International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-8 Issue-6C, April 2019

Improved Facial Recognition based Authentication approach to Secure Big Data

Shekhar Singh, Mayank Singh, Sandhya Tarar

Abstract*: When deploying biometric identification techniques over the massive data available on web for user authentication purposes, maintaining quality, security and integrity of confidential data are imperative. It is required to make sure the data is captured and stored over a trusted server and is readily available for authentication/ user identification without any interference. In this paper, facial recognition is used as a measure of biometric authentication to address the security issues in Big data. Discrete Wavelet Transform (DWT) is applied to normalize and de-noise the input image, in order to eradicate the unwanted variations preserved while storing the biometric data using traditional methods such as Principle Component Analysis (PCA). Following this, Gabor Filter bank is used to extract the facial features. Further, Expansive Discrete wavelet Transform (EDWT) is used to linearize the dimensional sub-space, using its high expansiveness to curb the number of features extracted from the facial data. The approach uses the spatial orientation of the processed image's high-frequency textural features to improve the accuracy of the trained data for overcoming the shortcomings which results in a 74% efficient algorithm which viably and feasibly achieves the objective of minimizing the expanse of features extracted.

Index Terms*: Biometric Identification, Big Data, Discrete Wavelet Transform (DWT), Facial Recognition.

I. INTRODUCTION

User authentication plays a key role in applications pertaining to digital security. Validating an individual's identity is of integral importance while accessing any confidential or personal data over the web. There are many techniques which help in order to attain this objective viz., password protection, pattern locking, pin protection, digital signatures, etc. Although, these means are quite useful in providing secure access to the user, these methods are still tangible to larceny thus, resulting in identity theft.

Passwords or pins are susceptible to being hacked and therefore require regular updating which still doesn't guarantee a sure shot protective measure against data leakage or data theft. Biometrics, on the other hand, due to its property of uniqueness and universality, provides a secure measure of data protection and therefore, successfully prevents from data leakage.

Authentication based on biometrics provides a convenient way to safeguard data as a user's biometric cannot be stolen, forgotten, altered, or replicated. No two individuals possess the same metrics hence, rendering it unique and universal. Different types of biometrics (palm print, fingerprint, iris, face, retina, voice, gait, signature, etc.) are present to suit various types of applications. The variety of biometrics can be used in accordance with the aim of the application. The choice of biometrics is dependent on the type of application it will be used for; although, there is no restriction over the use of any type of biometric traits.

Providing security benefits across the spectrum, from technical developers to end users, biometrics has come a long way today and is used ona large scale.A number of biometric technologies are available today, such as, the employee attendance system, passenger identity verification on airports and across LOCs for border control, corpse identification in forensics, patient verification at hospitals, citizen user identification, maintaining criminal track record, surveillance, national identification cards, voter ids, etc. Its accuracy and ease of use can render biometric identification as a viable measure of security detail. Although, time and environmental conditions affect the accuracy of biometric datain multiple form as noise or ambient illumination changes.

Technically, there exists two ways in which biometric identification can be obtained wherein, one pertains to preservingprivacy during query execution – the database containing the biometric data is stored unencrypted on a trustable server(Schneider &Wehrenberg, 2009) whereas, in the other one the biometric database is stored encrypted on an untrusted server (Chabanne&Kindarji, 2009).

In the literature, the proposed method CloudId (M. Haghighat, S. Zonouz, M. Abdel-Mottaleb, 2015),over the encrypted biometrics, a query is applied which results in a binary result, match or no match.This method is based on dimensionality reduction used for reducing the feature vector's length.In this, facial data is used as an example. The face image is captured by a camera at the time of query and is fed to a feature extractor. The number of features extracted is quite high. More the number of these extracted features,lower the system's performance.

Extraction of featuresin facial recognition is necessary for storing the biometric data in a biometric database such that at the time of matching (authentication), a result can be obtained as match or no match, effectively, based on the uniqueness of these features.

Revised Manuscript Received on December 22, 2018.

Shekhar Singh, CSE, ShriVenkateshwara University, Gajraula, Uttar Pradesh 244236, India

Mayank Singh, EECE, University of KwaZulu-Natal, Durban 4041, South Africa

Sandhya Tarar, CSE, School of Information & Communication Technology, Gautam Buddha University, Greater Noida, UP - 201312, India

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Feature extraction is crucial to obtain the essence based on which the matching formula can be derived at. These features are unique to an individual and thus can't be fiddled with.
On the other hand, the decades old method of dimensionality reduction, Principle Component Analysis (PCA) based on distance matching of Eigen values of trained data is considered most popular in real world applications.

Although, it preserves unwanted variations in the data due to varying facial expressions and surrounding illumination changes (Hespanha&Kriegman, 1997). This method is widely utilised in face recognition techniques as it finely reduces the length of the feature vector thus resulting in dimensionality reduction of the trained data although, it does have the above mentioned shortcomings.
Challenges like huge size with varying structures in real time processing regarding security of Big data can be effectively addressed by Biometric modalities. Need of hour is to ensure Big data security without affecting its quality. The issues in storing massive amount of data and big data implementation is storing and updating and analyzing such a huge data and preventing the unauthorized access of data to maintain the reliability in digital security. Biometric provides a strong security measure in order to ensure Big data security.

Therefore, this paper suggests a Hybrid methodology to overcome the major setback encountered while storing features distilled from the input face image wherein, both the Discrete Wavelet Transform and Gabor Filter bank are used as the directional transforms and therefore for deriving statistical features.

II. SYSTEM DESIGN AND MODELING:

A suitable feature extraction and dimensionality reduction technique is suggested, which can be used during biometric identification in order to provide successful user recognition in real time applications. As stated above, feature extraction is integral to any biometric identification system. Even so, the quantity of extracted featuresis highly dependent on the image quality as well. Lower quality results in reduced viability of the process whereas, higher quality results in a large number of features extracted thus, again hindering the system's performance.
In PCA, directions with largest variance are assumed to be most important. Commonly used to separate data points, PCA basically analyses the visual cluster. The data should be linear in order for PCA to work. For non-linear datasets, kernel PCA is used which requires a higher dimensional space, as seen in figure 1. The original concept on which PCA acts is by adding additional dimensions. Large dimensions increases the feature space as well.
The following paper suggests an approach which applies wavelet transform to work upon the minimal subspace, accessing the major sub-bands only, followed by Gabor filters application to extract the features and minimize their number for dimensionality reduction and increased system performance.

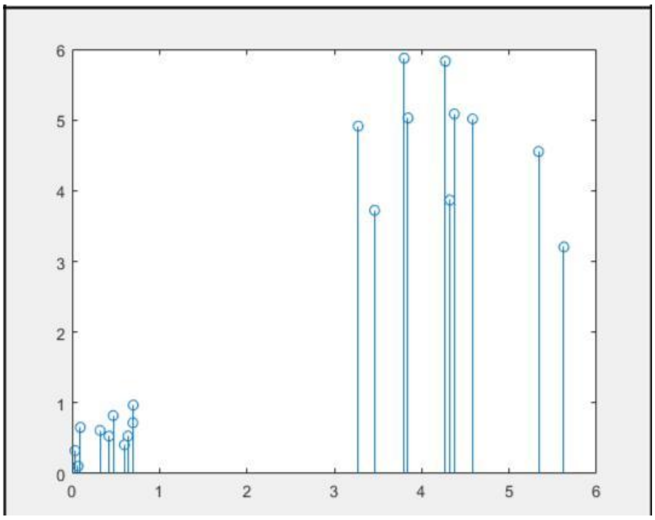

FIGURE 1. PCA DIVISION OF DIMENSIONS

Verifying any individual uniquely by evaluating one of their biological traits is referred to as biometric identification. Face recognition has always been a popular choice for biometric authentication. It is not only popular due to its ease with image processing but also due to the use of PCA (Principle Component Analysis), rendering it popular as a real world application. PCA is an ideal method for recognizing statistical patterns in data. The method can be easily applied to see if it is working, without needing to know how the process is working. The only disadvantage is that it preserves unwanted variations at the time of data capture. This is where the proposed approach throws some light upon the issue at hand. DWT powerfully provides sparse and efficient representations of an image. It is essentially a compression tool for images though directional analysis is not performed at arbitrary directions, unlike Gabor wavelets. Whereas, Gabor function processes images with respect to preferred orientations at arbitrary spatial frequencies. Further, it provides non-redundant information and offers directional selectivity which is higher. Thus, the sequential combination of the wavelet transform and Gabor filter leads to improved extraction of features from biometric images and their improved classification, rather than using any one of these alone. In this hybrid approach, DWT works as a high-frequency filter and extracts abrupt fluctuations in the texture of the image, and it also acts as an image compressor to reduce image dimensions. Next, Gabor filter bank is used for extracting the image features.
The methodology proposed herein uses both DWT and Gabor filters. Methodical representations of the image are provided by DWT and these images are then processed using Gabor filters to preferred orientation.

DWT works both as a frequency filter for extracting deviations in texture of images and also works as an image compressing engine for reducing dimensionality of images.

The two-dimensional DWT performs sub-band coding of an image. The LH, HL, and HH sub-bands first represent the image and encode its details in three directions and the LL sub-band provides its approximation. The continuous wavelet transform is described by:

$$\int_{-\infty}^{\infty} \psi(t)dt = \psi(0) = 0. \quad (1)$$

$$CWT_f(a,b) = \frac{1}{\sqrt{a}} \int_R \psi\left(\frac{t-b}{a}\right) f(t)dt. \quad (2)$$

whereφa,b(.) is the wavelet function, ais the scale, and b is the translation parameters. For an image it is applied to the row variable first followed by the column variable of the result obtained. At each step, there are two sub-images with the processed row or column having half the number of pixels. At the end, an M x N image is divided into 4 sub-images, each having M/2 x N/2 resolution with a preserved scale.

From the 1st level DWT, the 2nd level transform can be computed effectively. Through sampling *a* and *b* such that *W (a,b)* becomes that of a sequence, the transform can be obtained. In dyadic sampling, *a* and *b* corresponds to the power of 2 and multiples thereof respectively, and the wavelet coefficients are given as

= = (2⁻ , 2⁻)

Where j is the factor of discrete scale factor and k is the factor of discrete translation. The initial variable is replaced by 2− and b is replaced by2− . Decomposition through DWT is transformed into a filtering operation via sequence of filters that are high-pass and low-pass, as follows: where, coefficients is the approximation component and specifies the details component which are provided by the g(n) and h(n) impulse responses, respectively. It is then down-sampled by a factor of 2. Using row and column decompositions, 1D wavelet decomposition is extended to 2D objects.

This is followed by using the Gabor filters on an image which is decomposed into components having different scales and orientations therefore, capturing the visual properties like spatial localization, frequency and orientation selectivity. It is useful in extracting features from the face region detected. The Gabor function is given by:

Where, is the frequency of the sinusoid, is the orientation of the normal to the parallel stripes of a Gabor function, gives the standard deviation of Gaussian envelope, and is the spatial aspect ratio specifying the ellipticity of Gabor function. 40 Gabor filters in 05 scales and 08 orientations are used here. The feature vector's dimension is given by:

128 x 128 x 40 = 6,55,360,

which is down-sampled by a factor of 04 to obtain 40,650 features.This, in turn, on applying DWT + Gabor, reduces to 10,880 efficiently.

Gabor filters are advantageous due to their property of being invariant to rotation, scale, and translation. Moreover, agility against photometric disturbances (changes in illumination and noisy image) makes it quite effective in terms of reliability.

III. ALGORITHM (A):

The steps involved are as follows:

Step 01: For an image with dimensions M x N, wavelet transform is applied for image decomposition.
Step 02: The image is then decomposes into 4 sub-images of M/2 x N/2 resolution, LL, LH, HL, HH. **Step 03:** The HL sub-band is considered for feature extraction.
Step 04: Gabor filter bank is applied such that,
Step 05: Number of features extracted are *M*x*N*x*40.*
Step 06: This number is down-sampled by a factor of 4.
Step 07: The outcome obtained is the minimal number of features extracted.

In the proposed framework, a 5 times more expansive version of DWT is derived at, with tight frames and vanishing moments (Ivan W. Selesnick, IEEE, 2006). The expansive DWT is a combination of dyadic wavelets having two generators. The first wave's spectrum lies between the second wavelet's spectrums. The second wavelet, along with this, is translated by half instead of whole integers while constructing the frame, leading to an expansive wavelet transform which has intermediate scales and is shift – invariant.

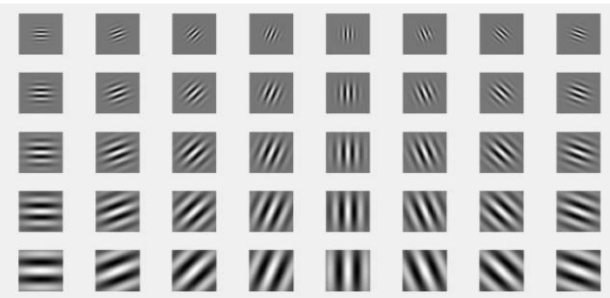

Figure 2. Gabor wavelets with 5 scales and 8 orientations

where (n), n ϵ Z, are the filters. Only the real values of the (n) function are considered. For generating a tight wavelet frame and developing a set of filters, conditions on filters are found. First, a low pass filter h_0 and a factorization of $H_0(z)$ are taken as a product P(Z)Q(Z). From this, h_1 and h_2 filters are determined. The wavelets' vanishing moments is obtained by examining the multiplicity of the zeroes of $H_1(z)$ and $H_2(z)$ transfer functions and factorizing the function $H_0(z)$. On doing this, following reconstruction conditions are obtained:

$$H_0 = \left(\frac{1+z^{-1}}{2}\right)^{K_0} A(z)$$

$$H_1 = z^{-\alpha}\left(\frac{1+z^{-1}}{2}\right)^{K_1}\left(\frac{1-z^{-1}}{2}\right)^{K_2} A(-1/z)(-z)^{-M}$$

$$H_2 = \frac{1}{\sqrt{2}}\left(\frac{1-z^{-1}}{2}\right)^{K_2} C(z)$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

IV. ALGORITHM (B):

The steps involved are as follows:

Step 1: For an image with dimensions M x N, EDWT is applied to decompose the image.
Step 2: Using EDWT, image sub-divides into 8 sub-images of M/2 x N/2 resolution.
Step 3: There are 8 sub - bands at each stage. The relative size of the sub – bands in total is derived as 15(N x N) where LB, BL, BB are of size N x N; LH, and BH are of N x 2N; HL, and HB are of 2N x N; and HH is of 2N x 2N. Here, H stands for 'high pass' and B stands for 'band pass'.
Step 4: Tight wavelet frames are obtained with minimized length for a specific number of vanishing points.
Step 5: The transformed image is then considered for feature extraction.
Step 6: Gabor filter bank is applied such that,
Step 7: Number of features extracted are *M* x *N* x *40.*
Step 8: This number is down-sampled by a factor of 4, *(M x N x 40) / (4 x 4).*
Step 9: The outcome obtained is the minimal number of features extracted.

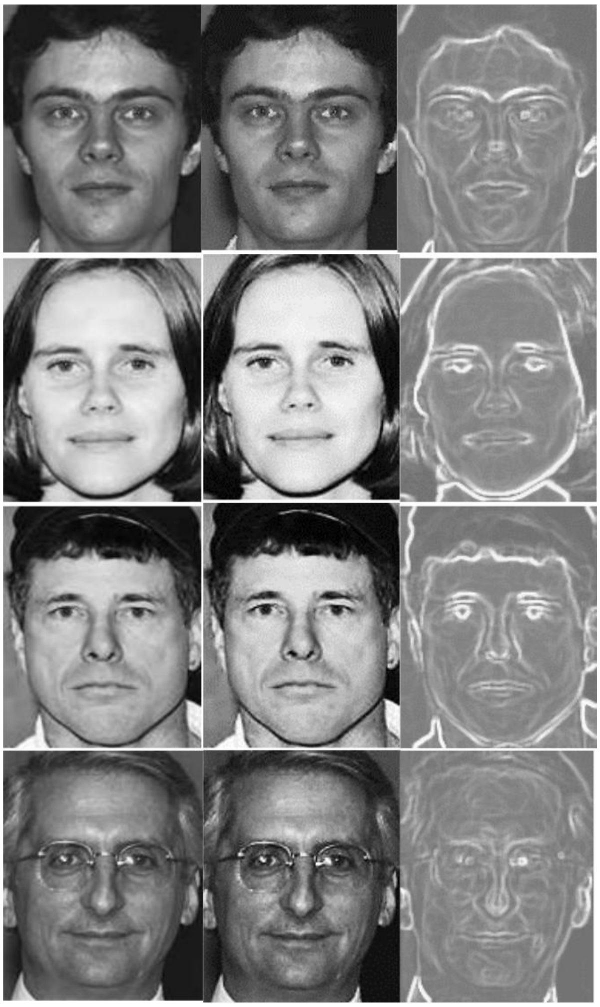

Figure 3. a) Original Image, b) Transformed Image (de-noised image) and c) filtered image

The critical data used by organizations, government and businesses are increasing with every passing day and thus it is keen to provide the security through biometrics, because biometrics provides the highest degree of security, privacy and accuracy due to the inherent biological features used of an individual. Security through face based recognition is extremely effective and very difficult to get compromised and thus an effective method to provide high degree of security to big data.

V. RESULTS AND DISCUSSION

It is observed that as proposed in CloudId (M. Haghighat, S. Zonouz, M. Abdel-Mottaleb, 2015), when an image with smaller dimensions is considered (AT & T's ORL database, image size: 92 x 112 pixels), the number of extracted features reduces significantly but when an image with higher dimensions (AR face database, image size: 576 x 768 pixels) is considered, there is a tremendous increase in the number of features extracted at the final stage, as depicted in figure 04. The drawback here is that while a large number of features would be difficult to accommodate, an image with very low dimensions will render it difficult to extract those features as it might get pixelated, making it difficult to recognize and authenticate the user during the biometric identification process. The variations encountered while extracting the features using CloudId are enlisted in Table 1.

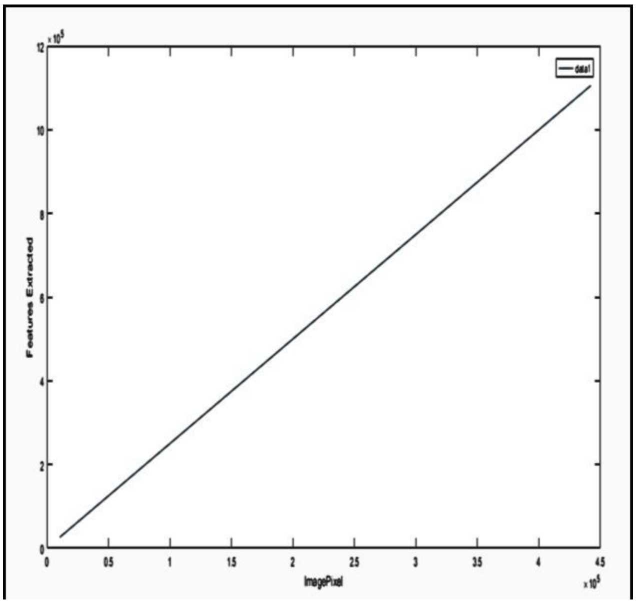

Figure 4: Increasing features extracted w.r.t. image pixels

In the proposed hybrid approach, a single core 64-bit machine is used with 4GB RAM and 1.6 GHz clock speed. The software MATLAB R2016b is used to analyse the results and perform the comparative study between the various techniques put to use herein. The FERET's database is used which contains images of 200 subjects with varying facial expressions, along with various others to perform a comparative study.

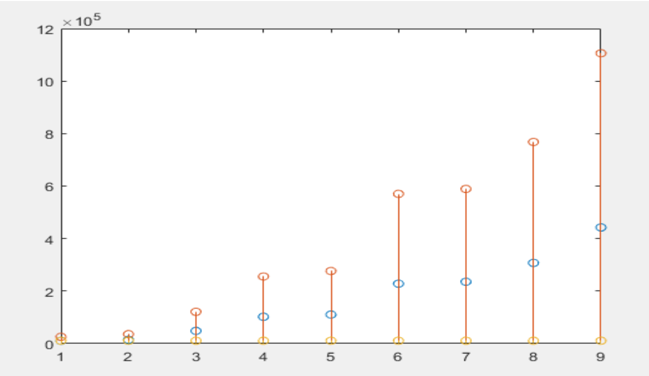

FIGURE 5 (A)

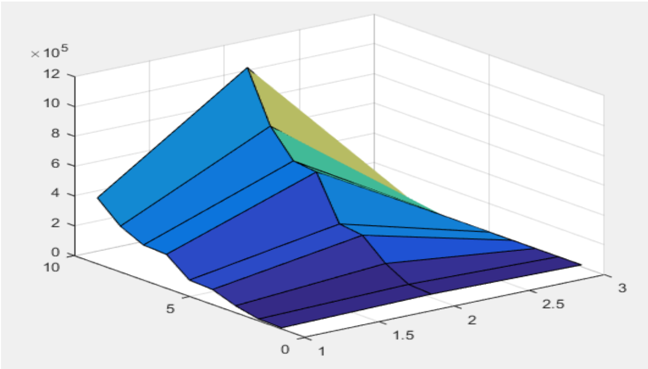

Figure 5 (b)

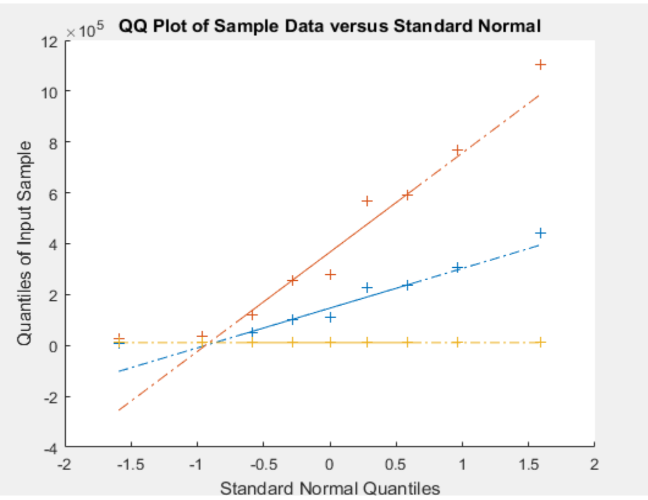

Figure 5 (c)

Fig 5: a) Increase in number of features extracted on increasing the image pixels, b) surf plot of the sampled data, and c) plot of sampled data versus standard normal

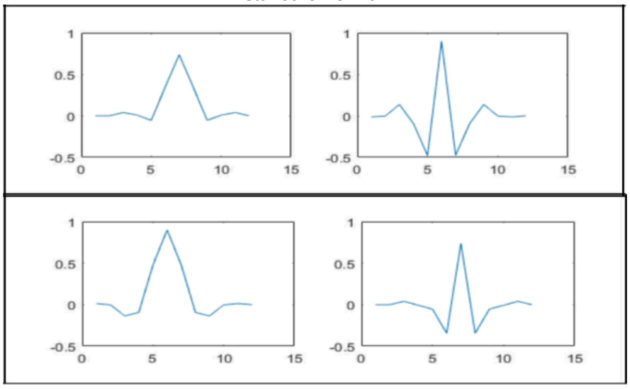

Figure 6. Vanishing moments with the tight frames

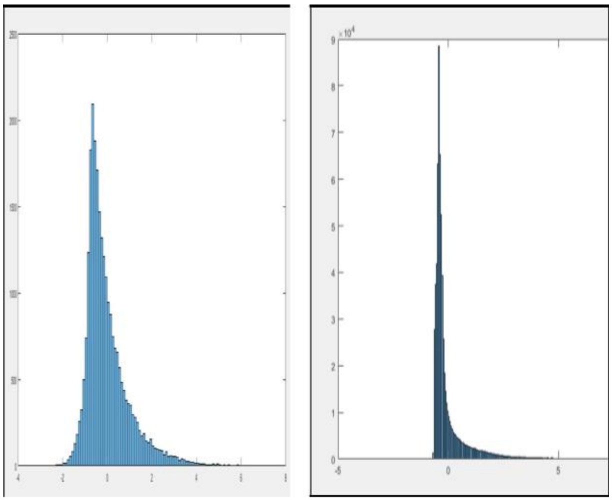

Figure 7. A) Histogram Of Data Sampled In The Cloudid Framework, B) Histogram Of Data Sampled In Proposedbig Dataframework

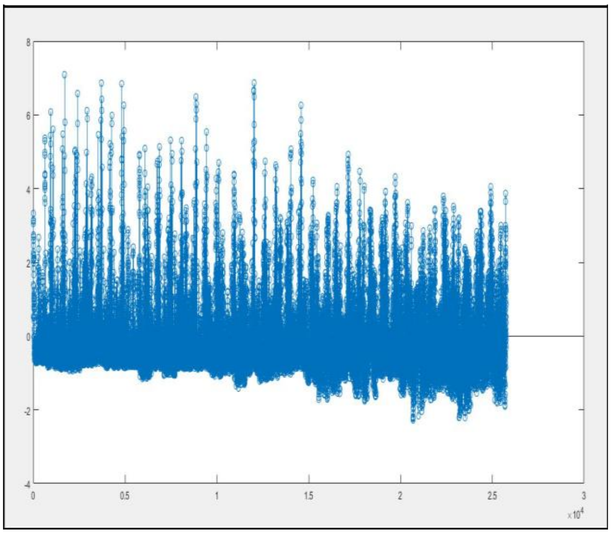

Figure 8. Sampling points observed in the proposed Big data framework

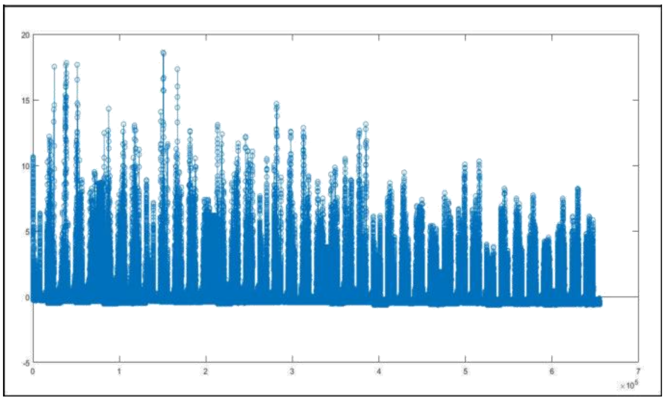

Figure 9. Sampling points observed in the proposed Big data framework

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 1: Number of features extracted on using images of different sizes

S.NO.	DATABASE USED	IMAGE SIZE	FEATURE REDUCTION (GABOR FILTER)
01.	FERET's	120 x 120	36,000
02.	AT & T's ORL	92 x 112	25,760
03.	AR Face Database	576 x 768	1,105,920
04.	Richard's MIT Database	480x640	768,000
05.	Yale Face Database	168x192	80,640
06.	Notingham Scans	438x538	589,110
07.	Nott-Faces Originals	288x384	276,480
08.	Aberdeen	432x528	570,240
09.	Stirling Faces	280x365	255,500
10.	UMIST Face Database	220x220	121,000

Table 1 shows an exponential increase in the number of features on increasing the size of the image, while extracting those features. This may pose deviations in deriving and storing the features for further processing resulting in, increased space and time complexity. The same can be observed through the sampled data plot given in figure 4.

Table 2: Number of Features extracted using a) only Gabor, and b) using DWT + Gabor

S.No.	DATABASE USED	IMAGE SIZE (PIXELS)	FEATURE REDUCTION (GABOR FILTER)	FEATURE REDUCTION (EDWT + GABOR FILTER)
01.	FERET's	120 x 120	36,000	9,150
02.	AT & T's ORL	92 x 112	25,760	9,150
03.	AR Face Database	576 x 768	1,105,920	9.150
04.	Richard's MIT Database	480x640	768,000	9,150
05.	Yale Face Database	168x192	80,640	9,150
06.	Notingham Scans	438x538	589,110	9,150
07.	Nott-Faces Originals	288x384	276,480	9,150
08.	Aberdeen	432x528	570,240	9,150
09.	Stirling Faces	280x365	255,500	9,150
10.	UMIST Face Database	220x220	121,000	9,150

Table 2 depicts results after implementing the proposed approach.Numbers of features extracted are fixated in totality to a constant 9,150 given any size of the user's image, using EDWT and Gabor filter, normalising the input image to zero mean, invariant to scale and rotation and having standard variance. The transformed image is compressed and de-noised, after which effective extraction of features for facial recognition takes place.

Using expansive DWT, which is 5 times more expansive, linear periodicity in sampled data is observed (figure 10) as compared to the one observed while using only Gabor filters (figure 9). Higher variations in irregularities of the intervals are seen in the data sampled as shown in figure 9. Thus, the suggested approach, as shown in figure 10, proves effective in obtaining linearity and accurate results with least photometric disturbances. Moreover, on training the data over the recognizer, correct results are obtained,i.e.,true recognition is derived at.

Therefore, combining EDWT with Gabor filter banks enhances the viability of the process and helps achieve the objective seamlessly.

The aim was to successfully recognize the registered individuals by disregarding unwanted variations in the data while performing the recognition process. For this purpose, we applied a composition of Gabor wavelet and discrete wavelet techniques which posed to be useful in improving the recognition accuracy.

The main objective to control the various types of deviations such as noise and light variations in the data is achieved successfully. Through adopting this measure it is made sure that the unwanted variations preserved during training the data are discarded and the recognition process takes place accurately, avoiding any external interference. The proposed framework can be executed on a system having a single core processor with an x64 or x 86 configurations. The memory utilisation is 699 MB and CPU utilisation is considerably low.

VI. CONCLUSION AND FUTURE WORK:

An approach for efficient dimensionality reduction is proposed for face recognition in order to extract minimal features and remove variations due to photometric disturbances during the recognition process. One of the biggest biometric databases used in Aadhar, a project run by Indian govt. consists of around 1.2 billion people database. This project is used to provide people access to the govt schemes offered for the welfare of needy along with the authentic access to the services and benefits to the right people proposed method can provide improved security on this Big data.

A combination of DWT and Gabor wavelet is used for the same. This approach results in a tremendous decrease in the number of extracted features, viz. 9,150, as compared to traditional methods. First, the 5 times expansive EDWT is applied on the input data for normalisation. It is normalized to standard variance and zero mean. Followed by this Gabor Filters are applied to extract the corresponding features effectively thus, resulting in required reduction in dimensionality.

The method utilises a single core processor with a configuration of x64 with a considerably low memory utilisation of 0.7 GBs, resulting in efficient and feasible performance. In future, other biometrics such as eye or mouth can be used for the same for which further experimentation will be required.

References:

1. Haghighat, Mohammad, SamanZonouz, and Mohamed Abdel-Mottaleb. CloudID: Trustworthy cloud-based and cross-enterprise biometric identification.Expert Systems with Applications. November, 2015. 7905-7916.
2. Selesnick, Ivan W. A higher density discrete wavelet transform.IEEE Transactions on Signal Processing. July, 2006. 3039-3048.
3. Haghighat, Mohammad, Mohamed Abdel-Mottaleb, and WadeeAlhalabi. Fully automatic face normalization and single sample face recognition in unconstrained environments.Expert Systems with Applications. July, 2016. 23-34.
4. Sadeghi, Ahmad-Reza, Thomas Schneider, and ImmoWehrenberg. Efficient privacy-preserving face recognition.International Conference on Information Security and Cryptology. Springer Berlin Heidelberg. December, 2009.
5. Bringer, Julien, HervéChabanne, and Bruno Kindarji. Error-tolerant searchable encryption.International Conference on. IEEE. June, 2010.
6. Lahmiri, Salim, and MounirBoukadoum. Hybrid discrete wavelet transform and Gabor filter banks processing for features extraction from biomedical images.Journal of medical engineering. March, 2013.
7. Belhumeur, Peter N., João P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection.IEEE Transactions on pattern analysis and machine intelligence. June, 1997. 711-720.
8. Jain, Anil K., Arun Ross, and SalilPrabhakar. An introduction to biometric recognition.IEEE Transactions on circuits and systems for video technology. 2004. 4-20.
9. Haghighat, Mohammad, Mohamed Abdel-Mottaleb, and WadeeAlhalabi. Discriminant correlation analysis for feature level fusion with application to multimodal biometrics.2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. July, 2016.
10. Jain, Anil K., and Ajay Kumar. Biometric recognition: an overview.Second Generation Biometrics: The Ethical, Legal and Social Context. Springer Netherlands. May, 2012. 49-79.
11. S. Tarar, and E. Kumar. Fingerprint image enhancement: iterative fast Fourier transform algorithm and performance evaluation. Int. J. Hybrid. Inf. Technology. July, 2013.
12. S. Tarar, and E. Kumar. Fingerprint Mosaicking Algorithm to Improve the Performance of Fingerprint Matching System. Computer Science and Information Technology. June, 2014.
13. S. Tarar, and E. Kumar. Design Paradigm and Risk Assessment of Hybrid Re-engineering with an approach for development of Re-engineering Metrics. International Journal of Software Engineering & Applications. January, 2012.
14. S. Zaidi, S.K. Singh, and S. Tarar.To evaluate the performance of fingerprint enhancement techniques. India Conference (INDICON), 2015 Annual IEEE, 2015.
15. H. Chauhan, and S. Tarar.Image Processing Edge Detection Technique using Iterative Enhancement Wavelet used for Traffic Control Problem. International Journal of Engineering Science. May, 2016.
16. S. Tarar, and C. Verma. Secure Random Sequence based Frequency Hoping Spread Spectrum Audio Watermarking. International Journal of Engineering Science. 2016.
17. S. Tarar, A.P. Singh, and S. Singh.Speech recognition approach: Desktop items activation with comparative analysis.International Conference of Computer Science and Information Technology (ICCSIT). October, 2010.