Biometric authentication data with three traits using compression technique, HOG, GMM and fusion technique

Balaka Ramesh Naidu a,*, Maddali Surendra Prasad Babu b

a Department of Information Technology, AITAM Engineering College, Tekkali, India
b Department of Computer Science & Systems Engineering, Andhra University, Vizag, India

This paper presents a three trait identification model called multimodal recognition system developed by using different traits like face, finger and voice (Babu and Naidu, 2014, 2016; Balaka and Surendra, 2017) [1–3]. This system provides more security when compare to existing works. Initially, all the traits are followed by pre-processed, extract features using Histogram of oriented gradients (HOG), then apply Gaussian mixture model (GMM) for finding probability density function (PDF) values and then combining these features by using Score level fusion. The result of these features considered as a trainee dataset. In verification process, each test image trait compare with the trainee dataset. This entire process of authentication is done by using machine learning based technique.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
How data was acquired
Traits are acquired with sensor and processed with HOG and GMM.

Data format
Tiff data, jpeg, audio data, authentication data

Experimental factors
Three traits: Face, Fingerprint and Audio format are targeted. Compression techniques, HOG, GMM and fusion on three traits. 150 Face, Finger and Audio format to be used in authentication purposes are extracted with this approach.

Experimental features
Extract three different biometric traits from each person. All the 150 traits showed that there are biometric recognition similarities.

Data source location
AITAM, TEKKALI, ANDHRA PRADESH, INDIA

Data accessibility
The data is provided with this article

Value of the data

- The data presented in this article are face, fingerprint and audio format are obtained from AITAM, TEKKALI. These are subjected to Compression, HOG, GMM and fusion has shown the high specificity.
- These data suggest that THREE traits are useful and robust method for providing information for various detections.
- Access to the raw face, fingerprint and audio format data allows researchers to perform further analysis based on their own computational algorithms.

1. Data

In this paper, consider the AITAM college dataset of three traits face, fingerprint and voice for experiments. The entire recognition system can be divided into five steps. In the step 1 collect various face images from AITAM College information shown in Fig. 1(a), respective fingerprints and voice signals are shown in Fig. 1(b) and (c). Complete dataset of AITAM College is presented in Appendix Table 1. In this article 150 face, finger and voice traits are collected. Initially, The finger data is collected using SecugenHamster Plus fingerprint scanner which gives images of size 260×300 pixels. These images are again compressed to 69×57 pixels without losing predominant information to match with the sizes of face data. The face data is collected with single image size of 580×560 pixels and then compressed to 69×57 pixels without losing predominant information as database management is also very important. The voice is collected from the same persons with audio length of 5 secs and sampling frequency of 48 KHz. Then it is downsampled to 5 times to reduce the database size. This voice is denoised and its power spectrum is obtained along with dominant frequency points. The data is taken from the staff members of AITAM Engineering College located in Tekkali, India. In this paper face and finger data is presented in Appendix I and voice data link is presented in Appendix II. The authentication system using face recognition is there since long time but the authentication system can be break down in some cases using morphing etc. So to make the system more secure, along with face recognition, fingerprints and voice are considered. Dealing with these three traits separately is common practice. In this paper, all these traits are converted into single entity using fusion techniques and treated it as final data for authentication. With this, the database can be utilized in more efficient way and the authentication system will become more robust. For this, the fingerprints and voices are collected from the same persons [1-3]. Only 3933 samples of the total audio signal is considered to match with the sizes of face and finger. The images considered here are in Tiff format and audio is in wav format. As it is difficult to present all the data, only few samples are presented in this article.

The face and fingerprint traits shown in Fig. 1(a) and (b) are pre-processed by different steps such as convert RGB to gray, remove noise and compressed using standard techniques PCA, KLT, DCT, HAAR. Fig. 1(c) is different voice signals of eight persons.
In the second step, extract the features of each trait. This is achieved by Histogram of oriented gradients. In this, cells can be either radial or rectangular shape and these are spread over 0–360° or 0–180° and it depends on whether the gradient is signed or unsigned.

Fig. 1. (a) Face dataset. (b) Fingerprint dataset. (c) Voice dataset.
Fig. 1. (continued)
In the third step, the Gaussian mixture model is applied for finding the Gaussian distribution for face shown in Fig. 2(a), fingerprint shown in Fig. 2(b) and voice shown in Fig. 2(c).

In the fourth step, combine the fingerprint, face and voice traits of PDF values by using score level fusion technique. This dataset is called training dataset (Fig. 3).

In the fifth step, test dataset is compared with the training dataset by using correlation method. If match is found then it is “accepted” else “not accepted”. This is achieved by following Eq. (1).

\[
\rho_{X,Y} = \frac{\text{Cov}_{AB}}{\sigma_X \sigma_Y}
\]

Where \( \mu \) is mean, Let us assume, A, B are two different image vectors. In the step1, calculate the mean of A (i.e., \( \mu_X \)) and B (i.e., \( \mu_Y \). In the step2, subtracts the mean value from the A matrix (i.e., A−\( \mu_X \)) and subtracts the mean value from the B (i.e., B−\( \mu_Y \). In the step 3, multiply both A_sub and B_sub and then calculate the mean value (i.e., Cov_{AB}=mean (A_sub×B_sub)). In the step 4, find standard deviation of both A and B. Finally, \( \rho_{X,Y} = \frac{\text{Cov}_{AB}}{\sigma_X \sigma_Y} \). If correlation coefficient result is “1” then highly correlated else “0” if it is not correlated.

2. Experimental design, materials and methods

The present model is designed to retrieve the quality of output image trait based on the principle is as follows: pre-processing which is discussed in Section 2.1 and extract features using HOG which is discussed in Section 2.2, Gaussian mixture model (GMM) process which is discussed in Section 2.3 is performed to predict the probability density values (PDF), Fast Fourier Transform for converting audio signal to digital which is discussed in Section 2.4, score level fusion (SCL) combines all the features of each individual traits which is discussed in Section 2.5. All these steps are implemented to generate a training dataset. To test the results, we consider the testing dataset such as face, fingerprint and voice compared with existing training dataset. Testing dataset is also generated by the same process as discussed. Using correlation both the trainee and test datasets are compared. Finally the result shows “user is authenticated as a genuine user” or “user is not genuine”.

2.1. Image pre-processing

The image pre-processing assessment is used to convert the RGB to Gray level shown in Fig. 4, removal noise shown in Fig. 5 and compression processes applied to digital images and result is shown in Fig. 6. In the context of image pre-processing, for example, such kind of assessment is used.
Fig. 2. (a) GMM face. (b) GMM fingerprint. (c) GMM for voice.
to improve the quality of the reconstructed image. In this paper pre-processing image quality assessment metrics are divided into three categories:

1) RGB to Gray level quality, where the target image is converted to tiff image format by using rgb2gray() method in MATLAB.
2) Apply median filter for removing noise in images; and
3) Apply lossy compression technique on both face and fingerprint for removing reluctant information in image. In this work, we propose the different standard compression methods like DCT, HAAR and KLT.

After removing the noise from the traits, then apply the compression technique which is widely used to reduce the storage space in the image. In this paper we propose the different compression methods like DCT, HAAR, KLT and PCA.

2.2. Histogram of oriented gradients (HOG)

HOG implementation is based on four steps [4,5]. These include Gradient calculation, Histogram of Gradients, Block normalization and Feature vector. The following steps for calculating the HOG descriptor for a 64x128 image are listed in Fig. 7(a) and (b).

Step 1: Gradient calculation: To calculate a HOG descriptor, first need to calculate the x horizontal and y vertical gradients, after calculation of all gradients, then calculate the histogram of gradients. This is easily achieved by filtering the image with the following kernels masks are 1x3 and 3x1

Step 2: Calculate histogram of gradients: In this step, the image is divided into 8x8 cells and a histogram of gradients (64 magnitudes and 64 directions i.e. 128 numbers) is calculated for every 8x8 cells. Histogram of these gradients will provide a more useful and compact representation and also very less noise. Convert these 128 numbers into a 9-bin histogram which can be stored as an array of 9 numbers.

Step 3: Block normalization: In the previous step, we created a histogram based on the gradient of the image. Gradients of an image are sensitive to overall lighting. If you make the image darker by dividing all pixel values by 2, the gradient magnitude will change by half, and therefore the histogram values will change by half. Ideally, we want our descriptor to be independent of lighting variations. In other words, we would like to “normalize” the histogram so they are not affected by lighting variations.

Step 4: Calculate the HOG feature vector: To calculate the final feature vector for the entire image patch, the 36x1 vectors are concatenated into one huge or massive vector. Size of the vector is calculated using step 4.1 and 4.2.

Step 4.1: In this step finds the number of block positions, for example the 16x16 blocks there are 7 horizontal and 15 vertical positions making a total of 7x15 = 105 positions.

Step 4.2: Each one (16x16 block) is represented by a 36x1 vector. Finally, combine all the blocks into one huge vector is a 36x105 = 3780 dimensional vector.

Fig. 7. (a) HOG features of face, (b) HOG features of fingerprint.
2.3. Gaussian mixture model (GMM)

Gaussian is typical symmetric bell shaped curve that is exactly similar parts facing each other or around an axis. Mixture model is probabilistic which assumes the underlying data to fit in to a mixture distribution. GMM is parametric representation of PDF based on sum of weighted multi-variant Gaussian distributions. This model commonly used in continuous measurements or features in biometric traits. GMM used in classification, signal processing, speaker recognition, Language identification and so on. In this paper, we propose Gaussian Mixture model is considered since most of the traits like facial, fingerprint and voice templates exhibit a pattern, which in nature resembles a normal distribution [6,7]. The following mathematical notation of Gaussian mixture model is (shown in Eq. (2)).

\[ P(x) = W_1P_1(x) + W_2P_2(x) + W_3P_3(x) + \ldots + W_nP_n(x) \]  

(2)

Where \( P(x) \) is mixture component, \( W_1, W_2, W_3 \ldots W_n \) is mixer weight or coefficient and \( P_i(x) \) is density function where \( i = 1, 2, 3 \ldots n \).

The most common distribution is the Gaussian (Normal) density function in which each of the components are the Gaussian distributions, each one with their own mean and variance parameters (shown in Eq. (3)).

\[ P(x) = W_1N(x|\mu_1, \Sigma_1) + W_1N(x|\mu_2, \Sigma_2) + W_1N(x|\mu_3, \Sigma_3) + \ldots + W_1N(x|\mu_n, \Sigma_n) \]  

(3)

where \( \mu_1, \mu_2, \mu_3, \ldots \mu_n \) are means, \( \Sigma_1, \Sigma_2, \Sigma_3, \ldots \Sigma_n \) are covariance matrix of individual components (PDF) and \( P_i(x) \) is result of density functions. The function of the Gaussian distribution is (shown in Eq. (4)):

\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]  

(4)

\( \mu \) is the mean or expectation of the distribution, \( \sigma \) is the standard deviation and \( \sigma^2 \) is \( \pi = 3.14151 \) variance and \( e = 2.1728 \).

The formula of Z score is (shown in Eq. (5)):

\[ Z = \frac{x-\mu}{\sigma} \]  

(5)

\( Z \) is the "z-score" that is (Standard Score) and \( x \) is the value to be standardized. A Gaussian mixture model is a weighted sum of \( M \) component Gaussian densities as given by the Eq. (6),

\[ p(x|\lambda) = \sum_{i=1}^{M} w_i g\left( \frac{x}{\mu_i}, \Sigma_i \right) \]  

(6)

where \( x \) is a D-dimensional continuous valued data vector (i.e. measurements or features), \( w_i , i = 1 \ldots M \), are the mixture weights, and \( g\left( x|\mu_i, \Sigma_i \right) \), \( i = 1 \ldots M \), are the component Gaussian densities. Each component density is a D-variant Gaussian function of the form is (shown in Eq. (7)).

\[ g\left( x|\mu_i, \Sigma_i \right) = \frac{1}{(2\pi)^{\frac{D}{2}}}\left| \Sigma_i \right|^\frac{1}{2} \exp \left\{ -\frac{1}{2} \left( x - \mu_i \right)' \Sigma_i^{-1} \left( x - \mu_i \right) \right\} \]  

(7)

With mean vector \( \mu_i \) and covariance matrix \( \Sigma_i \), the mixture weights satisfy the constraint that the \( \sum_{i=1}^{M} w_i = 1 \). The complete Gaussian mixture model is parameterized by the mean vectors (\( \mu_i \)), covariance matrix (\( \Sigma_i \)) and mixture weights (\( w_i \)) from all component densities.

2.4. Fourier transforms representation for voice signal

In this paper, develop a speaker recognition technology is used to find the dissimilarity among speakers and speech signals [8]. Generally, voice have individual trait for every person, therefore to recognize the person by their voice is more helpful technique in recent trends. This is achieved by Fast Fourier transform. Spectrum analysis of a ‘speech signal’ is the method of shaping the ‘frequency domain representation’ of a time domain signal and this method usually known as Fourier transform.
The Discrete Fourier Transform (DFT) is used to find out the frequency content of analog signals. And Fast Fourier Transform (FFT) is a competent process for calculating the DFT.

The Fourier transform of the function \( f(t) \) is the function \( F(\omega) \) where

\[
F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt
\]

and the inverse Fourier transform is

\[
f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega
\]

If \( f(t) \) is considered as a signal then \( F(\omega) \) is the signal's spectrum when a signal is discrete and periodic, we use discrete Fourier transform. Suppose our signal is \( a_n \) for \( n = 0, 1, \ldots, N-1 \). The discrete Fourier transform of \( a \), also known as the spectrum of \( a \) is

\[
A_k = \sum_{n=0}^{N-1} e^{-j\frac{2\pi k n}{N}} a_n
\]

This can be written

\[
A_k = \sum_{n=0}^{N-1} W_{N}^{kn} a_n
\]

\[
W_N = e^{j\frac{2\pi}{N}}
\]

and \( W_k^N \) for \( k = 0, \ldots, N-1 \) are called \( N \)th roots of unity. The sequence \( a_n \) is the inverse discrete Fourier transform of the sequence \( A_k \).

The formula for inverse DFT is

\[
a_n = \frac{1}{N} \sum_{n=0}^{N-1} W_{N}^{-kn} A_k
\]

The FFT is a fast algorithm for computing the DFT. To compute the DFT of an \( N \)-point sequence would take \( O(N^2) \) multiplies and adds. The FFT algorithm computes the DFT using \( O(N\log N) \) multiplies and adds.

2.5. Fusion

Fusion means combining different images into a single image with more informative than any of the individual input images and also suitable for both visual perception and further computer processing. In this paper, we combine three feature values such as face, fingerprint and voice. The solution to multimodal biometrics is the fusion of the different biometric data after feature extraction. Fusion can be implemented in three types like score level fusion, feature level fusion and decision level fusion [9]. In this paper we propose score level fusion is considered to fuse the biometric traits. Fusion of three traits results are presented in Appendix III from Fig. 8(a)–(h).

2.5.1. Score level fusion

Score level Fusion is used for finding the matching scores between different biometric traits [10]. The scores in biometric system may be similar or dissimilar. In score level fusion, the image is reduced into single piece as a match score or similarity score by a classifier and that classifier trains and test the input data. Trained and tested data are compared to find the required biometric trait. Let us consider \( (X_{ij}, Y_{ij}), (X_{ik}, Y_{ik}) \) and \( (X_{il}, Y_{il}) \) as pixels of three different images, where \( i \) indicate position of pixel and \( j, k, l \) indicates image number. Initially we compare \( (X_{ij}, Y_{ij}) \) and \( (X_{ik}, Y_{ik}) \) image vectors and a new fused image \( (X_{im}, Y_{im}) \) is formed based on the following condition. If \( X_{ij} > X_{ik} \) then \( X_{im} = X_{ij} \) else \( X_{im} = X_{ik} \).

If \( Y_{ij} = Y_{ik} \) then \( Y_{im} = Y_{ij} \) else \( Y_{im} = Y_{ik} \). In this case the maximum value among the pixels is considered as score. Secondly we compare \( (X_{im}, Y_{im}) \) and \( (X_{il}, Y_{il}) \) image vectors, a new fused image \( (X_n, Y_n) \) is formed based on the following condition. If \( X_{im} = X_n \) then \( X_n = X_{im} \) else \( X_n = X_{il} \) If \( Y_{im} > Y_{il} \) then \( Y_n = Y_{im} \) else \( Y_n = Y_{il} \).
Transparency document. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dib.2018.03.115.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.03.115.

References

[1] M.S.P. Babu, B.R. Naidu, A novel framework for JPEG image compression using baseline coding with parallel process, in: Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research, pp. 1–7, 2014.
[2] M.S.P. Babu, B.R. Naidu, Development of a biometric authentication system based on HAAR transformation and score level fusion, in: Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science, pp. 1092–1099, 2016.
[3] R.N. Balaka, P.B. Surendra, A novel biometric authentication system with score level fusion, Ann. Data Sci. (2017) 1–22.
[4] S.D. H.Hara, Gopala, K.M. Pillutla, Improved face recognition rate Using HOG features and SVM classifier, IOSR J. Electron. Commun. Eng. (IOSR-JECE) 11 (4) (2016) 34–44.
[5] (http://www.learnopencv.com/histogram-of-oriented-gradients).
[6] R.D. Kumar, A.B. Ganesh, S.S. Kala, Speaker identification system using Gaussian Mixture Model and Support Vector Machines (GMM-SVM) under noisy conditions, Indian J. Sci. Technol. 9 (19) (2016) 1–6.
[7] G. Michael, S. Martin, F. Schwenker, Ensemble Gaussian mixture models for probability density estimation, Comput. Stat. 28 (2013) 127–138.
[8] Md.S. Ali, Md.S. Islam, Md.A. Hossain, Gender recognition system using speech signal, Int. J. Comput. Sci. Eng. Inf. Technol. 2 (1) (2012) 1–9.
[9] M.Bagher A. Haghighat, Ali Aghagolzadeh, H. Seyedarabi, A non-reference image fusion metric based on mutual information of image features, Comput. Electr. Eng. 37 (2011) 744–756.
[10] G.L. Marcialis, F. Roli, L. Didaci, Multimodal fingerprint verification by score-level fusion: an experimental investigation, J. Intell. Fuzzy Syst. 24 (2013) 51–60.