\right|} \tag{6}$$

where S_B and S_T are two matrices representing between-class and total class scatter matrices respectively after the MLDP features of each face have gone through a kernel function. Once, GDA feature space is created based on MLDP features, MLDP features F of each face is projected on the feature space G_{GDA} as

$$F = G_{GDA}^T A. \tag{7}$$

Fig. 7 depicts a 3-D plot of MLDP-GDA features of training depth faces from different facial expressions. It indicates a very good maximization of scatterings among the between-class samples and minimization of scatterings among within-class samples. As depth face images of different classes are well-clustered as shown in the figure, GDA seems to be an appropriate choice for FER from depth video. Furthermore, MLDP-GDA features from each image in a facial expression video are augmented to represent the features for the video. For a depth video of length r, the MLDP-GDA feature vector Q can be obtained as

$$Q = F^1, F^2, \ldots, F^r. \tag{8}$$

Once MLDP-GDA features are obtained from all the facial expression depth videos, the next step is to train a DBN to recognize expressions.

4. Expression modeling

A DBN is adopted in this work for training and recognition of facial expressions using the MLDP-GDA features from the depth videos. DBN follows a stochastically training architecture where the objective function depends mainly on the learning purpose. Basically, a DBN is trained as a discriminative model to solve the classification problem. The key advantage of DBN is that it has the capability of learning features from raw inputs which makes it different from typical deep learning structures such as DNN. DBN achieves its learning objective by a layer-by-layer learning strategy where current level features are learned from previous layers. Generally, next-level features are believed to be more extensive than the current level. DBN has two phases: namely pre-training and fine-tuning. The former phase consists of layer-by-layer stack of Restricted Boltzmann Machines (RBMs). Once the DBN is pre-trained, the weights of the whole network are utilized in a typical fine-tuning algorithm.

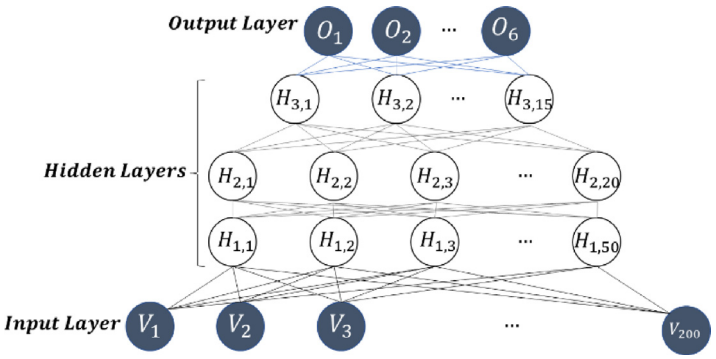

Fig. 8. Structure of a DBN with 200 input neurons, 50 neurons in hidden layer1, 20 neurons in hidden layer2, 15 neurons in hidden layer3, and 6 output neurons.

Fig. 8 depicts a sample DBN with one input layer, three hidden layers where RBM is used, and one output layer. Once the weights of the first layer RBM is trained then weights of the first hidden layer become fixed. Fixed weights of the first hidden layer are then utilized to train and adjust the weights of the second hidden layer's RBM. Following similar fashion, the third hidden layers' RBM is trained using the weights from previous layer. To adjust weights in each hidden layer, a contrastive divergence algorithm is applied in this work [26]. Contrastive divergence is the mostly used algorithm for training RBMs in DBN where it tries to determine approximation of gradient (i.e., accurate direction) information of log-likelihoods based on a Markov chain observed in previous step.

The training steps for a RBM starts with initialization of a bias vector H for the hidden layer and a weight matrix W. Generally, they are set to zero in the beginning. The following steps are used for training of the RBMs. After initialization, the binary state values in hidden layer H_1 are computed using (9). Then, state values of visible layer V_r are reconstructed from the binary state values of the hidden layer H_1 using (10). Then, the hidden layer H_r is re-adjusted using (11) and (12).

$$H_1 = \begin{cases} 1, & f\left(H + V_1 W^T\right) > r \\ 0, & otherwise \end{cases}. \tag{9}$$

$$V_r = \begin{cases} 1, & f(V + H_1 W) > r \\ 0, & otherwise \end{cases}. \tag{10}$$

$$H_r = f\left(H + V_r W^T\right). \tag{11}$$

$$f(t) = 1/(1 + exp(-t)). \tag{12}$$

Finally, weights of the current layer are computed based on a summation of weights multiplied by inputs in the previous layer. Once the pre-training process of the RBMs in the different layers is finished, a classical back propagation algorithm is adopted as supervised training for a fine-tuning process in DBN.

5. Experimental results

Six different facial expressions were recorded and used for training and testing in this work. The expressions were sad, surprise, anger, happy, disgust, and neutral. For each facial expression, 40 videos were recorded where each video consisted of ten frames. Then, a two-fold cross validation was done on videos of each expression. In each fold, 20 videos were considered for training set and other 20 for testing set. Finally, results of these two-fold experiments were accumulated to represent the recognition rate for each expression using the selected FER approaches. Thus, a total of 40 videos were used for training and 40 for testing each expression.

RGB camera-based FER experiments were conducted first. Experimental results after applying different FER approaches on RGB camera based videos are shown in Table 1. PCA with HMM on gray faces from RGB cameras generated the lowest recognition performance (i.e., 59.17% mean recognition rate). To check the recognition performance with a potentially better approach, local face feature extraction method ICA was combined with HMM to model time-sequential face features and obtained better recognition performance (i.e., 81.67% mean recognition rate). To focus more on local face features, illumination tolerable features like LBP was tried with HMM that obtained the mean recognition rate of 82.50%. LBP with HMM showed similar FER performance as ICA. Furthermore, a gradient information-based face feature extraction technique was adopted (i.e., LDP) and combined with HMM which achieved 85% mean recognition rate, which is better performance than PCA, ICA, and LDP. Then, the more robust feature extraction approach MLDP was combined with non-linearly supervised

Table 1
Facial expression recognition rates using RGB images with different approaches.

Approach	Expression	Recognition Rate	Mean
PCA with HMM	Anger	55%	59.17
	Happy	52.50	
	Sad	60	
	Surprise	62.50	
	Neutral	60	
	Disgust	65	
ICA with HMM	Anger	85	81.67
	Happy	77.50	
	Sad	80	
	Surprise	80	
	Neutral	87.50	
	Disgust	80	
LBP with HMM	Anger	80	82.50
	Happy	85	
	Sad	80	
	Surprise	82.50	
	Neutral	80	
	Disgust	87.50	
LDP with HMM	Anger	85	85
	Happy	80	
	Sad	87.50	
	Surprise	85	
	Neutral	85	
	Disgust	87.50	
MLDP-GDA with HMM	Anger	85	89.16
	Happy	87.50	
	Sad	90	
	Surprise	95	
	Neutral	87.50	
	Disgust	90	
MLDP-GDA with DBN	Anger	92.50	93.33
	Happy	95	
	Sad	92.50	
	Surprise	95	
	Neutral	92.50	
	Disgust	92.50	

classifier GDA and tried with HMMs which achieved 89.16% mean recognition rate. Furthermore, proposed MLDP-GDA features were adopted with deep learning approach (i.e., DBN) which showed the mean recognition rate of 93.33%, the highest performance in RGB camera-based experiments.

After getting inspiration from performances obtained via RGB camera-based experiments, the RGB camera was replaced by a depth camera and FER experiments were done under same settings as the RGB camera-based ones. Depth video-based FER experiments achieved better recognition performance than the RGB-based ones as shown in Table 2. At first, PCA-HMM was applied on depth images and achieved 62.50% mean recognition rate, which is the lowest in depth camera-based FER approaches. Like RGB camera-based experiments, we tried ICA with HMM on depth videos that achieved 83.75% mean recognition rate, indicating ICA to be superior to PCA on depth images of faces. More local feature extraction approaches were experimented on depth images for better FER. Then, LBP-HMM was tried that obtained the mean recognition rate of 84.58%. Thus, LBP showed similar performance as ICA on depth images of faces. Furthermore, pixel's edge strength-based approach LDP was applied with HMM that generated higher recognition rate than LBP i.e., 87.08%, shows superiority of LDP over LBP, ICA, and PCA features on depth faces. Then, we continued our FER experiments with more robust feature extraction approach i.e., the proposed MLDP-GDA features. Then, MLDP-GDA was tried with HMMs on depth videos, and it achieved 91.67% mean recognition rate. Finally, the proposed MLDP-GDA features were tried with DBN which showed superior performance over all other FER approaches by obtaining the mean recognition rate of 96.25%, the highest recognition performance among all approaches using both RGB and depth cameras.

6. Conclusion

A depth video-based robust facial expression recognition system has been investigated in this work by applying MLDP with GDA to represent expression features and DBN for training as well as testing. The proposed approach was compared against other traditional approaches where it proved its superiority over them by obtaining the maximum recognition rate of 96.25% whereas others could obtain 91.67% at the best. In future, we plan to focus on analyzing the proposed system in

Table 2
Facial expression recognition rates using depth images with different approaches.

Approach	Expression	Recognition Rate	Mean
PCA with HMM	Anger	55%	62.50
	Happy	52.50	
	Sad	65	
	Surprise	70	
	Neutral	65	
	Disgust	67.50	
ICA with HMM	Anger	82.50	83.75
	Happy	80	
	Sad	85	
	Surprise	87.50	
	Neutral	85	
	Disgust	82.50	
LBP with HMM	Anger	85	84.58
	Happy	82.500	
	Sad	85	
	Surprise	85	
	Neutral	85	
	Disgust	85	
LDP with HMM	Anger	85	87.08
	Happy	85	
	Sad	87.50	
	Surprise	90	
	Neutral	87.50	
	Disgust	87.50	
MLDP-GDA with HMM	Anger	92.50	91.67
	Happy	90	
	Sad	92.50	
	Surprise	95	
	Neutral	90	
	Disgust	90	
MLDP-GDA with DBN	Anger	97.50	96.25
	Happy	95	
	Sad	95	
	Surprise	97.50	
	Neutral	97.50	
	Disgust	95	

various deep learning-based dynamic cloud environments to support concurrent users in real time. Furthermore, we plan to improve our approach so that it can be applied for spontaneous facial expression recognition in regular daily environments.

Acknowledgement

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through research group no (RGP- 1437-35). This work is partially supported by The Research Council of Norway as a part of the Multimodal Elderly Care Systems (MECS) project, under grant agreement 247697.

References

[1] Baxter P, Trafton JG. Cognitive architectures for human-robot interaction. In: Proceedings of the 2014 ACM/IEEE international conference on human-robot interaction - HRI '14. ACM Press; 2014. p. 504–5.
[2] Tadeusz S. 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)*; 2016. p. 1069–75.
[3] Kim D-S, Jeon I-J, Lee S-Y, Rhee P-K, Chung D-J. Embedded face recognition based on fast genetic algorithm for intelligent digital photography. IEEE Trans Consum Electr 2006;52(3):726–34.
[4] Padgett C, Cottrell G. Representation face images for emotion classification. Advances in neural information processing systems, 9. Cambridge, MA: MIT Press; 1997.
[5] Mitra S, Acharya T. Gesture recognition: a survey. IEEE Trans Syst Man Cybern Part C 2007;37(3):311–24.
[6] Donato G, Bartlett MS, Hagar JC, Ekman P, Sejnowski TJ. Classifying facial actions. IEEE Trans Pattern Anal Mach Intel 1999;21(10):974–89.
[7] Buciu I, Kotropoulos C, Pitas I. ICA and gabor representation for facial expression recognition. In: Proceedings of the IEEE; 2003. p. 855–8.
[8] chen F, Kotani K. Facial expression recognition by supervised independent component analysis using MAP estimation. IEICE Trans Inf Syst 2008;E91-D(2):341–50.
[9] Hyvarinen A, Karhunen J, Oja E. Independent component analysis. John Wiley & Sons; 2001.
[10] Lee JJ, Uddin MZ, Kim T-S. Spatiotemporal human facial expression recognition using fisher independent component analysis and hidden Markov model. In: Proceedings of IEEE conf of eng med biol soc; 2008. p. 2546–9.
[11] Bartlett MS, Donato G, Movellan JR, Hager JC, Ekman P, Sejnowski TJ. Face image analysis for expression measurement and detection of deceit. Proceedings of the sixth joint symposium on neural computation 1999:8–15.

[12] Chao-Fa C, Shin FY. In: Recognizing facial action units using independent component analysis and support vector machine, 39; 2006. p. 1795–8.
[13] Calder AJ, Young AW, Keane J. Configural information in facial expression perception. J Exp Psychol 2000;26(2):527–51.
[14] Lyons MJ, Akamatsu S, Kamachi M, Gyoba J. Coding facial expressions with Gabor wavelets. In: Proceedings of the third IEEE international conference on automatic face and gesture recognition; 1998. p. 200–5.
[15] Uddin MZ, Lee JJ, Kim T-S. An enhanced independent component-based human facial expression recognition from video. IEEE Trans Consum Electr 2009;55(4):2216–24.
[16] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray scale and rotation invariant texture analysis with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002;24:971–87.
[17] Jabid T, Kabir MH, Chae O. Local directional pattern (LDP) a robust image descriptor for object recognition. In: Proceedings of the IEEE advanced video and signal based surveillance (AVSS); 2010. p. 482–7.
[18] Yu P, Xu D, Yu P. Comparison of PCA, LDA and GDA for palm print verification. In: Proceedings of the international conference on information, *networking and automation*; 2010. p. 148–52.
[19] Li W, Zhang Z, Liu Z. Action recognition based on a bag of 3d points. In: Proceedings of workshop on human activity understanding from 3D Data; 2010. p. 9–14.
[20] Li W, Zhang Z, Liu Z. Expandable data-driven graphical modeling of human actions based on salient postures. IEEE Trans Circ Syst Video Technol 2008;18(11):1499–510.
[21] Liu X, Fujimura K. Hand gesture recognition using depth data. In: Proceedings of international conference on automatic face and gesture recognition; 2004. p. 529–34.
[22] Mo Z, Neumann U. Real-time hand pose recognition using low-resolution depth images. In: Proceedings of IEEE conference on computer vision and pattern recognition; 2006. p. 1499–505.
[23] Breitenstein MD, Kuettel D, Weise T, Van Gool L, Pfister H. Real-time face pose estimation from single range images. In: Proceedings of IEEE conference on computer vision and pattern recognition; 2008. p. 1–8.
[24] Aleksic PS, Katsaggelos AK. Automatic facial expression recognition using facial animation parameters and multistream HMMs. IEEE Trans Inf Security 2006;1:3–11.
[25] Minsky M, Papert S. Perceptrons. An introduction to computational geometry, 165. Cambridge: M.I.T Press; 1969. p. 780–2.
[26] Hinton GE, Osindero S, The Y. A fast learning algorithm for deep belief nets. Neural Comput 2006:1527–54.
[27] Lv Y, Ma T, Tang M, Cao J, Tian Y, Al-Dhelaan A, et al. An efficient and scalable density-based clustering algorithm for datasets with complex structures. Neurocomputing 2016;171:9–22.
[28] Ma T, Rong H, Ying C, Tian Y, Al-Dhelaan A, Al-Rodhaan M. Detect structural-connected communities based on BSCHEF in C-DBLP. Concurr Comput 2016;28(2):311–30.

Md. Zia Uddin received his Ph.D. in Biomedical Engineering in February of 2011. He is currently working as a post-doctoral research fellow under Dept. of Informatics, University of Oslo, Norway. His researches are mainly focused on computer vision, image processing, artificial intelligence, and pattern recognition. He got more than 60 research publications including international journals, conferences and book chapters.

Mohammad Mehedi Hassan is currently an Associate Professor of Information Systems Department in the College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He received his Ph.D. degree in Computer Engineering from Kyung Hee University, South Korea. His research areas of interest are cloud computing, Deep learning, Internet of things, image processing and Big data.

Ahmad Almogren has received PhD degree in computer sciences from Southern Methodist University, Dallas, Texas, USA in 2002. He is now an Associate Professor at the college of computer and information sciences at King Saud University in Saudi Arabia. His research areas of interest include mobile and pervasive computing, computer security, sensor and cognitive network, and data consistency.

Mansour Zuair is currently an Assistant Professor in the Department of Computer Engineering and the Dean of College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He received his M.S. & Ph.D. degree in Computer Engineering from Syracuse University. His research interest is in the areas of computer architecture, Parallel Processing, Embedded Systems.

Giancarlo Fortino is an Associate Professor of Computer Science in the Dipartiment of di Informatica, Elettronica e Sistemistica (DEIS) of the Università della Calabria. Prof. Fortino's research is mainly focused on methodologies, frameworks and tools for programming distributed computing systems, distributed health analytics and cloud based health analytics.

Jim Torresen is a professor of computer science at the University of Oslo, Norway. His research interests include nature-inspired computing, adaptive systems, reconfigurable hardware, and robotics and their use in complex real–world applications. Torresen received a PhD in computer science from the Norwegian University of Science and Technology.