IOSR Journal of Engineering
e-ISSN: 2250-3021, **p**-ISSN: 2278-8719,
Vol. 2, Issue 12 (Dec. 2012), ||V4|| PP 24-31

A Novel Technique in stegnanalysis using S-DES and Back propagation Algorithm

Sudha .D[1], Mariadas Ronnie C.P [2]
[1]Asst. Professor, KMM College of Arts and Science, Trikkakkara
[2] Asst. Professor, KMM College of Arts and Science, Trikkakkara

Abstract:- **This paper proposes a new algorithm by improving and combining S-DES and Back Propagation in neural networks to identify the stego image in mails. Back propogation is ideal for simple pattern recognition and mapping tasks. This new algorithm checks the mail inbox for JPEG files. As the image is send across the network for transmission, compression is required. Compression is used to minimize the seeking time of the JPEG files. After compression, the files are stored in a large database for further processing. A sample image is taken from the database as cover channel which is used to hide the secret information. For each image, we embedded a random binary stream of different lengths using S-DES algorithm. The proposed research analyses the performance of the improved version of image steganalysis algorithms in corporate mails.**

Index Terms:— **S-DES Algorithm, Back propagation, Neural Networks, DWT, DCT, Steganalysis.**

I. INTRODUCTION

Steganography is a science of invisible communication or secret communication. The main goal of steganography is to hide the fact that the message is present in the transmission medium. The word steganography is derived from the Greek word *Covered Writing*. The information to hide is embedded in the cover image in such a way that the message is undetected and appearance of the resulted image is exactly the same as the original image. The goal of steganalysis is to detect hidden information from data with no knowledge about the steganography algorithm and its parameters. Current trend in steganalysis seems to suggest two extreme approaches: (a) No statistical assumptions about the image under investigation. (b) A parametric model is assumed for the image and its statistics are computed for steganalysis detection.

In ancient Greece, text was written on wax covered tablets. Demeratus wanted to notify Sparta that Xerxes intended to invade Greece. To avoid capture, he scraped the wax off of the tablets and wrote a message on the underlying wood. He then covered the tablets with wax again. The tablets appeared to be blank and unused so they passed inspection by entries without question. Another method was to shave the head of a messenger and tattoo a message or image on the messengers head. After allowing his hair to grow, the message would be undetected until the head was shaved again. Another common form of invisible writing is through the use of Invisible inks. Such inks were used with much success as recently as WWII. An innocent letter may contain a very different message written between the lines. Early in WWII steganographic technology consisted almost exclusively of invisible inks. Common sources for invisible inks are milk, vinegar, fruit juices and urine. A digital image is composed of pixels. Each pixel contain information as the intensity of the three primary colors Red, Green and Blue. The secret message is broken into bits by steganographic tools and embedded in the cover image. A password is used to extract the hidden message and is referred to as a stego key. The result of the processes is known as stego image. A typical image with 64 X 480 pixel and 256 colors can hide approximately 30 kilo bytes f information. A BMP or GIF format algorithm is preferred for choosing a cover image. JPEG compression algorithm uses floating point calculations t translate the picture into an array.

Discrete Wavelet Transform (DWT)

Spectral images usually have a similar global structure across components. However, different pixel intensities could exist among nearby spectral components or in the same component. This means that two kinds of correlations may be found in images: intraband correlation [10] among nearby pixels in the same component, and interband correlation among pixels across adjacent components. Interband correlation should be taken into account because it allows a more compact representation of the image by packing the energy into less number of bands, enabling a higher compression performance. There are many technologies which could be applied to remove correlation across the spectral dimension, but two of them are the main approaches for images: the KLT and the DWT Discrete Wavelet Transform [11] [12]. (DWT) is the most popular transform for image-based application. They have lower computational complexity, and they provide interesting features such as component and resolution scalability and progressive transmission. A 2-dimensional wavelet transform is applied to the original image in order to decompose it into a series of filtered sub band images. At the top left of the image is a low-pass filtered version of the original and moving to the bottom right, each component contains progressively higher-frequency information that adds the detail of the image. It is clear that the higher-frequency components are relatively sparse, i.e., many of the coefficients in these components are zero or insignificant. The wavelet transform is thus an efficient way of decorrelating or concentrating the important information into a few significant coefficients. The wavelet transform is particularly effective for still image compression and has been adopted as part of the JPEG 2000 standard and for still image texture coding in the MPEG-4 standard.

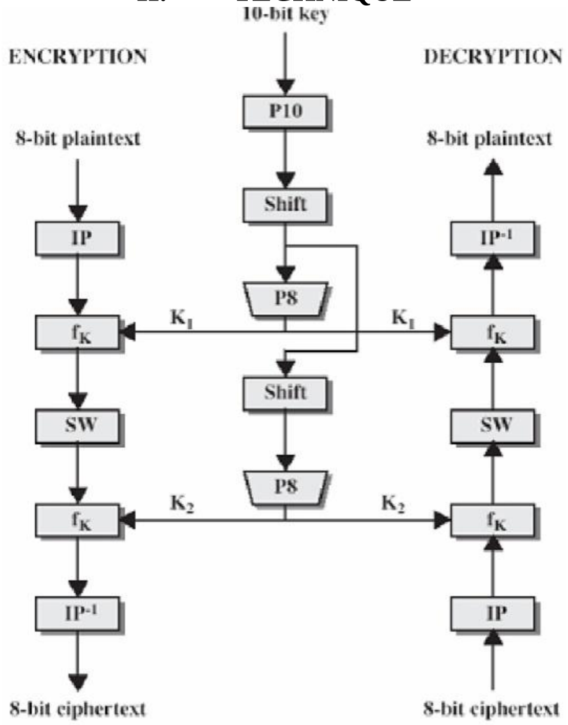
Motion Estimation Prediction

By Motion estimation, we mean the estimation of the displacement of image structures from one frame to another. Motion estimation from a sequence of images arises in many application areas, principally in scene analysis and image coding. Motion estimation obtains the motion information by finding the motion field between the reference frame and the current frame. It exploits temporal redundancy of an image sequence, and, as a result, the required storage or transmission bandwidth is reduced by a factor of four. Block matching [13] is one of the most popular and time consuming methods of motion estimation. This method compares blocks of each frame with the blocks of its next frame to compute a motion vector for each block; therefore, the next frame can be generated using the current frame and the motion vectors for each block of the frame. Block matching algorithm is one of the simplest motion estimation techniques that compare one block of the current frame with all of the blocks of the next frame to decide where the matching block is located. Considering the number of computations that has to be done for each motion vector, each frame of the image is partitioned into search windows of size H*W pixels. Each search window is then divided into smaller macro blocks of size, say, 8*8 or 16*16 pixels. To calculate the motion vectors, each block of the current frame must be compared to all of the blocks of the next frame with in the search range and the Mean Absolute Difference for each matching block is calculated. The block with the minimum value of the Mean Absolute Difference is the preferred matching block. The location of that block is the motion displacement vector for that block in current frame. The motion activities of the neighboring pixels for a specific frame are different but highly correlated since they usually characterize very similar motion structures. Therefore, motion information of the pixel, say, pi can be approximated by the neighboring pixels in the same frame. The initial motion vector of the current pixel is approximated by the motion activity of the upper-left neighboring pixels in the same frame.

Prediction Coding

An image normally requires an enormous storage. To transmit an image over a 28.8 Kbps modem would take almost 4 minutes. The purpose for image compression is to reduce the amount of data required for representing images and therefore reduce the cost for storage and transmission. Image compression plays a key role in many important applications, including image database, image communications, remote sensing (the use of satellite imagery for weather and other earth-resource application). The image(s) to be compressed are gray scale with pixel values between 0 to 255. There are different techniques for compressing images. They are broadly classified into two classes called lossless and lossy compression techniques. As the name suggests in lossless compression techniques, no information regarding the image is lost. In other words, the reconstructed image from the compressed image is identical to the original image in every sense. Whereas in lossy compression, some image information is lost, i.e. the reconstructed image from the compressed image is similar to the original image but not identical to it. The temporal prediction residuals from adaptive prediction are encoded using Huffman codes. Huffman codes are used for data compression that will use a variable length code instead of a fixed length code, with fewer bits to store the common characters, and more bits to store the rare characters. The idea is that the frequently occurring symbols are assigned short codes and symbols with less frequency are coded using more bits. The Huffman code can be constructed using a tree. The probability of each intensity level is computed and a column of intensity level with descending probabilities is created. The intensities of this column constitute the levels of Huffman code tree. At each step the two tree nodes having minimal probabilities are connected to form an intermediate node. The probability assigned to this node is the sum of probabilities of the two branches. The procedure is repeated until all branches are used and the probability sum is 1.Each edge in the binary tree, represents either 0 or 1, and each leaf corresponds to the sequence of 0s and 1s traversed to reach a particular code. Since no prefix is shared, all legal codes are at the leaves, and decoding a string means following edges, according to the sequence of 0s and 1s in the string, until a leaf is reached. The code words are constructed by traversing the tree from root to its leaves. At each level 0 is assigned to the top branch and 1 to the bottom branch. This procedure is repeated until all the tree leaves are reached. Each leaf corresponds to a unique intensity level. The codeword for each intensity level consists of 0s and 1s that exist in the path from the root to the specific leaf.

II. TECHNIQUE



S-DES Algorithm

The **Data Encryption Standard** (DES) [9] is a block cipher that was selected by the National Bureau of Standards as an official Federal Information Processing Standard (FIPS) in 1976 It is based on a symmetric-key algorithm that uses a 56-bit key. DES is now considered to be insecure for many applications. This is chiefly due to the 56-bit key size being too small and the Electronic Frontier Foundation collaborated to publicly break a DES key in 22 hours and 15 minutes. **Simplified DES**, developed by Professor Edward Schaefer of Santa Clara University, is an educational rather than a secure encryption algorithm. It has similar properties and structure to DES, with much smaller parameters.

Figure shown above illustrates the overall structure of the simplified DES, which we will refer to as SDES. The S-DES encryption algorithm takes an 8-bit block of plaintext (example: 10111101) and a 10-bit key as input, and produces an 8-bit block of cipher text as output. The S-DES decryption algorithm takes an 8-bit block of cipher text and the same 10-bit key used to produce that cipher text as input, and produces the original 8-bit block of plaintext. S-DES encryption (decryption) algorithm takes 8-bit block of plaintext (cipher text) and a 10-bit key, and produces 8-bit cipher text (plaintext) block. Encryption algorithm involves 5 functions: an initial permutation (IP); a complex function f_K, which involves both permutation and substitution and depends on a key input; a simple permutation function that switches (SW) the 2 halves of the data; the function f_K again; and finally, a permutation function that is the inverse of the initial permutation (IP^{-1}). Decryption process is similar. The function f_K takes 8-bit key which is obtained from the 10-bit initial one two times. The key is first subjected to a permutation P10. Then a shift operation is performed. The output of the shift operation then passes through a permutation function that produces an 8-bit output (P8) for the first sub key (K1). The output of the shift operation also feeds into another shift and another instance of P8 to produce the 2nd sub key K2.

We can express encryption algorithm as superposition:

$$IP^{-1} \circ f_{K_2} \circ SW \circ f_{K_1} \circ IP$$

or

Cipher text= IP^{-1} ($f_{K_2}(SW(f_{K_1}(IP(pla\text{int}\ ext)))))$

Where,

$$K_1 = P8(Shift(P10(key)))$$

$$K_2 = P8(Shift(Shift(P10(key))))$$

Decryption is the reverse of encryption:

Plaintext= IP^{-1} ($f_{K_1}(SW(f_{K_2}(IP(ciphertext)))))$

Example

P10 = { 3, 5, 2, 7, 4, 10, 1, 9, 8, 6}

P8 = { 6, 3, 7, 4, 8, 5, 10, 9}

K = 10100 00010

P10 10000 01100

LS-1 00001 11000 -> P8 -> K1 = 1010 0100

LS-2 00100 00011 -> P8 -> K2 = 0100 0011

Data Embedding:

Data embedding is accomplished by different techniques

1. LSB
2. DCT
3. DFT
4. Wavelet

LSB (Least Significant bit): e LSB (Least Significant Bit), which replaces the least significant bits of samples selected to
hide the information.

Least significant bits (LSB) is an approach to embedding information in image file. The simplest steganographic techniques embed the bits of the message directly into least significant bit plane of the *cover-image* in a deterministic sequence. Modulating the least significant bit does not result in human-perceptible difference because the amplitude of the change is small. Least significant bit carries the least information about host signal. It is known that people cannot able to notice a change in this bit. It's assumes, that LSB is noise and it's possible to embed information by replacing the host signal least significant bits by bits of secret message. The main drawback is high sensitivity to small distortions in the container. The advantages of LSB are its simplicity to embed the bits of the message directly into the LSB plane of *cover-image* and many techniques use these methods

Discrete cosine Transform (DCT)

DCT is a mechanism used in the JPEG compression algorithm to transform successive 8×8-pixel blocks of the image from spatial domain to 64 DCT coefficients each in frequency domain. It separates the image into parts of differing importance. It transforms a signal or image from the spatial domain to the frequency domain. It can separate the image into high, middle and low frequency components. DCT is used in steganography as- Image is broken into 8×8 blocks of pixels. Working from left to right, top to bottom, the DCT is applied to each block. Each block is compressed through quantization table to scale the DCT coefficients and message is embedded in DCT coefficients. It is well known that the discrete cosine transform (DCT) coefficients of an image that are used widely as a message carrier have a generalized Gaussian distribution [8]
The basic operation of the DCT is as follows:

- The input image is N by M;
- f(i,j) is the intensity of the pixel in row i and column j;
- F(u,v) is the DCT coefficient in row k1 and column k2 of the DCT matrix.
- For most images, much of the signal energy lies at low frequencies; these appear in the upper left corner of the DCT.
- Compression is achieved since the lower right values represent higher frequencies, and are often small - small enough to be neglected with little visible distortion.
- The DCT input is an 8 by 8 array of integers. This array contains each pixel's gray scale level;
- 8 bit pixels have levels from 0 to 255.

Back Propagation

Most people would consider the Back Propagation network to be the quite essential part of Neural Networks. Actually, Back Propagation 1,2,3 is the training or learning algorithm rather than the network itself. These are called *Feed- Forward* Networks or *Multi-Layer Perceptrons (MLPs)*. Back Propagation network learns by example. Giving the algorithm examples of what you want the network to do, it changes the network's weights so that, when training is finished, it will give you the required output for a particular input. Back

Propagation networks are ideal for simple Pattern Recognition and Mapping Tasks. As just mentioned, to train the network, you need to give it the examples to obtain what we want. Information in stego image can be retrieved by S-DES algorithm. With our method, the use of neural network is the key technique. In this method the cover image is divided into group of blocks and all these blocks are generated by S-DES method. Now the embedder adjust a neural network weights with desired hidden bit code from the collection of both cover image and stego-image folder. This is common for both embedding and extracting the hidden information. We can use supervised learning of the neural network for learning. We embed the secret message within a cover image by using XOR neural network learning model. A new algorithm is developed to screen the JPEG images from the Email attachments and this will store the same in the hard disk in a separate folder. Steganography is used to hide the cipher text obtained in the above step into an image. The main purpose of steganography is encoding and decoding. The inputs of the steganography are Cover data, Plain text and the key value. An image file contains the binary representation of the colour or light intensity of each picture element comprising the image. Each pixel is represented by three bytes representing the intensity of three primary colours red, green, blue (RGB) respectively. A typical 640x480 pixel image using a palette of 256 colours would require a file about 307 KB in size whereas a 1024x768 pixel image would result in 2.36 MB file. JPEG uses lossy compression on the other hand. In this the expanded image is very nearly the same as the original but not an exact duplicate. Jpeg can be used for stego applications because it is more common to embed data in the GIF and BMP files.

Information in stego image can be retrieved by S-DES algorithm. With this method, the use of neural network is the key technique. In this method the cover image is divided into group of blocks and all these blocks are generated by S-DES method. Now the embedded adjust a neural network weights with desired hidden bit code from the collection of both cover image and stego-image folder. This is common for both embedding and extracting the hidden information. We can use supervised learning of the neural network for learning. We embed the secret message within a cover image by using XOR neural network learning model. There are many structures of Artificial Neural Network including Percepton, Adaline, Madaline, Kohonen, Back Propagation and many others. Probably, Back Propagation Artificial Neural Network is the most commonly used, as it is very simple to implement and effective. In this work, we will deal with Back Propagation Artificial Neural Network has an excellent capability to simulate any nonlinear relation, so we make use of neural network to classify images . Once the network is trained, it will provide the desired output for any of the input patterns. In practice, this is the main limitation of neural network applications. And many new algorithms claimed fast convergence were developed. In this paper a single parameter dynamic search algorithm is used to accelerate network train. Each time only one parameter to be searched to achieve best performance, so this learning algorithm has a better improvement than other old algorithms. We set the number of this network's input as features, and node number of hidden level is set to be 40, and output is either yes or no. The network keeps training all the patterns repeatedly until the total error falls to some pre-determined low target value and then it stops. Note that when calculating the final error used to stop the network (which is the sum of all the individual neuron errors for each pattern) you need to make all errors positive so that they add up and do not subtract (an error of -0.5 is just as bad as an error of +0.5).

Algorithm

1. Divide the image into 8×8 blocks.
2. If a block has atleast one pixel with a gray value 0 or 255 then remove that. Then total no of block be T
3. Extract the quantization matrix *Q* from all *T* .If all the elements of *Q* are ones, the image was not previously stored as JPEG and our steganalysis method does not apply (exit this algorithm). If more than one plausible candidate exists for *Q*, the steps 4−6 need to be carried out for all candidates and the results that give the highest number of JPEG compatible blocks will be accepted as the result of this algorithm.
4. For each block *B* calculate the quantity *S*
5. If $S>16$, the block *B* is not compatible with JPEG compression with quantization matrix *Q*. If $S\leq16$, for each DCT coefficient QD_i' calculate the closest multiples of $Q(i)$, order them by their distance from QD_i', and denote them $q_p(i)$, $p=1$, …. For those combinations, for which the inequality (4) is satisfied, check if expression (5) holds. If, for at least one set of indices $\{p(1), …, p(64)\}$ the expression (5) is satisfied, the block *B* is JPEG compatible, otherwise it is not.
6. After going through all *T* blocks, if no incompatible JPEG blocks are found, the conclusion is that our steganalytic method did not find any evidence for presence of secret messages. If, on the other hand, there are some JPEG incompatible blocks, we can attempt to estimate the size of the secret message, locate the message-bearing pixels, and even attempt to obtain the original cover image before secret message embedding started.

Performance Analysis and Experiment Results

From the measured statistics of training sets of images with and without hidden information, our destination is to determine whether an image has been hidden information or not. Artificial Neural Network has the ability to adapt, learn, generalize, cluster or organize data. There are many structures of Artificial Neural Network including, Percepton, Adaline, Madaline, Kohonen, Back Propagation and many others. Probably, Back Propagation Artificial Neural Network is the most commonly used, as it is very simple to implement and effective. In this work, we will deal with Back Propagation Artificial Neural Network Neural network has an excellent capability to simulate any nonlinear relation, so we make use of neural network to classify images. In this paper we take use of BP neural network to train and simulate images. This BP neural network uses three levels: Input level, Hidden level and Output level. In neural network, the important issue is the slow of convergence. In practice, this is the main limitation of neural network applications. And many new algorithms claimed fast convergence were developed. In this paper a single parameter dynamic search algorithm is used to accelerate network train. Each time only one parameter to be searched to achieve best performance, so this learning algorithm has a better improvement than other old algorithms. We set the number of this network's input as features, and node number of hidden level is set to be 40, and output is either yes or no. A typical Back Propagation ANN is as depicted below. The black nodes (on the extreme left) are the initial inputs. Training such a network involves two phases. In the first phase, the inputs are propagated forward to compute the outputs for each output node. Then, each of these outputs are subtracted from its desired output, causing an error [an error for each output node]. In the second phase, each of these output errors is passed backward and the weights are fixed. These two phases is continued until the sum of [square of output errors] reaches an acceptable value.

III. FIGURES

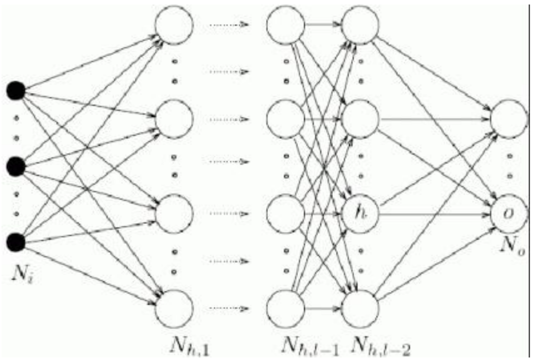


Figure 1: Back Propagation Network

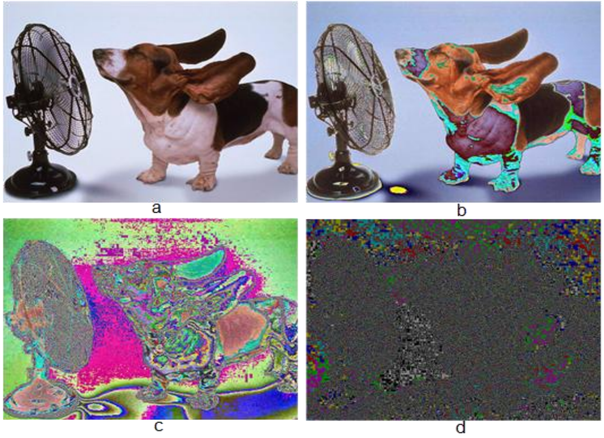


Figure 2: {a – host signal (8 bits of each color component); b – 7 significant bits; c – 5 significant bits; d – least significant bit}

IV. CONCLUSIONS

With the help of this algorithm we can check the images in the mails for the stego contents. S DES is used for standard encryption and the Back Propagation nueral network had given the training for checking all the patterns and it identify whether any stego content is hidden in the image. Once the network has been trained, it should be able to recognize not just the perfect patterns, but also corrupted or noisy versions. Each carrier media has its own special attributes and reacts differently when a message is embedded in it. Steganalysis algorithms have been developed in a manner specific to the target stego image and the algorithms developed for one cover media are generally not effective for a different media. This paper concludes that it is possible to design efficient web search algorithms to detect covert messages in corporate emails. The cover image was taken from the image database. The image was originally in JPEG format in 680x480 resolutions. Since a BMP image was also required for the evaluation, a second image in BMP format was generated using the same JPEG image. Once both the cover images have been obtained, the proposed method generates the secret code for both the images were created. The encrypted image thus obtained was steganographically concealed in the carrier image.

Image + secrets = Cover file steganography document.
The compression ratio and detection ratio of stego content is also analyzed. By analyzing the images in the sampled database the probability of occurrences of images with stego content in the corporate mails is zero.

ACKNOWLEDGEMENT

The authors would like to thank the management and members of the Dept. of IT, Karnataka State Open University, Dept. of CITE, M S University, for various algorithms used in this research, and all the people who helped in preparing and carrying out the experiments and data collection.

REFERENCES

1). Ahmed Ibrahim, 2007, Steganalysis in Computer Forensics, Security Research Centre Conferences, Australian Digital Forensics Conference, Edith Cowan University.

2). Chandramouli, R., 2002, A Mathematical Approach to Steganalysis, Proc. SPIE Security and Watermarking of Multimedia Contents IV, California.

3). Geetha ,S., Siva, S. and Sivatha Sindhu, 2009, Detection of Stego Anomalies in Images Exploiting the Content Independent Statistical Footprints of the Steganograms, Department of Information Technology, Thiagarajar College of Engineering, Madurai, , Informatica(25–40).

4). Greg Goth, 2005, Steganalysis Gets Past the Hype, IEEE, Distributed Systems Online 1541-4922 © 2005 Published by the IEEE Computer Society Vol. 6, No. 4.

5). Sujay Narayana and Gaurav Prasad, 2010, Two new approaches for secured image Steganography using cryptographic Techniques and type conversions, Department of Electronics and Communication,NITK,Surathkal, INDIA

6). Liu Shaohui, Yao Hongxun, and Gao Wen, 2003, Neural network based steganalysis in still images, Department of Computer Science, Harbin Institute of Technology, ICME.

7). Niels Provos, and Honeyman, P.,2007, Detecting steganographic content on the internet. Retrieved from http://www.citi.umich.edu/u/provos/papers/detecting.pdf

8). Samir K Bandyopadhyay, and Debnath Bhattacharyya, 2008, A Tutorial Review on Steganography, University of Calcutta, Senate House, 87 /1 College Street, Kolkata, UFL & JIITU.

9). P. T. Anitha1 , M. Rajaram2 and S. N. Sivanandham3,2011, Analysis of Detecting Stegnography contents in Corporate mails,IJRRECE

10). Z. Figov, K.Wolowelsky, and N. Goldberg, L. Bruzzone, Ed., "Co-registration of hyperspectral bands," *Image Signal Process. Remote Sens. XIII. Proc. SPIE*, vol. 6748, pp. 67480s-1–67480s-12, 2007.

11). Simple fast and adaptive image compression algorithm.,. *Roman Starosolskil,. Dec 20,* 2006, 2007, 37(1):65-91, DOI: 10.1002/spe.746

12). Recent trends in image compression and its applications., *IEEE member*
M.A Ansari and R.S Anand. XXXII NATIONAL SYSTEMS CONFERENCE, NSC 2008, December 17-19, 2008

13). R.Moorhead, S.Rajala .Motion-compensated interframe coding., *Proc. ICASSP*, pp. 347-350, 1985.

14). The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS Marcelo J. Weinberger and Gadiel Seroussi Hewlett-Packard Laboratories, Palo Alto, CA 94304, USA Guillermo Sapiro Department of Electrical and Computer Engineering University of Minnesota, Minneapolis, MN 55455, USA.

Authors

Sudha .D completed M.Tech Degree from Karnataka State Open University (KSOU), India in 2012, received MCA degree from Cochin University College of Engineering (CUCEK), Kuttanadu, Cochin Universty of Science and Technology (CUSAT), in the Year 2005. Her research interest include Digital Image Processing , AI (Artificial Intelligence) and Data Mining.

Mariadas Ronnie C.P received M.Tech CIT Degree from M.S. University, Coimbatore India in 2011, completed M.Phil IT degree from M.S. University in the year 2012 and received MCA degree from Bharathiar University, Coimbatore in the year 2001. His research interest includes Digital Image Processing, Artificial Intelligence and Data Mining.