




Research Article

Iranian kinect face database (IKFDB): a color-depth based face database collected by kinect v.2 sensor



Seyed Muhammad Hossein Mousavi¹  · S. Younes Mirinezhad¹

Received: 9 December 2019 / Accepted: 21 December 2020 / Published online: 7 January 2021

© The Author(s) 2020 

Abstract

This study presents a new color-depth based face database gathered from different genders and age ranges from Iranian subjects. Using suitable databases, it is possible to validate and assess available methods in different research fields. This database has application in different fields such as face recognition, age estimation and Facial Expression Recognition and Facial Micro Expressions Recognition. Image databases based on their size and resolution are mostly large. Color images usually consist of three channels namely Red, Green and Blue. But in the last decade, another aspect of image type has emerged, named “depth image”. Depth images are used in calculating range and distance between objects and the sensor. Depending on the depth sensor technology, it is possible to acquire range data differently. Kinect sensor version 2 is capable of acquiring color and depth data simultaneously. Facial expression recognition is an important field in image processing, which has multiple uses from animation to psychology. Currently, there is a few numbers of color-depth (RGB-D) facial micro expressions recognition databases existing. With adding depth data to color data, the accuracy of final recognition will be increased. Due to the shortage of color-depth based facial expression databases and some weakness in available ones, a new and almost perfect RGB-D face database is presented in this paper, covering Middle-Eastern face type. In the validation section, the database will be compared with some famous benchmark face databases. For evaluation, Histogram Oriented Gradients features are extracted, and classification algorithms such as Support Vector Machine, Multi-Layer Neural Network and a deep learning method, called Convolutional Neural Network or are employed. The results are so promising.

Keywords Depth image · Kinect sensor V.2 · Color-depth based database · Facial micro expressions recognition · Support vector machine (SVM) · Convolutional neural network (CNN)

1 Introduction

Before the appearance of depth images, there were just two dimensions for calculating a color digital image, but depth images or 2.5-Dimensional (2.5-D) images added a third dimension to calculation which led to making 3-Dimensional (3-D) form of objects. Recently, with emergence of depth sensors like Microsoft Kinect [1], working on 3-D applications has gained ease of use. Thanks to its cheap price and high quality, it is widely used in a lot of

fields. Academic arena is one of the fields depth image science has had a significant impact on. For a better performance, a simultaneous development in both hardware and software in the field of computer vision is necessary.

Here are some major applications of depth images: 3-D modeling and reconstruction [2], augmented reality [3], industry [1], medicine [4], Human-Computer Interaction (HCI) [5, 51], Robotics [6] and more. Identity recognition using fingerprints and iris is so precise, but it is limited to recognizing human identity. Face gives more details about

✉ Seyed Muhammad Hossein Mousavi, mosavi.a.i.buali@gmail.com; S. Younes Mirinezhad, y.mirinezhad93@basu.ac.ir | ¹Department of Computer Engineering, Bu Ali Sina University, Hamadan, Iran.



humans; details such as age [7], gender [8], ethnicity or race [9], identity [10], Facial Expression Recognition (FER) and Facial Micro Expressions Recognition (FMER) [11, 12] and some sicknesses [13] with enough precision. Due to kinect's nature, one of its applications is in detection and recognition of objects; such as face. Due to the novelty of depth images, there is a shortage of data in this field. There are proper Red, Green, Blue- Depth (RGB-D) face databases available, but few, and mostly embody East Asian and European face type. Also, there is a lack of micro facial expressions in existing ones.

For covering Middle-Eastern face type and micro facial expressions, a face database was created in which proper lighting and 40 subjects were employed.

1.1 Face detection

The process in which system attempts to differentiate face object from other ones in a specific scene is called face detection [58]. Also, cropping techniques are employed to extract the face object out of the scene for further processing. A number of famous face detection methods such as ellipse fitting [14] or viola and jones face detection algorithm [15, 45] could be used. For the purpose of this paper, viola and jones algorithm is employed on face detection in color images. In the section III This algorithm is demonstrated in details. Face detection is the initial phase in face recognition [56, 57], gender recognition, facial expression recognition and other similar activities. Figure 1 represents some examples of face detection from color and depth images [16, 17]. As shown in Fig. 1, Viola and Jones face detection algorithm is applied on color images [16]. And in depth image, after discovering the tip of the nose which is closest pixel value to the sensor, some type of segmentation with specific threshold is applied for cropping just the face muscles [17].

1.2 Facial expression recognition (FER)

This action happens after face detection and extraction procedure. There are 7 main types of human facial expressions such as Joy or happiness, Sadness, Surprise, Neutral, Disgust, Fear and Anger [18]. These expressions represent our internal emotions which in turn manifest on our faces. In order to taxonomize movements of human facial muscles appearing on face, which is developed by Carl-Herman Hjortsjö [19] and later published by Paul Ekman and Wallace V. Friesen in 1978 [20], the Facial Action Coding System or FACS was made. They further developed the FACS and published a significant update on it in 2002 [21]. Slightest change of individual facial muscles is encoded using FACS, even smallest instant changes in the appearance of the face. By employing FACS [22], It is possible for programmers to manually code nearly any anatomically possible facial expression. It happens by splitting it into a number of Action Units (AU) and their temporary segments that makes the expression. With combining these action units, it is possible to make any possible facial expressions on the face. It is possible to say that facial expression is the other name for emotions, because the facial expression of a blind person is almost similar to a normal person. Figure 6 shows 7 main facial expression from proposed IKFDB RGB-D database. Facial expressions last between 0.5 to 4 s on the face [22].

1.3 Facial micro expression recognition

Facial micro expressions are completely similar to the facial expressions, except the time of accruing. Facial micro expression happens in 1/3, 1/5 or 1/25 s [23, 24]. So, for recording these very fast actions, employment of sensors with 30–90 frame per second is necessary. Figure 7

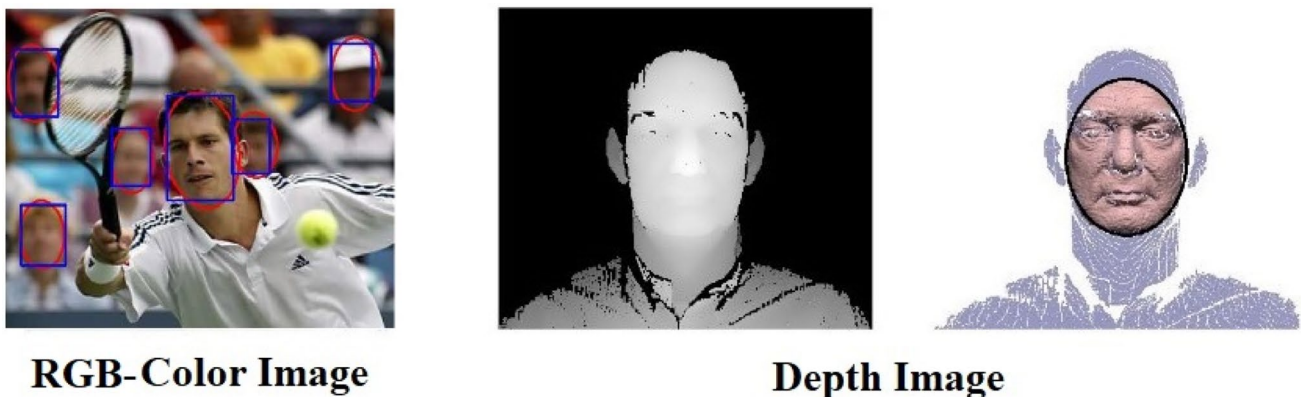


Fig. 1 Some examples of face detection from RGB and Depth images [16, 17]

displays several micro expressions from proposed IKFDB RGB-D database.

1.4 Depth image and sensors

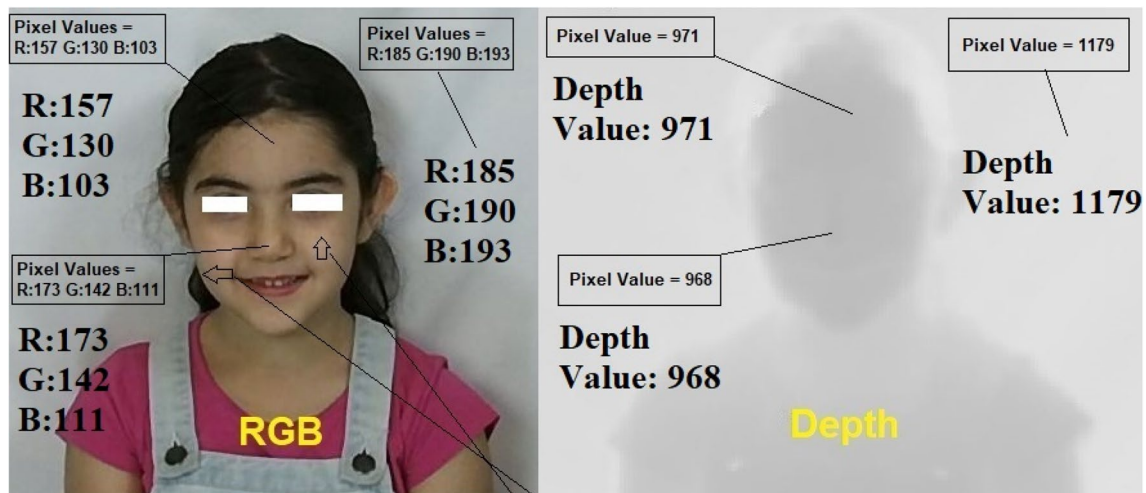
Depth images mostly are used in range or distance detection. There are several depth sensors, most of which use infrared spectrum to sense the distance between an object and sensor. Kinect is one of the most popular depth sensors in the world, especially among researchers. Kinect is a Microsoft product, and right now second version of it is available. For more information about Kinect sensors refer to [1, 46, 47]. Depth images are so popular in 3-D applications. Here we use it to increasing the accuracy of facial expressions and facial micro expressions recognition along with color image. Depth image versus color image structure is shown in Fig. 2. As depth image in Fig. 2 suggests, as distance between subject and sensor gets higher, pixel value increases, which means longer distance. Closest distance between subject and sensor is tip of the nose with pixel value of 968. And the farthest point with a value of 1179 belongs to curtain behind the subject. Kinect's depth image Unit is in millimeter. So, curtain has almost 1-m distance from the Kinect sensor. RGB or color image has its famous applications known to all of us. Also, corresponding action units for facial muscles based on FACS system [22] is determined in the image (it is same in the depth image).

There are VI main sections in the paper, which section I explains the fundamentals; section II investigates and describes prior works on making RGB-D face database in recent years; section III discusses proposed face database in details. The chosen name for the database is Iranians Kinect Face Database (IKFDB). Section IV includes achieved experiments results using feature extraction and classification algorithms along with comparison results with other similar databases in the same condition. Section V includes discussion on the work and results. Finally, section VI concludes with conclusion and suggestions, discussing acquired results and future work.

2 Prior work

In this section some famous face databases are discussed and summarized. Works are from color and color-depth fields and most of them have facial expressions application. All of the available facial micro expressions databases are color type and proposed work is the first proper color-depth type of facial micro expressions databases.

The Japanese Female Facial Expression (JAFFE) database is one of the oldest and the most cited color-based face database in this area. JAFFE consists of 10 subjects and 212 Gy images along with 7 main expressions in 256*256 dimensions; all images are gray level. This database is the result of Lyons, Michael, et al. efforts in 1998 [25].



Facial Action Coding System or FACS for Happiness = 6+12, Which Action Units for Facial Muscles Are as Follow: 6= Cheek Raiser Using Orbicularis Oculi (Pars Orbitalis) Muscle and 12= Lip Corner Puller Using Zygomaticus Major Muscle

Fig. 2 Color image versus depth image structure and corresponding action units (Subject belongs to proposed IKFDB database- subject has 7 years old—subject presents happy or joy expression)

In the other hand we can mention The Karolinska Directed Emotional Faces (KDEF) [26] database, achieved in the same year as JAFFE by Lundqvist, D., Flykt, A., and Öhman, A. Just like JAFFE this database is not depth based, but colorful and not gray level. Having 7 main facial expressions and 70 male and female subjects in an age range of 20–30 along with higher number of images and higher image size dimensions than JAFFE, which are 5041 color images in 862×762 dimensions, has made this database so appealing to many researchers. This database is recorded using Pentax LX sensor.

One of the most famous facial micro expressions databases is called SMIC database. This color-based database consists of 16 subjects and 4 micro expressions in 640×480 video frame size. There are 164 videos, each consists of 1000 frames. This database is recorded using PixelLINK PL-B774U sensor, and it is the result of Li, Xiaobai, et al. [27] research in 2013.

In 2014 Min, Rui, Neslihan Kose, and Jean-Luc Dugelay collected an RGB-D face database and named it EURECOM. They used Kinect sensor V.1 to record the data [28]. The database consists of 52 subjects (14 females and 38 males) for just 3 expressions, and it could also be used in face recognition. All 1248 samples are cropped to 256×256 image size for all color and depth images which facilitates usage for end user.

Another mentionable color-based micro expression database is CASEM database [29]. This one is recorded using BenQ M31-GRAS 03K2C for 7 main micro facial expressions in 2013. There are 195 video files in 60 frames in two different sizes of 1280×740 (original) and 640×480 (cropped) versions for 19 subjects. This database is collected by Yan, Wen-Jing, et al.

Qu, Fangbing et al. [53] made the second version of CASME database called CASME 2 in 2017. This database is made for macro and micro facial expressions recognition purposes and just like its first version, it is a color-based database. CASME 2 includes 300 cropped spontaneous macro-expression, and 57 micro-expression samples. The database is recorded with the same quality and sensor as used in CASME.

Also, there are remarkable color-based micro facial expressions databases such as: Polikovskiy's database [12], USF-HD [30] and YorkDDT database [31].

2.1 Some remarkable color-depth or RGB-D face database

In 2012, Hg, R. I., et al. succeeded to design a color-depth based face database serving specially for facial expressions recognition using Kinect V.1 [32]. There were 13 subjects in the experiment; having 2960 color and depth images. Their database includes 5 facial expressions of

neutral, happiness, surprise, anger and sadness. Color image has a dimension of 421×351 and depth image a dimension of 640×480 . The database is called VAP-RGB-D. Another database with the name of VAP-RGB-D-T, introduced in [33] a paper in 2014, which includes thermal data along with color and depth. The difference with VAP-RGB-D was having 4 expressions and more subjects to 51. Thermal data was recorded using AXIS Q1922 sensor. Data is in the form of frames and includes 46,360 color, depth and thermal images. Final image dimension for end users is 640×480 for color data, and for depth and thermal data 384×288 dimension is considered. As amount of data and data type increases, recognition accuracy increases too.

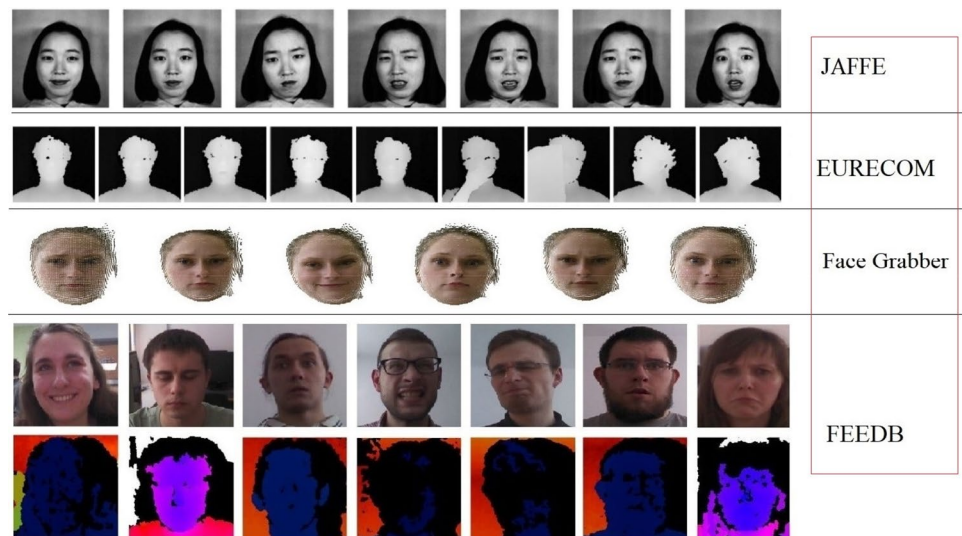
Also, Curtin face database with 7 main expressions [10] and FEEDB database with 33 facial expressions [34] are notable; both of which use Kinect V.1 as recording device. They are recorded in color-depth or RGB-D form, but FEEDB have some weaknesses such as: camera positioning, lighting and background problems. Curtin face database is almost perfect. Both are recorded in 2013.

In 2016, Merget, Daniel et al. [35] made a facial expressions database using Kinect V.2 sensor, having all 7 main expressions. The database is called Face Grabber and consists of 40 subjects -just like proposed database- but 33 males and 7 females. It could be used in different fields of face applications. It has some problems which are fixed in the proposed one; problems such as: very busy background which leads face detection algorithms to wrong detections and also limited age range. But overall, it is an appropriate face database. There are 67,159 frames of color and depth images in the database in which raw data is in 1920×1080 size for color images, and 512×424 for depth ones. They are also able to provided 3-dimensional model of the depth images. Figure 3 illustrates some of the samples from mentioned face databases in color and depth form.

KaspAROV RGB-D video face database [54] is the result of Chhokra, Pawas, et al. research, conducted in 2018. This database was created for multiple facial analyses such as: Pose, illumination, expression, sensor interoperability, resolution, distance and spectrum variations. They used both Kinect sensors versions 1 and 2. 108 subjects were employed. Database consists of 432 videos and 117,831 color and depth images.

Another valuable research in this area belongs to Turan, Cigdem, Karl David Neergaard, and Kenneth KM Lam in 2019 [55]. They present a new multimodal facial expression database, named Facial Expressions of Comprehension (FEC), made of the videos recorded during a computer-mediated task in which each trial consisted of reading, answering and feedback to general knowledge true and false statements using Kinect sensor version 2.

Fig. 3 Some of the samples from some mentioned RGBD face databases in color and depth form



3 Proposed database

As said earlier, according to the shortage of RGB-D face database for FER and FMER for Middle-Eastern face type, and novelty of this subject, it was decided to create one. Also, there are a lot of bugs and recording errors in available databases except few. In order to fix problems like noisy background, weak lighting, few age range, bad camera positioning and weak expressions in available databases and supporting Middle-East race face type, there was a need to create a better version. It is recorded with the stronger sensor (Kinect V.2). For instance, FEEDB database clearly indicates such weaknesses. But databases Like Face Grabber, VAP and EURECOM indicate a better overall standard. EURECOM has just 3 facial expressions and VAP just 5; nevertheless, their performance on face recognition is promising. Face Grabber and Curtain face databases do not indicate a significant lack. The only problem-as with other depth face databases- is their weak performance on supporting FMER. This proves another reason to improve the existing face RGB-D databases. Others like JAFFE, KDEFSMIC, and CASME are color databases and do not supporting depth data, but they are almost perfect in color type.

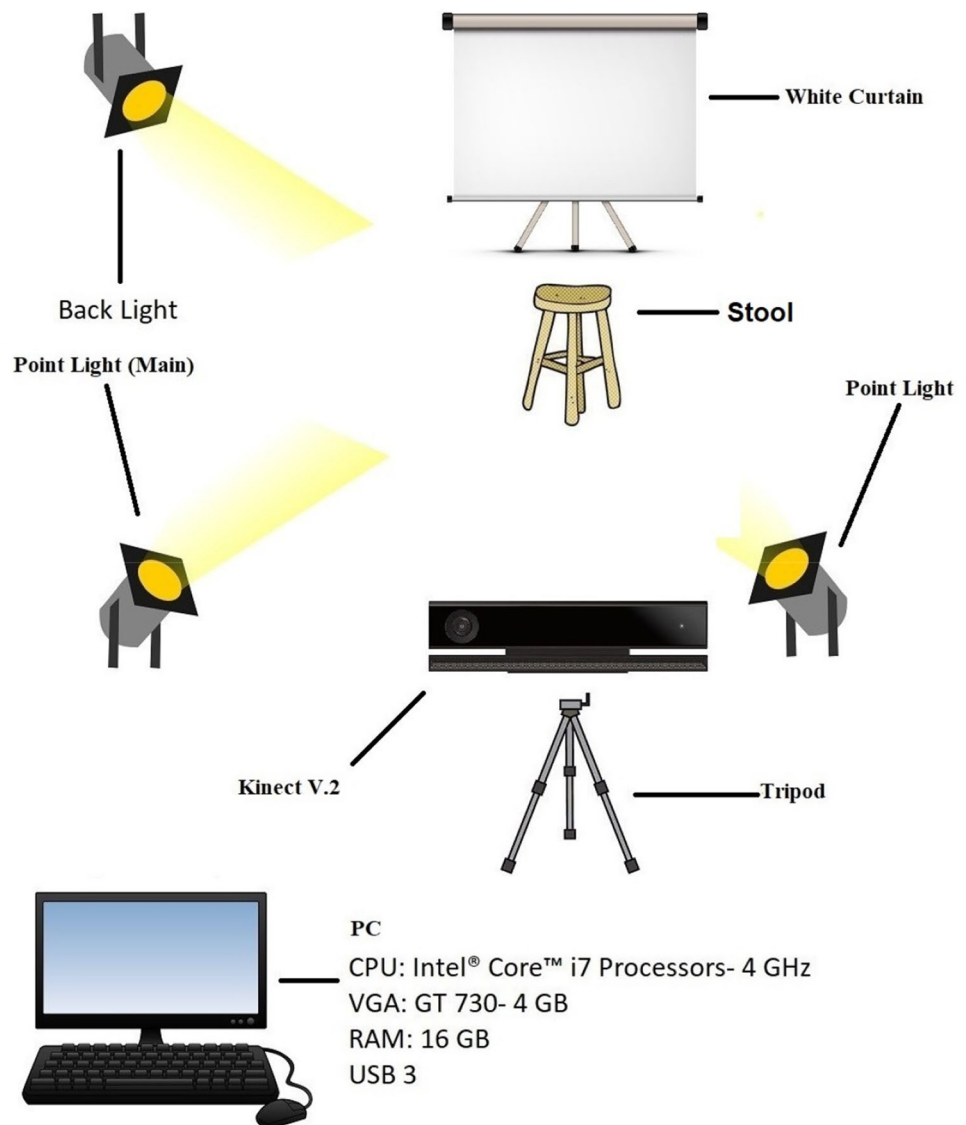
The proposed Iranian Kinect Face Database (IKFDB) consists of 40 subjects in age range of 7–70 (16 female—24 male). More than 100,000 color and depth frames are recorded using Kinect V.2 sensor in 30 fps. Due to the high number of subjects, good variety in gender and slight expressions in initial frames, database could be used in different fields of face analysis such as face recognition, gender recognition, age estimation, facial expressions and facial micro expressions recognition. Database is prepared in two different forms of raw and cropped. In raw version there are color and depth images in 1920×1080 and 512×424 image sizes respectively; and in cropped version

they are respectively in 256×400 and 128×200 image sizes. All the subjects are located in a range of 0.8–1.2 m from sensor. Also, pitch and yaw actions are considered for better face recognition purpose (recognition from any angle). It was requested from all the subjects to perform the 7 main expressions during 150–250 frames from slight to strong expressions for covering micro expressions in the first 50 frames. Obviously, the last 50 frames cover the whole expressions.

In order to explain the merits and demerits of IKFDB, some points are highlighted. IKFDB covers color-depth image data for FMER purposes. Based on our investigation, IKFDB covers Middle-Eastern face type for the first time (as an independent RGB-D database). It covers an age range of 7–70, including different genders. It has applications in other face analysis purposes; by converting 2.5-Dimensional (2.5-D) depth image to 3-Dimensional model or mesh (3-D), face data could be analyzed using 3-D algorithms better and fixing available databases bugs such as noisy background, weak lighting, few age range, bad camera positioning and weak expressions. Demerits: IKFDB covers only up to 30 fps of data as it is the best performance of Kinect V.2 sensor. There are few subjects wearing glasses or having beard. The number of subjects is 40, which could be even more.

Figure 4 shows the recording environment and equipment. Figure 5 represents some samples of the databases in pitch and yaw forms in color and depth. Subjects are asked to change head angle till 45 degree in right, left, top and bottom sides. Figure 6 shows some samples of database, performing 7 main facial expressions in color and depth forms. This figure also shows the variety of ages and genders in the database. Figure 7 presents 4 samples from proposed database, performing 4 micro expressions of anger, disgust, surprise and happiness in color and

Fig. 4 Recording environment and tools



depth forms. Figure 8 shows the relation between subjects and their ages. Figures 9 show gender distribution, age distribution and wearing glasses-beard distributions in subjects respectively.

3.1 Recording environment and tools

See Figs. 4, 5, 6, 7, 8 and 9

4 Experiments and results

Proposed database has advantages of number of samples, covering both color and depth image types and covering both FER and FMER tasks in comparing with other employed databases for comparison purposes in the Table 2 which helps to act properly in learning process

of different classification algorithms. In validation section, some processing for extracting features and simple classification task takes place. After pre-processing stage, face detection and extraction tasks on color images using Viola & Jones algorithm [36] will be done. For extracting face out of depth image, [45] method is used. After extracting faces out of color and depth images, it is time to change the structure of depth image. Somehow color feature extraction methods could be applied on them. This step is performed using a new method as presented in Fig. 14. Feature extraction simply occurs using Histogram of Oriented Gradient (HOG) algorithm [37] and after fusing extracted features from color and depth images, it is time to use Lasso feature selection to reduce data for classification task. Finally, each expression is labelled and final matrix is ready for classification using Support Vector Machine (SVM) [38], Multi-Layer Neural Network (MLNN) [50] and

Fig. 5 Pitch and Yaw

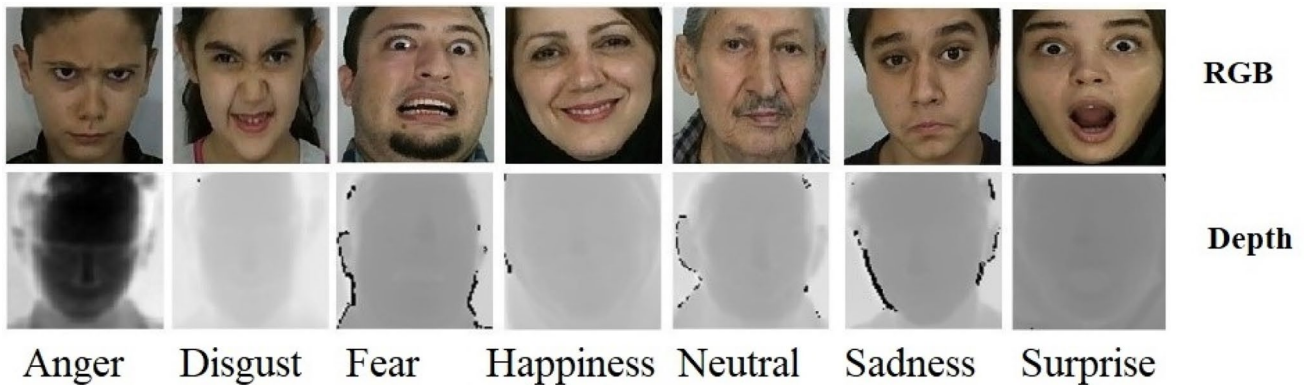
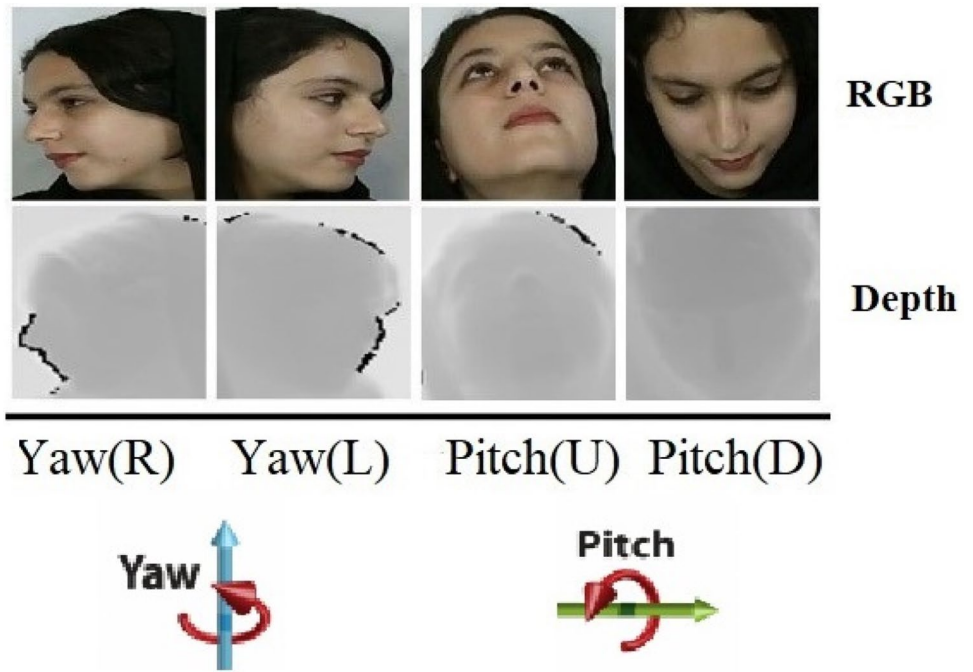


Fig. 6 Proposed IKDFB samples, performing 7 main facial expressions

Convolutional Neural Network (CNN) [50] algorithms. Figure 10 illustrates the flowchart of the research's structure.

4.1 Histogram of oriented gradients (HOG)

Histogram of Oriented Gradients (HOG) is one of the best local feature extractions and descriptors methods in image processing field. This feature is fast and robust, and works well on Depth images. This method is edge based and similar to Scale-Invariant Feature Transform descriptors (SIFT) [52], and shape contexts, but differs is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy [37].

The basic rule in the histogram of oriented gradients descriptor is that local object figure and the shape inside of an image can be demonstrated by the distribution of direction of edges. It starts with dividing the image into a number of small connected regions called cells, and for pixels of cells, a gradient histogram direction is assembled. The concatenation of these histograms is the descriptor. In order to improving the accuracy, normalizing the contrast of the local histograms by calculating a measure of the intensity along a big region of the image is necessary. This is called a block, and then it uses this value to normalize all cells within the final block.

There are two main factors of gradient magnitude and gradient direction in hog. Gradient direction shows the direction of moving from black to white gray levels in

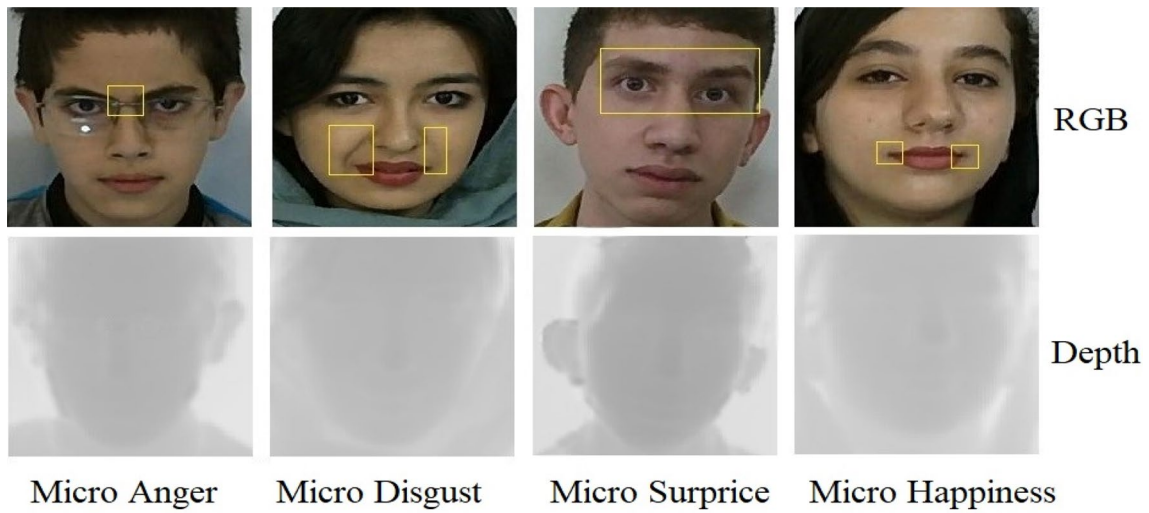


Fig. 7 Proposed IKFDB samples, performing 4 facial micro expressions

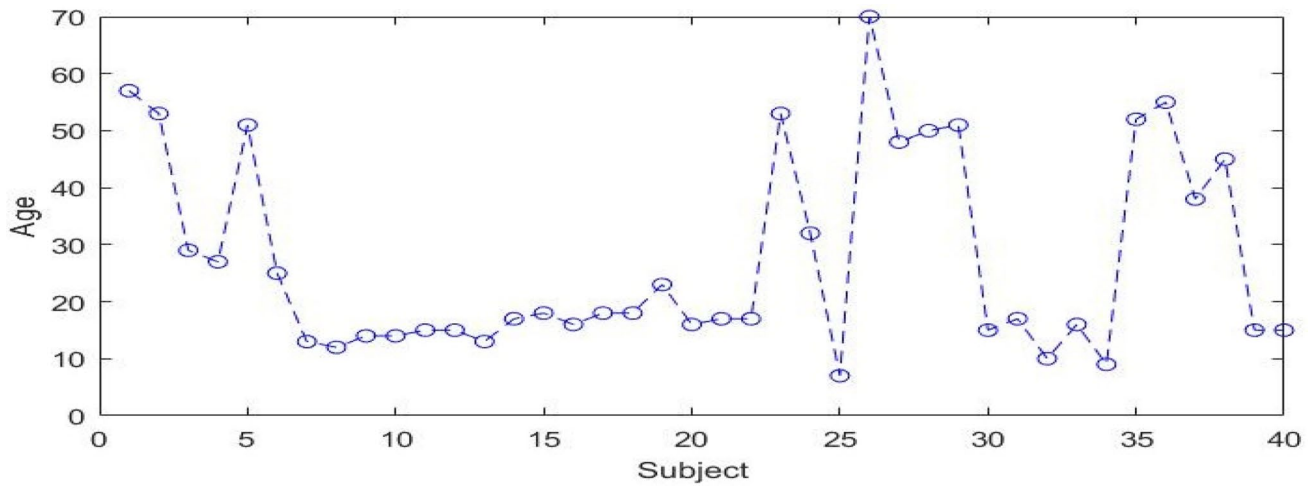


Fig. 8 Proposed IKFDB, subject versus age relation

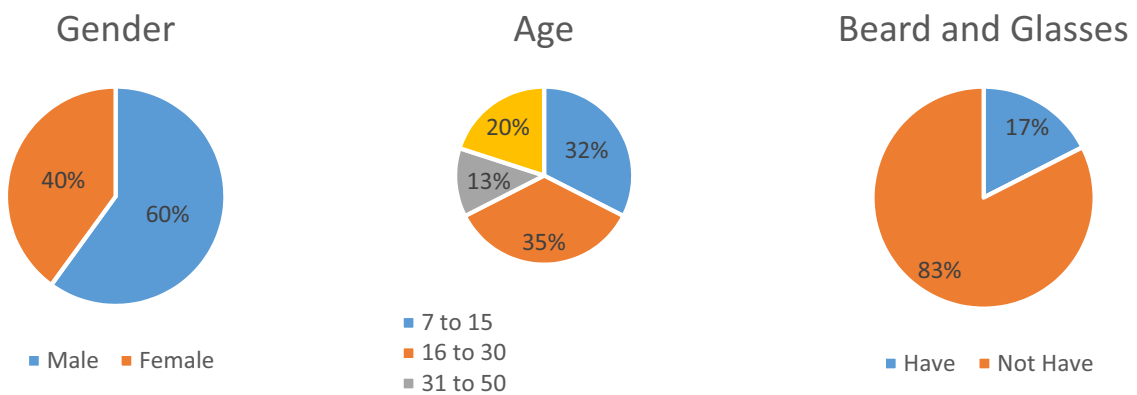


Fig. 9 Gender, age and beard-glasses distribution in proposed database

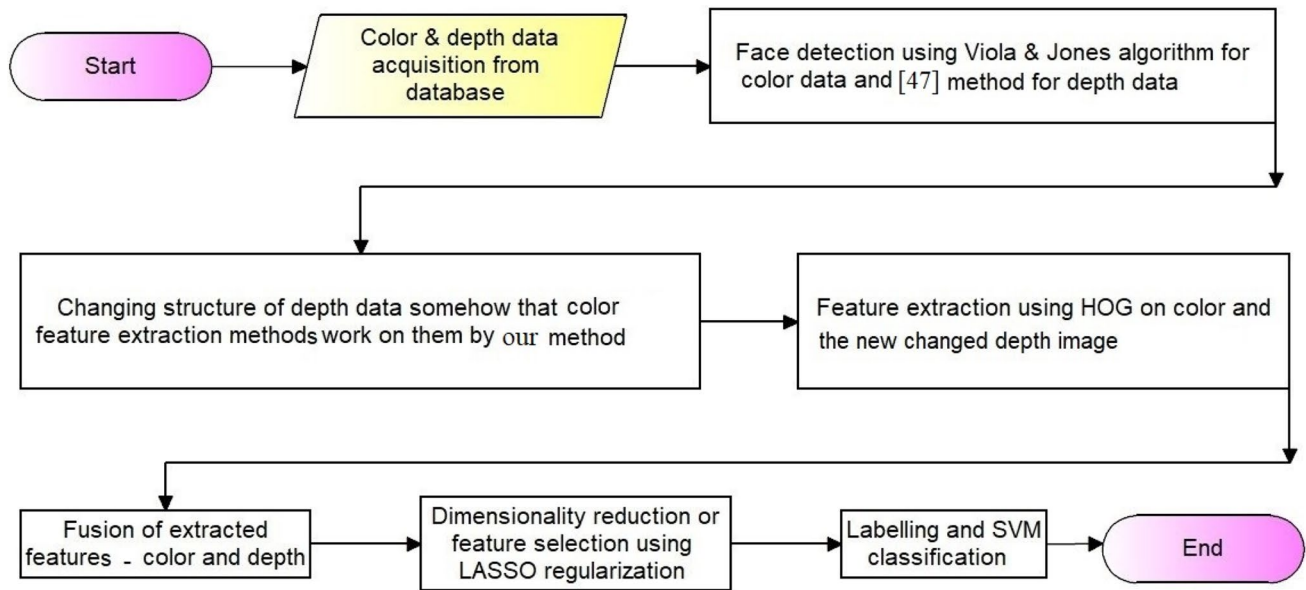


Fig. 10 Flowchart of the research's structure

gradient, and magnitude represents the power of each directions. Figure 11 shows the gradient magnitude and direction on a sample. As contrast in gradient increases per pixel, directions grow stronger. Gradient direction is calculated using 0–180 or 0–360 degree. It is possible to achieve gradient from traditional edge detection filter.

As the main goal of the paper is to introducing a new face database and other methods from preprocessing to classification is just for representing the finale accuracy, some of the most famous algorithms employed in order to do it. HOG is famous, robust and fast, also it works on depth images very well, so it is decided to use it in the feature extraction step. Figure 12 shows an edge detection sample for usable for HOG and gradient direction example. Finally, each patch's histogram will be placed in a bin to make final feature vector. Table 1 shows the main parameters for HOG.

4.2 Viola and jones face detection algorithm for RGB images

This algorithm [36] is almost one of the most accurate and fastest face detection algorithms and has been around for a long period of time in facial image processing. This algorithm is so robust and could be employed for depth images with lower accuracy as well. This algorithm is an object detection algorithm and could be used for any learned object but that is mainly used for face. The algorithm has four stages of 1. Haar Feature Selection, 2. Creating an Integral Image, 3. Adaboost Training and finally Cascading Classifiers.

After face detection phase, face can be extracted by cropping techniques. For more information and details about this algorithm refer to [36].

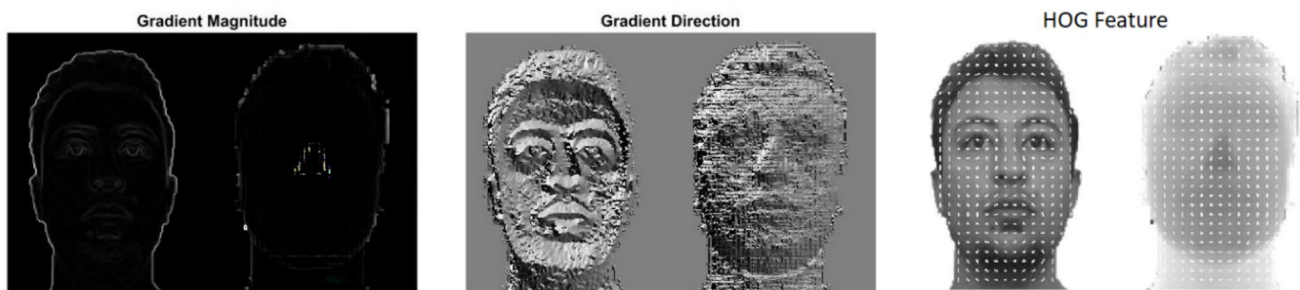


Fig. 11 Gradient magnitude and direction on a sample (color and depth)

Fig. 12 Edge detection sample usable by HOG (left), gradient direction (right)

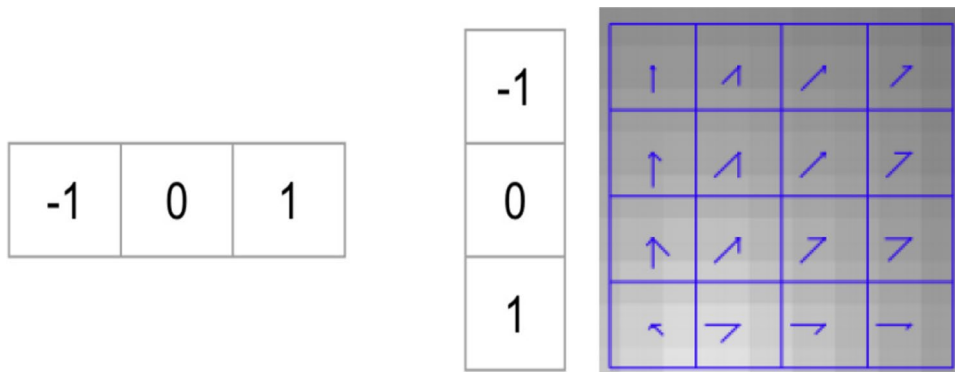


Table 1 HOG parameters

Window size (pixels)	$S_x \times S_y$	Cell size (pixels)	$P \times P$
#Cells in block	$c \times c$	Block size (pixels)	$P_c \times P_c$
Block stride (pixels)	(L, L)	#of bins per cell	N
#Cells in window	$CX \times CY$ ($CX = S_x/p, CY = S_y/p$)		
#Blocks on window	$BX \times BY$ ($BX = (S_x - pc + L)/L, BY = (S_y - pc + L)/L$)		

4.3 Depth image face detection

A new interesting method to extracting face out of depth images is used. Following shows the steps: 1. Finding closest pixel to the sensor with smallest value (nose tip). 2. Rectangular face cropping to a specific threshold and saving it into a temporary and second matrixes. 3. Applying standard deviation filter [39] to second matrix and using ellipse fitting [14] technique to fit ellipse on face. 4. Selecting pixels inside the temporary matrix based on second matrix calculations and removing other pixels. 5. Finally, some percentage of the image sides could be removed to increase accuracy. Figure 13 represents that in a visual form.

4.4 Deforming face to color structure

HOG feature works well on depth images, but it works based on gradient directions. The problem is that depth images are so smooth in original form. For finding a way to sharpen edges, a method to change depth image structure is presented which it gives possibility to use color features like HOG on them with higher accuracy. In this method, Gaussian filter with $\sigma = 2$ is applied to the depth image and then, system searches for the pixels with different pixel values with previous and next pixels in horizontal and vertical directions; and replaces found pixel values with 1 and others to 256. So, there is such mesh type 3-D color image and also in 2 dimensions which gives possibility to use color image features on them. Using this technique may increases recognition accuracy up to 40%, which is incredible (Fig. 14).

4.5 LASSO regularization and support vector machine (SVM)

Lasso method [40] is a developed version of Regression method [41]. The process of presenting additional information in order to solve an unpleasant problem or to prevent over fitting is called regularization [42]. The regularization term is adding to the loss function. In classification, Loss functions are computationally feasible. Loss functions presenting the price paid for inexactitude of predictions in classification tasks [43]. Regularization can be stimulated as a method to improve the generalizability for a model which is learned [44]. One of the regularization techniques is Lasso to do linear regression. Lasso had a final term which constrains the size of the estimated coefficients. Therefore, it simulates the ridge regression and also, it is a shrinkage estimator which it produces coefficient estimates that are biased to be a small amount. In the other hand, a lasso estimator can have smaller value of Mean Squared Error (MSE) than an ordinary least-squares estimator for new data.

In Lasso when the penalty term increases, it makes more coefficients to zero, unlike ridge regression. Which it makes sense that lasso estimator is a much smaller model, which includes fewer predictors. So, Lasso could be a good replacement for stepwise regression and dimensionality reduction techniques.

For a given value of λ , a nonnegative parameter, lasso solves the problem

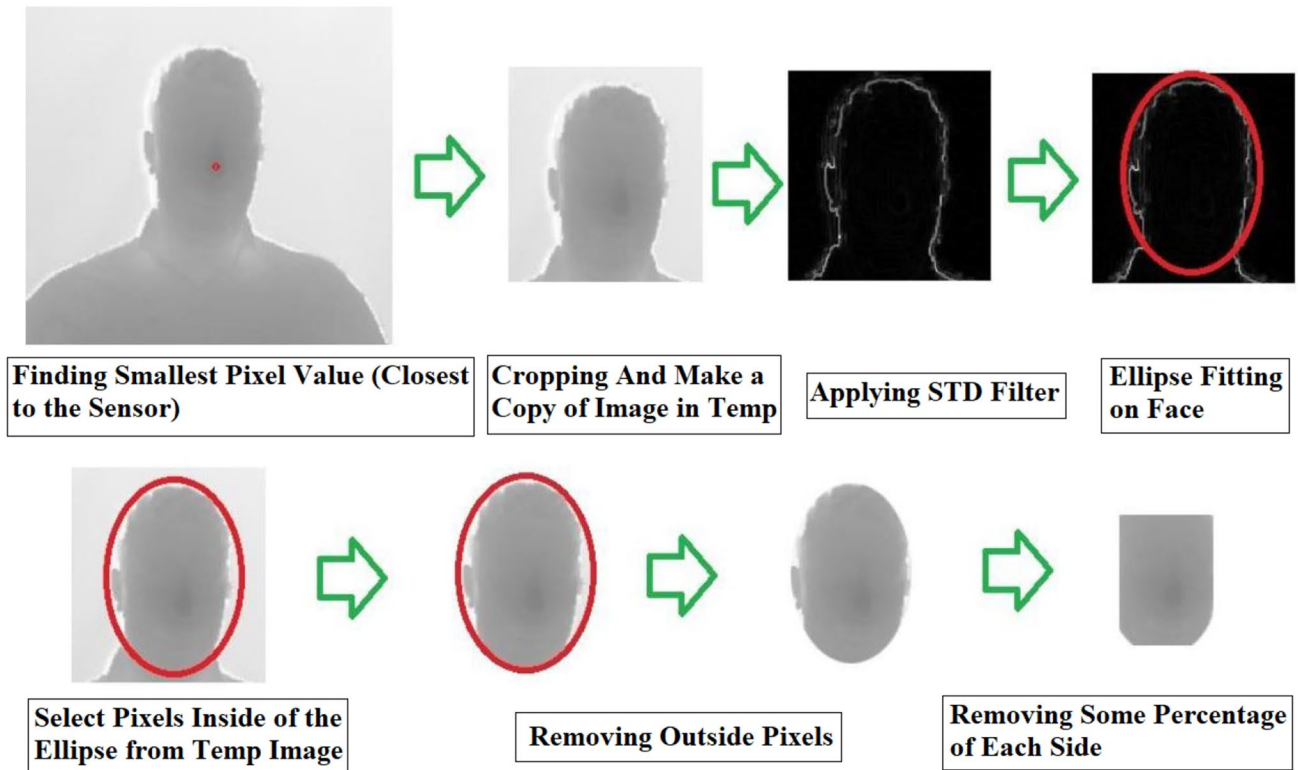


Fig. 13 Employed depth face detection and extraction algorithm procedure [45]

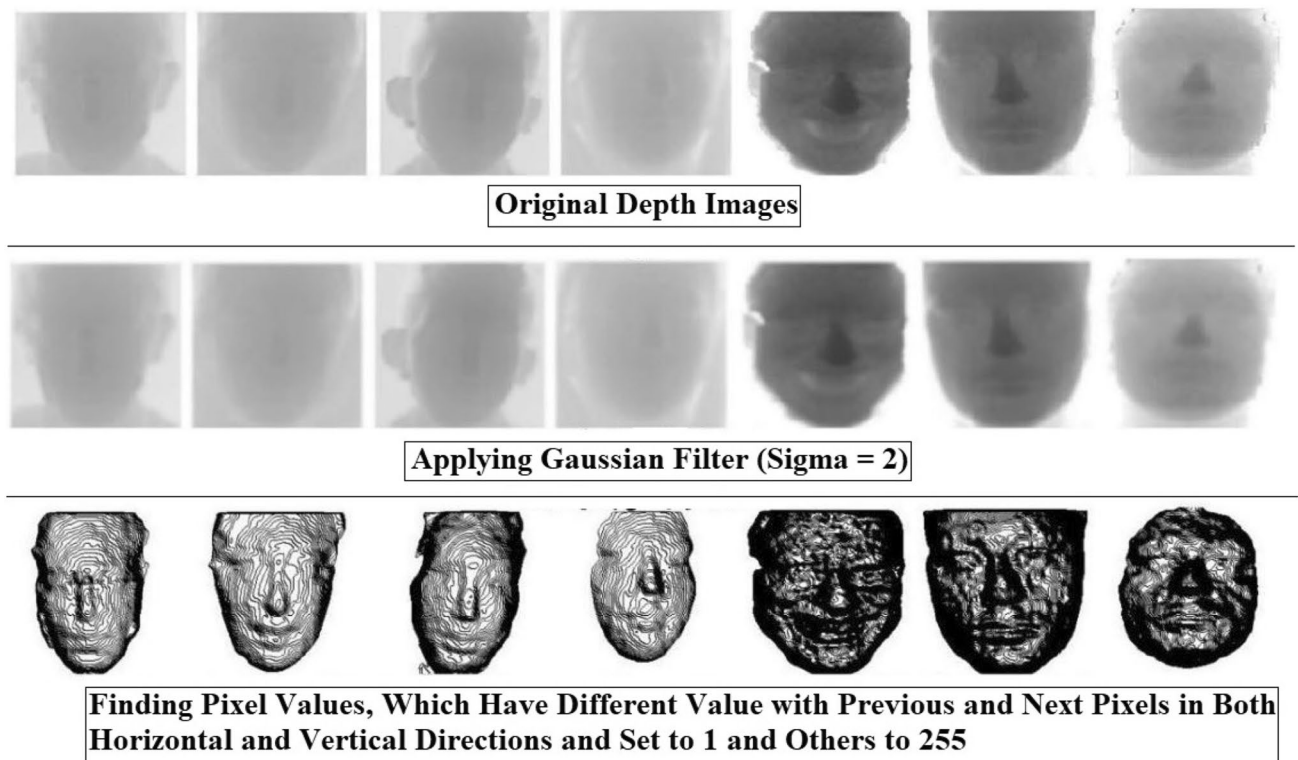


Fig. 14 Proposed method to changing the structure of depth image for processing using color features

$$\min_{\beta_0, \beta} \left(\frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - X_i^T \beta)^2 \right) + \lambda \sum_{j=1}^p |\beta_j|$$

- N is the number of observations.
- y_i is the response to observation i .
- x_i is data, a vector of p values at observation i .
- λ is a positive regularization parameter corresponding to one value of Lambda.
- The parameters β_0 and β are scalar and p -vector respectively.

As λ increases, the number of nonzero components of β decreases. The lasso problem involves the L^1 norm of β .

SVM is a famous classification algorithm in Artificial Intelligence (A.I), and researchers of the field are familiar with its fundamentals like finding support vectors and margin between them to make separating line or hyper plane between support vectors based on weight and bias to classify. For more information refer to [38]. Fig-ure 15 illustrates a simple linear model of SVM.

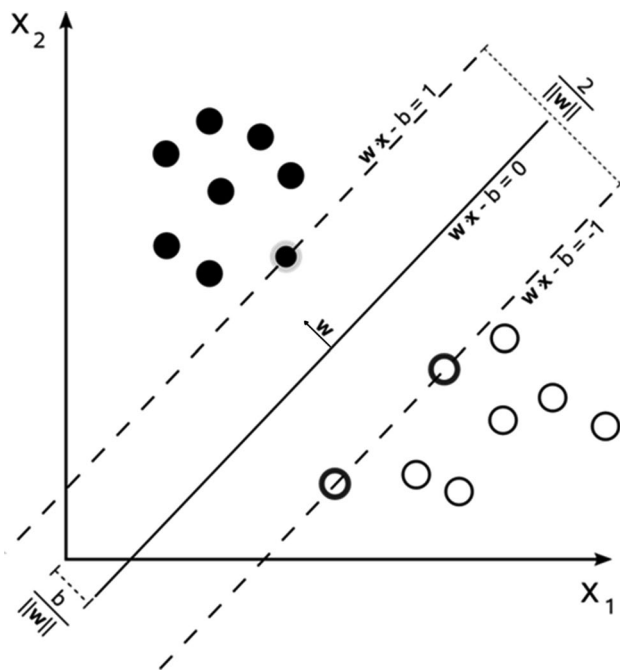


Fig. 15 A simple linear model of SVM

4.6 Deep learning and convolutional neural network (CNN)

One of the newest areas in machine learning is Deep Learning, pushing Machine Learning closer to original goals of Artificial Intelligence [48, 49].

There are different deep learning techniques such as fully connected and Convolutional Neural Network (CNN) [50] methods. They gained popularity in image processing, and deep learning is currently the state of the art for detecting what an image is, or what image matrix includes.

CNN structure is as based on 1. Convolution 2. Pooling 3. Convolution 4. Pooling 5. Fully Connected Layer and the Output.

The act of taking the original data in order to create feature maps is called convolution. Pooling is down-sampling, using max-pooling technique where it selects a region and taking the maximum value of it. This maximum value is the new value for the entire region. Fully Connected Layers are common neural networks, where all nodes are in fully connected form and the convolutional layers are not fully connected like a traditional neural network.

For more information about Multi-Layer Neural Network (MLNN), Shallow Neural Network and Deep Neural Network (deep learning) refer to [50]. CNNs are often used in object and scene detection and segmentation. CNNs learn directly from image matrix which eliminates the feature extraction step. A convolutional neural network may have tens or hundreds of layers; each layer may learn to detect different features of an image. Filters are applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer. The filters can start as very simple features, such as brightness and edges, and increase in complexity to features that uniquely define the object as the layers progress. Filters are applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer.

Table 2 presents classification results on all databases for two main purposes of FER and FMER using three main classification methods of SVM, MLNN and CNN. Table 3 contains results on proposed RGBD-IKFDB face database for FER purpose as confusion matrix and Table 4 represents same for FMER purpose using SVM algorithm. Tables 5 and 6 present the same results but with MLNN (feed forward NN) algorithm and 100 hidden layers. Finally, Tables 7 and 8 represent the same method on our database but this time using CNN algorithm.

Table 2 Classification results on proposed IKFDB and other face databases

CASME (%)	VAP RGB-D (%)	Face grabber (%)	IKFDB (%)
Facial Expression Recognition(FER)			
SVM			
RGB=-	RGB=93	RGB=88±3	RGB=89
Depth=-	Depth=88±2	Depth=88±1	Depth=80±3
Total=-	Total=89±1	Total=91	Total=90
Facial Micro Expression Recognition(FMER)			
SVM			
RGB=63±2	RGB=-	RGB=61±2	RGB=72±2
Depth=-	Depth=-	Depth=60±2	Depth=65±2
Total=63±2	Total=-	Total=62±1	Total=72
Facial Expression Recognition(FER)			
MLNN			
RGB=-	RGB=95±1	RGB=91±1	RGB=93±2
Depth=-	Depth=90±2	Depth=88±1	Depth=85±2
Total=-	Total=91±1	Total=89±1	Total=92±3
Facial Micro Expression Recognition(FMER)			
MLNN			
RGB=71±3	RGB=-	RGB=65	RGB=75±1
Depth=-	Depth=-	Depth=63±3	Depth=70±3
Total=71±3	Total=-	Total=64±1	Total=76
Facial Expression Recognition(FER)			
CNN			
RGB=-	RGB=97±2	RGB=94±1	RGB=95
Depth=-	Depth=93±3	Depth=90	Depth=90±2
Total=-	Total=94±1	Total=92±2	Total=95±1
Facial Micro Expression Recognition(FMER)			
CNN			
RGB=74±1	RGB=-	RGB=69	RGB=86±1
Depth=-	Depth=-	Depth=68±1	Depth=81±1
Total=74±1	Total=-	Total=68	Total=85±1

5 Discussion

As mentioned earlier, there are similar databases for European face type, but due to the shortage of Middle-Eastern face type database, creating a Middle-Eastern

face type color-depth based database for FER and FMER seemed essential. Figure 16 represents comparison results using three classification algorithms on all four databases. Classification results using different classification algorithms on IKFDB and based on each expression and micro expression is presented in Fig. 17. Results of employed method on proposed IKFDB and three other databases show a promising performance of IKFDB on FER and FMER (Table 2). Surely this system works on face and gender recognition too. Table 2 shows just test results on each database and there are two databases with lack of information in one of two parts (having FMER or being Depth-based). CASME database is color type but supports micro expressions, so there is no facial expressions and depth results for it. Total micro expressions achieved by employed method on this database is satisfactory. As mentioned before, CASME supports 7 main micro expressions. It has to be mentioned that metric system is based on accuracy. Having recognition accuracy of 63, 71 and 74% for SVM, MLNN and CNN individually on FMER, was one of our weak results. VAP RGB-D database is color—depth type and does not supports micro expressions, and with a total recognition accuracy of 89, 91 and 94% for SVM, MLNN and CNN on FER returned acceptable results. Face Grabber database is a perfect face database with 7 main facial expressions and micro expressions, created in 2016 and it is employed in the paper very well. Having total recognition accuracy of 91, 89 and 92% for FER and 62, 64 and 68% for FMER made our method on this database functional. And finally, our proposed IKFDB database in both FER and FMER achieved 90% and 72% recognition accuracy for SVM, 92% and 76% for MLNN, 95% and 85% for CNN respectively, which is quite promising. Also, some additional experiments for face recognition and face detection purposes are done on proposed IKFDB. For face recognition purpose and for SVM, MLNN and CNN algorithms, recognition accuracies are 99.2%, 99.7% and 100.0% respectively. Face detection for all algorithms indicates 100.0% accuracy on the IKFDB.

Table 3 Confusion matrix of proposed RGBD-IKFDB face database for FER purpose (SVM)

Neutral	85%	10%	-	2%	3%	-	-
Happiness	2%	96%	-	-	2%	-	-
Anger	2%	1%	89%	5%	-	3%	-
Sadness	3%	-	6%	85%	-	5%	1%
Surprise	-	1%	-	-	98%	-	1%
Disgust	-	-	4%	4%	-	91%	1%
Fear	2%	-	4%	4%	1%	2%	87%
-	Neutral	Happiness	Anger	Sadness	Surprise	Disgust	Fear

Table 4 Confusion matrix of proposed RGBD-IKFDB face database for FMER purpose (SVM)

Neutral	65%	9%	2%	4%	7%	6%	7%
Happiness	9%	71%	–	–	13%	–	7%
Anger	4%	–	68%	16%	3%	9%	–
Sadness	6%	–	11%	67%	2%	10	4%
Surprise	4%	12%	–	–	80%	–	4%
Disgust	7%	1%	13%	9%	–	70%	–
Fear	2%	8%	1%	3%	12%	2%	72%
–	Neutral	Happiness	Anger	Sadness	Surprise	Disgust	Fear

Table 5 Confusion matrix of proposed RGBD-IKFDB face database for FER purpose (MLNN)

Neutral	92%	5%	–	–	3%	–	–
Happiness	–	97%	–	–	3%	–	–
Anger	3%	–	87%	5%	–	5%	–
Sadness	1%	–	4%	89%	–	4%	2%
Surprise	–	1%	–	–	99%	–	–
Disgust	–	–	–	3%	–	93%	3%
Fear	3%	–	–	–	3%	4%	90%
–	Neutral	Happiness	Anger	Sadness	Surprise	Disgust	Fear

Table 6 Confusion matrix of proposed RGBD-IKFDB face database for FMER purpose (MLNN)

Neutral	74%	10%	–	–	12%	–	4%
Happiness	7%	80%	1%	2%	8%	–	2%
Anger	3%	–	77%	8%	2%	8%	2%
Sadness	4%	2%	5%	75%	4	6%	4%
Surprise	5%	11%	–	–	82%	–	2%
Disgust	2%	4%	8%	3%	1%	78%	4%
Fear	4%	7%	2%	6%	4%	5%	72%
–	Neutral	Happiness	Anger	Sadness	Surprise	Disgust	Fear

Table 7 Confusion matrix of proposed RGBD-IKFDB face database for FER purpose (CNN)

Neutral	96%	4%	–	–	–	–	–
Happiness	–	99%	–	–	1%	–	–
Anger	1%	–	94%	4%	–	1%	–
Sadness	1%	–	2%	95%	–	2%	–
Surprise	–	–	–	–	100%	–	–
Disgust	–	–	4%	4%	–	92%	–
Fear	2%	3%	–	–	4%	–	91%
–	Neutral	Happiness	Anger	Sadness	Surprise	Disgust	Fear

Table 8 Confusion matrix of proposed RGBD-IKFDB face database for FMER purpose (CNN)

Neutral	85%	4%	–	–	6%	–	5%
Happiness	4%	88%	–	1%	5%	–	2%
Anger	3%	–	85%	7%	–	3%	2%
Sadness	7%	–	4%	86%	–	3%	–
Surprise	1%	9%	–	–	90%	–	–
Disgust	–	1%	6%	8%	–	83%	2%
Fear	3%	3%	2%	2%	6%	6%	78%
–	Neutral	Happiness	Anger	Sadness	Surprise	Disgust	Fear

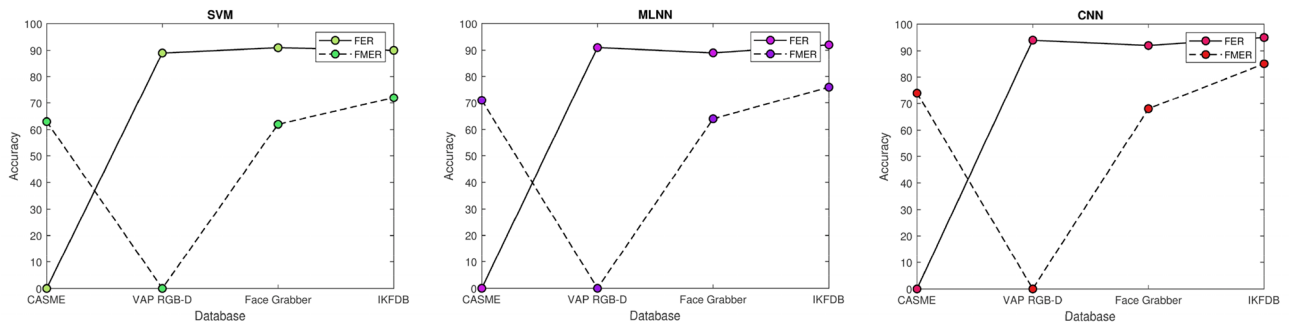


Fig. 16 Comparison results using different classification algorithms on all databases using employed method based on Table 2

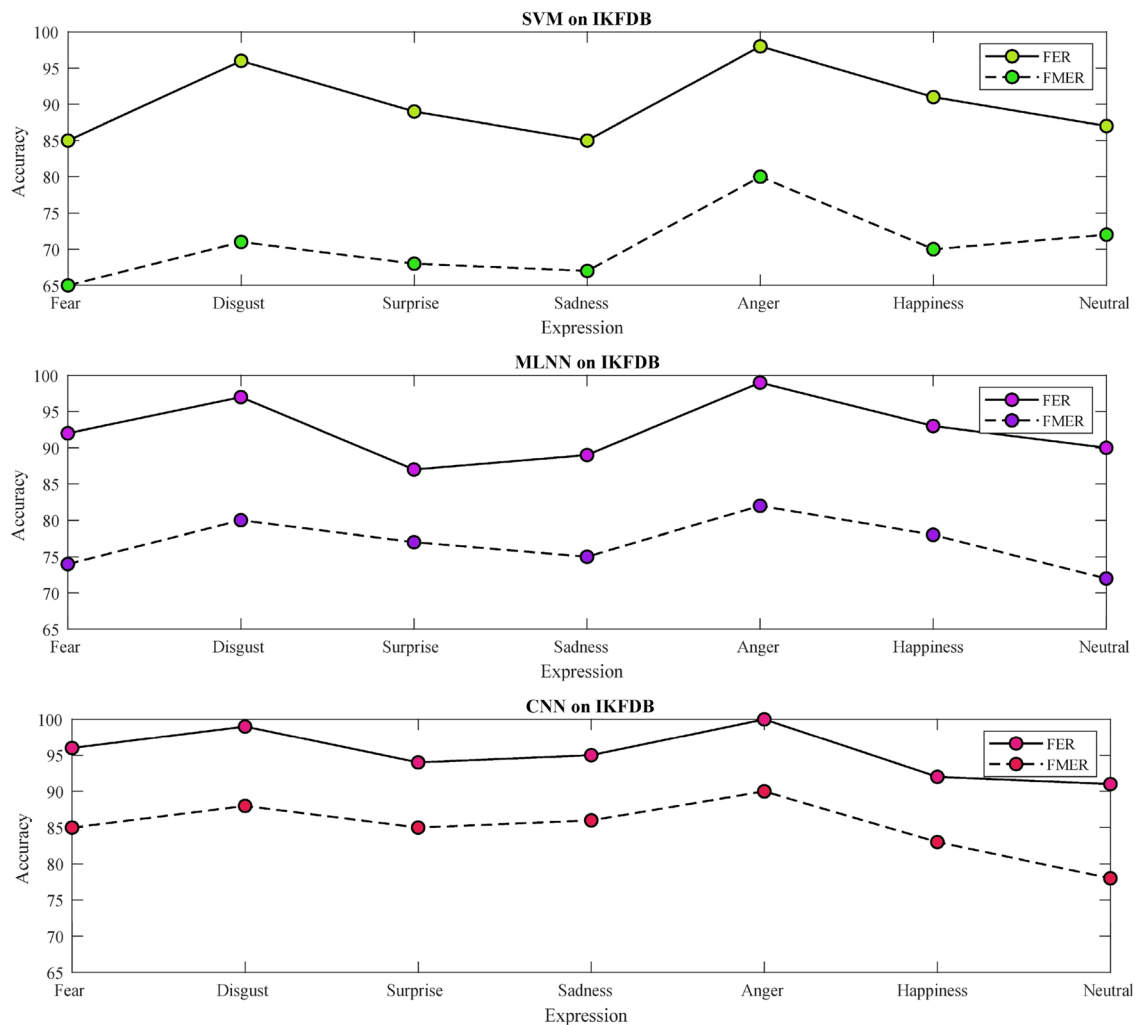


Fig. 17 Result on each expression and micro expression individually using three classification algorithms on IKFDB

6 Conclusion

The shortage of RGB-D FMER face databases have been a problem for researchers in the field of face analysis. Considering existing bugs in available RGB-D face databases and the shortage of Middle-Eastern face type database for researchers and necessity of making such RGB-D face databases, a new face database covering these problems was collected. The main purpose of the paper is to introduce a database in color and depth form and investigating its applicability on FER and FMER purposes. In order to conforming the validity of the databases, several experiments on features extraction and classification has been done. Also, a robust face extraction method out of depth images is used which changes the structure of depth image into color image, in order to using color image feature extraction methods on depth images. IKFDB indicated almost a perfect performance in FER and FMER using CNN in comparison with other similar databases. Finding subjects with a wide variety of age range and sex types was the main challenge of the study. We managed to overcome this limitation, using subjects from an English language school in Tehran when they were at break. Some of future works are increasing the number of expressions to 12 or more and using 2 or more sensors for fusion of images to have better accuracy. Also, testing IKFDB for face recognition and age estimation purposes is recommended. It is possible to convert 2.5-Dimensional depth images to 3-Dimensional models in order to analyze subjects by 3-D algorithms. Hope this database would play a constructive role in future studies in this field.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal rights Research involving Human Participants. All the subject participants contributed in the IKFDB experiment, have granted their consents in EULA form. Also, as the database includes few children involved, recording process was done with complete informed consent of the parents and under supervision of their English language school teacher. The database is recorded in an English language school at Rudehen—Tehran.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended

use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Zhang Z (2012) Microsoft Kinect sensor and its effect. *IEEE Multimedia* 19(2):4–10
- Mohammadkhani, Mahsa. Kinect 3D reconstruction. <https://slideum.com/doc/3391910/mahsam.ir>
- Meng, Ma, et al (2013) "Kinect for interactive AR anatomy learning." *Mixed and Augmented Reality (ISMAR), IEEE International Symposium on. IEEE, 2013*
- Casas, Xavier, et al (2012) "A Kinect-based Augmented Reality System for Individuals with Autism Spectrum Disorders." *GRAPP/IVAPP*
- Arun K, Zorn C, and Joseph LaViola J (2013) "Poster: Real-time markerless kinect based finger tracking and hand gesture recognition for HCI." *3D User Interfaces (3DUI), 2013 IEEE Symposium on. IEEE*
- El-laithy Riyadh A, Jidong H, and Michael Y (2012) "Study on the use of Microsoft Kinect for robotics applications." *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION. IEEE*
- Zhuang X et al (2008) "Face age estimation using patch-based hidden markov model supervectors." *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE*
- Huynh T, Min R, and Jean Luc D (2012) "An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGBD face data". *Asian Conference on Computer Vision Springer, Berlin*
- Boutellaa E et al (2015) On the use of Kinect depth data for identity, gender and ethnicity classification from facial images. *Pattern Recognit Lett* 68:270–277
- Li Billy YL, et al (2013) "Using kinect for face recognition under varying poses, expressions, illumination and disguise." *Applications of Computer Vision (WACV), 2013 IEEE Workshop on. IEEE*
- Sandbach G et al (2012) Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image Vis Comput* 30(10):683–697
- Polikovskiy S, Yoshinari K, and Yuichi O (2009) Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In: *3rd international conference on imaging for crime detection and prevention (ICDP 2009)*, p 16
- Cox JL, Holden JM, Sagovsky R (1987) Detection of postnatal depression: development of the 10-item edinburgh postnatal depression scale. *Br J Psychiatry* 150(6):782–786
- Hsu R-L, Abdel-Mottaleb M, Jain AK (2002) Face detection in color images. *IEEE Trans Pattern Anal Mach Intell* 24(5):696–706
- Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vision* 57(2):137–154
- Sun X, Wu P, and Hoi SCH (2017) "Face detection using deep learning: An improved faster rcnn approach. *Neurocomputing* 299: 42-50
- ter Haar FB, and Veltkamp RC (2008) "3D face model fitting for recognition." *European Conference on Computer Vision. Springer, Berlin, Heidelberg*
- Tian Y-I, Takeo K, Cohn JF (2001) Recognizing action units for facial expression analysis. *IEEE Trans Pattern Anal Mach Intell* 23(2):97–115

19. Carl-Herman H (1969) Man's face and mimic language. Studen litteratur, Sweden
20. Ekman P, Friesen W (1978) Facial action coding system: a technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto
21. Ekman P, Friesen WV, Hager JC (2002) Facial action coding system: the manual on cd rom. A Human Face, Salt Lake City
22. Jihun H et al (2011) Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *J Neurosci Method* 200(2):237–256
23. Wu Qi, Shen X, Xiaolan Fu (2011) The machine knows what you are hiding: an automatic micro-expression recognition system *Affective Computing and Intelligent Interaction*. Springer, Berlin, pp 152–162
24. Ekman P (2009) Lie catching and microexpressions. *Philos Decept* 1:118–133
25. Lyons M et al (1998) "Coding facial expressions with gabor wavelets." *Automatic Face and Gesture Recognition*, 1998. Proceedings. Third IEEE International Conference on IEEE
26. Lundqvist D, Flykt A, and Öhman A (1998) The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9
27. Li X, et al (2013) "A spontaneous micro-expression database: Inducement, collection and baseline." *Automatic face and gesture recognition (fg)*, 10th IEEE international conference and workshops on. IEEE, 2013.
28. Min R, Kose N, Dugelay J-L (2014) Kinectfacedb: a kinect database for face recognition. *IEEE Trans Syst Man Cybern: Syst* 44(11):1534–1548
29. Y Wen-Jing, et al (2013) "CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces." *Automatic face and gesture recognition (fg)*, 2013 10th IEEE international conference and workshops on. IEEE
30. S Matthew et al (2011) "Macro-and micro-expression spotting in long videos using spatio-temporal strain." *Automatic Face and Gesture Recognition and Workshops (FG 2011)*, 2011 IEEE International Conference on. IEEE,
31. Warren G, Schertler E, Bull P (2009) Detecting deception from emotional and unemotional cues. *J Nonverbal Behav* 33(1):59–69
32. Hg RI et al (2012) "An rgb-d database using microsoft's kinect for windows for face detection." *Signal Image Technology and Internet Based Systems (SITIS)*, 2012 Eighth International Conference on. IEEE
33. Nikisins O et al (2014) "RGB-DT based face recognition." *Pattern Recognition (ICPR)*, 2014 22nd International Conference on. IEEE,
34. Szwoch M (2013) "FEEDB: a multimodal database of facial expressions and emotions." *Human System Interaction (HSI)*, 2013 The 6th International Conference on. IEEE
35. D Merget, T Eckl, M Schwörer, P Tiefenbacher, and G Rigoll (2016) "Capturing Facial Videos with Kinect 2.0: A Multithreaded Open Source Tool and Database", in *Proc. WACV*, IEEE
36. Viola P, and Michael J (2001) "Rapid object detection using a boosted cascade of simple features." *Computer Vision and Pattern Recognition*, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE
37. William FT, and Michal R (1995) "Orientation histograms for hand gesture recognition." *International workshop on automatic face and gesture recognition*. Vol. 12
38. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
39. <https://www.mathworks.com/help/images/ref/stdfilt.html>
40. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Royal Stat Soc Ser B (Methodol)* 58:267–288
41. Friedman J, Hastie T, Tibshirani R (2001) *The elements of statistical learning*, vol 1. Springer series in statistics, New York
42. Bühlmann P, Van De Geer S (2011) *Statistics for high-dimensional data: methods, theory and applications*. Springer Science and Business Media, NY
43. Rosasco L et al (2004) Are loss functions all the same? *Neu Comput* 16(5):1063–1076
44. Abu-Mostafa YS, Magdon-Ismail M, Lin H-T (2012) *Learning from data*, vol 4. AMLBook, New York
45. Mousavi SMH (2018) A new way to age estimation for rgb-d images, based on a new face detection and extraction method for depth images. *Int J Image Gr Signal Process* 10(11):10
46. MSH Mousavi, and Surya Prasath VB (2019) "On the Feasibility of Estimating Fruits Weights Using Depth Sensors.", 4th International Congress of Developing Agriculture, Natural Resources, Environment and Tourism of IranAt: Tabriz Islamic Art University In cooperation with Shiraz University and Yasouj University, Iran
47. MSH Mousavi, V Lyashenko, and S Prasath (2019) "Analysis of a robust edge detection system in different color spaces using color and depth images." *Компьютерная оптика* 43(4)
48. <http://deeplearning.net/>
49. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
50. Nielsen MA (2015) *Neural networks and deep learning*, vol 25. Determination press, San Francisco
51. Mousavi MSH, and N Aghsaghloo (2018) "Using Genetic Programming for Making a New Evolutionary Artwork, Based on Human-Computer Interactions for Autism Rehabilitation.", The third International Conference on Intelligent Decision Science (IDS 2018)At: Tehran-Iran
52. Zhang L, and H Ma (2019) "Dense Scale Invariant Feature Transform-Based Quality Assessment for Tone Mapping Image." 2019 International Conference on Electronical, Mechanical and Materials Engineering (ICE2ME 2019). Atlantis Press
53. Qu F et al (2017) CAS (ME)²: A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition. *IEEE Trans Affect Comput* 9(4):424–436
54. Chhokra P et al (2018) Unconstrained kinect video face database. *Inf Fus* 44:113–125
55. Turan C, KD Neergaard, and KKM Lam (2019) "Facial Expressions of Comprehension (FEC)." *IEEE Transactions on Affective Computing*
56. Hassaballah M, Saleh A (2015) Face recognition: challenges, achievements and future directions. *IET Computer V* 9(4):614–626
57. Scherhag U et al (2019) Face recognition systems under morphing attacks: A survey. *IEEE Access* 7:23012–23026
58. Hassaballah M, Kenji M, Shun I (2013) Face detection evaluation: a new approach based on the golden ratio ϕ . *Signal Image Video Process* 7(2):307–316

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.