9th Iranian Joint Congress on Fuzzy and Intelligent Systems
March 2-4, 2022, Bam, Kerman, Iran

Seven Staged Identity Recognition System Using Kinect V.2 Sensor

1st Seyed Muhammad Hossein Mousavi [1], 2nd Atiye Ilanloo [2]

[1] Independent Scientist, Tehran, Iran
[2] Faculty of Humanities- Psychology, Islamic Azad University-Rasht, Gilan, Iran

Abstract— **By employing artificial intelligence techniques and algorithms such as color and depth image processing, signal processing, machine learning, evolutionary algorithms and fuzzy systems, an identity recognition expert system with approximate recognition accuracy of 99% is proposed. Available identity recognition systems mostly are in three stages which may lead to some security problems, so it is decided to make a robust system. Proposed system uses Kinect Version 2 sensor in order to conduct 7 main stages of recognition. The system includes following stages of recognition and estimation which are, face and voice recognition, finger print recognition, iris recognition, gesture recognition, sex detection and age estimation. By adding macro lens to the sensor, recognition accuracy for fingerprint and iris increases significantly. All efforts on this project were to achieve the highest potential out of available techniques. The system is learning based and has high precision and could be well used in industrial purposes. By installing macro lens on Kinect sensor, the system could compete with other expensive identification systems. It has to be mention that proposed system works well in the pure darkness ass Kinect sensor supports the infrared spectrum.**

Keywords- Expert system; Security; Kinect sensor; Identity recognition; Image and signal processing; Macro lens

I. INTRODUCTION

As technology enhances, possibility of using full potential of it gets higher. A lot of organizations and even personal sites need to have a proper security system which mostly uses image processing devices like cameras. Also, most of the use just color sensor, not depth. But it could be developed to a higher degree when a security breach could occur. Obviously using more security technologies and devices helps to the rising problem. As it mentioned, available identity recognition systems are mostly in three stages. It is mostly enough, but making more robust system is always sensible. Proposed system uses Kinect sensor version 2 [1] and high level of artificial intelligence algorithms and techniques such as image and signal processing [2], data mining [3], machine learning [4], fuzzy sets [5] and evolutionary algorithms [6] in order to make the final results as better as possible. System performs voice recognition [7], face recognition [8], fingerprint recognition [9], iris recognition [10], gesture recognition [11], sex detection and age estimation [12] tasks. Also, a macro lens is installed on the color sensor in order to increase the receiving details. In the other hand depth sensor of the Kinect, overcomes the different lighting environment for the final use.

II. PRIOR RELATED RESEARCHES

Akihiro Machida in 2006, succeeded to invent a touchless fingerprint recognition system which was assisted by a remote control under the patent number of US7116805B2 [13].

In the field of sex detection and age estimation Yamazaki, Kazuhiro, et al, could invent a system which employed spatial and frequency domains features to perform the task. This patent is registered under the US10204266B2 serial in 2011 [14].

Another age estimation method which uses Kinect sensor is belong to SMH, Mousavi in 2018. System is real time based and so fast which works by just a single image even in pure darkness [12].

Anter Abozaid and et al, made a system for face and voice recognition purposes in 2018. They used eigenface and Principal Component Analysis (PCA) features and classified their systems by Gaussian Mixture Model (GMM) [15], Artificial Neural Network (ANN), and Support Vector Machine (SVM) [8].

Another research belong to face recognition is for Mousavi, Seyed Muhammad Hossein, and S. Younes Mirinezhad in 2021, which they used Convolutional Neural Networks [16] to train their system [17].

Peng, Chang, et al, made an ultrasonic fingerprint recognition system in 2021 for the first time. Basically, they used signal processing instead of image [9].

For iris recognition task, Muthazhagan, Balaji, and Suriya Sundaramoorthy employed a deep learning algorithm called CNN on their systems, achieving high recognition accuracy [10].

Fenglin Liu, et al succeed to make a gesture recognition system using Kinect sensor and RGB network with high accuracy in 2019 [11].

III. Proposed Identity Recognition Method

A. Advantages

The system has 7 stages of estimation, detection and recognition by color and depth data using Kinect V.2 Sensor. As the system employs infrared sensor, some stages could perform in pure darkness. It is an expert system, means that it could be replaced for human and never gets tired or error. It is cheaper than its similar types as it is using cheap sensor. Fingerprint recognition is touchless and does not need to professional tool for keeping it. Also, macro lens increases the final recognition accuracy for iris and fingerprint, significantly. As depth data is known as 2.5-Dimensional (2.5-D), it could be easily converted to 3-Dimensional (3-D) model, which is an advantage. If one of the steps does not work with enough accuracy, system get into halt situation. Also, it has ease of use for final user and has pretty convenient speed.

B. Usage

This expert system could be used in security and criminology systems, law schools, defense industry research, museum preservation, museum security of banknotes and ingots in banks, storage of the latest available technologies that are not available to the public, storage of the above documents Confidential, secret and top secret that should not be disclosed, where the latest achievements of medical science are kept, such as deadly viruses that should not be taken out of the building, and so on. In general, in any place where the building configuration is solid, has valuable objects or documents, and the number of visits and people visiting is low.

C. Steps

System starts with acquiring color and depth data from Kinect sensor and system announces “Please stand in front of the sensor (1 meter)”. Second step is belonged to gesture, voice, face recognitions, sex detection and age estimation. In gesture recognition step, body will be extracted from depth data and Gabor features [17, 21] will be extracted in the frequency domain [18, 21] from the data. Voice recognition is consisted of noise removal and signal normalization by median filter followed by extracting Mel-Frequency Cepstral Coefficients (MFCCs) features [2] in the frequency domain. Face recognition however has some preprocessing steps of background removal and face extraction by viola and jones algorithm [19] in color and depth data followed by applying multi directional edge detection (canny) [20] and closing morphological operation plus edge sharpening by unsharp masking [20]. Then the data sends for CNN for final classification. Age estimation and sex detection steps have the same preprocessing stage of face recognition step, but for age estimation Local Phase Quantization LPQ [21] features are extracted in frequency domain from color and depth data and normalizes in the range of 0 to 1. Also, for sex detection step, Speeded Up Robust Features SURF [21] are extracted from color data in spatial domain. Now, all extracted features from all steps goes for dimensionality reduction step conducted by Lasso regularization algorithm [21] for removing outliers and faster classification speed. Here in classification step final data of gestures and sex goes for classification by Deferential Evolution-ANFIS [22, 24] classification algorithm and other steps simply use SVM. Now it is time for fingerprint and iris recognition. System announces “Please hold your finger in 3–4-centimeter distance from sensor” and data acquisition in color and depth starts. For fingerprint recognition, CNN is used and for iris recognition another announcement “Please hold your right eye in 3–4-centimeter distance from the sensor”, system says. Afterwards, Histogram of Oriented Gradient (HOG) [17, 21] features got extracted from color data followed by shallow neural network classification algorithm. Finally expert system has to check if all stages done successfully and matches the database or not. If yes, identity recognition completed and electronic lock gets unlocked followed by halting the system. If not, system returns to stage 1 for reidentification. Figure 1 represents experiment environment and setup. Figure 2 shows the flowchart of proposed method.

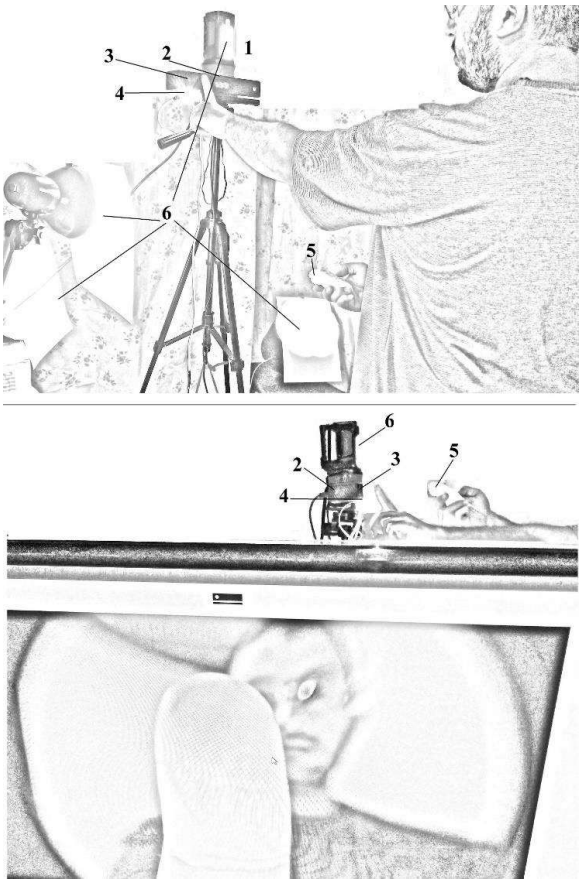

Figure 1. Experiment enviroment

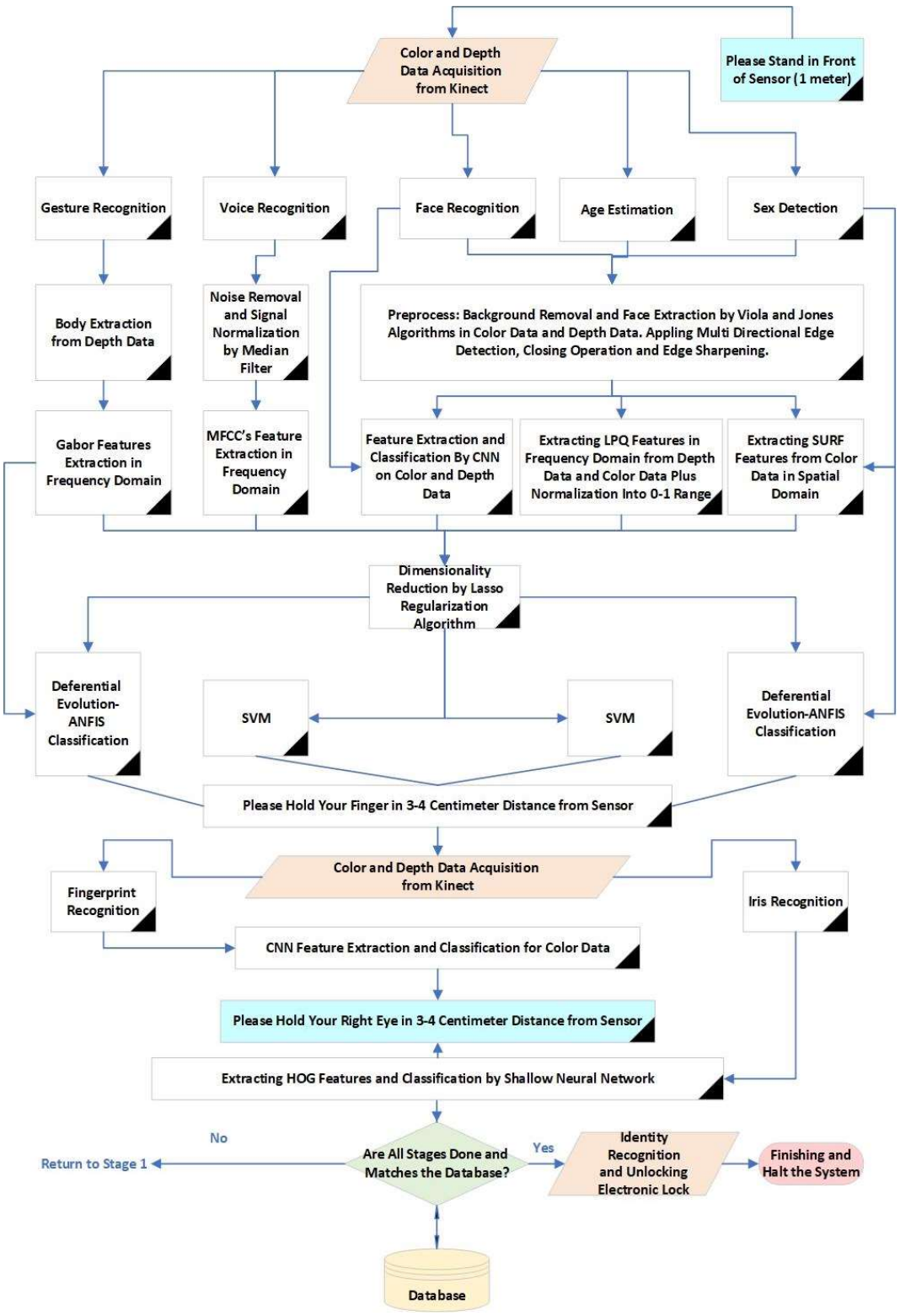

Figure 2. Proposed method flowchart

Figure 3 shows equipment used in the experiment. Also, Figuree 4 represents some steps of fingerprint and iris recognition steps. Figures 5 and 6 present different steps of the process (part 1 and 2).

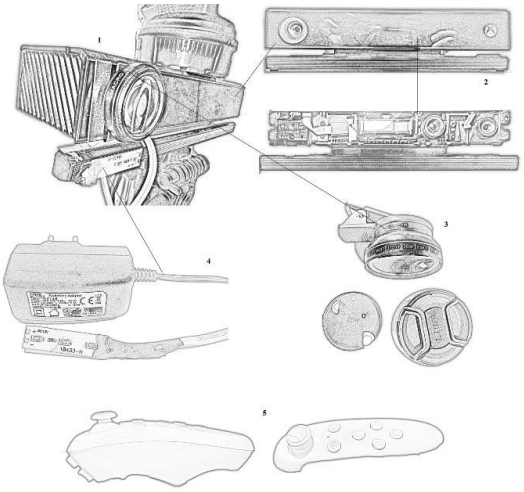

Figure 3. Experiment's Equepments

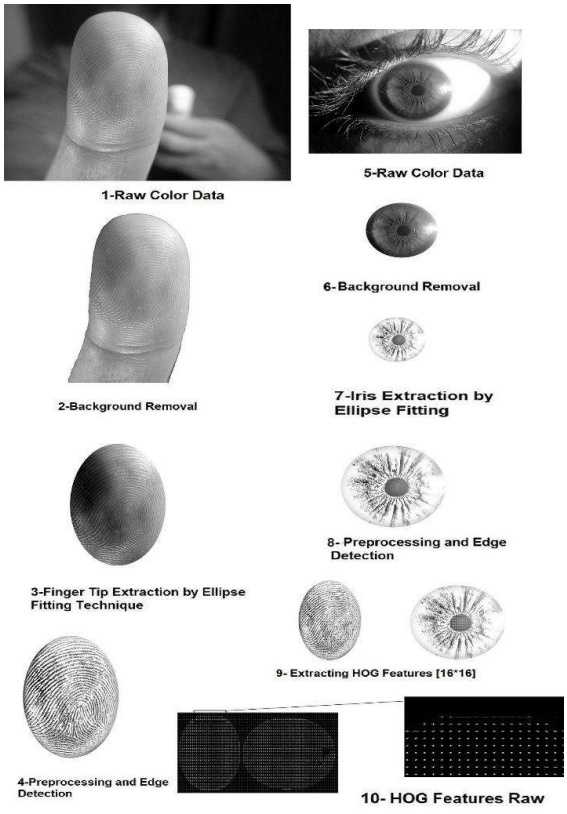

Figure 4. Some steps of fingerprint and iris recognition

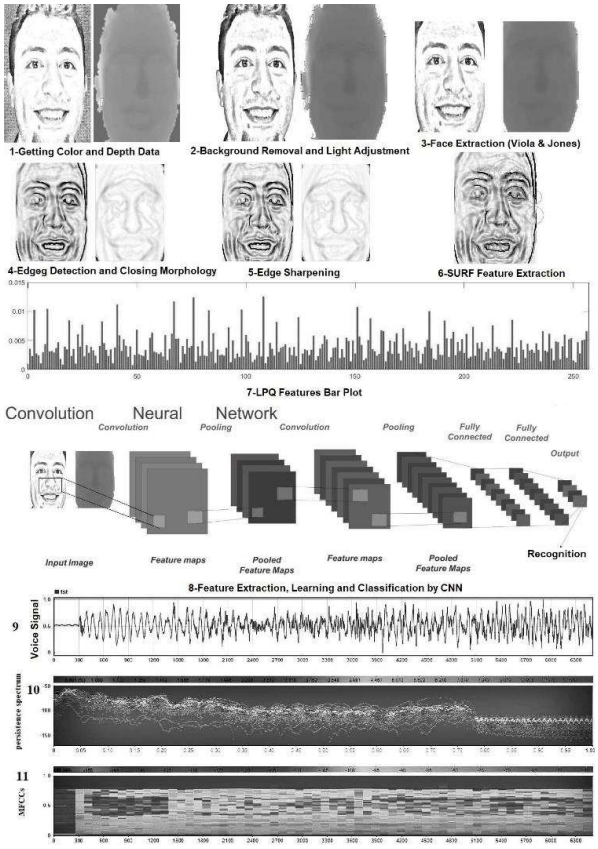

Figure 5. Different steps of the process (part 1)

IV. Validation and Results

Experiment conducted on 6 subjects which their color and depth data of face, gesture, fingerprint, iris plus voice data is collected as a database for the system. Each subject has unique 5 second command for unlocking in voice part. Classification accuracy for CNN, SVM and DE-ANFIS for each subject and in whole is calculated and visible in Table I. Also, accuracy, precision and recall [23, 25] values are calculated for more details on the system's performance and is presented in the Table II. Figure 7 represents accuracy, precision and recall structure. Table III, IV and V are presenting confusion matrix for SVM, CNN and DE-ANFIS classification results, respectively

TABLE I. Classification Results

	SVM	CNN	DE-ANFIS
Subject 1	98.12 %	99.60 %	98.03 %
Subject 2	98.99 %	99.09 %	97.52 %
Subject 3	98.14 %	99.73 %	97.47 %
Subject 4	97.98 %	99.16 %	97.63 %
Subject 5	99.11 %	99.98 %	98.19 %
Subject 6	98.68 %	98.90 %	98.63 %
Whole	**98.50 %**	**99.41 %**	**97.91 %**

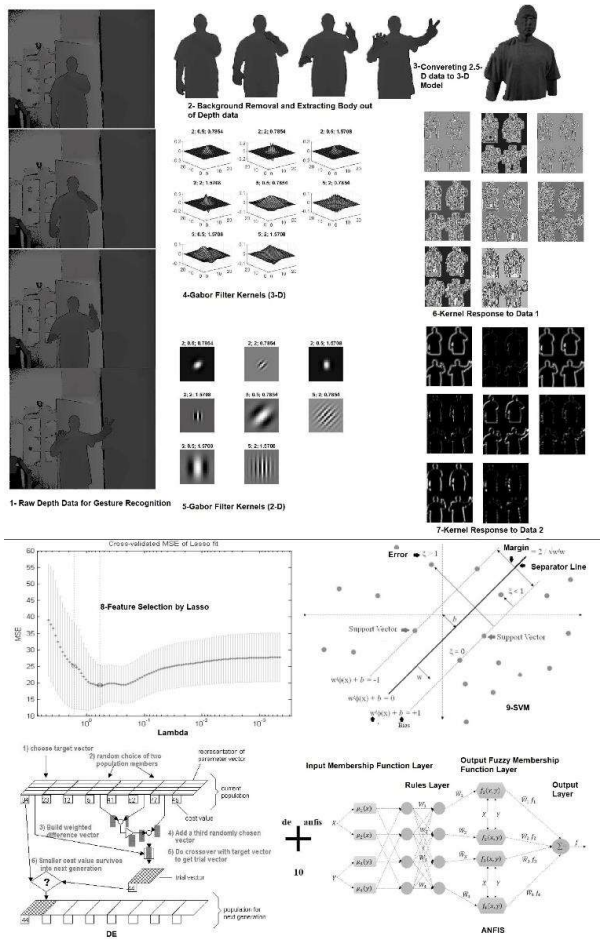

Figure 6. Different steps of the process (part 2)

$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \text{ or } \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \text{ or } \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive + True Negative}}{\text{Total}}$$

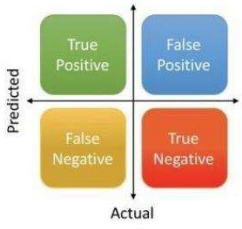

Figure 7. Accuracy, precision and recall

TABLE II. ACCURACY, PRECISION AND RECALL

Threshold	Accuracy	Precision	Recall
30	**99.12 %**	**99.19 %**	**98.73 %**
50	99.01 %	99.07 %	98.52 %
80	98.94 %	99.00 %	98.38 %

TABLE III. SVM CONFUSION MATRIX

	S 1	S 2	S 3	S 4	S 5	S 6
S 1	98 %	-	1	-	1	-
S 2	-	98 %	2	-	-	-
S 3	1	1	98%	-	-	-
S 4	-	-	1	97 %	2	-
S 5	1	-	-	-	99 %	-
S 6	-	-	-	-	2	98 %

TABLE IV. CNN CONFUSION MATRIX

	S 1	S 2	S 3	S 4	S 5	S 6
S 1	99 %	-	1	-	-	-
S 2	1	99 %	-	-	-	-
S 3	-	-	99 %	1	-	-
S 4	-	-	-	99 %	-	1
S 5	-	1	-	-	99 %	-
S 6	1	-	-	-	-	98 %

TABLE V. DE-ANFIS CONFUSION MATRIX

	S 1	S 2	S 3	S 4	S 5	S 6
S 1	98 %	-	-	-	-	2
S 2	-	97 %	1	1	-	1
S 3	1	1	97 %	1	-	-
S 4	-	-	-	97 %	3	-
S 5	-	-	-	-	98 %	2
S 6	-	-	2	-	-	98 %

As it is clear in Table I, all three classification algorithms achieved pretty acceptable recognition accuracy by the system. However, best result belongs to CNN, weakest to DE-ANFIS and SVM in the middle. Figure 7 shows accuracy, precision and recall's structure. Figure 8 presents the classification result for classifiers and Figure 9 represents accuracy, precision and recall results in graphical form.

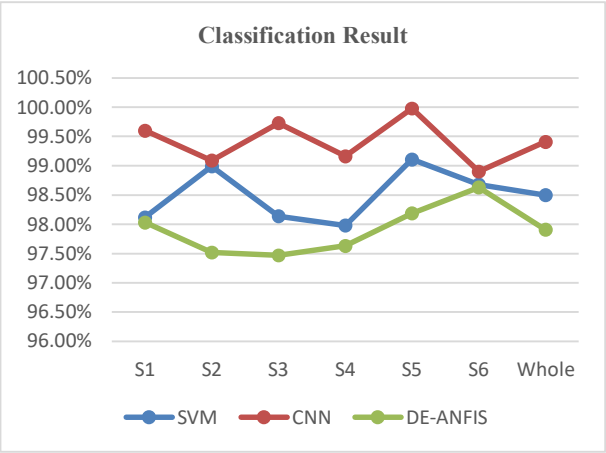

Figure 8. Classification Bar Plot

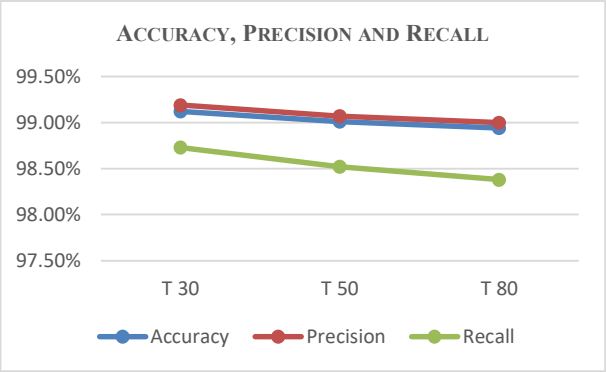

Figure 9. Accuracy, Precition and Recall Bar Plot

V. CONCLUSION AND SUGGESTIONS

Getting as much as possible potential that artificial intelligence offers in this expert system, made one of the most precise identity recognition systems for security systems as cheap as possible. Combining different algorithms in variety of tasks plus employing a suitable sensor brought the performance into a level higher. Overall performance of the system is very desirable and enhanced the era of learning-based security and surveillance systems. It is suggested to employ the system with other classification algorithms such as K-Nearest Neighborhood (K-NN), Tree and ensemble. Also, it is suggested to increase the number of subjects in the database. Extracting Scale-invariant feature transform (SIFT) and Local Binary Patterns (LBP) features and adding 3-D face recognition capability is of future works.

REFERENCES

[1] 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).

[2] Mousavi, Seyed Muhammad Hossein, VB Surya Prasath, and Seyed Muhammad Hassan Mousavi. "Persian classical music instrument recognition (PCMIR) using a novel Persian music database." 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE). IEEE, 2019.

[3] Dezfoulian, Mir Hossein, et al. "Optimization of the Ho-Kashyap classification algorithm using appropriate learning samples." 2016 Eighth International Conference on Information and Knowledge Technology (IKT). IEEE, 2016.

[4] Mousavi, Seyed Muhammad Hossein, S. Younes MiriNezhad, and Atiye Mirmoini. "A new support vector finder method, based on triangular calculations and K-means clustering." 2017 9th International Conference on Information and Knowledge Technology (IKT). IEEE, 2017.

[5] 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.

[6] Mousavi, Seyed Muhammad Hossein, Vincent Charles, and Tatiana Gherman. "An Evolutionary Pentagon Support Vector Finder Method." Expert Systems with Applications 150 (2020): 113284.

[7] Hu, Hao-Chun, et al. "Deep Learning Application for Vocal Fold Disease Prediction Through Voice Recognition: Preliminary Development Study." Journal of Medical Internet Research 23.6 (2021): e25247.

[8] Abozaid, Anter, et al. "Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion." Multimedia Tools and Applications 78.12 (2019): 16345-16361.

[9] Peng, Chang, Mengyue Chen, and Xiaoning Jiang. "Under-display ultrasonic fingerprint recognition with finger vessel imaging." IEEE Sensors Journal 21.6 (2021): 7412-7419.

[10] Omran, Maryim, and Ebtesam N. AlShemmary. "An iris recognition system using deep convolutional neural network." Journal of Physics: Conference Series. Vol. 1530. No. 1. IOP Publishing, 2020.

[11] Liu, Fenglin, et al. "Kinect-based hand gesture recognition using trajectory information, hand motion dynamics and neural networks." Artificial Intelligence Review 52.1 (2019): 563-583.

[12] 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 (2018): 10.

[13] https://patents.google.com/patent/US20060280346

[14] https://patents.google.com/patent/US20130089245A1/en

[15] Mousavi, Seyed Muhammad Hossain. "A New Clustering Method Using Evolutionary Algorithms for Determining Initial States, and Diverse Pairwise Distances for Clustering.", International Journal of Mechatronics, Electrical and Computer Technology (IJMEC), 9.31, (4098-4110), 2019.

[16] Hossein, Seyed Muhammad, et al. "Utilizing SURF Features and KLT Tracking Algorithm in Augmented Reality (AR), Using Kinect V. 2 with the Aim of Autism Therapy." Proceedings of the National Conference on Computer, Information Technology and Applications of Artificial Intelligence, Ahvaz, Iran. Vol. 5. 2020.

[17] Mousavi, Seyed Muhammad Hossein, and S. Younes Mirinezhad. "Iranian kinect face database (IKFDB): a color-depth based face database collected by kinect v. 2 sensor." SN Applied Sciences 3.1 (2021): 1-17.

[18] Gonzalez, Rafael C., and Paul Wintz. "Digital image processing(Book)." Reading, Mass., Addison-Wesley Publishing Co., Inc.(Applied Mathematics and Computation 13 (1977): 451.

[19] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Vol. 1. Ieee, 2001.

[20] Mousavi, S. Muhammad Hossein, and Marwa Kharazi. "An edge detection system for polluted images by gaussian, salt and pepper, poisson and speckle noises." 4th National Conference on Information Technology, Computer & TeleCommunication. 2017.

[21] MOUSAVI, SEYED MUHAMMAD HOSSEIN. "A SPATIAL AND FREQUENCY BASED METHOD FOR MICRO FACIAL EXPRESSIONS RECOGNITION USING COLOR AND DEPTH IMAGES. Journal of Software Engineering & Intelligent Systems 6 (1), 17, 2021".

[22] Mousavi S. M. H., MiriNezhad S. Y., Lyashenko V. An Evolutionary-Based Adaptive Neuro-Fuzzy Expert System as a Family Counselor before Marriage with the Aim of Divorce Rate Reduction // Conference: 2nd International Conference on Research Knowledge Base in Computer Engineering and IT, At Tehran-Iran. – 2017. – P. 1-10.

[23] Mousavi, Seyed Muhammad Hossein, and S. Younes Mirinezhad. "Fingerprint Recognition Out of Object's Surface Using Ordinary Cameras with the Aid of Magnifier." Recall (without magnifier), 3rd National Conference on Computer, information technology and applications of artificial intelligence, ahvaz, iran, (2020), 30.89.6: 88-3.

[24] Razavi-Termeh, Seyed Vahid, Kourosh Shirani, and Mehrdad Pasandi. "Mapping of landslide susceptibility using the combination of neuro-fuzzy inference system (ANFIS), ant colony (ANFIS-ACOR), and differential evolution (ANFIS-DE) models." Bulletin of Engineering Geology and the Environment 80.3 (2021): 2045-2067.

[25] Polyvyanyy, Artem, et al. "Monotone precision and recall measures for comparing executions and specifications of dynamic systems." ACM Transactions on Software Engineering and Methodology (TOSEM) 29.3 (2020): 1-41.