Multi-Agent System to Assist Pupils in the Primary Stage

Sarah Mohamed Mohamed Ali, A. E. E. Elalfi and A.F. Elgamal

Abstract
There is an increasing number of children in primary and pre-schools with overcrowding in the areas where educational institutions exist. Children may be exposed to many risks such as accidents, loss, and fainting caused by any chronic diseases. Therefore, there is a huge need to recognize the pupils to know their information which enables us to help them. This paper presents a multi-agent system for the mobile device to recognize pupils. The system consists of three agents; image recognition agent, voice recognition agent, and knowledge base agent. Image recognition agent captures the pupil image with a mobile camera then detects his face and recognizes him. Voice recognition agent records the voice using a mobile microphone, extracts features, and decides the child's identity. The knowledge base agent receives information from a user, compares it with facts, and applies rules to decide the pupil's identity. Every mentioned agent takes some actions after the decision like show personal data of pupils on screen or suggesting first aids for some cases. The system used algorithms such as Eigenfaces, Gray-Level Co-occurrence Matrix (GLCM) for image processing, Mel-Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC) for sound processing. The system was implemented on a smartphone device with good and accurate results.

Keywords: Face Recognition, Speaker Recognition, Eigenfaces and GLCM Algorithms, Knowledge Base, Face Detection, Mobile Apps.

1 INTRODUCTION
With the prevalent usage of mobile devices and their applications, many fields of innovation caused the spread of mobile technology with a very interesting impact [1]. Companies and individuals widely use mobile devices to access their data and networks, in addition to their services and products. Personal mobile devices have spread and become popular, so the user looks forward to a full range of services available through the mobile Internet [2]. Smartphones and mobile applications have become an integral part of our daily life [3]. The field of face recognition has been promoted rapidly in the last decade. The reason for this development is its highly sought-after applications; entertainment, information security, surveillance, and law enforcement are just a few areas where face recognition technology is desired [4]. There are many methods of authentication via mobile phones, such as facial recognition through the camera, fingerprints, and voice identification through the microphone, gait or activity recognition via the accelerometers and/or gyroscopes, and gesticulate recognition by the camera or the accelerometer and gyroscope [2]. As a result of the tremendous increase in the computational capacity of mobile devices in the last decade, therefore, complex facial recognition systems can be portable [5]. In recent years, face recognition systems have an important role in the security systems of many applications such as passport/identification card (ID) authentication, home safety, Criminal examination, monitoring, enterprise security, and government events security [6]. Biometric systems use the behavioral and physiological features of a person to determine his identity and these biometrics are various like a fingerprint, face images, retina, iris, and voice [7]. Identifying a person with his voice as efficiently as forensic quality is challenging. Generally, voice recognition, like all other biometric features, is used to distinguish between people based on their voiceprint [8]. As a result of the need to automate the identification process for persons, there has been a great development of biometric identification systems recently [9].

Building mobile applications that can identify people with voice and image is a great technique for solving a lot of problems and providing all information related to them in a short time. This study aims to build a multi-agents system implemented on the mobile device which can identify primary stage children by their face image and voice to give the necessary data to provide urgent assistance for them. This paper
is organized as follows: Section 2 an overview of some relevant works that addressed identification systems on mobile devices. The proposed system architecture is illustrated in Section 3. Section 4 shows the discussion of experimental results. Finally, the conclusion and references are presented.

2 RELATED WORK

This section reviews several research studies about the identification systems and methods. All of these systems have used biometric techniques, a human can be identified by his biological characteristics and physiological. The most used, famous, and successful biometrics are face and voice. Many studies have been done in this field; the following section will show some of them. A study was done by Tugberk Duman in the field of face recognition, to improve proof of concept for intelligent glass and face recognition aiming to test the feasibility of the boarding pass free travel for premium passengers. In this research, boarding pass-free travel refers to the ability of airports to identify travelers through their faces at any stage of the inspection. The study showed promising results considering that the face recognition systems require a strong database that needs at least four to five photos when it comes to access control [10].

Another study was carried in the field of voice recognition by Hairol Nizam Mohd, et al. The voice recognition system was created to identify an administrator's voice by a software application that was programmed using the MATLAB language. The audio signal is captured and recorded, and then this signal is compared with other signals stored in a database, Mel-Frequency Cepstral Coefficients (MFCC) method was used and the voice biometric system depended on the signal word recognition. From the result of testing the system, it successfully recognizes the specific user’s voice and rejected other users' voices. [11].

Ishtar S. Jadhav, et al have performed human identification using face and voice recognition by merging both face and speech which can extraordinarily develop the accuracy of recognition as compared to the single biometric identification for security system development. The results showed that the Eigenface approach for Face Recognition and Mel-Frequency Cepstral Coefficients (MFCCs) for the voice recognition process is fast and simple which works well under a constrained environment [12]. Most of the previous studies used both image and sound systems to identify the people on the computer and some of them were mentioned in the previous part. In the last few years, some studies and researches have been carried out to determine the identity on mobile devices in many applications and for different purposes. Al rashed, H. H presented a study to build an Android application which can identify faces. Processing methods for the face photos were restricted to the mobile device capabilities. The preprocessing steps were grayscale conversion, then resizing the images to 128 x 128 pixels to decrease the computational processes on the image. The results showed that having a face recognizer on a mobile device can be used in conjunction with other facilities such devices can offer [13].

Weihao GAO developed an embedded system application that can take the face image and identify it by comparing it and the faces stored in the system database. This system is designed to work on smartphones and tablets with an Android operating system. By the study results, it has been proved that the recognition rate is over 90% using proper strategy [14]. Pinjie Ye et al have developed a mobile face identify authentication system. The system takes the face image, then verifies it with web services. The results showed that the solution was implemented and run on Note3 with 30fps of face detection, and its correct rate is above 90% [15]. Mrs. Madhuram M, et al have created a real-time face recognition system that used An Internet Protocol (IP) camera and image set algorithm. Open CV and Python programming were used to develop the system. The system consists of three parts: Detection module, training module, and recognition module [16].

Shonal Chaudhry and Rohitash Chandra implemented a face detection and recognition system for the visually impaired. They used mobile computing. A server-based support system supported this mobile system. Experimental results showed good performance with an accuracy of up to 94% in conditions where adequate lighting is present and lower performance in poorly-lit areas. [17].

All the mentioned studies have applied face recognition and authentication on mobile phones. Subsequently, increasing the accessibility, ease of use, and spread of these systems due to their development on mobile platforms. Other studies used sound to recognize the identity on mobile devices, for example, Gebremedhin T. Abraham has developed a real-time application to detect and recognize user activities based on audio signals on mobile phones. The audio recognition applications make mobile phone smarter and develop mobile features which make users more satisfied [18].

Angelos Pillos, et al, suggested a simple approach that relieves some of the privacy concerns and assesses the implemented real-time environmental sound recognition system on Android mobile devices. Their solution depended on performing the processing operations and the identification together on the mobile device. Also, they explored various feature extraction algorithms which helped in improving the accuracy of identifying sounds. Experimental results show improvement over state-of-the-art [19]. The identification of a person through his voice within a forensic quality context is challenging. Mohammed Alageeb, et al suggested a method for forensic speaker recognition for the Arabic language; the King Saud University Arabic Speech Database was used for getting experimental results. Mel-Frequency Cepstral Coefficients were applied for feature extraction and the Gaussian mixture model-universal background model was used for speaker modeling. Low, Equal Error Rates (EER), within noisy environments were resulted from their approach with very short test samples [20].

731
3 PROPOSED SYSTEM ARCHITECTURE

In this study, a multi-agent system has been proposed on the mobile device to identify the pupil's identity with some of his physical characteristics such as image, voice, and personal information. This system depends on three types of agents:

- Image Recognition Agent (Face and Landmarks)
- Voice Recognition Agent
- Knowledge Base Agent

Agents Interaction

We can overcome the defects and errors that may appear if only one agent is used for the identification process, by using more than one agent. We can reduce the error rate and improve the system accuracy; agents can work together and cooperate with each other as follows:

1-in case of the child is lost:

- When user chooses the image agent to identify the unknown child’s identity, the agent processes the unknown face image, extracts the features, matches it with the stored dataset images, takes a decision about the child identity, shows his image, and name on the screen, then asks user whether the output image is like the child in front of him. When the answer is yes, the image agent has successfully performed the identification process without any error. The image agent does the appropriate actions, such as offering a help, presenting the child data and calling parents.
- If the answer is no or the face is not identified, then the sound agent processes the unknown recorded voice, recognizes the unknown voice then shows his data and image. Sound agent asks user again if the output results are true. When answer is yes, sound agent takes the appropriate actions.
- when the answer is no, then the knowledge base agent collects all necessary information, applies knowledge base rules and finally determines the child identity.

2-in case of accident or fainting:

- If face image was deformed, the system will go directly to the knowledge base agent to identify the unknown child.
- If face image is obvious, the system will go to the image agent to perform the identification steps as mention before.
- If the result of identification process is not identified or the output image is different from reality, the agent sends the results to the knowledgebase agent to take the necessary steps to determine the identity of the child.

we notice that agents within this system interact with users and with each other in order to achieve a common goal, which is recognizing the unknown child in order to help him. when one of the system agents fails, other agent can do the task, so, this increases the reliability and robustness of the system.

3.1 Image Recognition Agent

This agent is composed of three main parts:

- The first part is a sensor, it is a mobile camera that captures a pupil’s photo.
- The second part which is responsible for image processing and classification, locates the face in the image and gets rid of the background, then extracts the distinctive features for each face image with any image processing algorithm to obtain the image pattern. The agent makes comparisons and matches between the anonymous image and the images stored in the dataset. Finally, the face recognition agent can help in determining the pupil identity and makes the necessary processes. Then decides about the unknown face identity.
- The third part is actions, the agent takes a set of actions after identifying the unknown face as follows:
  - Show all personal information on the screen like name, address, and others.
  - Call parents or staff phone number
  - Suggest the proper first aids to help chronic disease cases.

3.1.1 Face and Landmarks Recognition

Face recognition is one of the most popular methods of automated recognition for its ease and speed compared to other methods. Face recognition steps are image preprocessing, feature extraction, training data, classification, and matching. These steps will be discussed in the following sections:

Step 1. Image Preprocessing

In this study, the image preprocessing was done for face and landmarks as follows:

- Face Detection: The location of the face in the image is determined and detected by the Google vision technique, we can integrate the google vision library in our applications which have classes to detect a face and other features like nose, mouth and eyes. The FaceDetector class is used in this system for detecting the face and all other features. In the following figure the face is detected as follows:

  ![Face Detection by Google Vision](image)

- Face Rotation and Alignment: To make face alignment, an angle is calculated based on the right and left eye points located in the face image to rotate the face. The angle is calculated based on the following equation [7]:

\[
\phi = \text{atan} \left( \frac{y_2 - y_1}{x_2 - x_1} \right) \times \frac{180}{\pi}
\]  

(1)
Where $\emptyset$ is the rotation angle and $(y_1, x_1)$ $(y_2, x_2)$ are the left eye and right eye coordinates, we calculate the distance between the left eye and right eye points. We use the coordinates of left eye $(y_1, x_1)$ and right eye $(y_2, x_2)$ in the above equation to compute the required angle to rotate the face.

- Crop a face and Scaling: The face is cut from the image and the background is eliminated. After the face is detected in the image, the left, top, right and bottom points are set, a rectangle is drawn on the face area, then this area is cut from image.

- Grayscale Conversion: The colored face image is converted to grayscale. The preprocessing steps are shown in figure 2:

![Image](image1)

**Fig. 2 Face Image Preprocessing**

In this study, face features were used to identify the child identity, such as the left and right eye, nose, and mouth. The feature vectors are extracted from features images. The preprocessing steps are applied to the facial landmark images as well, but with a slight difference in the last step. The left or right eye image is only cut from the face image by certain calculations based on the distance between the left and right eye and the distance equation was mentioned previously. The preprocessing for left and right eyes is shown in figure 3:

![Image](image2)

**Fig. 3 Left and Right Eye Image After Preprocessing**

By Google Vision technology the face location was determined in the image, then the face width was measured. By the ratio of the nose width to the face width, nose width is calculated. The relationship between the nose width and the rest of the facial features such as the eye and the base of the nose, the top left, right, and bottom points were calculated which were used for cropping the nose area. Then the nose area was cropped with great precision as shown in figure 4:

![Image](image3)

**Fig. 4 Nose Image After Preprocessing**

After calculating the face width, the mouth width is computed. Mouth width and the rest of features such as nose were used to compute the top left, right and bottom points which used for cropping the mouth area very accurately as shown in figure 5:

![Image](image4)

**Fig. 5 Mouth Image After Preprocessing**

**Step2: Training Data**

The images were stored in the storage of the mobile device. Eigenface and GLCM algorithms were used to extract the distinguishing characteristics from each image in the dataset which contains images. These images are trained to be used in the matching and identification process. The mentioned algorithms will be discussed in the following sections.

**Step 3. Feature Extraction**

In this step, the distinctive features for the face image (a query face) are extracted by image algorithms, in this study, two methods of algorithms were used to increase the overall quality, recognition rate, and the system efficiency the following methods will be discussed in more details:

- Eigenfaces Algorithm
- GLCM Algorithm

**A. Eigenfaces Algorithm**

Eigenfaces refers to an appearance-based approach for face recognition. It captures the variation in the data set of face images which is later used to convert and match images or individual persons [22]. The following steps explain eigenfaces calculation:

1. Each image consists of a matrix of pixels; each pixel represents a number. Each training image becomes a matrix $I$ and $I_1$, $I_2$......$I_M$ are the images and $M$ is the number of it. Matrix $I_i$ is converted to a vector of values $T_i$ [23].
2. The average face is calculated as the following equation from [23- 25]:

\[ \Psi = \frac{1}{M} \sum_{i=1}^{M} T_i \]  

(2)

Where the sum of each vector \( T_i \) is calculated, and then the sum is divided by the number of images \( M \) which gives the vector \( \Psi \) representing the average.

3. Each vector of the image will be subtracted by the average face to compute the vector \( \emptyset_i \) as shown in the following equation from [23, 26, 27]:

\[ \emptyset_i = T_i - \Psi \]  

(3)

Where:
- \( i=1 \) to \( M \)
- a difference matrix \( A = \emptyset_1, \emptyset_2, \emptyset_3, \ldots, \ldots, \emptyset_M \) will be computed which keeps only the distinguishing features for face images and removes the common features.

4. The covariance matrix is computed, it is from [26, 27] as the following equation:

\[ C = \frac{1}{M} \sum_{n=1}^{M} \emptyset_n \emptyset_n^T = AA^T \]  

(4)

Where \( C \) is calculated by multiplying \( A \) matrix with \( A \) transpose called \( A^T \).

5. Then compute the eigenvalues and eigenvectors from the covariance matrix, but this matrix is very big size \( N^2 \times N^2 \) and this will cause a huge computation so the calculation of covariance matrix will be by equation [28] like this:

\[ L = A^T X A \]  

(5)

Where \( L \) is a matrix of size \( M \times M \) which gives the same eigenvectors and eigenvalues but with reduced dimension.

6. The eigenvectors of \( C \) (matrix \( U \)) can be obtained by using the eigenvectors of \( L \) (matrix \( V \)) by the following equation which is from [25, 29]:

\[ U_i = A V_i \]  

(6)

Where \( U_i \) the eigenface with the order of \( i \). \( A \) is the matrix of the images vectors after subtracting the average face and \( V_i \) is the eigenvector with the order of \( i \). The Eigenfaces are by equation [29]:

\[ \text{Eigenfaces} = U_1, U_2, U_3, \ldots, U_M \]  

(7)

Where \( i=1 \) to \( M \).

7. Instead of using \( M \) eigenfaces, the highest \( K <= M \) is chosen as the eigenspace. Then the weights of eigenvectors \( W_i \) to represent the query image in the eigenface space, is given by the equation [29]:

\[ W_i = U_i^T \times \emptyset_{\text{query}} \]  

(8)

Where \( i=1, 2, 3, \ldots, K \), \( \emptyset_{\text{query}} \) is the vector of query image after subtracting the average face.

8. The weights \( \Omega \) can be represented by a column vector [25]:

\[ \Omega = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ \cdots \\ W_k \end{bmatrix} \]  

(9)

B. Gray Level Co-occurrence Matrix (GLCM) Algorithm

The gray level co-occurrence matrix (GLCM) is extracted from an image by computing how pairs of pixels with particular values and in a special spatial relationship occur in an image, then the statistical formulas will be applied to extract the feature vector from this matrix to be used for classification and matching processes [30]. These features are generated by calculating the features of each one of the co-occurrence matrices obtained by using the directions 0°, 45°, 90°, and 135°, then averaging these four values [31].

To extract the feature vectors some equations have been used. The following equations contrast, homogeneity, coloration, and energy equations are from [32].

\[ \text{Contrast} = \sum_i \sum_j (i-j)^2 P[i,j] \]  

(10)

Where \( M, N \) are the dimension of the image \( P[i,j] \) for \( i=1, 2, 3, \ldots, M, j=1, 2, \ldots, N \). \( P[i,j] \) is a two-dimensional function and it is composed of \( M \) pixels in the vertical direction and \( N \) pixels in the horizontal direction. \( i \) and \( j \) are horizontal and vertical co-ordinates of the image.

\[ \text{Homogeneity} = \sum_i \sum_j \frac{P[i,j]}{1 + |i-j|} \]  

(11)

Where \( M, N \) are the dimension of the image \( P[i,j] \) for \( i=1, 2, 3, \ldots, M, j=1, 2, \ldots, N \). \( i-j \): the absolute value for subtracting \( j \) from \( i \) value.
Correlation = \sum_{i} \sum_{j} \frac{(i-\mu)(j-\mu)P[i,j]}{\sigma^2} \tag{12}

Where \( \mu \) and \( \sigma \) are the mean and standard deviation of image respectively.

Energy = \sum_{i} \sum_{j} (P[i,j])^2 \tag{13}

Where M, N are the dimension of the image \( P[i,j] \) for \( i=1, 2, 3 \ldots M, j=1, 2, \ldots N \).

Variance and inertia equations are from \[33\]. Shade, entropy and prominence equations are from \[33,34\]. They are shown as follows:

Entropy = -\sum_{i} \sum_{j} P[i,j] \times \log(P[i,j]) \tag{14}

Where \( \log(P[i,j]) \) is the logarithm of \( P[i,j] \) which is a two-dimensional function and it is composed of M pixels in the vertical direction and N pixels in the horizontal direction. M and N are the dimension of the image.

Variance = \sum_{i} \sum_{j} (i-\mu)^2 P[i,j] \tag{15}

Interia = \sum_{i} \sum_{j} (j-\mu)^2 P[i,j] \tag{16}

Where \( i \) and \( j \) are horizontal and vertical co-ordinates of the image and \( \mu \) is the mean of an image which is calculated by the following formula \[30\]:

\[
\mu = \frac{\sum_{i} \sum_{j} P[i,j]}{n} \tag{17}
\]

Where \( n \) the total number of pixels in an image.

Shade = \sum_{i} \sum_{j} \{i+j-\mu_x \mu_y\}^3 \times P[i,j] \tag{18}

prominence = \sum_{i} \sum_{j} \{i+j-\mu_x \mu_y\}^4 \times P[i,j] \tag{19}

Where \( \mu_x, \mu_y \) are the means of probability matrix \( P[i,j] \) along row wise \( x \) and column wise \( y \). The following formulas are used to compute these means \[33\] as shown:

\[
\mu_x = \sum_{i} \sum_{j} i \ P[i,j] \tag{20}
\]

\[
\mu_y = \sum_{i} \sum_{j} j \ P[i,j] \tag{21}
\]

Where \( i \) and \( j \) are horizontal and vertical co-ordinates of the image.

**Step 4: Classifications and Matching**

This is the last step in face recognition. After obtaining the characteristics (features vector) of an anonymous image as well as images stored in the database, the process of comparison and matching is done. The matching process depends on one of the classification algorithms such as the Euclidean distance, then a decision is taken about the identity of the unknown image. Euclidean distance formula \[35\] is shown as follow:

\[
d(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \tag{22}
\]

Where:

\( d(p,q): \) The Euclidean distance between two points \( x \) and \( y \) or a distance value.

\( p \) is the point of the extracted features vector for the query image and \( q \) is the corresponding point of the trained image features vector, these vectors are extracted in the feature extraction step as mentioned previously. For example, the contrast point of a trained image is subtracted from the corresponding contrast point of the test image and all the other features also.

\( i=1,2,3,4\ldots n. \)

\( n \) is the number of the vector points for each image.

To make the matching process the above equation is used to get the difference between the feature vector for the query image and each trained image vector.

**3.2 Voice Recognition Agent**

This agent is a software agent composed of three main parts:

- The first part is a sensor; it is a mobile mic that records a pupil’s voice.
- The second part which is responsible for sound processing and classification, after recording the sound by the mobile mic, the agent extracts the distinctive features for each recorded voice with any sound processing algorithm to obtain the sound pattern. The agent makes comparisons and matches between the anonymous voice and the stored voices in the dataset. Finally, the voice agent can know who is the speaker and make the necessary process for help.
- The third part is actions, the agent takes a set of actions as we mentioned before with the image agent.
3.2.1 Speaker Recognition

The voice is one of the physical features that distinguish one person from another. Consequently, many security systems use the voice for matching and identification processes. Sound processing is divided into two types, the first depends on speech recognition and the other depends on voice recognition. Studies have shown high accuracy in the processes of identification for people by their voices. In this study, a sound processing algorithm is used to extract the distinctive characteristic of the recorded voice regardless of the speech itself. Voice agent can recognize pupil using the following sound processing steps:

Step 1: Sound Preprocessing

Some pre-treatment operations are applied to the audio signal in order to get rid of noise and silence through the following steps:

1- Analogue to Digital Transformation:
   The analog audio signal is converted into a digital signal Pulse-Code Modulation (PCM).

2- Voice Detection
   Some statistical equations such as mean and standard deviation are applied to classify the acoustic signal into spoken and non-spoken.
   - Compute the mean and standard deviation of the first N samples of the given utterance. The following formulas were used to check the audio signal, and they are from [36]:
     \[
     \mu = \frac{1}{N} \sum_{i=1}^{N} x(i)
     \]
     \[
     \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x(i) - \mu)^2}
     \]
     Where:
     - \(x(i)\) is the first samples of the given utterance.
     - \(\mu\) is the mean.
     - \(N\) is the number of first samples of audio signal

   - Apply Mahalanobis distance function from the first sample to the last sample of the speech recording. In each sample and check whether distance greater than 3 or not and if the distance > 3 then it is treated as voice as shown in the following formula [36]:
     \[
     \frac{x - \mu}{\sigma} > 3
     \]
     Where \(x\) is the sound sample.

   - Mark each frame to be voiced or unvoiced frame.

3- Silence Removal

After dividing the audio signal into some frames then, determining the voiced and unvoiced values and frames, selecting the frames which consist only of ones and the silent values are deleted then a matrix of the acoustic signal values is obtained.

Step 2: Feature Extraction

At this step, the distinctive features of the recorded sound (sound pattern) are extracted by sound processing algorithms, two algorithms were used for extracting features from voice: MFCC and LPC. The two methods were used in this study to recognize the identity of the person through his voice, not on what he says words or talks, these algorithms depend on human hearing perception. We compared between the two algorithms in extracting features, the overall system quality, recognition rate, the time taken for identification.

The following steps are applied to audio signal for the two algorithms:

- Pre-emphasis
  A filter is used to reduce the noise and achieve the balance between low and high-frequency values by equation [37,38]:
  \[
  y[n] = x[n] - 0.95x[n - 1]
  \]
  Where:
  - \(n\) is the number of samples in each frame.
  - \(y[n]\) is the output signal and \(x (n) = \) input signal.

- Frames
  The digital audio signal is divided into a number of frames \(N\). Each frame is from 20-40 msec.

- Hamming Windowing
  Convert the discontinuous signal to a connected signal by applying the hamming window equation [38].
  \[
  y[n] = x[n] * w[n]
  \]
  Where:
  - \(w [n] : \) the windowing signal [40] that is shown below:
    \[
    w[n] = 0.54 - 0.46 \cos \left( \frac{2\pi n}{N - 1} \right) \quad 0 \leq N \geq N - 1
    \]

A. Mel Frequency Cepstral Coefficients (MFCC)

After applying the hamming window, the following steps are calculated to extract the MFCC features as follow:

- Fast Fourier Transform
  Convert each frame of \(N\) sample from the time domain to the frequency domain. Spectral estimation is computed for each frame by applying Discrete Fourier Transform (DFT) to produce spectral coefficients with this equation [39].

736
\[ X[k] = \sum_{n=0}^{N-1} y(n) e^{-2\pi \frac{jk}{N}} \quad 0 \geq n, k \geq N - 1 \quad (29) \]

Where:

- \( X[k] \) are the spectral coefficients and \( y[n] \) is the framed speech signal, \( N \) is the number of discrete frequencies.

- **Mel Filter Bank Processing**

  The frequency range in the FFT spectrum is very wide and voice signal does not follow the linear scale. A set of triangular filters that are used to approximate the frequency resolution to the human ear or Mel scale as the following equation [40,41]:

  \[ \text{Mel}[f] = 2595 \times \log_{10}(1 + f/700) \quad (30) \]

  Where, the equation is used to compute the Mel for given frequency \( f \) in HZ.

- **Discrete Cosine Transform**

  Discrete Cosine Transform (DCT) process converts the log Mel spectrum to the time domain. The result of the conversion is called Mel Frequency Cepstrum Coefficient. The set of coefficients is called acoustic vectors [38]. Coefficients are the voice feature vectors; they will be used in the classification and matching process. Figure 7 shows the steps of the MFCC calculations as shown:

  ![Fig. 7 LPC implementation Steps [42]](image)

- **B. Linear Predictive Coding (LPC)**

  LPC technique can supply estimation of poles of vocal tract transfer function. Samples in LPC can be approximated as past samples linearly. To generate the LPC features-emphasize process should be applied to the input speech. The output of pre-emphasizer is used as the input to frame blocking where the signal is blocked into frames of \( N \) samples. To eliminate the signal discontinuity at the beginning and end of every frame, the windowing step is applied. Hamming frame is an example of typical frame. After windowing, the autocorrelation process is done and the highest autocorrelation value gives the order of LPC analysis [42].

\[ \hat{s}(n) = \sum_{k=1}^{p} a_k s(n - k) \quad (31) \]

Where:

- \( \hat{s}(n) \): is the predicted sample at time \( n \).
- \( s(n) \): is the speech sample at time \( n \)
- \( a_k \): \( k = 1, 2, 3, \ldots \) \( p \) are the predictor coefficients.

The prediction error \( e(n) \) is defined as

\[ e(n) = s(n) - \hat{s}(n) \quad (32) \]

The mean square of the prediction error over an analysis frame of \( N \) samples is given by:

\[ E = \sum_{n=0}^{N-1} e^2(n) \quad (33) \]

Where:

Minimizing \( E \) with respect to the set of predictor coefficients results in a set of \( p \) normal equations. The predictor coefficients are obtained by solving this set of \( p \) normal equations.
Step 3: Training Data

In this step, a set of sounds is recorded for a number of pupils in one of the primary schools. Each pupil's voice is recorded in many cases (normal - anger - fear - happiness - sadness - disgust). These recorded sounds are then placed in a dataset as a sound file path and the files themselves are placed in the internal storage of the mobile device.

Sound processing algorithm is then applied to the recorded files to extract the characteristics of each recorded sound (acoustic vector) and stored them in a data file, then, the stored features will be used in matching operations.

Step 4: Classifications and Matching

This is the last step in voice recognition. After obtaining the characteristics (features vector) of an anonymous sound as well as voices stored in the dataset, the process of comparison and classification is done. Euclidean distance was used too for the matching process. Equation (22) was used for classification as mentioned before.

3.2 Knowledge Base Agent

This agent aims to integrate with the two mentioned agents to overcome some problems such as incomplete information. The face image of the unknown child may be incomplete or distorted as a result of accidents or defects in photography and lighting. Also, the system may be unable to obtain the child's voice in some cases like fainting therefore we need this agent to help in solving these problems.

The agent asks the user a set of questions. Obtained answers are compared with the facts stored in the knowledge base. Then agent deduces the identity of the unknown child and also provides all the information related to this child such as his address, phone number, father's name and school.

The knowledge base consists of a set of facts and rules. The facts are saved in a dataset with two tables implemented by SQLite on android studio. The first table saves physical data like face shape, face color, hair style, eye size, mouth shape and others. The second saves other data like the accent of speech, hobby, sensitive type, body disabilities and chronic diseases. The following table shows an example for facts located at the system.

<table>
<thead>
<tr>
<th>Table 1: Facts in the Knowledge Base</th>
</tr>
</thead>
<tbody>
<tr>
<td>child_1</td>
</tr>
<tr>
<td>Name</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Face shape</td>
</tr>
<tr>
<td>Face color</td>
</tr>
<tr>
<td>Face mark</td>
</tr>
<tr>
<td>Hair style</td>
</tr>
<tr>
<td>Hair color</td>
</tr>
<tr>
<td>Hair cover</td>
</tr>
<tr>
<td>Eye size</td>
</tr>
<tr>
<td>Eye color</td>
</tr>
<tr>
<td>Nose shape</td>
</tr>
<tr>
<td>Mouth shape</td>
</tr>
<tr>
<td>Body shape</td>
</tr>
<tr>
<td>Body shape</td>
</tr>
<tr>
<td>Body shape</td>
</tr>
<tr>
<td>Hair length</td>
</tr>
<tr>
<td>Accent type</td>
</tr>
<tr>
<td>Walking type</td>
</tr>
<tr>
<td>Hobby</td>
</tr>
<tr>
<td>Sensitive type</td>
</tr>
<tr>
<td>Brothers number</td>
</tr>
<tr>
<td>Body disability</td>
</tr>
<tr>
<td>Favorite food</td>
</tr>
<tr>
<td>Love subject</td>
</tr>
<tr>
<td>Hate subject</td>
</tr>
<tr>
<td>Chronic disease</td>
</tr>
<tr>
<td>Panic type</td>
</tr>
<tr>
<td>Favorited program</td>
</tr>
<tr>
<td>School problem</td>
</tr>
</tbody>
</table>

The knowledge base has a form of production rules. Those rules are in the form of sequences with a structure as follows: If <conditions> then actions. The inference engine in the proposed system examines these rules based on the stored facts and then decides about the identity of the unknown child, also gives all the information related to the unknown child as shown in the following figure:
4. SYSTEM IMPLEMENTATION

To implement the proposed system on mobile devices, java language was used for programming, android studio as an Integrated Development Environment (IDE), and SQLite data on android for creating a data table. The system or application is suitable for all and smart devices and supports a variety of connectivity technologies. The proposed system on mobile has been developed by using the appropriate Software Development Kit (SDK). The mobile system application was applied on a smartphone (Sony Xperia Z3) with the following specifications:

- OS: Android 4.4.4 (Kit Kat).
- CPU: Quad-core 2.5 GHz Krait 400
- Main camera: 20.7 MP, f/2.0, 25mm (wide)
- Speakers: with stereo speakers
- Battery: Non-removable Li-Ion 3100 mAh battery
- Display Resolution: 1080 x 1920 pixels.

The first screen of the user interface of the implemented system contains two buttons, the user can choose from them to specify the child case, when a user selects the LOST button, the next screen will be shown to choose one of two options: Identify by Image, identify by voice methods as in the following figure:

![Fig. 8 The Proposed System Main Screens](image)

When the user selects BY IMAGE button, the next screen will be shown to choose any of the two methods, pressing any of the two previous methods, a screen will appear to select the recognition method; by the whole face or by the landmarks like eyes, mouth, and nose as shown in the following figure:

![Table 2: Knowledge Base Rule](image)
When pressing **Face Recognition** button, the following screen appears:

As shown in figure 10, the user presses **Select Photo** button to select the unknown child image from the stored images then the face image will be shown. Then, the user presses the **Recognize** button, to start matching, then the unknown face is identified. The recognized child image will be shown with his related details.

Eyes, noses, and mouth images are cut from the pictures and the distinguishing features are extracted from them. Then the image agent compares the feature vector of the query image to other stored features to identify a pupil as shown in figure 11:

User can use voice to identify the unknown child by pressing the voice button then, the following screen will appear:
In figure 12, when pressing the SELECT_AUDIO button, agent will get the recorded voice stored in the music folder. By pressing RECOGNIZE button, the agent will start recognition process. Back to figure 8, when pressing Accident/fainting button, two options will be available for user; face image is clear or deformed.

When user press Face image is deformed button, the face image is incomplete or unclear, then user can use a smart knowledge base to identify the child's identity. The system graphical user interface (GUI) provides screens display a set of multiple-choice questions to answer some questions about the pupil, then the knowledge base agent compares the answers to the stored information and decides about the unknown child identity as shown in the following figure:

5. EXPERIMENTAL RESULTS

The proposed system has been applied to a set of pictures and voices for primary school pupils. Two datasets have been built, the first for faces and the second for voices. In this study, we compared the results for applying GLCM and eigenfaces methods for identifying the child according to his face and features image. Also, all results for applying system with MfCC and LPC algorithms were compared.
The confusion matrix equations were used to analyze results, measure the recognition rate, and also, to specify the accuracy and efficiency of the system in identifying people by their voice and face image. Some performance measures were used to judge the proposed system results such as accuracy, Precision, recall, f-measure, and specificity.

5.1 Face Images Results

A set of about 100 photos was taken for pupils in a primary school in Egypt by mobile camera 20.7 MP. The pictures were divided into two groups, the first is used for training and the second is dedicated to testing by the proposed system. The training set contains 42 pictures and 6 classes or pupils in various situations with different facial expressions, including anger, sadness, fear and laughter. All images are frontal with a fixed dimension: 210w x 200h. The following figure shows a group of student pictures stored in the dataset:

In this study, two algorithms have been used Eigenfaces and GLCM for face recognition. The image recognition agent picks the unknown child picture, detects and crops the query face image as shown in the following figure:

Then recognizes him, in this case the pupil was recognized as Hossam, the following figure shows this:

The following table shows the comparison results between the two methods as follow:
After applying the proposed system to the tested pictures, the results between the two methods were compared. Table 3 shows the final results of applying the proposed system on the tested pictures for identifying the unknown children, two algorithms have been used.

The results were calculated by comparing the values of the proposed system and the expert to know how many times the proposed system results matched the expert. This was done by confusion matrix calculation. Accuracy, recall and specificity equations were calculated.

According to the table, the value of accuracy of the proposed system by the eigenfaces method is 94% and the accuracy by the GLCM method is 95 %. So, the recognition rate by the proposed system with the two algorithms is very high and efficient. We also notice that, the precision value for the two methods of the proposed system is 87%. This means that the proposed system has a very good ability to recognize the child with a very acceptable ratio.

From the previous results shown in table3, anyone can easily notice that both GLCM and eigenfaces methods give very close and satisfying results and the proposed system is capable of identifying the unknown child by his frontal facial image. The main difference between the two algorithms that eigenfaces method uses a grayscale image while the GLCM method uses the colored face image, but the two methods give results that are almost identical to the front face images.

But the time taken to do the pupil identification process is faster in the case of using the GLCM method than the Eigenfaces method. So GLCM method is better than the other method in terms of speed and performance.

### 5.2 Face Landmark Results

A set of photos was taken for primary school pupils, then the face landmarks like eyes, nose, and mouth were cut from the face image and this was done in the preprocessing stage as mentioned before. By the same way, the results of applying the system with the two methods on face features images like nose, mouth and eyes were compared in table3.

The pictures of eyes, mouth, or nose were divided into two groups, the first group was used for training and the second was dedicated to testing by the proposed system. Image recognition agent crops left eye and right eye image from face image for the query pupil as shown in the following figure:

**Fig. 17 Left and Right Eye Images After Cropping**
The agent applies image preprocessing then recognizes the pupil and in this case the pupil is Ahmed.

Table 3 shows the results of applying the proposed system on facial features images and results of identity recognition by face landmarks as we see in the left eye results, the accuracy value is 0.943 or 94% for the GLCM method and 0.872 or 87% for eigenfaces method. This means that the recognition rate of the proposed system using the left eye with the GLCM algorithm is better. In right eye results, the accuracy value is 86% for the GLCM method and the eigenfaces method.

We also notice that the precision values for the two methods of the proposed system are very close. Image recognition agent crops mouth and nose images from the face also as shown in figure 18:

\[
\text{Table 3 shows nose image results. The accuracy value is 0.823 for the GLCM method and 0.87 for the eigenfaces method. We also notice that the precision value for the two methods of the proposed system is 0.553 and 0.70 respectively. This ratio is lower than the precision value of using the face image as a whole by the two methods also. In the same way, mouth results show the accuracy value is 82% for the GLCM method and 87% for the eigenfaces method. The specificity value of the two methods of the proposed system is 0.89 and 0.92 respectively.}
\]

After comparing the system recognition results for both the face image as a whole and its landmarks, we found that the accuracy ratio is very high and gives a very satisfactory rate with simple errors in the case of using the face image as a whole. But in the case of mouth, nose, and eye images, the results are less accurate, and the ability to identify people also.

The two methods give very high efficiency in extracting the distinctive features and considering photography defects and lighting conditions. Also, there are no significant differences between the results of the GLCM and the eigenfaces methods. But they vary in how to extract the features from the image itself. Eigenfaces method completely deals with all the images in the dataset to extract the features from each image or a holistic method. GLCM method extracts the features from each image separately using statistical equations. The above table shows the time taken for the preprocessing of the dataset images and recognition process by the GLCM and eigenfaces methods. As we can see the run time is shorter in the case of the GLCM method than the eigenfaces method, and this makes it more flexible and effective. The following figure shows the recognition rate and as follows:
5.3 Voice Results

A set of voices was recorded for primary school students, then the sound files were divided into two groups, the first is used for training and the second is dedicated to testing. The sound dataset contains voices for the pupil in various situations, including anger, sadness, fear, and laughter. The format of the sound file was a .wav format as shown in figure 21:

The proposed system was applied to the query voice by the voice recognition agent to identify the pupil identity. The MFCC and LPC methods were used. After applying the system, the output results were evaluated by the confusion matrix as shown in the following table:

<table>
<thead>
<tr>
<th>Voice results</th>
<th>MFCC</th>
<th>LPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.96363</td>
<td>0.95533</td>
</tr>
<tr>
<td>Recall</td>
<td>0.823</td>
<td>0.837</td>
</tr>
<tr>
<td>Precision</td>
<td>0.807143</td>
<td>0.7457143</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.979873</td>
<td>0.97416789</td>
</tr>
<tr>
<td>F. Measure</td>
<td>0.7975758</td>
<td>0.76738095</td>
</tr>
<tr>
<td>Processing Time</td>
<td>29 sec</td>
<td>30 sec</td>
</tr>
</tbody>
</table>

According to the previous table, the accuracy value of the proposed system for the MFCC and LPC methods is 96%, which means that there are no significant differences between the results of the MFCC method and the LPC method. The recognition rate of the proposed system by the sound agent is precise and high by the two methods. The precision value of the proposed system is 80%, but by the LPC method is 0.75%. From results, the proposed system can recognize the pupils with an acceptable ratio based on their voiceprint.

The above table shows the time taken for preprocessing the dataset voices and the recognition process for query voice by the MFCC and LPC methods also as we can see the run time is for two is very close, so there is no high difference between the two methods. The following figure shows the recognition rate and as follows
6. CONCLUSIONS

With the technological development in artificial intelligence (AI) and smartphone technologies, image, sound, and knowledge base techniques were employed to develop integrated systems to help children or pupils in necessary cases. In this paper, a multi-agent system was developed for the mobile device to identify the pupil's identity based on their face, voice, and a knowledge base containing information about them to assist in critical cases like loss or fainting. The system was applied to a set of primary school pupils. Experimental results showed that the proposed system can identify the pupil's identity by his frontal facial image and his voice with good accuracy and high recognition rate.

7. FUTURE WORK

In this paper, a multi-agent system for mobile devices has been developed to identify children’s identity. Some biometric features such as voice and image were used. From this point, it is also possible to develop other systems based on fingerprints or others biometric to improve the accuracy results for these systems. The current system can also be developed to work online from the server, where all related data and information can be stored. This will work more efficiently, especially in the case of the huge datasets which may be used in these systems. This system can also be developed to help in recognizing people under the ruins or in disaster cases.

REFERENCES


[38] Lindasalwa Muda et al., 2010, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," journal computing, volume 2, issue 3, march 2010, ISSN 2151-9617.


[40] Lallouani Bouchakour and Mohamed Debyeche., 2018, "MFCCs and Gabor Features for Improving Continuous Arabic Speech Recognition in Mobile Communication Modified," International Conference on Advanced Aspects of Software Engineering Page 116 ICAASE.


dependent speaker recognition using machine learning
algorithms," International Journal of Engineering Science and
Technology, Vol. 2(10), 2010, 5150-5157