A Spatial and Frequency Based Method for Micro Facial Expressions Recognition Using Color and Depth Images

Seyed Muhammad Hossein Mousavi

Independent Scientist, Tehran, Iran

Email: mosavi.a.i.buali@gmail.com

Abstract:

Human face states the inner emotions, thoughts and physical disorders. These emotions are expressed on the face via facial muscles. The estimated time through which a facial expression occurs on the face is between 0.5 to 4 seconds, and a micro expression between 0.1 to 0.5 seconds. Obviously, for the purpose of recording micro expressions, obtaining videos frames between 30 up to 200 frame per second is essential. This research uses Kinect V.2 sensor to get the color and depth data in 30 fps. Depth image stores useful 2.5-Dimentional information from skin wrinkles which is the main key to recognize even slightest micro facial expressions. Experiment starts with splitting color and depth images into facial parts, and after applying preprocessing techniques, features extraction out of both type of data in spatial and frequency domain takes place. Some of the features which are used in this study are Histogram of Oriented Gradient (HOG), Gabor Filter, Speeded Up Robust Features (SURF), Local Phase Quantization (LPQ), Local Binary Pattern (LBP). Non dominated Sorting Genetic Algorithm II (NSGA-II) feature selection algorithm applies on extracted features to have faster learning process and finally selected features are sent to neuro-fuzzy and neural network classifiers. Proposed method is evaluated with the benchmark databases such as, Eurecom Kinect Face DB, VAP RGBD-T Face, JAFFE, Face Grabber DB, FEEDB, and CASME. Also, the proposed method is compared with other similar methods and Convolutional Neural Network (CNN) method on mentioned databases. The results are really satisfactory, and it indicates classification accuracy improvement of proposed method versus other methods.

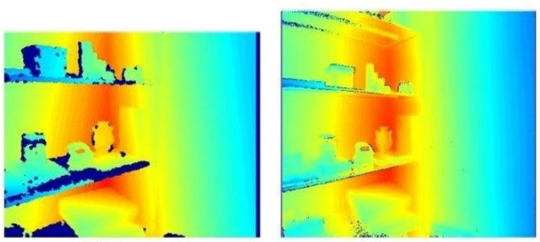
Key Words: Micro facial expressions recognition; Kinect sensor; Depth data; Spatial and frequency domain; Evolutionary feature selection; Neuro-fuzzy classifier

1. INTRODUCTION

Image processing sensors are affected our daily lives. They are employed in vehicle control systems, security, entertainment, market places, art [3], army, psychology, medicine, agriculture [4] and even gaming industry and … [1, 2]. One of its usage is a subcategory of psychology. It is face analysis and two important subcategory of face analysis are called Facial Expressions Recognition (FER) [5] and Micro Facial Expressions Recognition (FMER) [6] as they express human emotions via facial muscles. As micro expressions are hard to recognize in color images and Depth images [14] or 2.5-Dimentional (2.5-D) images [7] bring more details from surface of any object, Kinect sensor V.2 [8] is employed in this experiment. By converting 2.5-D images from point cloud space into 3-Dimentional (3-D) space, it is possible to get all face wrinkles which are the micro expression this paper intended to recognize in their slight appearance level on the face. All depth sensors could detect distance between object and the sensor by projecting infrared points to the object and receiving them in the origin place for example as millimeter. It has to mention that they have different technologies to achieve this. Those which uses infrared spectrum, could work perfectly in the pure darkness condition. Some of the famous depth sensors are Kinect, Asus Xtion, Minolta, Inspeck Mega Capturor and etc. Table 1 shows some of these sensors and their specifications. Figure 1 presents Kinect first and second generations differences. Also Figure 2 represents Kinect V1. And V2 beside each other.

Table 1
Famous depth sensors and their specifications

SENSOR	TYPE	RESOLUTION IN MM	WORKING DISTANCE IN M	PRICE IN $
MINOLTA	3-D Laser Scanning	0.041 – 0.22	2.5	25000
3DMDFACE	Vision Cameras	< 0.2	-	10 K – 20 K
CYBERWARE 3030 RGB/PS	Low-Intensity Laser Light Source	0.08 – 0.3	0.35	72000
INSPECK MEGA CAPTURER II	Structred Light	0.7	1.1	Not Availble
KINECT V.1	IR laser Emitter	1.5 – 0.5	0.5 – 4.5	Not Availble
KINECT V.2	Time of Flight	-	0.5 – 8	149.99
SOFTKINETIC DS325	Diffused Laser	1.4 at 1 mdistance	0.15 – 1	259
STRUCTURE	IR Structured Light	0.5 – 30	3.5	379
PRIMESENSE CARMINE	IR Laser Emitter	0.1 – 1.2	3.5	Not Availble
ASUS XTION PRO LIVE	IR Laser Emitter	-	0.8 – 3.5	169.99
INTEL REALSENSE	Structured Light	< 1	0.2 - > 10	99 - 399



(a) Kinect v1 (b) Kinect v2

	Kinect v1		Kinect v2	
	Resolution [Pixel × Pixel]	Frame Rate [Hz]	Resolution [Pixel × Pixel]	Frame Rate [Hz]
color	640 × 480	30	1920 × 1080	30
depth	640 × 480	30	512 × 424	30
infrared	640 × 480	30	512 × 424	30

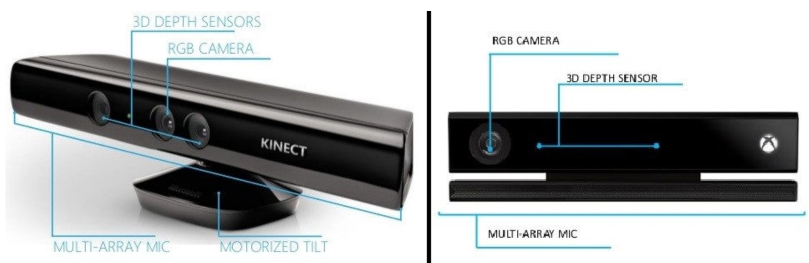
(a)

Kinect V1 vs Kinect V2

Feature	Kinect for Windows 1	Kinect for Windows 2
Color Camera	640 x 480 @ 30 fps	1920 x 1080 @ 30 fps
Depth Camera	320 x 240	512 x 424
Max Depth Distance	~4.5 M	~4.5 M
Min Depth Distance	40 cm in near mode	50 cm
Horizontal Field of View	57 degrees	70 degrees
Vertical Field of View	43 degrees	60 degrees
Tilt Motor	yes	no
Skeleton Joints Defined	20 joints	25 joints
Full Skeletons Tracked	2	6
USB Standard	2.0	3.0
Supported OS	Win 7, Win 8	Win 8-8.1 (WSA)
Price	$299	TBD

(b)

Fig 1. Kinect V.1 and V.2 differences. Constant scene recording (a) and main differences (b) [8]



Kinect V.1 **Kinect V.2**

Fig 2. Kinect V1 and V2 [8]

Color images are 2-Diomentional (2-D) in different color spaces such as Red Green Blue (RGB), YIQ, CIELAB, YCbCr, CMYK [9] and etc. But Kinect depth images are stored in a 2-D matrix which each cell value represents the distance between object and the sensor in millimeter and in the range of 0 to 255 and that's why depth image is called 2.5-D image. Depth image is visible to human eye as a gray image and as blacker the pixel is, the closer distance between that pixel and sensor indicates. Figure 3 shows a recorded Kinect sample in the experiment in both color and depth modes. In the figure, left image is color image in gray level form and right image is depth data.

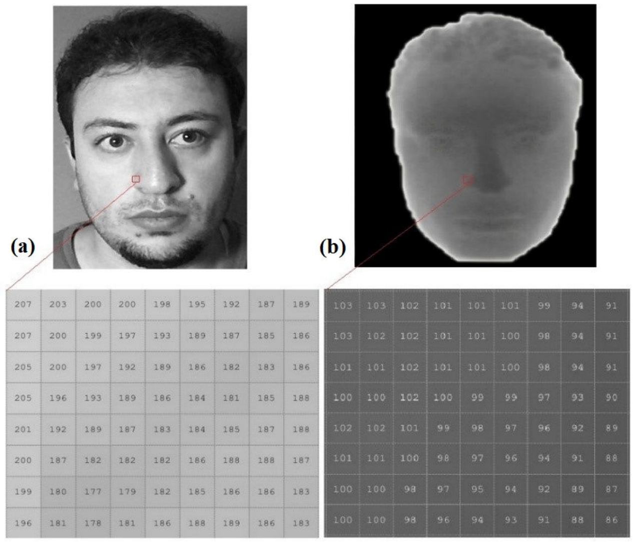


Fig 3. A recorded sample in the experiment using Kinect V.2 sensor. Color (a) and Depth (b)

1.1 Facial expressions recognition and micro facial expressions recognition

In order to explain the Facial Expressions Recognition (FER) [5] and Facial Micro Expressions Recognition (FMER) [6], it is needed to explain face detection and recognition first. If a system could distinguish the face objects out any other objects in a digital image, then this system is called face detection system. Now if a system could distinguish a specific identity by face in a bunch other face images, then the system is called face recognition system. Each human face, despite of gender, age and race could express seven main expressions in general. Expression recognition states which type of emotion subject is in, out of seven main emotions. It is mentionable that other emotions or expressions are combinations of these seven emotions, or it can be said combination of facial muscle which are employed to express seven main emotions. These seven main emotions or face expressions are joy or happiness, sadness, anger, surprise, disgust, fear and neutral. This paper is intended to classify these expressions and micro version of them from face object out of color and depth images. Figure 4 represents neutral expression and from left to right in 2-D color, 2.5-D depth and 3-D point cloud forms.

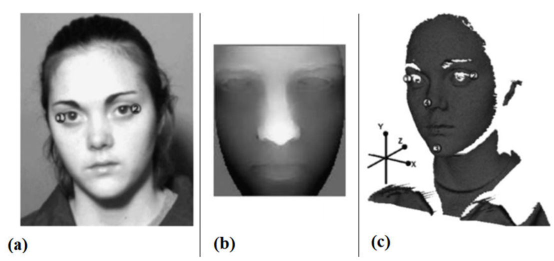


Fig 4. 2-D color (a), 2.5-D depth (b), 3-D point cloud (c) [10]

Facial features are using to determine race, gender, mood, age and etc. Some of these features are permanent like bone structure, skin texture color and some of them are temporary like cosmetics, glass, bear and facial muscle exercises. In total facial muscles are the main factor for facial expressions appearance. Also, facial parts have important impact on a facial expression, like mouth, eyes and nose.

It is better to weight each face element to have better result in final recognition accuracy. Eyes and mouth due to have higher effect, should have more weight than other parts. Figure 5 shows some facial elements or parts and facial elements weighting.

There are factors which should be considered in facial expressions recognition experiment such as: face pose, environmental light intensity changes and face blocking. Figure 6 presents these factors. Also, Figure 7 shows all seven main facial expressions along with their 3-D model.

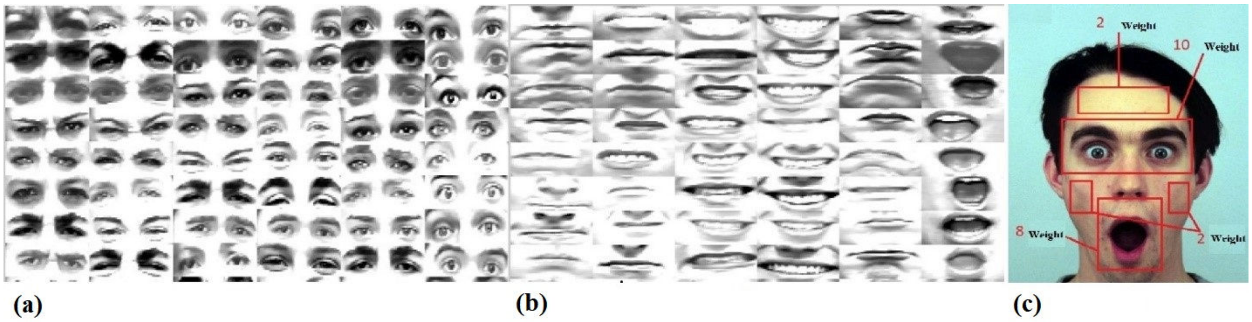


Fig 5. Face part eye (a), face part mouth (b) and facial parts weighting (c) [40]

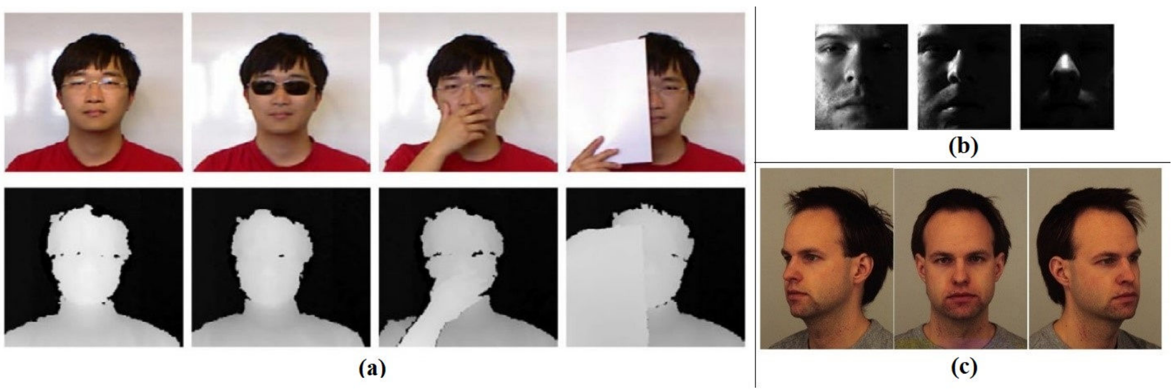


Fig 6. Face blocking (a) [11], environmental light intensity changes (b) and face pose (c) [40]

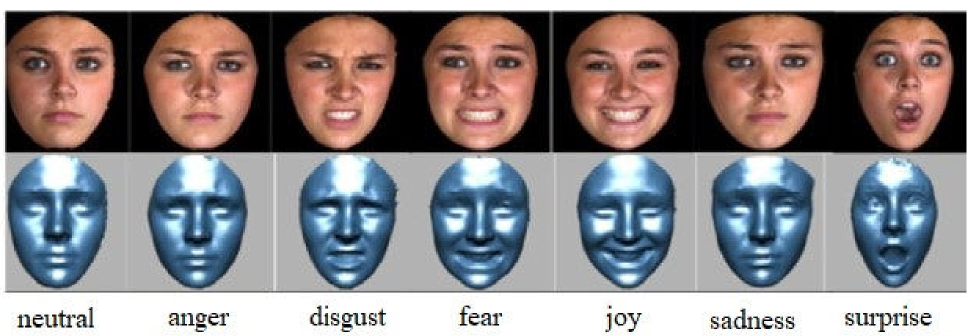


Fig 7. Seven main facial expressions [12]

1.2 Facial Action Coding System

Facial Action Coding System (FACS) [13] is the best way for coding human facial muscles movement known. Each action consisted of a muscle movement which with combining them, making every expression is possible. FACS is made in 1978 and included 44 action units and during time increased till 51 action units in 2002 by scientists. Action units 1 to 7 are related to upper parts of the face and others to lower parts. First 20 action units along with their specifications or muscle descriptions are described in the Table 2.

Table 2
First 20 actions unit in FACS

AU NUMBER	FACS NAME	MUSCULAR BASIS
0	Neutral face	-
1	Inner brow raiser	frontalis (pars medialis)
2	Outer brow raiser	frontalis (pars lateralis)
4	Brow lowered	depressor glabellae, depressor supercilii, corrugator supercilii
5	Upper lid raiser	levator palpebrae superioris, superior tarsal muscle
6	Cheek raiser	orbicularis oculi (pars orbitalis)
7	Lid tightener	orbicularis oculi (pars palpebralis)
8	Lips toward each other	orbicularis oris
9	Nose wrinkle	levator labii superioris alaeque nasi
10	Upper lip raiser	levator labii superioris, caput infraorbitalis
11	Nasolabial deepener	zygomaticus minor
12	Lip corner puller	zygomaticus major
13	Sharp lip puller	levator anguli oris (also known as caninus)
14	Dimple	buccinator
15	Lip corner depressor	depressor anguli oris (also known as triangularis)
16	Lower lip depressor	depressor labii inferioris
17	Chin raiser	mentalis
18	Lip pucker	incisivii labii superioris and incisivii labii inferioris
19	Tongue show	-
20	Lip stretcher	risorius w/ platysma

By considering action units and combining them, it is possible to get all expressions possible on human face. Below just some action unit's combination and their related expressions are mentioned. Figure 8 presents some expressions based on FACS.

- 26+25+5+2+1= Surprise
- 6+12 = Joy
- 17+15+1= Sadness
- 25+10+9+7+4 = Anger

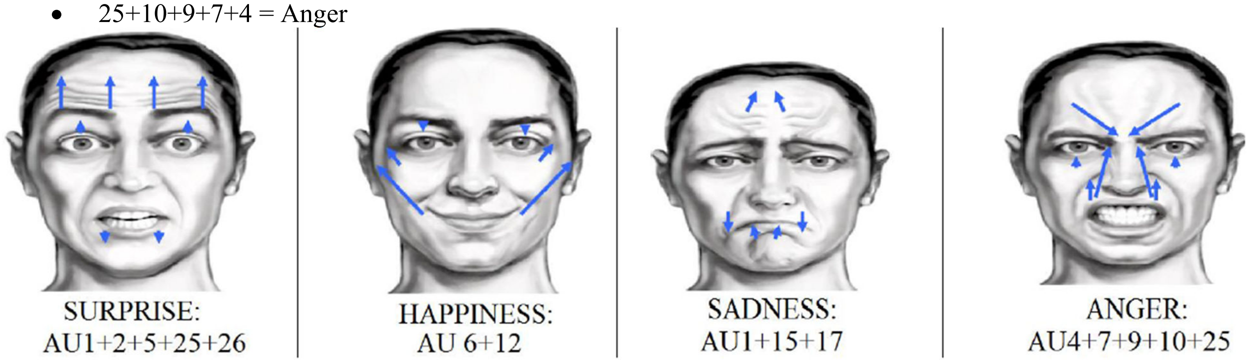


Fig 8. Four famous expressions based FACS [15]

2. PRIOR RELATED WORKS

As prior work section is relay on databses and verity of them, Table 3 presents these database in detailes. In order to increase the visiblity of the paper and saving space, prior related work setion is summerized in Table 4. This helps to have more space for proposed method and validation sections and decreses confusion in final reader.

Table 3
FER and FMER color and depth based databses along with their detailes

DATABASE	SAMPLES	SENSOR	USAGE	DATA TYPE	DIMENSIONS	EXPRESSIONS	YEAR	REF
Eurecom Kinect Face	14 male and 38 females	Kinect V.1	FER – face recognition	1248 Color + depth images	RGB= 256*256 Depth= 256*256	3 expressions of neutral, joy and surprise	2014	[11]
VAP RGB-D Face	13	Kinect V.1	FER – face recognition	2960 color + depth images	RGB= 351*421 Depth= 480*640	4 expressions of neutral, joy, anger, surprise and sadness	2012	[16]
VAP RGB-D-T Face	51	Kinect V.1 And AXIS Q1922	FER – face recognition	46360 color + depth + thermal images	RGB= 480*640 Depth= 480*640 Thermal= 288*384	4 expressions of neutral, joy, anger and surprise	2014	[17]
Curtin Face	25	Kinect V.1	FER – face recognition	5000 color + depth images	RGB= 480*640 Depth= 480*640	7 main expressions	2013	[18]
FEEDB	50	Kinect V.1	FER - FMER – face recognition	30 color and depth videos	RGB= 480*640 Depth= 480*640	33 facial expressions	2013	[19]
Face Grabber	33 male and 7female	Kinect V.2	FER - FMER – face recognition	67159 color + depth images	RGB= 2080*1920 Depth= 424*512	7 main expressions	2016	[20]
SMIC	16	Pixel INK PL-B774U	FMER	164 video files in 100 frame fps	640*480	4 expressions	2013	[21]
CASME	19	BenQ M31 GRAS-03K2C	FMER	195 video files in 60 fps	640*480 1280*780	**7** expressions	2013	[22]
Polikovsky's	10	Grasshopper	FMER	200 video files in 200 fps	640*480	13 expressions	2009	[23]
USF-HD	100	-	FER - FMER	56 video files in 30 fps	1280*780	6 expressions	2011	[24]
YorkDDT	9	-	FER	30 fps	640*480	18 expressions	2009	[25]
JAFFE	10	-	FER – face recognition	212 gray images	256*256	7 main expressions	1998	[56]

Table 4
Prior related works for FER and FMER

DATABASE	AUTHOR(S)-YEAR	FEATURE(S)	CLASSIFIER	ACCURACY	USAGE	COMMENT	REF
JAFFE	Wei-LunChao 2015	LPQ+ (es-LBP)	SVM	94.88	FER	Expression specific local binary pattern	[26]
KDEF	Elgarrai, Zineb 2016	Gabor Filter	HMM	88.6	FER	Dimensionality reduction Fisher's Analysis	[27]
Eurecom Kinect Face	Ijjina, Earnest Paul 2014	CNN	CNN	87.9	FER	-	[28]
VAP RGB-D Face	Hg RI, Jasek P 2012	PCA	PCA	92.88	FER	-	[29]
VAP RGB-D-T Face	Oliu Simon, Marc 2016	LBP + Haar + Hog	Weighted Nearest Neighbor Classifier (WNNC)	95.7	FER	LBP+ Haar + Hog = HOGOM	[30]
FEEDB	Mariusz Szwoch 2015	Local Features	KNN	50.0	FER	Using 25 samples and 9 expressions	[31]
Face Grabber	Sen Yuan, Xia Mao 2017	exponential elastic preserving projections (EEPP),	exponential elastic preserving projections (EEPP),	83.0	FER	Single faced	[32]
Politkovskaya's	S Polikovsky-2009	3D-Gradients orientation histogram		80%		Results are based on average of facial parts	[33]
YorkDDT SMIC	T Pfister-2011	LBP-TOP	Multiple Kernel Learning (MKL)	%71.4 %71.5		Weaker results using RF and SVM classifiers are achieved.	[34]
Cohn and Kaneda's (CK)	Wu, Qi-2011	Gabor Filter	Gentle SVM	%85.42		Core i5 650 system with 4GB memory	[35]
CASME	SJ Wang-2014	Discriminant Tensor Subspace Analysis (DTSA) and Extreme Learning Machine (ELM)		47%		Just micro expressions	[36]
SMIC	Huang, Xiao Hua-2015	LBP-TOP	SVM	%57.93		Using Temporal Interpolation Model (TMP) to normalizing each video to 10frames	[37]
SMIC, CASME	X Huang-2016	Spatiotemporal Completed Local Quantization Patterns (STCLQP)		75.31%		CASME= 68.93	[38]
CASME	Zheng, Hao-2017	2D Gabor filter	SVM	71.19%		CASME II= 64.88	[39]

3. PROPOSED METHOD

System starts with data acquisition from stream online or offline input. Then, face detection and extraction take place using Viola and Jones algorithm [41] for both depth and color images. Some of these extractions are presented in Figure 9 on some samples of Internet based, KDEF [40] and FEEDB [19] databases.

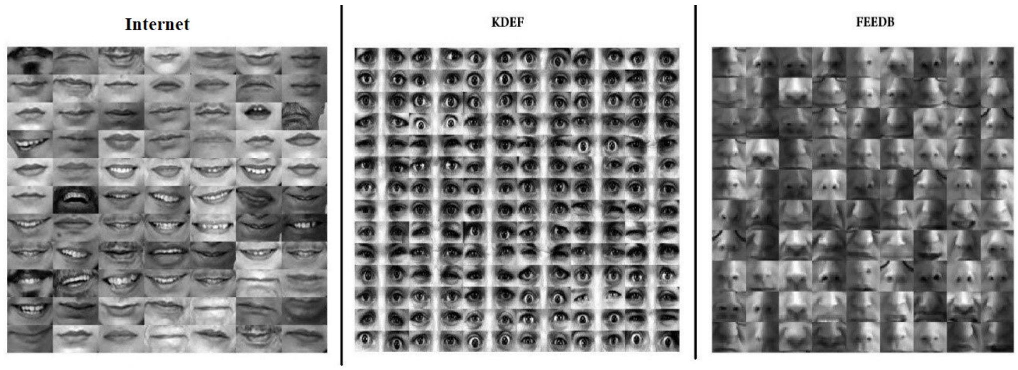


Fig 9. Extracted facial features from few databases in color mode

Third step is to split input color and depth data into facial parts of mouth, eyes and nose for each subject. Forth step is consisting of preprocessing operations. As this step has high of importance in achieving better results, it needs to be done using the best algorithms. Steps are as follow:

Median Low pass filter applies on facial parts followed by unsharp mask filter which is a high pass filter to have smoothen edges from inside and sharpened edges from outside. Histogram equalization fixes the brightness and illumination levels, especially in-depth image. Closing morphological operation is very important to get rid of any unwanted holes [1] [62]. Canny edge detection finds the best edges possible for feature extraction step [42]. The following step consists of extracting spatial and frequency domains features from color and depth images. Figure 10 represents the proposed method workflow.

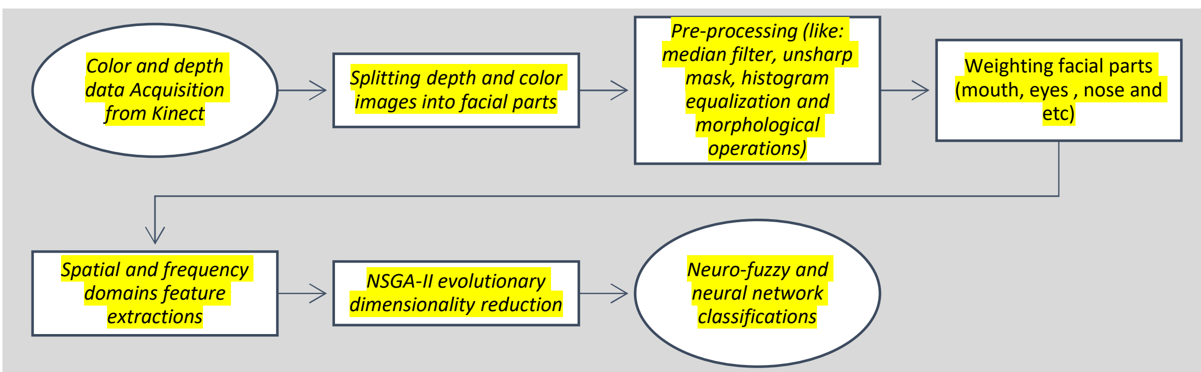


Fig 10. Proposed method flowchart

Features of Histogram of Oriented Gradient (HOG) [43], Gabor Filter [44], Speeded Up Robust Features (SURF) [45] are extracted from color images and features of Local Phase Quantization (LPQ) [46], Local Binary Pattern (LBP) [47] from depth images. In order to decrease runtime and getting rid of unnecessary outlier data and also increasing recognition accuracy, NSGA-II evolutionary dimensionality reduction algorithm [48] is employed. In the last step and for classification, two robust classification algorithms of Artificial Neural Network (feed forward) [49] and neuro fuzzy [50] classifiers are employed for better results.

Image features are based on texture, appearance, illumination and edge. Now each feature provides specific information which is different for its related application. With combining these features, it is possible to cover all aspects and features of the scene, which this paper is intended to do it.

3.1 LBP

Local Binary Pattern (LBP) [47] is a color and texture-based feature and it is very nice feature for texture analysis. This feature introduced as a 3*3 rectangle for start and has good resistance against different illumination levels. So, it is used to reduce the effect of illumination changes in the experiment.

In dealing with face analysis that each face is different with another, local and texture-based features like LBP are so useful. Obviously with adding more features to the final feature vector, learning time increases which a solution is made for this purpose to have as higher accuracy as possible along with as lowest runtime speed as possible. LBP is one of the most famous local features which is using in different illumination conditions. That is why this feature is used as all databases does not have the same illumination levels. Figure 11 shows the performance steps of LBP algorithm.

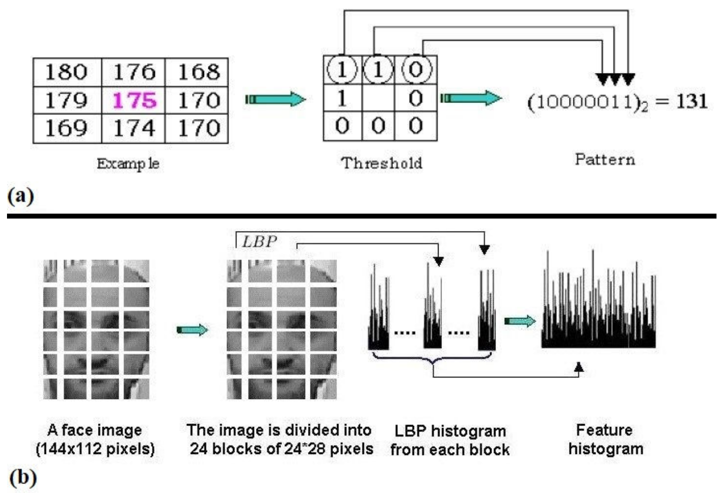


Fig 11. LBP algorithm workflow (a) and applying on a sample (b)

3.2 HOG

There is another type of features which are based on edge, place and angle of the pixels. It is possible to extract these features using image gradients. They are Histogram of Oriented Gradient (HOG) [43] features. These features are local, just like LBP. As these features are perfect to extract face wrinkles edges, it is rationale to employed it.

In edge-based features which are possible to get by gradient of the image, useful information is extracted from angles and position of the connected pixels. HOG features are in horizontal, vertical and diagonal directions. HOG features are extracting from blocks with different sizes. These blocks have two values of magnitude and direction. Magnitude determines the scale of the block and direction determines the path which that specific edge follows. Figure 12 shows HOG feature gradient magnitude and direction on a sample.

3.3 LPQ

If the image has unwanted smoothing, it is needed to use frequency domain features. As a lot of images in different databases have a lot of blurring or average filtering, it is rational to add frequency domain features to fix blurring effect in final feature vector. Local Phase Quantization (LPQ) [46] feature is used on depth images in this research as this feature is just like HOG on color images and extracts huge amount of edge features.

LPQ is a local feature in frequency domain based on Fourier transform system [51]. Blurring effect in magnitude and phase of frequency domain has different effect. Phase channel could deactivate low pass smoothing filters which might be in some of the images. This feature is perfect to use on depth images. Figure 13 represents the process of LPQ algorithm workflow.

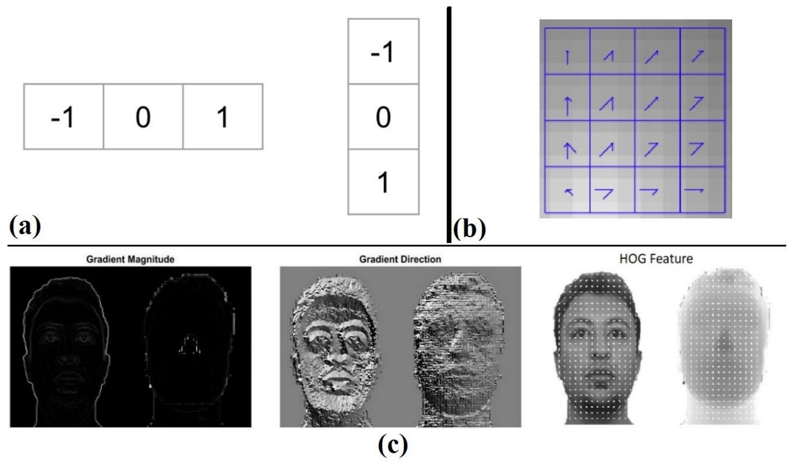


Fig 12. Gradient kernel (a), gradient directions (b) and gradient magnitude and directions on a sample (c)

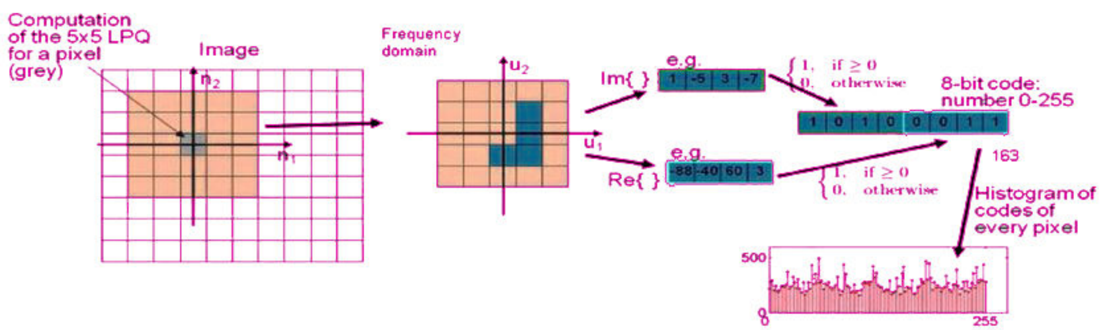


Fig 13. LPQ algorithm workflow

As it is clear in the Figure 14, LPQ method has robust performance in dealing with low pass gaussian smoothing filter. In the figure, standard deviation is (left side) and 0.5-1.5 (right side).

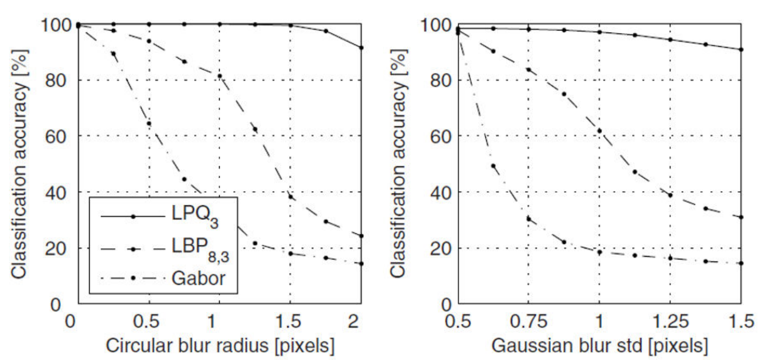


Fig 14. Blurring effect with different sigma on Gabor filter, LBP and LPQ algorithms

3.4 Gabor Filter

Gabor filters [44] are so common in face analysis applications. This feature reveal face wrinkles very well. These features are not sensitive to rotation, resizing and illumination changes. Gabor filter is based on texture just like LBP and is robust against low pass filters. Gabor filters widely used in texture analysis and edge detection. This filter is linear and local. Gabor filter convolution core is based on an exponential linear function in a gaussian one. If they be adjusted very well, they could have very precise performance. They have great response into sudden changes which makes these filters very good in face analysis. Their main advantages are in change in illumination, rotation and resizing. Below re Gabor filter parameters.

- **Sigma**

  Standard deviation which is used in gaussian function. Sigma shows the changes width in the wave form.
- **Theta**

  Wavelet direction angle. The most important parameter and determines to which features should be responded.
- **Lambda**

  The size of sine wave length.
- **Gamma**

  How much elliptical wavelet is and 1 means a circular gaussian function.
- **Psi**

  Phase changes during time.

Gabor filter complex, real and imaginary equations are as follow:

$$Complex = g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda}+\psi\right)\right) \quad (1)$$

$$Real = g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\psi\right) \quad (2)$$

$$Imaginary = g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x'}{\lambda}+\psi\right) \quad (3)$$

Where

$$x' = x\cos\theta + y\sin\theta \quad (4)$$

$$y' = x\sin\theta + y\cos\theta \quad (5)$$

Figure 15 show two images in color and depth form and in the frequency domain which a gaussian filter with sigma=3 is applied on them. As it is clear in this figure, low pass filtering in depth image and in frequency domain has weaker effect compared with color image. Amplitude and phase spectrums in frequency domain has slight change in depth image in the figure, which means using frequency domain features on depth image is a rational effort.

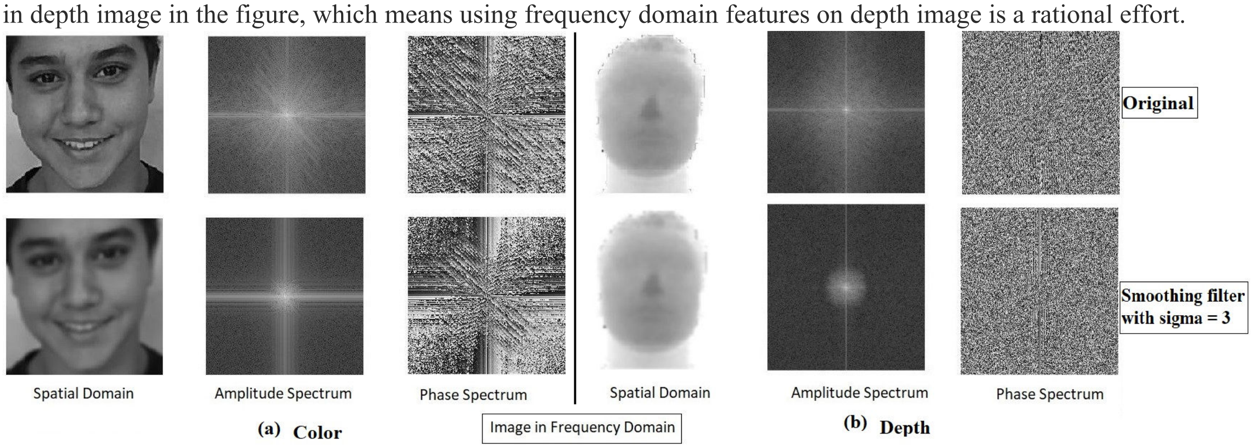


Fig 15. Blurring effect on color (a) and depth (b) images in the frequency domain with similar sigma amount.

Figure 16 shows Gabor filter in different frequency and directions in 2-Dimentional (2-D) and 3-Dimentional (3-D) forms (a), Gaussian kernel in Gabor filter with 30 Degree of sin (b) and Gabor filter with wave length of 8 and in directions of 0, 22, 45, 67, 90, 112, 135 and 180 degrees.

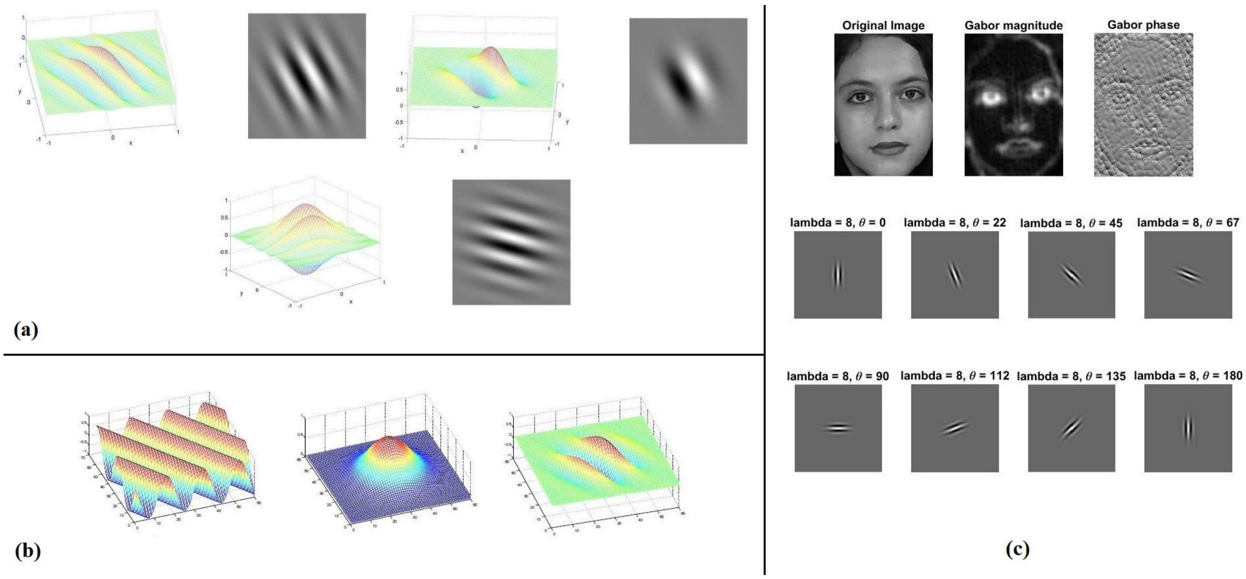


Fig 16. Gabor filter in different freqency and directions (a), Gabor filter in sin waveform in 30 degree (b), Gabor filter with wveleanght of 8 in different directions on a sample color image

3.5 SURF

Speeded Up Robust Features (SURF) [45] is a feature detector and descriptor algorithm. SURF is so fast algorithm and has great resistance against rotation. First, image integral calculates just like Harr method. Then, feature points using Hessian algorithm [52] will be found. Making scale space is the third step. Determining maximum point is next step. Finally feature vector will be make using preview step. SURF is the advanced version of SIFT [53] but it is faster multiple times. Figure 17 represents LBP, LPQ, HOG, SURF and Gabor filter features (machine understanding) on a sample in color and depth forms. Figure 18 illustrates proposed methods steps in visual form.

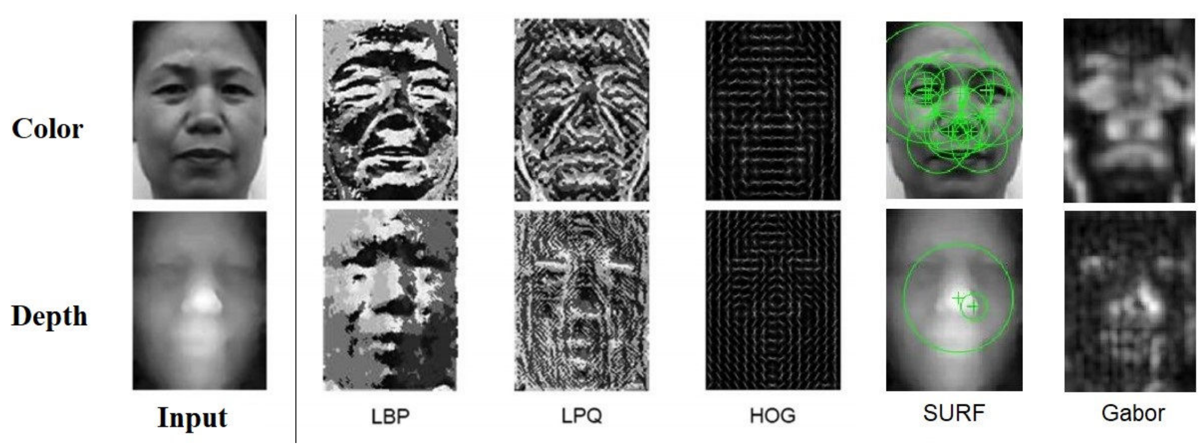


Fig 17. Appling 5 main features on a sample face image in color and depth forms, showing machine understanding of different features in spatial and frequency domains [54]

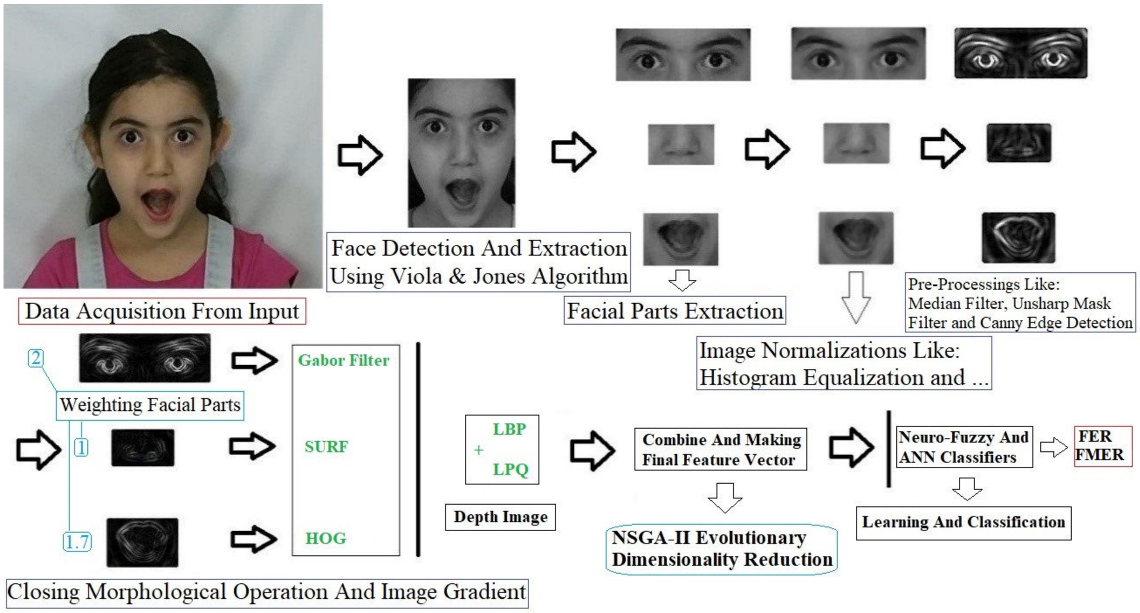


Fig 18. Proposed method workflow on a sample image

3.6 NSGA-II Evolutionary dimensionality reduction:

In FMER task, usually there are feature vectors which contain discriminative and non-discriminative data. As the task is about micro expressions, and slight details has high of importance, so selecting the best features and removing others is a very important step. In Multi Objective optimization using Evolutionary Algorithms (MOEA) [59], Non dominated Sorting Genetic Algorithm (NSGA) [55] is one of the most famous, as it is a type of genetic algorithm. But it has high complexity, lack of elitism and for choosing optimal parameters. So, it was decided to modified it and Non dominated Sorting Genetic Algorithm II (NSGA-II) [48] was made which has better sorting algorithm, having elitism and also sharing parameters not need to be chosen. This paper employed this great algorithm in feature selection step, as it using its nondominated sorting approach which provides the best final features. According to the research, it is the first usage of NSGA-II feature selection in FMER for depth images. Figure 19 shows the chart of NSGA-II implementation used in this paper. For more information about this method, it can be referred to [48].

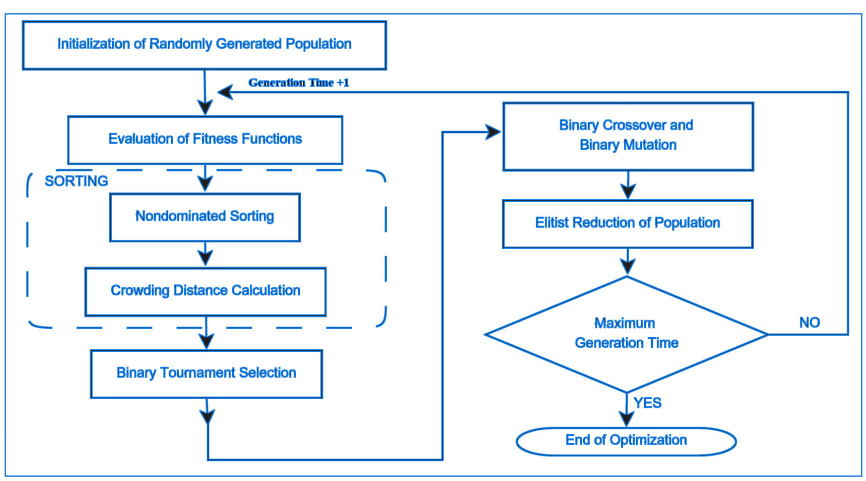


Fig19. Flow chart of NSGA-II implementation

4. VALIDATION AND RESULTS

For validating the final results, Neuro-fuzzy [50] and Artificial Neural Network classifiers [49] are employed. Also, for determining the robustness of the proposed system, most famous and new similar methods in FER and FMER tasks would be compared with propose system on same databases. Also, the selected databases for validation which together cover both FER and FMER tasks in color and depth image types are Eurecom Kinect Face DB [11], VAP RGBD-T Face [17], JAFFE [56], Face Grabber DB [20], FEEDB [19], and CASME [22]. Also, the mentioned databases validated using Conventional Neural Network (CNN) [57] which feature extraction would be done by the algorithm itself to have better understanding of machine understanding of image versus human. Windows 10 64-bit operating system along with MATLAB R 2019 b software are used for getting the final evaluations results. Also, the hardware setup for processing is as follow: Intel Core I-7 4790-K CPU 4.00 GHz, 32 GB of RAM, NVIDIA GeForce GTX 1050 2GB. Data for both classifiers are divided to 70 % for training and 30 % for testing.

Experiment's parameters for NSGA-II feature selection used in this research are listed in Table 5. Table 6 represents characteristics of all 6 face databases which are selected to be used in validation section in order to comparison purpose. Figure 20 some samples of these 6 databases in different expressions.

Table 5

Simulation parameters for feature selection

NSGA-II PARAMETERS	***PARAMETER VALUE***
POPULATION SIZE	300
NUMBER OF ITERATIONS	400
CROSSOVER PROBABILITY	0.8
MUTATION PROBABILITY	1/d (d = 3403)
CROSSOVER METHOD	Binary crossover
MUTATION METHOD	Binary mutation
SELECTION METHOD	Tournament selection
POOL SIZE	200
POPULATION SIZE	2

Table 6

Selected databases characteristics for validation

DATABASE	**SAMPLES**	**SENSOR**	**USAGE**	**DATA TYPE**	**DIMENSIONS**	**EXPRESSIONS**	**YEAR**
Eurecom Kinect Face	14 male and 38 females	Kinect V.1	FER – face recognition	1248 Color + depth images	RGB= 256*256 Depth= 256*256	3 expressions of neutral, joy and surprise	2014
VAP RGB-D-T Face	51	Kinect V.1 And AXIS Q1922	FER – face recognition	46360 color + depth + thermal images	RGB= 480*640 Depth= 480*640 Thermal= 288*384	4 expressions of neutral, joy, anger and surprise	2014
FEEDB	50	Kinect V.1	FER - FMER – face recognition	30 color and depth videos	RGB= 480*640 Depth= 480*640	33 facial expressions	2013
Face Grabber	33 male and 7female	Kinect V.2	FER - FMER – face recognition	**67159** color + depth images	RGB= 2080*1920 Depth= 424*512	7 main expressions	2016
CASME	19	BenQ M31 GRAS-03K2C	FMER	195 video files in 60 fps	640*480 1280*780	7 main expressions	2013
JAFFE	10	-	FER – face recognition	212 gray images	256*256	7 main expressions	1998

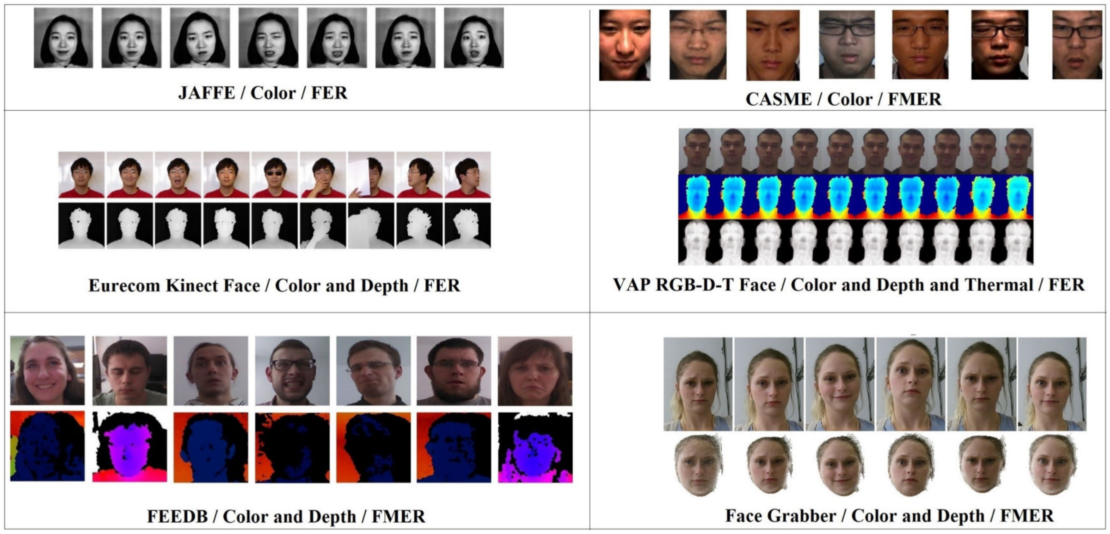


Fig 20. Samples from selected databases in different expressions

4.1 Neuro-fuzzy classifier:

Fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1. It is employed to handle the concept of partial truth, where the truth value may range between completely true and completely false [58]. Fuzzy sets are often defined as triangle or trapezoid-shaped curves, as each value will have a slope where the value is increasing, a peak where the value is equal to 1 (which can have a length of 0 or greater) and a slope where the value is decreasing.

An adaptive Neuro-fuzzy Inference System or Adaptive Network-based fuzzy Inference System (ANFIS) is a kind of artificial neural network that is based on Takagi–Sugeno fuzzy inference system [50, 60]. The technique was developed in the early 1990s. Since it integrates both neural networks and fuzzy logic principles, it has potential to capture the benefits of both in a single framework. Its inference system corresponds to a set of fuzzy IF–THEN rules that have learning capability to approximate nonlinear functions. Hence, ANFIS [60] is considered to be a universal estimator. Figure 21 represents neuro-fuzzy system important parts.

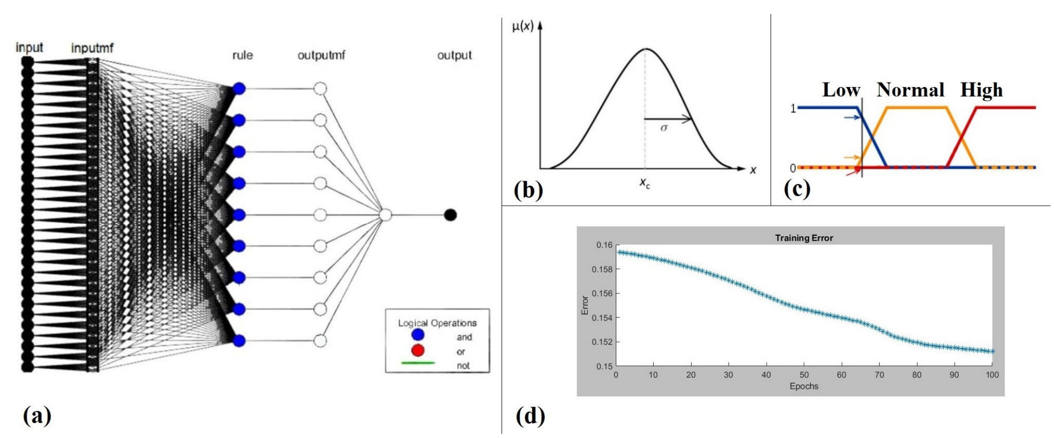


Fig 21. Neuro fuzzy network structure (a), Gaussian membership function (b), Training fis using hybrid learning in 100 epochs (c), Fuzzy model and linguistics variables and sets

4.2 ANN classifier:

Artificial Neural Networks (ANN) [49] are very good at solving pattern recognition problems. A neural network with enough neurons can classify any data with optional accuracy. They are suitable for complex decision boundary problems having many variables. Multilayer networks fix the classification task for non-linear sets, using hidden layers, that neurons are not directly connected to the final element. Also, multilayer neural networks mostly employ the log-sigmoid transition function. Here conjugate gradient back propagation algorithm [61] is used for training process along with 50 hidden layers. Train and test data are dividing to 70% and 30 % respectively.

Table 7 presents acquired results on all databases using proposed method for FER and FMER tasks by expressions, micro expressions and total. Table 8 shows comparison table for other similar methods versus proposed method on same databases for FER and FMER tasks. The last column is runtime speed for proposed method in second. For more information about each method, Table 4 would give enough details. Figure 22 illustrates Table 7 values (except total cells) in graphical form

Table 7

Acquired results on all databases using proposed method for FER and FMER

DATABASE	USAGE	JOY	SADNESS	ANGER	SURPRISE	DISGUST	FEAR	NEUTRAL	TOTAL
EURECOM	FER	98.13 %	-	-	96.77 %	-	-	94.23 %	96.37 %
VAP RGBD-T	FER	97.67 %	-	93.90 %	98.81 %	-	-	97.64 %	97.00 %
FEEDB	FMER	81.35 %	79.63 %	88.93 %	91.65 %	71.25 %	79.34 %	80.02 %	81.73 %
FACE GRABBER	FMER	91.57 %	90.69 %	88.97 %	97.72 %	93.16 %	96.39 %	92.24 %	92.96 %
JAFFE	FER	100.00 %	97.91 %	97.98 %	100.00 %	98.11 %	99.92 %	99.97 %	99.12 %
CASME	FMER	85.61 %	78.72 %	75.19 %	90.28 %	86.37 %	87.72 %	86.67 %	84.36 %

Table 8

Comparison results

DATABASE	USAGE	METHOD	CNN	PROPOSED	RUNTIME
EURECOM	FER	[28] = 87.90 %	97.12 %	96.37 %	0.72 sec
VAP RGBD-T	FER	[30] = 95.70 %	96.27 %	97.00 %	0.61 sec
FEEDB	FMER	[-] = -	87.90 %	81.73 %	0.87 sec
	FER	[31] = 50.00 %	-	-	-
FACE	FMER	[] = -	91.17 %	92.96 %	0.62 sec
GRABBER	FER	[32] = 83.00 %	-	-	-
JAFFE	FER	[26] = 94.88 %	99.01 %	99.12 %	0.67 sec
CASME	FMER	[39] = 71.19 %	88.97 %	84.36 %	0.79 sec

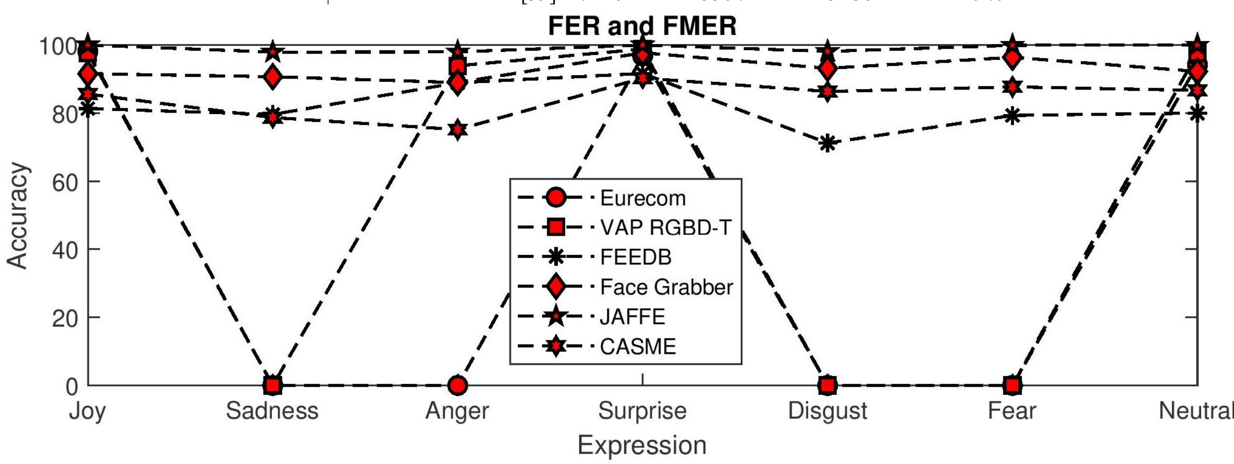


Fig 22. Proposed method results on all databases

- Result Discussion

In Table 7, the blank cells mean that database does not contain that specific expressions. Also, in Table 8 and for FEEDB and Face grabber databases, comparison methods are for FER purposes. But as FMER recognition is higher level of recognition, so just FMER acquired results are used in cells. It has to be mentioned that comparison methods did not determined any FMER results for FEEDB and Face grabber databases in their papers. As it is clear in Table 8, proposed method has better results versus other methods (except CNN) which shows the robustness of the system. Also, a sidelong experiment for all databases using CNN shows that CNN is not better than proposed method in all databases, but in most of them had better performance. Proposed method shows better performance on VAP RGB-D-T and Face Grabber databases versus CNN.

5. CONCLUSION, SUGGESTIONS AND FUTURE WORKS

Having combining of spatial and frequency domain features from color and depth images, it is possible to recognize micro facial expressions with high precision. Also, using evolutionary algorithms in order to select most reliable features and removing outliers is a smart action in such highly feature based systems. Having perfect compatibility with pure darkness increased the application of the system and pushed the limits. Validation section emphasizes on promising results of the proposed system versus most of the other methods, except CNN. It is suggested to use other infrared sensor like Kinect which supports more than 5-meter cover and more than 30 fps recording capability to achieve more robust results. Also, using evolutionary segmentation could be useful in pre-processing step. Applying proposed system on more and new FMER depth-based databases, having more than 7 facial expressions and recognizing faces with bear and glasses make the future works.

- **References**

[1] Gonzalez, Rafael C., Richard Eugene Woods, and Steven L. Eddins. Digital image processing using MATLAB. Pearson Education India, 2004.

[2] Plataniotis, Konstantinos N., and Anastasios N. Venetsanopoulos. Color image processing and applications. Springer Science & Business Media, 2013.

[3] Mousavi, Seyed Muhammad Hossein, and Narges Aghsaghloo. "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, نمجنا, نارهت تاقیقحت مولع دحاو یملاسادازآ هاگشناد-ناریا یزاف یاه متسیس, https://www. civilica. com/Paper-IDS03-IDS03_107. html. 2018.

[4] Mousavi, Seyed Muhammad Hossein, and VB Surya Prasath. "On the Feasibility of Estimating Fruits Weights Using Depth Sensors.", 4 th International Congress of Developing Agriculture, Natural Resources, Environment and Tourism of Iran (2019), Tabriz, Iran.

[5] Kumari, Jyoti, R. Rajesh, and K. M. Pooja. "Facial expression recognition: A survey." Procedia Computer Science 58.1 (2015): 486-491.

[6] Wang, Su-Jing, et al. "Micro-expression recognition using dynamic textures on tensor independent color space." 2014 22nd international conference on pattern recognition. IEEE, 2014.

[7] Min, Rui, Neslihan Kose, and Jean-Luc Dugelay. "Kinectfacedb: A kinect database for face recognition." IEEE Transactions on Systems, Man, and Cybernetics: Systems 44.11 (2014): 1534-1548.

[8] Wasenmüller, Oliver, and Didier Stricker. "Comparison of kinect v1 and v2 depth images in terms of accuracy and precision." Asian Conference on Computer Vision. Springer, Cham, 2016.

[9] Mousavi, Seyed Muhammad Hossein, Vyacheslav Lyashenko, and Surya Prasath. "Analysis of a robust edge detection system in different color spaces using color and depth images." Компьютерная оптика 43.4 (2019).

[10] Abate, Andrea F., et al. "2D and 3D face recognition: A survey." Pattern recognition letters 28.14 (2007): 1885-1906.

[11] Min, Rui, Neslihan Kose, and Jean-Luc Dugelay. "Kinectfacedb: A kinect database for face recognition." IEEE Transactions on Systems, Man, and Cybernetics: Systems 44.11 (2014): 1534-1548.

[12] Yin, Lijun, et al. "A 3D facial expression database for facial behavior research." 7th international conference on automatic face and gesture recognition (FGR06). IEEE, 2006.

[13] Hamm, Jihun, et al. "Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders." Journal of neuroscience methods 200.2 (2011): 237-256.

[14] Mousavi, Seyed Muhammad Hossein. "A New Way to Age Estimation for RGB-D Images, based on a New Face Detection and Extraction Method for Depth Images." International Journal of Image, Graphics and Signal Processing 10.11 (2018): 10.

[15] Grigorios, Kalliatakis. "Towards an automatic intelligible monitoring of behavioral and physiological metrics of user experience: head pose estimation and facial expression recognition." Master of Science, Department of Applied Informatics and Multimedia, School of Applied Technology, Technological Educational Institute of Creece (2015).

[16] Hg, R. I., et al. "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, 2012.

[17] Nikisins, Olegs, et al. "RGB-DT based face recognition." Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014.

[18] Li, Billy YL, et al. "Using kinect for face recognition under varying poses, expressions, illumination and disguise." Applications of Computer Vision (WACV), 2013 IEEE Workshop on. IEEE, 2013.

[19] Szwoch, Mariusz. "FEEDB: a multimodal database of facial expressions and emotions." Human System Interaction (HSI), 2013 The 6th International Conference on. IEEE, 2013.

[20] Merget, Daniel, et al. "Capturing facial videos with Kinect 2.0: A multithreaded open source tool and database." Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016.

[21] Li, Xiaobai, et al. "A spontaneous micro-expression database: Inducement, collection and baseline." Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013.

[22] Yan, Wen-Jing, et al. "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, 2013.

[23] Polikovsky, Senya, Yoshinari Kameda, and Yuichi Ohta. "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor." (2009): 16-16.

[24] Shreve, Matthew, et al. "Macro-and micro-expression spotting in long videos using spatio-temporal strain." Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on. IEEE, 2011.

[25] Warren, Gemma, Elizabeth Schertler, and Peter Bull. "Detecting deception from emotional and unemotional cues." Journal of Nonverbal Behavior 33.1 (2009): 59-69.

[26] Chao, Wei-Lun, Jian-Jiun Ding, and Jun-Zuo Liu. "Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection." Signal Processing 117 (2015): 1-10.

[27] Elgarrai, Zineb, et al. "Robust facial expression recognition system based on hidden Markov models." International Journal of Multimedia Information Retrieval 5.4 (2016): 229-236.

[28] Ijjina, Earnest Paul, and C. Krishna Mohan. "Facial expression recognition using kinect depth sensor and convolutional neural networks." Machine Learning and Applications (ICMLA), 2014 13th International Conference on. IEEE, 2014.

[29] Hg, R. I., et al. "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, 2012.

[30] Nikisins, Olegs, et al. "RGB-DT based face recognition." Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014.

[31] Szwoch, Mariusz, and Paweł Pieniążek. "Facial emotion recognition using depth data." Human System Interactions (HSI), 2015 8th International Conference on. IEEE, 2015.

[32] Yuan, Sen, and Xia Mao. "Exponential Elastic Preserving Projections for Facial Expression Recognition." Neurocomputing (2017).

[33] Polikovsky, Senya, Yoshinari Kameda, and Yuichi Ohta. "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor." (2009): 16-16.

[34] Pfister, Tomas, et al. "Recognising spontaneous facial micro-expressions." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

[35] Wu, Qi, Xunbing Shen, and Xiaolan Fu. "The machine knows what you are hiding: an automatic micro-expression recognition system." Affective Computing and Intelligent Interaction (2011): 152-162.

[36] Wang, Su-Jing, et al. "Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine." Neural processing letters 39.1 (2014): 25-43.

[37] Huang, Xiaohua, et al. "Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015.

[38] Huang, Xiaohua, et al. "Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns." Neurocomputing 175 (2016): 564-578.

[39] Zheng, Hao. "Micro-Expression Recognition based on 2D Gabor Filter and Sparse Representation." Journal of Physics: Conference Series. Vol. 787. No. 1. IOP Publishing, 2017.

[40] Lundqvist, Daniel, Anders Flykt, and Arne Öhman. "The Karolinska directed emotional faces (KDEF)." CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet 1998 (1998).

[41] Viola, Paul, and Michael Jones. "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, 2001.

[42] Iqbal, Bilal, et al. "Canny edge detection and Hough transform for high resolution video streams using Hadoop and Spark." Cluster Computing 23.1 (2020): 397-408.

[43] Manikonda, Santhosh KG, and Dattatraya N. Gaonkar. "Islanding detection method based on image classification technique using histogram of oriented gradient features." IET Generation, Transmission & Distribution (2020).

[44] Tadic, Vladimir, et al. "Comparison of Gabor Filter Bank and Fuzzified Gabor Filter for License Plate Detection." Acta Polytechnica Hungarica 17.1 (2020): 1-21.

[45] Baig, Fahad, et al. "Boosting the performance of the BoVW model using SURF–CoHOG-based sparse features with relevance feedback for CBIR." Iranian Journal of Science and Technology, Transactions of Electrical Engineering 44.1 (2020): 99-118.

[46] Freitas, Pedro Garcia, et al. "Image quality assessment using BSIF, CLBP, LCP, and LPQ operators." Theoretical Computer Science 805 (2020): 37-61.
[47] Khan, Kashif Ahmad, et al. "A hybrid Local Binary Pattern and wavelets based approach for EEG classification for diagnosing epilepsy." Expert Systems with Applications 140 (2020): 112895.
[48] Wang, Penghong, et al. "A Gaussian error correction multi-objective positioning model with NSGA-II." Concurrency and Computation: Practice and Experience 32.5 (2020): e5464.
[49] Frolov, Nikita, et al. "Feed-forward artificial neural network provides data-driven inference of functional connectivity." Chaos: An Interdisciplinary Journal of Nonlinear Science 29.9 (2019): 091101.
[50] Jafarifarmand, Aysa, and Mohammad Ali Badamchizadeh. "Real-time multiclass motor imagery brain-computer interface by modified common spatial patterns and adaptive neuro-fuzzy classifier." Biomedical Signal Processing and Control 57 (2020): 101749.
[51] Fathy, Alaa, et al. "On-chip parallel Fourier transform spectrometer for broadband selective infrared spectral sensing." Microsystems & Nanoengineering 6.1 (2020): 1-9.
[52] Zhang, Yangjing, et al. "An efficient Hessian based algorithm for solving large-scale sparse group Lasso problems." Mathematical Programming 179.1 (2020): 223-263.
[53] Gardezi, Akber, et al. "Enhanced target recognition employing spatial correlation filters and affine scale invariant feature transform." Pattern recognition and tracking XXX. Vol. 10995. International Society for Optics and Photonics, 2019.
[54] Boutellaa, Elhocine, et al. "On the use of Kinect depth data for identity, gender and ethnicity classification from facial images." Pattern Recognition Letters 68 (2015): 270-277.
[55] Deb, Kalyanmoy, et al. "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II." International conference on parallel problem solving from nature. Springer, Berlin, Heidelberg, 2000.
[56] Lyons, Michael, et al. "Coding facial expressions with gabor wavelets." Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. IEEE, 1998.
[57] Neelapu, Ramesh, Golagani Lavanya Devi, and Kurapati Srinivasa Rao. "Deep learning based conventional neural network architecture for medical image classification." Traitement du Signal 35.2 (2018): 169-182.
[58] Mousavi, Seyed Muhammad Hossein, et al. "A PSO fuzzy-expert system: As an assistant for specifying the acceptance by NOET measures, at PH. D level." 2017 Artificial Intelligence and Signal Processing Conference (AISP). IEEE, 2017.
[59] Mousavi, Seyed Muhammad Hossein, Vincent Charles, and Tatiana Gherman. "An evolutionary Pentagon Support Vector finder method." Expert Systems with Applications 150 (2020): 113284.
[60] Jang, J-SR. "ANFIS: adaptive-network-based fuzzy inference system." IEEE transactions on systems, man, and cybernetics 23.3 (1993): 665-685.
[61] Heravi, Ahmad Reza, and Ghosheh Abed Hodtani. "A new correntropy-based conjugate gradient backpropagation algorithm for improving training in neural networks." IEEE transactions on neural networks and learning systems 29.12 (2018): 6252-6263.
[62] Mousavi, Seyed Muhammad Hossein, and Vyacheslav Lyashenko. "Extracting old persian cuneiform font out of noisy images (handwritten or inscription)." 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP). IEEE, 2017.