Volume 5, Issue 4, 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

The Steganography and Biometric Based Online Voting System

Shweta A. Tambe*, Prof. P. S. Topannavar
Department of Electronics and Telecommunication,
JSPM's ICOER, Pune, India

Abstract— Now a days, internet is very important part of our day to day life. To avoid the integrity of the election process, it is necessary to use the online voting system. The online voting system helps to manage the elections easily and securely. In this paper the Steganography based online voting System by using biometric security is present. The steganography and cryptography methods are used to provide a biometric as well as password security to the voters account. The proposed system also detects whether the voter is correct person or not. For that purpose the proposed system uses voters fingerprint image as cover image and embed voter's secret data into the image using steganography. This proposed method produces a stego image which is equal to the original fingerprint image only. On the whole there are changes in the original fingerprint image & stego image but they are not visible by human eye.

Keywords— Biometric, Cover, Fingerprint, Online Voting, Password, Steganography, Security.

I. INTRODUCTION

Nowadays, internet is very important part of our day to day life. Because of internet we have a rapid entrance in the world by means of text, audio, video, images etc. These forms of data illustration can be used for variety of purposes such as message or image sending, for any type of important message in the form of audio or video etc.

But as the internet is a simple way of information exchange some drawbacks are also there associated to the internet is known as cyber crime. One can get the important information from internet and misuse it. To avoid such cases there are different kinds of keeping data safe like Cryptography, Steganography, and Watermarking. All these are the digital forms of information hiding. Digital media is widely used because it is easy to store, transfer and duplicate the data with good quality. As mentioned above Cryptography is a method of storing & transmitting data such that only intended persons will be able to read and develop it. It is a way of encoding the data which can't be decoded easily.

Steganography is an art of hiding secret data into another data so that it can't be detected by anyone. The benefit of steganography as compare to cryptography is that the secret data does not attract concentration to itself. Steganography process is most excellent for the media with large size. Media files such as image, audio, video are the best option to transmit stego images transmission because of their large size. The secret data can be a signature, a company logo, a password or ATM PIN etc.

Digital watermarks are the electronic versions of the watermarking system. Watermarking is a means of hiding data into digital content in order to identify its owner. Watermarking can be used both in transmission and for data usage. Digital watermarks cannot be removed or changed. This feature makes them a very important tool when fighting for the copyright on the Web.

An election is an official process by which person chooses an individual to hold all kind of public issues. The elected person should satisfy all necessary needs of common people so the system of whole country works properly. The main requirements of election system are like authentication, speed, accuracy, and safety. The voting system should be speedy so the valuable time of voters as well as the voting system conductors will be saved. Accuracy means the whole system should be accurate with respect to result. Safety involves the secure environment around the election area so that voters will not be under any force. In online voting system main aim is to concentrate the focus on security of voters account. For any type of voting system following points must be taken into consideration. This can include confusing or misleading voters about how to vote, violation of the secret ballot, ballot stuffing, tampering with voting machines, voter registration fraud, failure to validate voter residency, fraudulent tabulation of results, and use of physical force or verbal intimation at polling places. If online voting system works well then it will be a good progress over the current system.

Rest of the paper is organized as follows. In the next section II shows the different methods used in this paper. Section III shows the development of proposed system. Section IV shows the experimental results and finally section V concludes this paper.

II. METHODS USED

This section describes the proposed methods which are used in this paper.

A. Discrete Wavelet Transform

The fundamental idea of the DWT for a one dimensional signal is as follows. A signal is split into two parts, generally high frequencies and low frequencies. The edge components of the signal are largely restricted in the high frequency part. The low frequency part is split again into two parts of high and low frequency. This process is continued until the signal

has been entirely decomposed or stopped by the application at hand. For compression and watermarking application, usually no more than five decomposition steps are computed. Furthermore, from the DWT coefficients, the original signal can be reconstructed. The reconstruction process is called the inverse DWT (IDWT).

The two-dimensional wavelet transform that we describe can be seen as a one-dimensional wavelet transforms along the x and y axis. Mathematically the wavelet transform is convolution operation, which is equivalent to pass the pixel values of an image through a low pass and high pass filters. DWT involves decomposition of image into frequency channel of constant bandwidth. This causes the similarity of available decomposition at every level.

DWT is implemented as multistage transformation. Level wise decomposition is done in multistage transformation. At level 1: Image is decomposed into four sub bands: LL, LH, HL, and HH where LL denotes the coarse level coefficient which is the low frequency part of the image. LH, HL, and HH denote the finest scale wavelet coefficient. The LL sub band can be decomposed further to obtain higher level of decomposition. This decomposition can continues until the desired level of decomposition is achieved for the application. The watermark can also be embedded in the remaining three sub bands to maintain the quality of image as the LL sub band is more sensitive to human eye. The DWT of a signal x is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response 'g' resulting in a convolution of the two:

$$y[n] = (x * g)[n] \sum_{k=-\infty}^{\infty} x[k]g[n-k] \quad \text{……..……(1)}$$

A signal is decomposed and constructed by DWT, for this it uses different filter banks satisfying the perfect reconstruction (PR) condition. The decomposition of signals is given as below

$$y_{HIGH}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n+1-k] \quad \text{…..……...…….....(2)}$$

$$y_{LOW}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k] \quad \text{……..…….…….(3)}$$

The general discrete wavelet transform (DWT) of filter analysis tree for three levels of independent DWTs are shown in Fig.1.

Discrete wavelet transform is a multi-resolution decomposition of a signal. 1 level DWT involves applying a low pass and a high pass filters along the columns and then the rows, respectively. In two dimensional applications, for each level of decomposition DWT is applied in vertical direction, followed by applying the DWT in the horizontal direction. After the first level of decomposition, there are 4 sub-bands: LL1, LH1, HL1 and HH1.

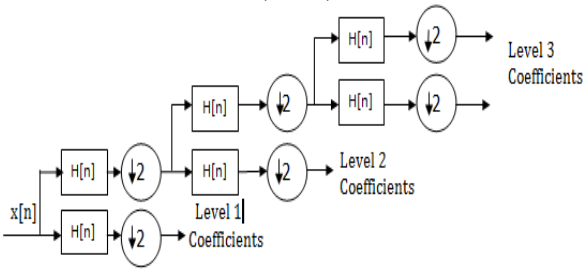

Fig.1. Filter analysis tree

For each successive level of decomposition, the LL sub-band of the previous level is used as the input. Each tile component undergoes three levels of decomposition. This results in 10 sub-bands per component. LH1, HL1, and HH1 contain the highest frequency bands present in the image tile, while LL3 contains the lowest frequency band. The three-level DWT decomposition is depicted in Fig.2.

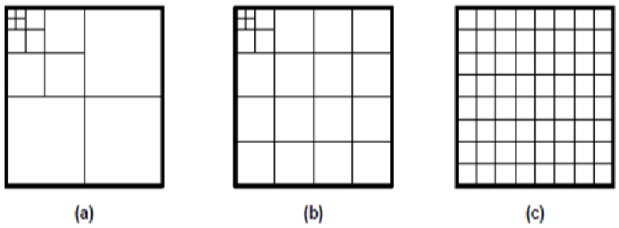

Fig.2. Three Popular Wavelet Decomposition Structures on Image: a) Pyramid, (b) Spacl, (c) Wavelet

B. Haar transform

The Haar transform was proposed in 1910 by a Hungarian mathematician Alfred Haar. The Haar transform is one of the earliest transform functions proposed.

Nowadays, several definitions of the Haar functions and numerous generalities as well as some alterations were published and used. One of the best alterations, which were introduced, is the lifting scheme. We can call it as forward as well as reverse lift.

These transforms have been applied, for instance, to spectral techniques for various valued logic, image coding processes, edge extraction processes, etc. Over the past few years, a variety of influential and sophisticated wavelet–based schemes for image compression, as discussed later, were established and implemented.

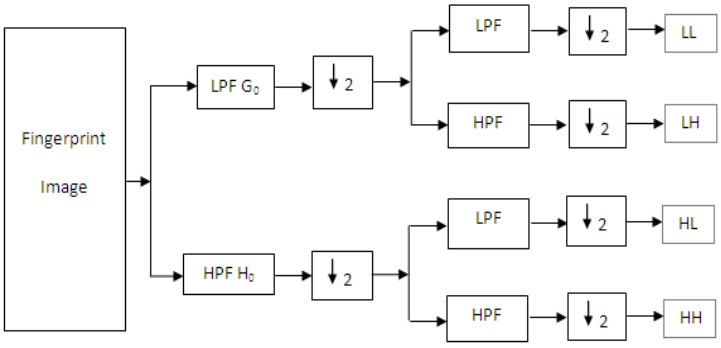

Fig.3. A 1 Level Haar Transform Representation

The Haar transform uses Haar function for its basis. The Haar function is an ortho-normal, rectangular as well as bi-orthogonal. Compared to the Fourier transform basis function which only differs in frequency, the Haar function varies in both scale and position.

The Haar transform serves as a prototype for the wavelet transform, and is closely related to the discrete Haar wavelet transform.

C. Principal component analysis (PCA)

Principal component analysis (PCA) is one of the most precious results from applied linear algebra. PCA is used in abundance in all forms of analysis - from neuroscience to computer graphics. As it is a simple, non-parametric method of extracting related information from puzzling data sets. With minimum extra effort PCA provides a roadmap for how to decrease a complex data set to a lower dimension to expose the sometimes hidden, simplified dynamics that often underlie it. Following are some fundamental properties which we require to study about PCA.

1) Statistics:

The topic of statistics is based on the idea that we have the big set of data, and we want to analyze it in terms of the relations between the individual points in that data set. We are going to look at some of the procedures we can do on a set of data, and what they tell us about the data itself.

2) Mean:

A mean can be defined as the ratio of addition of all the pixel values present in the image to the total number of pixels of image.For example: If an image is having the pixel values such as:

Pixel values = {5, 7, 4, 8, 9, 1, 6, 3, 7}

Then the mean value of that image can be given as

Mean = (5+7+4+8+9+1+6+3+7) / 10

= 5

3) Standard deviation:

The Standard Deviation (SD) of a data set is a measure of how spread out the data is. In English the definition of the Standard Deviation is: "The average distance from the mean of the data set to a point". The method to calculate SD is to calculate the squares of the distance from each data point to the mean of the set, add them all, divided by n-1, and take the positive square root. Where 'S' is the usual symbol for standard deviation of a sample.

$$S = \frac{\sqrt{\sum_{i=1}^{n} (Xi - \bar{x})^2}}{n-1} \quad \text{.....(4)}$$

Where, X is mean of the signal defined by the formula:

$$\bar{x} = \sum_{i=1}^{n} Xi \quad \text{.....(5)}$$

n = number of dimensions.

4) Variance:

's2' is the common symbol for variance of a sample. Variance is a measure of the deviation from the mean for points in one dimension e.g. heights. Both standard deviation & variance measurements are procedures of the spread of the data. Standard deviation is the most familiar measure, but variance is also used. The reason why we have introduced variance in addition to standard deviation is to offer a platform for covariance.

Variance is another measure of the spread of data in a data set identical to the standard deviation, given by the formula:

$$S^2 = \frac{\sum_{i=1}^{n} (Xi - \bar{x})^2}{(n-1)} \quad \text{.....(3)}$$

5) Covariance:

Variance and Covariance are the measure of the "spread" of a set of points around their center of mass (that is mean). Covariance is a measure of how much each of the dimensions varies from the mean with respect to each other. Covariance is measured between two dimensions to see if there is a relationship between the two dimensions e.g. number of hours studied & marks obtained.

The covariance between one dimension and itself is the variance. Standard deviation & variance these are purely 1 dimensional. Covariance is always measured between 2 dimensions. If you calculate the covariance between one dimension and itself, you get the variance. So, if you had a 3-dimensional data set (a, b, c), then you could measure the covariance between the 'a' and 'b' dimensions, the 'b' and 'c' dimensions, and the 'a' and 'c' dimensions. Measuring the covariance between 'a' and 'a' or 'b' and 'b' or 'c' and 'c' would give us the variance of the 'a', 'b' and 'c' dimensions respectively.

The formula for covariance is:

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^{n} (Xi-\bar{x})(Yi-\bar{y})}{(n-1)} \quad \text{……..…….(6)}$$

6) Covariance matrix:

A positive way to get all the probable covariance values between all the different dimensions is to compute them all and put them in a matrix. If you have an 'm'-dimensional data set, then the matrix has 'm' rows and 'm' columns and each entry in the matrix is the outcome of computing the covariance between two separate dimensions. For example: the entry on row 3, column 1, is the covariance value calculated between the 3rd dimension and the 1st dimension.

A useful way to get all the possible covariance values between all the different dimensions is to calculate them all and put them in a matrix. So, the definition for the covariance matrix for a set of data with n dimensions is:

$$C^{n\,xn} = (Ci,j,\ Ci,j = \text{cov}(\text{Dimi}, \text{Dimj})) \quad \text{…….…….(7)}$$

Where, Cnxn is a matrix with n rows and n columns and Dimx is the xth dimension.

As an example, following is the covariance matrix for an imaginary 3 dimensional data set, using the usual dimensions x, y and z.

Then, the covariance matrix has 3 rows and 3 columns, and the values are this:

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix} \quad \text{…….. (8)}$$

7) Eigenvalues and Eigenvectors:

Eigenvectors can only be found for square matrices. Not each square matrix has eigenvectors. All the eigenvectors of a matrix are perpendicular, that means at right angles to each other, no issue how many dimensions you have. Another word for perpendicular is orthogonal. This is essential because it means that we can express the data in terms of perpendicular eigenvectors, instead of expressing them in terms of the 'x' and 'y' axes. We will use this in PCA. Eigenvectors and Eigen values always come in pairs.

III. DEVELOPMENT OF SYSTEM

This section describes the proposed Online Voting System scheme. Fig.4. shows a detailed block diagram of the proposed system.

The proposed block diagram basically consists of three steps namely, database creation, data embedding, online voting and recognition.

a) Database Creation

For database creation a voter committee should be appointed. Committee member job is to collect the data from each person.

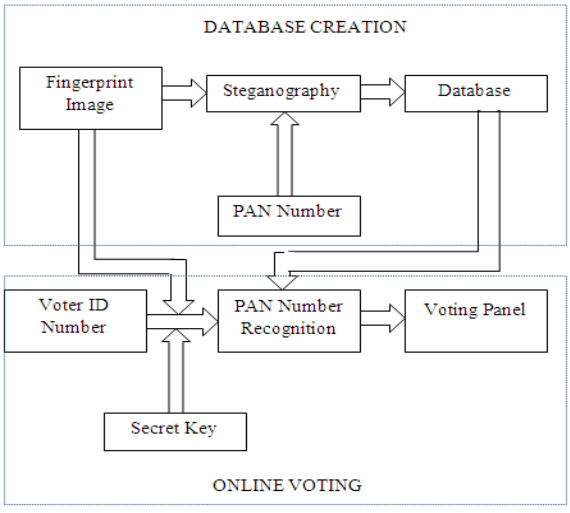

Fig.4.Block diagram of proposed system

Every voter should have the voter account identification number to maintain the account, PAN number for voter authentication & a secret key as a password or cross verification of the database.

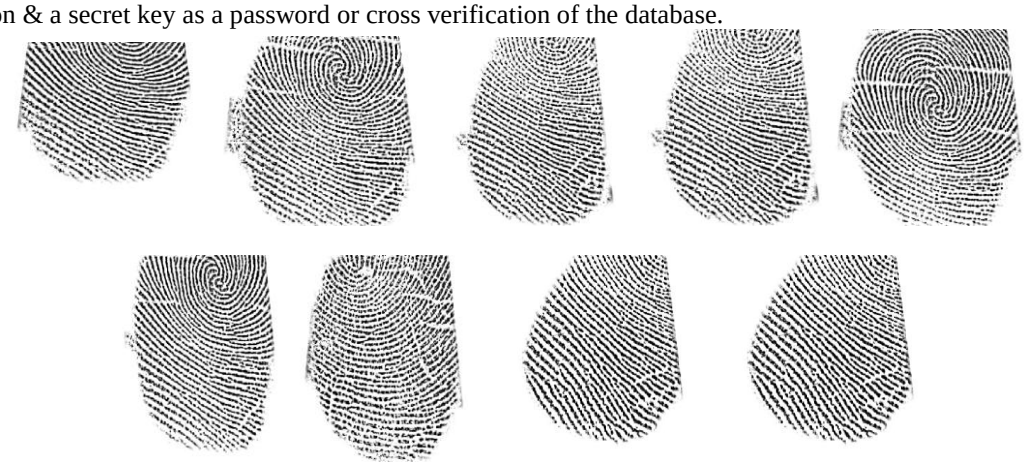

Fig.5. Database of Fingerprint Images

As shown in the fig4. the fingerprint image block takes the fingerprint image of voter as an input. PAN number block accepts the personal identification number as an input. Steganography block performs steganography on the personal authentication number. Thus a stego image is saved as database image. Two different aspects in data hiding systems are of great concern. They are capacity and security. Capacity means the amount of data that can be hidden in the cover object; security means an eavesdropper's failure to detect hidden information.

The fingerprint image should be plain as shown in fig.5. which will act as cover image after data hiding. So the cover image for each voter is its own fingerprint image only. Prior to the least significant bit insertion, system uses discrete wavelet transform. In discrete wavelet transform the fingerprint image is transformed from spatial domain to frequency domain.

The PAN number is embedded into fingerprint image with the help of least significant bit insertion method & thus steganography is performed. The combination of fingerprint image & PAN number is nothing but a stego image is produced with the help of LSB insertion technique. It is assumed that, embedding message in this way is not going to destroy the information of the original image to a great extent. A secret key is separately provided to each voter along with the PAN number. Voter should remember that, in order to use it at the time of online voting. After completion of all the steps thus the database creation of the voter is complete. This task will be performed for each person.

b) Data Embedding

In case of embedding the secret data it is essential to take the input fingerprint image. The input fingerprint image acts as the cover image for data hiding. A haar transform is applied to the columns of fingerprint image. In haar transform a discrete wavelet transform is carried out. In discrete wavelet transform the input fingerprint image is first transformed column wise into low and high level of frequency components.

Then again a discrete wavelet transform is applied to the rows of the same fingerprint image. In this case the rows are transformed into two sub bands which are low and high level frequency components. Thus finally the whole image gets converted into four sub bands. Those bands are LL, LH, HL, and HH.

The main job is to hide the 16 digit PAN number with the help of least significant bit insertion technique. In the above section we have studied in detail about the least significant bit insertion. In this technique we can change one or two or three or all of the four least significant bits.

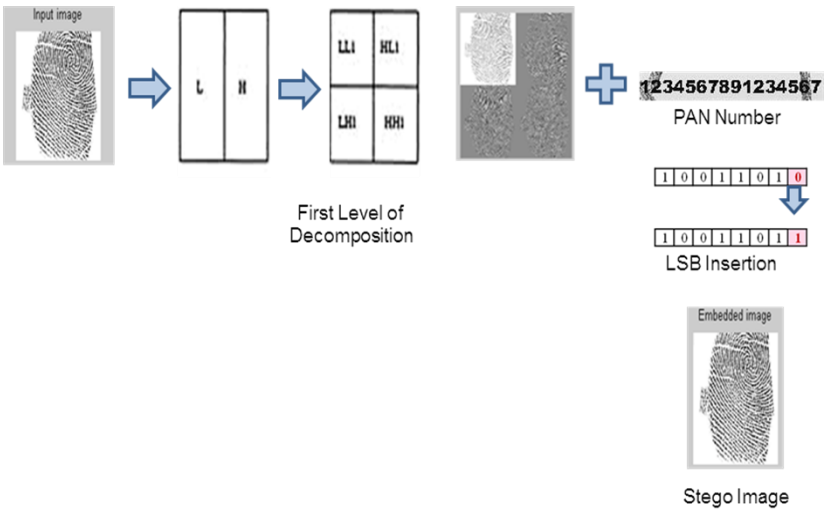

Fig.6. Secret Data Embedding

In this paper we are replacing all the four least significant bits. We need two pixels to hide one secret bit. For example if we want to hide a digit 1, then the ASCII value of 1 is 49. We have to convert 49 into its equivalent binary format. So we can represent 49 in binary format as 0011 0001. Hence we have to embed these binary bits into the fingerprint image.

c) Online voting

Online voting is the main process of the whole concept. All the safety precautions that we have carried out must get satisfied at the time of online voting. In the present voting system the voter is first asked for the identity proof such as driving licence or PAN card or adhar card etc. Then they will cross check the voter's photo with their database photo. Finally the person will allow for the voting.

At the time of online voting a voter is first asked for voter's account authentication number. With the help of authentication number the voter's election account will be opened. Then voter is asked to give the fingerprint image as input image to the system. Voter can give the fingerprint image with the help of fingerprint scanner machine. Finally voter is asked to give the secret key. A secret key can be entered with the help of any kind of input device like keyboard.

If the secret key is correct then the PAN number recognition is carried out with the help of (DWT) reverse discrete wavelet transform. Reverse discrete wavelet transform (DWT) is applied to the embedded or stego fingerprint image which is saved as the database image in order to get the embedded PAN number. Then the voter is asked to enter the PAN number. After comparing both the PAN numbers if the match is found then the voter is an authenticate person & can cast a vote.

d) Recognition Process

In fingerprint recognition process the system must identify the fingerprint which a voter has entered at the beginning of database creation process. The result of embedded process is a stego image. Recognition process includes extraction of the PAN number from the stego image. For recognition purpose reverse discrete wavelet transform is applied to embedded fingerprint database image. In this, system first takes the embedded fingerprint image as input.

Then a discrete wavelet transform is applied to it again. In the first step it is applied to the column. Because of which the image is divided into two sub-bands namely 'L' & 'H'.

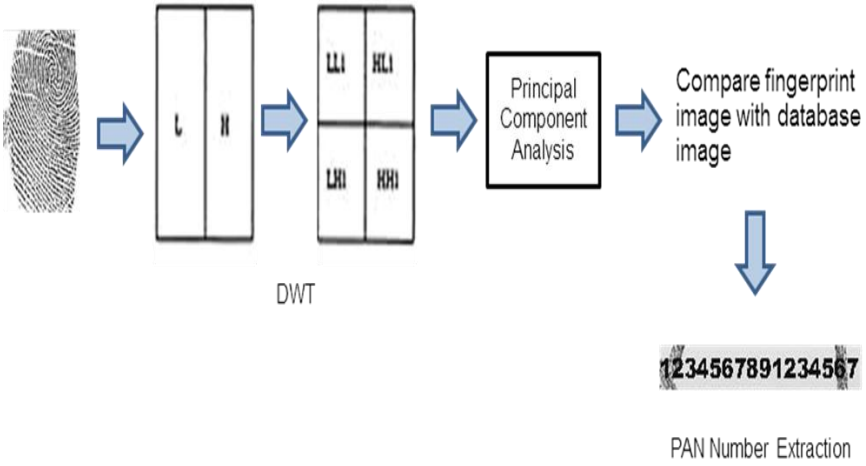

Fig.7. Block diagram of Recognition Process

Afterwards a discrete wavelet transform is applied to the rows. At this stage the image get divided into four sub-bands. Those sub-bands are LL, LH, HL, and HH.Principle Component Analysis is then used for fingerprint recognition. Principle component analysis is a way of identifying the patterns of fingerprint image in order to highlight their similarities & differences. Principle Component Analysis features extraction requires first calculating the mean of image pixel values. It is nothing but the addition of all the pixel values & then dividing it by the total number of pixels.

Then finding the co-variance matrix & eigen values of the image. Then finally comparison is done between input fingerprint image & embedded image to find out the match using euclidean distance. If match is found between database image & test image then the voter is authorized person.

IV. EXPERIMENTAL RESULTS

The detail steps of proposed Online Voting System scheme are described in this section. In this proposed method we will initiate the work with starting the database creation. In order to create the database we accept the fingerprints from different voters as shown in the fig. 8.

First of all take the fingerprint image from voters to create the database. The fingerprint image taken from the voter is shown in fig.8. Then we will perform discrete wavelet transform so that the image will get divided into four frequency bands. After performing DWT the resulting output is shown in the fig.9.

Out of four sub-bands we are hiding the secret message which is PAN number into LL sub-band as shown in the fig.10.

After performing the LSB insertion technique PAN number will be hidden in to the LL sub-band. And the final image produced after data hiding is called as stego image. The stego image is shown in the fig.11.

Fig.8. Input fingerprint image for database creation of voter 123.

Fig. 9. Discrete Wavelet Transform Results.

Hence the database is generated for each voter which includes the fingerprint of voter, PAN number, secret key, & a stego image. This is the prerequisite or basic requirement for online voting system. This complete data will be kept safe & hidden at server side. Now for actual online voting, the voter has to enter the fingerprint image online.

That means from a machine different than the server machine. This machine is called as client machine. So client machine system will ask to enter the fingerprint of voter.

Fig.10. Image with low frequencies.

Fig.11. Stego Image.

Fig.12. PAN number

Now this image is compared with the database input fingerprint image at server side. If the match is found the server will displayed which says that the person is authorized as the fingerprint match is found. Then system will ask voter to enter the secret key for authentication purpose.

After entering a four digit secret key the system will again check the match between the entered secret key & the database secret key. Again if the match is found then the system will proceed further. But if the secret key goes wrong then it will display the following message on the screen.

If the secret key is correct then the system will ask to enter the PAN number. We will see the window to enter the PAN number.

After entering a four 16 digit PAN number the system will again check the match between the entered PAN & the database PAN number. Again if the match is found then the system will proceed further.

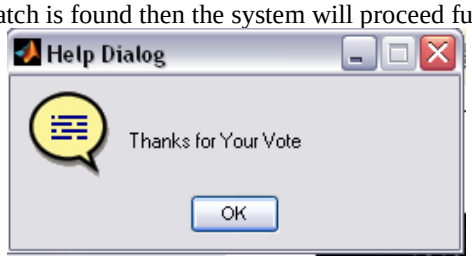

Fig.13. System shows end of voting process.

After putting your vote the voter has to submit it with the help of push button “Submit” provided. Once voter submit the vote system will display a message that thanks for your vote.

V. CONCLUSION

This system uses account identification number to maintain the voters account, fingerprint image as biometric security, PAN number for authentication & secret key for cross verification of the database. Thus the system provides a multilevel security and adequate proof of authenticity which is the advantage over the earlier election system. The entire work also offers a reduced man power as well as time to conduct voting.

The tool used for developing this proposed work is image processing and wavelet toolbox of MATLAB for "reading" the images and for performing the DWT operations.

In this project we have established a multilevel security for online voting system. But as we know where the word online comes into picture there are chances of online attacks also. It is similar to the online banking system.

So the future work is focused on establishing robustness towards attacks. Those attacks are nothing but online attacks. Robustness is an essential parameter used to express the strength of the steganography process towards attacks. If the external attacks which tend to degrade the hidden information, it shows that the embedding algorithm is not robust

adequate to withstand the attacks. The strength of the embedding process is determined by revealing the stego image to various attacks. Different kinds of attacks such as addition of noise like salt & paper noise & Gaussian noise, compressing the image, rotation of image, & scaling the image, cropping the image etc.

There are several procedures to determine the strength of the embedding algorithm. It can be expressed using PSNR, Correlation Coefficient etc. In our project we have worked with peak signal to noise ratio % percentage of bits modified. So the future work is concentrated on making the system more robust by including the parameter correlation coefficient as well as chances of attacks.

It is very difficult to keep in mind a 16 digit PAN number. So in future for simplicity we can store it on the paper in a code format. This can be done by using aadhar card & storing the PAN on it in the form of barcode. So a barcode scanner can check whether the PAN given as input is correct or not. So there is no difficulty in remembering the PAN.

As a future work, multi-biometrics measure or features can also be used to implement online voting system. In order to provide more security we can also include other human biometric parameters such as facial image or distance between the eyebrows etc. If we think about the present population of our country it will be growing rapidly. So as a future work one should create a system which will take less time.

REFERENCES

[1] Shivendra Katiyar, Kullai Reddy Meka, Ferdous A. Barbhuiya, Sukumar NandiV. X. Alfonso and J. Tompkins, "Online Voting System Powered By Biometric Security Using Steganography,"2011 Second International Conference on Emerging Applications of Information Technology

[2] Linu Paul, Anilkumar M.N., "Authentication for Online Voting Using Steganography and Biometrics," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 10, December 2012

[3] Lindsay I Smith, "A tutorial on Principal Components Analysis," February 26, 2002

[4] T. Morkel 1, J.H.P. Eloff 2, M.S. Olivier 3 "An Overview of Image Steganography"

[5] Sutaone, M.S. and Khandare, M.V., "Image based steanography using LSB insertion technique", IEEE WMMN, pp. 146-151, January 2008.

[6] J.Samuel Manoharan,Dr.Kezi C.Vijila, A.Sathesh "Performance Analysis of Spatial & Frequency Domain" (4); Issue (3)

[7] Vijay kumar sharma, 2vishal shrivastava, "A steganography algorithm for hiding image in image by improved lsb substitution by minimize detection," Journal of Theoretical and Applied Information Technology 15th February 2012. Vol. 36 No.1

[8] Johnson, N.F. Jajodia, S., "Exploring Steganography: Seeing the Unseen", Computer Journal, February 1998.

[9] K B Raja, Venugopal K R and L M Patnaik, "A Secure Stegonographic Algorithm using LSB, DCT and Image Compression on Raw Images", Technical Report, Department of Computer Science and Engineering, University Visvesvaraya College of Engineering,Bangalore University, December 2004.

[10] P. Kruus, C. Scace,M. Heyman, and M. Mundy. (2003), "A survey of steganography techniques for image files." Advanced Security Research Journal.

[11] Jessica Fridrich, Miroslav Goljan, and Rui Du, "Detecting LSB Steganography in Color and Gray-Scale Images" State University of New York, Binghamton.

[12] Kshetrimayum Jenita Devi "A Sesure Image Steganography Using LSB Technique and Pseudo Random Encoding Technique" May 2013

[13] J. Brassilet et al., "Document Marking and Identification using Both Line and Word Shifting," Proc. Infocom95, IEEE CS Press, Los Alamitos, Calif., 1995.

[14] Bloisi, D. and Iocchi, L., "Image based Steganography and Cryptography", In Proc. of 2nd Int. Conf. on Computer Vision Theory and Applications (VISAPP), pp. 127-134, 2007.

[15] V. Licks and R. Jordan, "On Digital Image Watermarking Robust to Geometric Transformations," Proceedings of 2000 International Conference Image Processing (ICIP 2000), Vol. 3, pp. 690-693, 2000.

[16] Nikolaidis, Pitas, "Robust Image Watermarking the Spatial Domain", International Journal of Signal Processing", Vol.66, Issue 3, pp. 385 – 403, 1998.

