A Two Layer Approach to Image Authentication and Encryption through ECC & Voice Features (MFCC, Pitch Value)

Vikas Pardesi, N. S. Raghava

Abstract- Speech Processing is an area in which we can find such unique features (Mel Frequency Cepstrum Coefficients, Pitch value, zero crossing rates etc.) in voice segment for recognition of any individual and pre-processing for further synthesis. In this paper we are presenting a simplified approach to image authentication with MFCC (Mel Frequency Cepstrum Coefficients) and Pitch Value and image Encryption through Elliptic Curve Cryptography. Because of ECC great advantages (small key size, no solution to discrete logarithmic problem, less time consuming encryption, infinite time taken for brute force attack) for handheld, portable devices. Applying MFCC and Pitch information with various methods on various encrypted images which is encrypted by Elliptic Curve Cryptography and at the receiver side we do reverse process of this approach for authentication and decryption of image .With this approach we can authenticate an image through voice segment which is advantageous because speech is a natural way to interact with people, Not required to sit and work with a keyboard and finally no specific training is required for end users.

***Index Terms*— Elliptic curve cryptography, MFCC, Pitch value, public key cryptography, Speech Processing**

I. INTRODUCTION

Today, in this information era we are just totally dependent on the computers, internet and such other information technology, here we are talking about security. Security is a very small term to understand but it is very difficult to implement a better security approach in transferring of the data. In area of information security [1] there are many algorithms and techniques which is using for secure transferring of the data as data encryption standard [1], elliptic curve cryptography [2] and RSA [1] technique. Apart from this if we are talking about speech processing [3] area then it should be no hyperbole that it is a very vast area for research. Speech is biometric for human being that is freely available to every individual.

Speech of any individual has many properties such as it is a unique feature of every individual, processing of speech is very easy, easy input given method (no need of computer keyboard for inserting data).

In this paper we are describing our two layer secure approach to image security through Elliptic curve cryptography and voice features (MFCC [4] & Pitch [4] value), first we will see how a voice can be helpful for processing of images in secure environment.

Manuscript received April, 2013

Vikas Pardesi, Dept. of Information Technology, Delhi Technology University, Delhi, India.

N. S. Raghava, Dept. of Information Technology, Delhi Technology University, Delhi, India.

Any human being speech or voice has many unique properties or we can say features such as Mel Frequency Cepstrum Coefficients [4], Pitch of voice [4], fundamental frequency [4], fundamental formants [4], linear predictive coefficients [4],
Cepstrum coefficients, line spectral Pairs, 2-D, 3-D-spectrogram etc. now our main motive is to find the Mel frequency Cepstrum coefficients and pitch value through best approach.

Mel frequency Cepstrum can be defined as the short time power spectrum [16] of a speech signal. It can be calculated as linear cosine transform of the log power spectrum on a nonlinear Mel scale of frequency. Mel frequency Cepstrum coefficients are the unique features that can be acquired from the voice segment.

$$MelFrequency = 2595 * \log(1 + \frac{F}{700})$$

Where F is the linear frequency. Procedure and method is given in separate section in detail to calculate Mel frequency Cepstrum coefficients.

Pitch of any human being's voice is also unique feature. Voice speech is generated when the excitation comes from a periodic pulse train generated by vocal cords [15]. These vocal cords vibrate with their natural frequency of vibration like a tuning fork and generate pulses at regular interval. A speech signal consists of different frequencies which are harmonically related to each other in the form of a series. The lowest frequency of this series is known as the fundamental frequency or pitch frequency. Pitch calculation can be done with four method (1) Autocorrelation method [6], Average Magnitude difference method [6], these methods were in time domain (2) Parallel processing approach [6], (3) pitch period using spectral domain [6] or frequency domain, (4) pitch period using Cepstrum domain. For example With Auto-correlation method

$$Rxx = \lim_{N \to w} \frac{\sum_{i=0}^{2N} a(i) * a(i+k)}{2N+1}$$

Where $a(i)$ is the sample number and k is the interval and N is the total number of samples, R_{xx} is the resultant.

Now we show a brief introduction to encryption and decryption of image. In encryption method we are taking those points which are satisfying the elliptic curve, from these points we take one point AF(x, y) which is affine point and a base point BP(x, y). Base point is the smallest (x, y) point i.e., satisfying the Elliptic curve $y^2=x^3+ax+b$ where $4a^3+27b^2 \neq 0$. Each value of 'a' and 'b' can give a different curve. We take one affine point and then we take each pixel value one by one onto image. We perform point multiplication operation with two entity i.e., pixel value and affine point $S=P_I*AF(x, y)$.

Where P_I is pixel value. In Elliptic Curve Cryptography point multiplication is the basic operation which consists of two operation point addition & point doubling. After multiplication operation we get a point cipher text as {N*BP(x, y)} where N is the random number and BP is the base point and (x, y) one point we get after multiplication. Ultimately we get our Cipher Text CT_T= {(x1, y1), (x2, y2)} Where (x1,y1) we got from N and BP multiplication and (x2, y2) is (S + N*PK_B) where S=P_I*AF and P_I is scalar value i.e. pixel value and AF is affine point. And N is a random number and PK_B is public key of user B so at last we got encrypted message in co-ordinate format, namely (N*BP, S + N*PK_B).

In Decryption of the encrypted Image we use the private key of the user B i.e. PRK_B on the first element i.e. N*BP. This will be subtracted from the other term (S+N*PK_B) to recover the S. In last we apply Discrete Logarithmic problem [7] to recover our pixel value i.e., P_I.

Due to using discrete logarithmic approach it is also advantageous that no solution till now found for breaking this. Biggest use of the ECC in security due to it's smallest key.

Structure of this paper follows these points-

- Section 2 describes the concept of finding the MFCC coefficients from a voice segment.
- Section 3 describes that how we can find the pitch information through a voice and different-2 techniques for finding pitch information.
- Section 4 describes two layer approach i.e., main concern of this paper, states that how encryption is done through Elliptic Curve Cryptography and results of implementation.
- Section 5 describes that how image authentication works through MFCC coefficients and pitch information.
- Section 6 states its application and future work.
- Section 7 is separated for conclusion and discussion.

II. CONCEPT OF EXTRACTING THE MFCC FEATURE THROUGH VOICE SEGMENT

In Auditory System as a filter bank gives a clear indication that critical band filters must be formed but the shape and the structure of the filter is not clear. Two commonly used frequency scales they are bark scale [13] and Mel scale [13], here we describes about Mel frequency and Mel scale. Basis of the Mel scale is pitch perception and filter is used in this i.e., triangular filter. The scale is linear below 1000 Hz and non-linear (logarithmic) above 1000 Hz. We can convert normal frequency to Mel frequency with formula Mel frequency=2595*log (1+linear frequency/700).

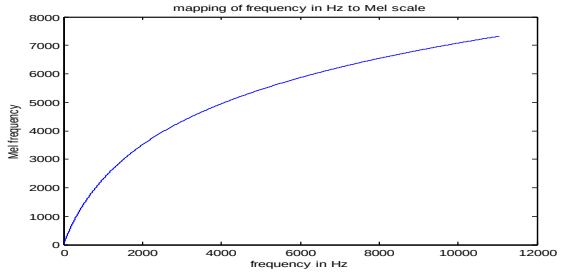

Fig-1.1 Mapping of frequency in Mel Scale

Now Mel frequency Cepstrum can be described as the short time power spectrum of a speech segment, which can be computed as linear cosine transform of the log power spectrum on a non-linear Mel scale of frequency. MFCC are the coefficients we got in the MFC representation. Now consider about the difference about Cepstrum [9] and MFC, in MFC frequency band are equally spaced on the Mel scale. Mel scale approximates the human auditory response very much closed to than the linearly-spaced frequency bands used in the case of spectrum. Mel scale wrap up the frequency and allows better representation similar to the human auditory system.

A. Procedure for calculating the Mel frequency Cepstrum coefficients:

- Take the FFT [14] (fast Fourier transform of a windowed signal).calculate its squared magnitude i.e., power spectrum.
- Pre analysis the spectrum to approximate the unequal sensitivity of human hearing at different frequencies.
- Combined the power spectrum within the overlapping critical band filter response. This integration is being done through triangular overlapping windows called Mel filters. This reduces the frequency sensitivity over the original spectral estimate; basically on higher frequencies are analyzed because of the wider band.
- Compress the spectral amplitude by taking log. Optionally the combination of log power spectrum [8] may be done.
- Take the IDFT (inverse discreet cosine transforms). This gives the cepstral coefficients.
- Perform spectral smoothing this is achieved by the truncating the spectrum, lower 12 or 14 coefficients are used out of 20 or more coefficients.

B. On the basis of these steps flow chart for better understanding of extraction method-

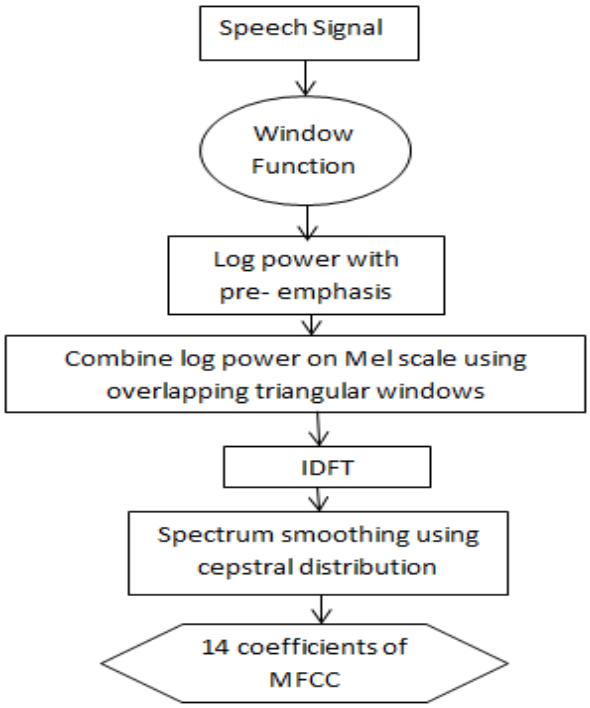

Fig-1.2 Flow Chart For Extracting MFCC

C. Algorithm-

Step-1 open the audio voice segment with FOpen()command;

Step-2 Read this voice segment with Fread ();

Step-3 fix the number of samples as you required and find FFT of this segment of voice.

Step-4 Find the log of the previous step output.

Step-5 for 1 to sample number Convert linear frequency to Mel frequency

Step-6 using function melcepst() for finding the Mel frequency Cepstrum coefficients.

Generally our speech does not follow linear way so this natural frequency we have to convert into Mel frequency or in any kind of frequency on which we can do analysis or synthesis of our speech signal. Above we discussed how can we convert our normal frequency to Mel frequency and also that how can we find the Mel frequency Cepstrum coefficients.

III. FIND PITCH VALUE FROM VOICE SEGMENT

There are different frequencies in the speech signal which are periodically related to each other in a manner. The lowest frequency of this series is known as the pitch frequency or fundamental frequency. Basic frequency of the vibrations of the vocal cords is called pitch frequency. vocal cords generate this frequency in the form of periodic excitation passes via the vocal tract filter and become convolved with the impulse response of the filter to produce a voice or speech signal because speech is a convolved signal. The number of samples after which the waveform repeats itself will reveal the pitch period in terms of number of samples. We can also find the time corresponding to these samples and calculate the inverse of time period then we get frequency in Hz.in time domain main theory is used for pitch detection i.e., find the similarity between the original speech and it's shifted speech. There are many methods for finding the pitch they are as follows.

1. Time Domain Approach

A. Autocorrelation Method for finding pitch Frequency:

Autocorrelation is the correlation of a signal with itself. We can say that this is a measure of the similarity between samples as a function of the time separation between them. It can be considered as a mathematical tool to find repeating patterns and their periods. Autocorrelation methods need at least two pitch periods to detect pitch.

$$R_{xx} = \lim_{N \to w} \frac{\sum_{i=0}^{2N} a(i) * a(i+k)}{2N+1}$$

Where *a(i)* is the i^{th} sample number and *N* is total number of samples and *K* is the interval and R_{xx} is the result.

B. Average Magnitude Difference Function:

We make the difference signal D by delaying the input speech by various amounts, subtracting the delayed waveform from the original, and summation of the magnitude of the difference between sample values. Finally we take the average of the difference function over the number of samples. Difference signal is always zero at delay zero.

$$AMDF\ (k) = \frac{\sum_{i=1}^{N} abs[a(i) - a(i+k)]}{N}$$

Where *a(i)* is the current sample and *N* is the total number of samples and *k* is the interval.

2. Frequency Domain Approach

A. FFT (fast Fourier transform) based method in Spectral domain:

First we have to know about spectral domain so we studied in above approach that are time domain approach and now we discuss about Frequency domain approach, in frequency domain we work with speech spectrum so it is called spectral domain. Generally there are four method for finding pitch in frequency domain they are as follows (a) FFT based Method, (b) Harmonic peak detection method, (c) spectrum similarity method, (d) spectral autocorrelation method. Here we discuss only one i.e., FFT based method. In this FFT is a very fast algorithm for computing the discrete Fourier transform. When we say take K=130 point DFT, we will get 130 DFT coefficients. The n^{th} DFT coefficients X (n) of any sampled signal x (k) is given by

$$X(\mathrm{n}) = \sum_{n=0}^{K-1} \mathrm{x(k)(W)}^{nk} \quad \text{Where W is } W = e^{-j4\frac{\pi}{2}/N}$$

Algorithm-

Step-1 Take the voiced Segment of speech containing N samples.

Step-2 Take a N-samples point FFT and plot it.

Step-3 Track First Frequency as fundamental frequency.// resolution of the fundamental frequency measurement is achieved by the number of points in the FFT. Say a N point FFT, the frequency resolution will improve to (basic frequency)/N. Using FFT is not sure that it gives exact pitch , it depends on the DFT domain resolution.

Step-4 fundamental frequency is the FFT point number (first peak) multiplied by the frequency resolution.

Step-5 Note the position of the first peak and multiply with step (4).

Step-6 END.

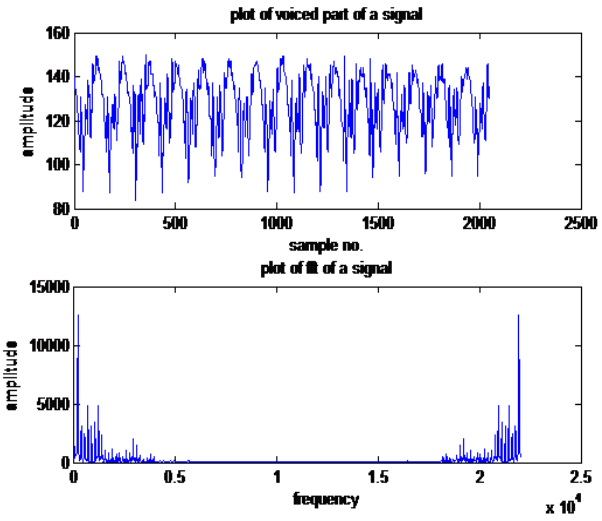

Fig-1.3 Plot of voiced part of a signal and FFT of that signal

For example we took a speech segment and first we separate the voiced part and unvoiced part of the signal then we find th pitch value through Fast Fourier Transform. Above given graph is showing plot of voiced part of the signal and plot of FFT of the voiced part of the signal. From our example we find our first peak occurs at 16^{th} position and has a value of 12560. Resolution frequency is 10.80 Hz so fundamental frequency is 10.8*16 i.e., 172.26 Hz. It is a valid pitch frequency and further we will use in our implementations.

IV. TWO LAYER APPROACH AND IMPLEMENTATION RESULTS

The main area of interest of this paper is here because here we have two layer approach and applied this on an image in secure transferring over network. We have to find Mel frequency Cepstrum coefficients with the above given formulas. So here we take one speech segment

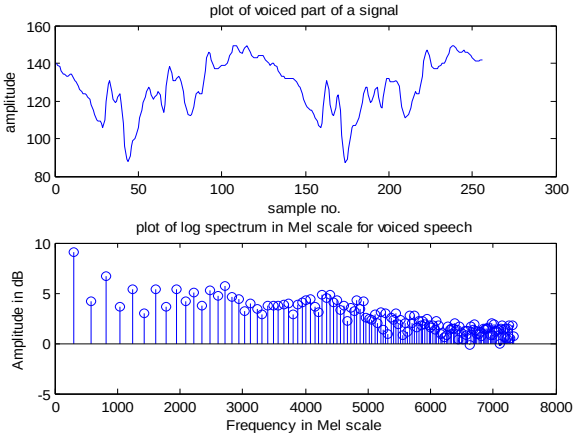

Fig-1.4 Plot of Voiced part of signal

We took one speech segment and find its voiced part first and then we find the log spectrum in Mel scale frequency. We got above given plot of that result. Now we have to find the MFCC for voiced speech [25] then below given graph is showing that 14 coefficients we are getting and they are plotted s follows.

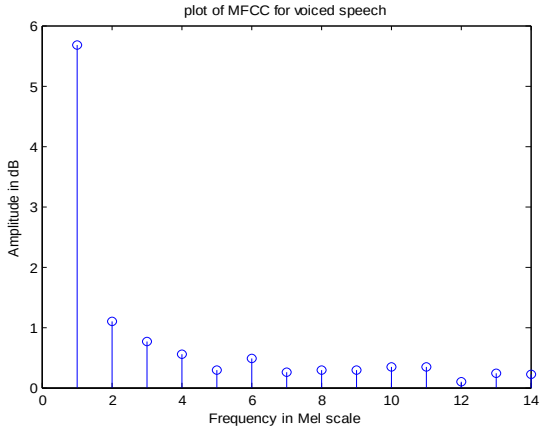

Fig-1.5 Plot of MFCC for voiced part of Speech

We got array of 14 coefficients they are as follows 5.6999, 1.0993, 0.7545, 0.5487, 0.2899, 0.4800, 0.2525, 0.2825, 0.2831, 0.3480, 0.3459, 0.1003, 0.2398 and 0.2271. if we plot a graph for this then the graph will show actual flow of the data in voiced part.

We can also find these coefficients from MATLAB command 'melcepst' from voice box, but you have to first install that tool box into your MATLAB software. Now we have to perform operation and apply these coefficients and pitch value to an image with this approach-

$$R = \sum_{i=1}^{14} \left(\frac{Xi}{14}\right) * \mathrm{P}$$

Where X_i is the coefficients and P is the pitch value and R is the resultant. Apply this resultant in a 2D image given below.

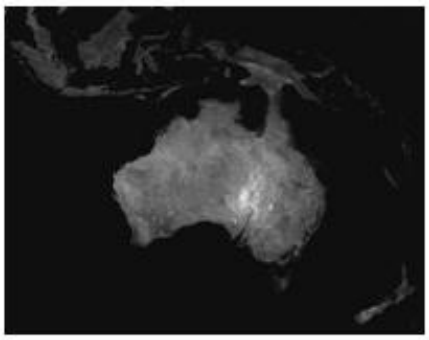

Fig-1.6 Original Image

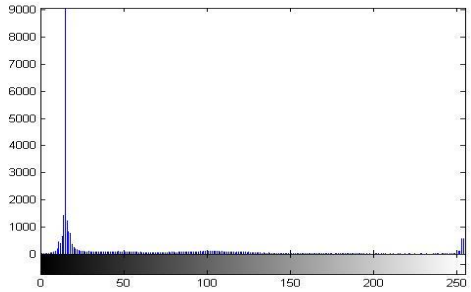

Fig-1.7 Histogram of Original Image

For applying on each pixel of the image we call a function

MFCConIMAGE (A[N][M], R)

```
Step-1 for (i=0 to N-1)
       {
          For (j=0 to M-1)
            {
              A[i] [j] =A [i] [j]*R;
            }
       }
```

Step-2 Traverse on each pixel value of the image to locate the pixel value.

Step-3 End;

After apply this approach we get below shown result.

Where i and j is the pixel values at x axis and y axis respectively. R we got from above mentioned formula. After applying MFCC value and pitch value our image get distracted and will be noisy that is also good for secure transferring over network. Here result of image is purely black and no one can identify any thing in the resultant image. Further we will find it's histogram for see the changes after performing operation.

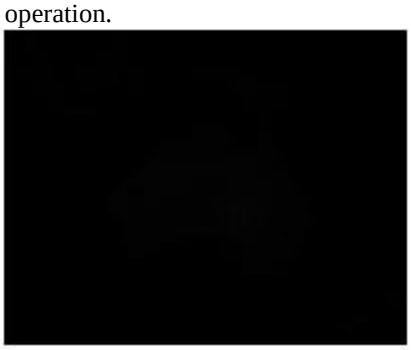

Fig-1.8 MFCC coefficients and pitch value applied on Image

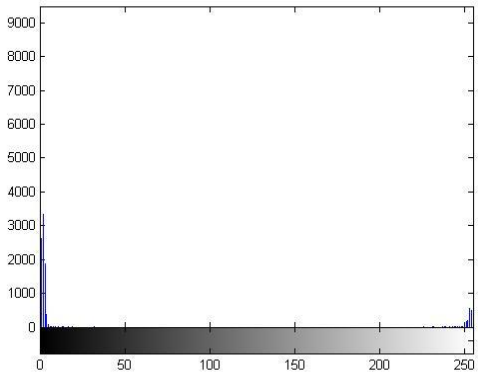

Fig-1.9. Histogram of above image after applied MFCC and pitch value to image

Now we have to apply ECC (Elliptic Curve Cryptography) on this output image and then we transfer over network, so now task is how to encrypt the image with ECC Encryption so we have to understand some points about ECC Encryption they are as follows-

- ECC is a public key algorithm which works on private key and public key.
- ECC designed and mainly used for handheld devices and small devices which are low in memory and resources.
- ECC successfully implemented first by Koblitz and miller [10].
- ECC encryption uses elliptic graph, onto this graph which points are generated that are taken for encryption.
- Main use of ECC is for those devices which can work only in constrained environment.

There is two operations involved in ECC which are (1) Point Multiplication [11] which internally consists two operation i.e., point addition and point doubling, in point doubling it doubles the point which is of same kind with point doubling formula i.e., say one point is $T(X_T, Y_T)$

$L(X_L, Y_l)=2T(X_T, Y_T)$

$X_L=S^2-2X_T \bmod p$

$Y_L=-Y_T+S(X_T-X_L) \bmod p$

$S=(3X_T^2+a)/(2Y_T) \bmod p$

Where S is the tagent at point T and a is one of the parameter chosen with the elliptic curve and p is a prime number.

Now in point addition it is used for addition of two points with point addition formula i.e., say two points $P(X_P, Y_P)$ and

$Q(X_Q, Y_Q)$ then $I(X_I, Y_I)$

$X_I= S^2-X_P-X_Q \bmod p$

$Y_I= -Y_P+S(X_P-X_I) \bmod p$

$S=(Y_P-Y_Q)/(X_P-X_Q) \bmod p$

Where S is the tagent passing through P and Q and p is a prime number.

ECC point generation

According to elliptic curve $y^2=(x^3+ax+b) \bmod p$ where $4a^3+27b^2\neq0$ we have to generate points that lies on elliptic curve. Let p=39, a=-1, b=1 on which 'a' and 'b' value satisfying above equation. We do encrypt so first of all we generate all points that satisfies the elliptical curve follow this function **GeneratePoints (a, b, p)**

Step 1 take x=0 or any other positive integer

Step 2 loop until x<p

I. $Y^2=(x^3+ax+b) \bmod p$

II. If y^2 is perfect square

Print(x, square root (y))

Else

x=x+1;

Step 3 End

where p is a prime number, x and y are co-ordinates.

ECC Encryption

Step 1-For all P_I (pixel value)

Find $S=P_I *AF_M$ //*P_I is constant, AF_M is random Affine-point in elliptic curve. *//

Step 2- Find $PK_B=PRK_B*BP$ //*B_P is the base point Of Elliptic curve, PRK_B is the private key. *//

Step 3- End;

Encrypted data= (N*BP, $S+N*PK_P$)

ECC decryption

Follow this method for decryption,

Let N*BP be the first point and ($S+N*PK_B$) be the second point PRK_B *N*BP=PRK_B *first point

Calculate $S=S+N*PK_B-PRK_B*N*BP$

Calculate the scalar value of 'P_I' from S_M using discrete logarithmic problem [7]. Where PK_A, PRK_A and PK_B, PRK_B is the public key and private key of user A and B respectively.

Fig-1.10 After ECC decryption of Image

In decryption of the ECC we will use the Discrete logarithmic Problem, after applying on each pixel we get all the decrypted 2D-matrix i.e., image. Below histogram is shown of above image and we can analyze this histogram with original image's histogram. It is approximately same.

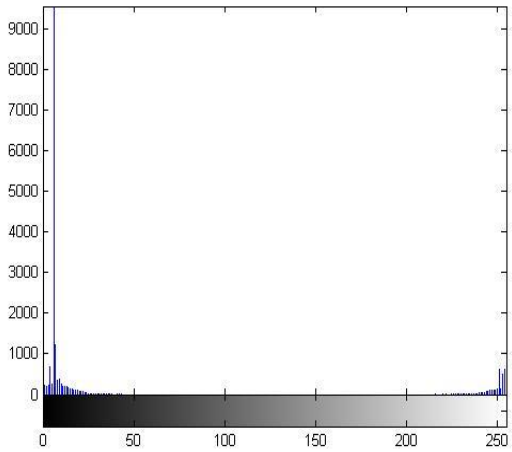

Fig-1.11 Histogram of Decrypted mage

Above shown image is the decrypted image after Discrete Logarithmic problem [7]. Now this image we got at the

receiver side but we have to follow one operation also i.e., remove the MFCC coefficients from encrypted image then we got our real image.

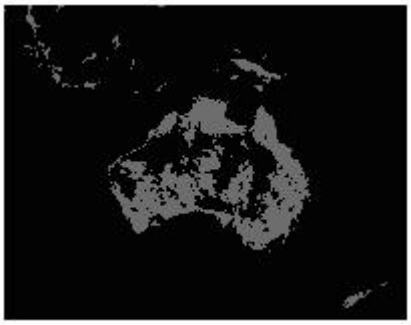

Fig-1.12 Remove MFCC and Pitch from Original Image

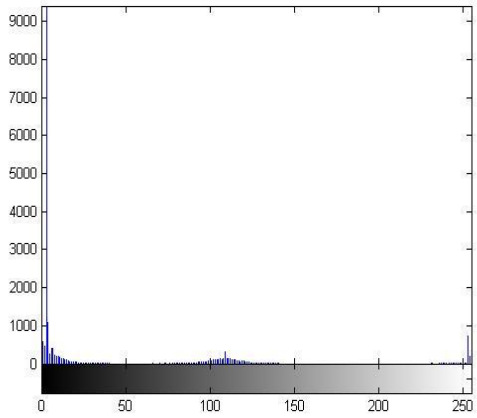

Fig-1.13 Histogram of Received Image after removing MFCC and Pitch.

We just apply reverse operation for removing the MFCC and pitch value. Now we can analyze our result image maximum feature of the image can be identified and this method is better working for satellite image further more operation can be performed on the resultant image if we want to enhance the image.

V. IMAGE AUTHENTICATION THROUGH MFCC

As we know very well use of images in now days is very much, in area of cryptography, in area of forensic science, in daily life and images is the best option to hide the data and transfer over network. Main concern in today's life is to secure the images and authentication of genuine images is very much. We have to protect our image from hackers and authenticate the sender is genuine or not and also we have to maintain the integrity of the data. In information security there are many algorithms which are used for authentication such as digital signature, digital certificate etc. in this paper our new idea is we are protecting the image from MFCC coefficients.

According to the paper we are using one sample voice segment and we find the MFCC coefficients and apply on image and then we are encrypting the image through MFCC and we all know very well that each human's voice is unique and MFCC also used for speaker verification so at the receiver side when we find same MFCC coefficients from same voice then we will remove MFCC feature from the image and then we can verify our image is that it is coming from genuine user or not. If anyone alters the message in between the network then real image will not recover. Some malfunctioned image we will get and then that image will not be accepted so with this we can also maintain the message integrity. So with MFCC coefficients we can do image authentication and message integration.

VI. APPLICATIONS & FUTURE WORK

Cryptography has a very great advantage in information security area that except of the expert no one can do harm to the information and securely we can transfer our confidential data to other party. Elliptic Curve Cryptography is a new technology which came in existence since 1965 by Koblitz [10] and miller. A large number of researches is going on ECC but ECC great advantage is its smaller key size compare to other algorithms such as DES (Data Encryption Standard) and RSA (Rivest- Shamir- Adelman). ECC works on small device in a better way compare to huge machines, ECC works on those device which are in constrained environment or use resources in a limit. We can make use of MFCC in making Elliptic curve digital signature through MFCC coefficients and we also can make Elliptic Curve Cryptography Digital certificate Through MFCC. These certificates can also be used further by the third party and also used by many users in different-2 encryption algorithm.

Future work related to decrypted image and image recovered after removing the MFCC coefficients from image, we can further perform image enhancement techniques such as histogram equalization and contrast stretching and noise removal from an image. Further this image can be recovered in a better way by applying the digital image processing [12]. Further if we analyze the histogram of the each image then one by one histogram of the original image and decrypted image is approximately same. Drawback is exactly we are not getting original image but further we can also do many operation with received image for improvement.

VII. CONCLUSION

Proposed approach in this paper works better for satellite images and can better work for text data and further we can also us MFCC coefficients in other encryption algorithm because it is the unique feature from human characteristics and freely available from every user. Here we mean to say each individual can use their voice for authentication of the data and this is more secure in public network. Generally in biometrics speech is available easily and also can detect and synthesized in any condition and in any environment but there is a drawback i.e., noise. If your recording of voice is in noisy environment then it will be a little bit more complex to detect easily the data. With use of speech user need not required keyboard but recording device quality must be better in recording manner. Furthermore work is going on speech processing. We are also just improving our implementation in better way.

REFERENCES

[1] William Stallings,"Cryptography & Network Security", Printce Hall, 5th Edition.

[2] Alessandro Cilardo, Luigi coppolino, Nicola Mazocca, and Luigi Roman, "Elliptical Curve Cryptography Engineering", Proceedings of the IEEE, vol. 94, no 9, pp. 395-406, Feb. 2006 .

[3]"SpeechProcessing",dea.brunel.ac.uk/cmsp/home…/chapter13-speech%20processing.pdf.

[4] en.wikipedia.org/wiki/Speech_processing.

[5] G. Khare, M. Kulkarni, "Generation of excitation signal in voice excited linear predictive coding using discrete cosine transform", TENCON 2005, IEEE region 10, page 1-4.

[6] S.S. Upadhyay , "Pitch detection in time and frequency domain", ICCICT 2012,IEEE, page 1-5.
[7] Adam J. Albirt, "Understanding & Applied Cryptography and Data Security" CRC press, Pearson
[8] G. Malherpe, O. Mesde and H. Riz, "Acoustic synthesis and methodology for improving cochlear implant speech processing techniques", Proc. Of the 25th annual international conference of IEEE *EMBS*, 2003
[9] H.Nagahama, Y.Miyanaga and N.Ohtsuki, "An adaptive speech analysis for speech recognition system", Proc. Of IEEE ISPACS 1999, pp.745-748, Dec. 1999.
[10] N. Kolbitz, Elliptic Curve Cryptosystems, Mathematics of Computation, vol.48, 1987, pp.203-209.
[11] Certicom website http://www.certicom.com/index.php? action =ecc_tutorial, home.
[12] Rafael C. Gonzalez, "Digital Image Processing", Third Edition.
[13] http://en.wikipedia.org/wiki/Mel_scale
[14] http://en.wikipedia.org/wiki/FFT
[15] H. Teffahi, "Relationship between control parameters and outputs in the two massmodel of vocal cords", ICMCS , 2011 IEEE page 1-5.
[16] H.Hoge "A parametric representation of short time power spectrum based on the acoustic properties of the ear", ICASSP 1984, speech and signal processing page 49-51.

Mr. Vikas pardesi, completed his B.Tech from AJAY KUMAR GARG ENGINEERING COLLEGE, Ghaziabad (U.P), INDIA. He currently pursuing M.Tech from DELHI TECHNOLOGICAL UNIVERSITY, DELHI, INDIA. His area of interests are Digital Image Processing, Speech Processing, Graph Theory, Data Structure.

Mr. N. S. Raghava Currently serving DELHI TECHNOLOGICAL UNIVERSITY, DELHI as ASSOCIATE PROFESSOR in Department of Information Technology since 1998. He has a very vast experience in ANTENNA theory. His area of interests are Signal Processing, Wavelet theory, Antenna theory, Speech processing.