ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, No 5, May 2013

Secure Data Transmission using Watermarking and Image Compression

Dr. Ajit Singh, Meenakshi Gahlawat

***Abstract*— In today's digital world, exchange of information is been held electronically. Therefore there arises the great need for secure transmission of the concerned data. Various practices like cryptography, watermarking, compression etc. are common since past few years. All these techniques were proved to be excellent in their respective work regarding security. In this paper a new approach has been presented to provide security at enhanced level. Here the two techniques namely watermarking and compression are combined together to enhance the level of security for data transmission purpose. Here the watermarking using DCT technique is been combined with image compression technique using improved adaptive Huffman encoding. The improved adaptive Huffman coding technique is been based on Huffman algorithm. This new compression algorithm not only reduces the number of pass required to encode the data but at the same time reduces the storage space in comparison to adaptive Huffman and static Huffman respectively.**

***Index Terms*—Digital watermarking, Image Compression, Improved adaptive Huffman, Spatial domain**

I. INTRODUCTION

Watermarking is one of the most important aspect related with information hiding. It is a process that is been used to embed some kind of information inside a guest content: the guest file can be a multimedia content such as picture, audio or video. It is basically used for the purpose of copyright protection and owner authentication. Digital Watermarking technique gets its name from watermarking, which is very common since past several years. Digital watermarking is a technique that provides solution to the many longstanding problems related with copyright of digital data that can be detected or extracted later to make out some statement about the data. This information can be textual data about the author, its copyright, etc; or it can be an image itself. The information that needs to be hidden is embedded by manipulating the contents of the digital data, allowing someone to identify the original owner, or in the case of illicit duplication of purchased material, the buyer involved. These digital watermarks remain inviolate under the conditions related with transmission/ transformation, allowing one to protect the ownership rights in digital form. Digital watermarking has become an active and important area of research, and development and commercialization of watermarking techniques is being deemed essential to help address some of the challenges faced by the rapid proliferation of digital content. Digital watermarking came to be in great demand when sharing information on the Internet became a usual practice. [1], [3]

Manuscript received May, 2013.

Dr. Ajit Singh, Department of Computer science and Engineering, School of Engineering and sciences, BPSMV, Khanpur Kalan, Sonepat, India

Meenakshi Gahlawat, Department of Computer science and Engineering, School of Engineering and sciences, BPSMV, Khanpur Kalan, Sonepat, India

A. Classifications of Watermarking

1) Visible Watermarks

Visible watermarks are those watermarks which can be easily perceived by the viewer, and clearly identify the owner. The visible watermarks are viewable to the normal eye such as bills, company logos and television channel logos etc. This type of watermarks can be easily viewed without the requirement of any mathematical calculation but at the same time these embedded watermarks can be destroyed easily. [3]

2) Invisible Watermarks

Invisible watermarks are those watermarks that cannot be perceived by human eyes. This type of watermark is not visible in the watermarked image without degradation of image or data. Invisible watermark may be any logo or any signature**.** Most research currently focuses on invisible watermarks, which are imperceptible under normal viewing conditions. [3]

B. Requirements of Watermarks

The major requirements of digital watermarking are:

1) Transparency:

The watermark that is been embedded should not degrade the original image quality. And in rare case if any distortions are visible in the image it tends to degrade the commercial value of the image. [1]

2) Robustness:

Robustness is simply the notion of how much can be done to the watermarked image in the form of attacks (deliberate and otherwise), such that the watermark can still successfully be extracted from that altered image. In general, a more robust watermark is preferred to one that is less so. [1]

3) Capacity or Data Load:

This quantity describes the maximum amount of data that can be embedded into the image to ensure correct removal of watermark during extraction. [1]

C. GENERAL WATERMARKING SYSTEM

The digital watermarking system essentially consists of a watermark embedder and a watermark detector. The watermark embedder inserts a watermark onto the cover signal and the watermark detector detects the presence of watermark signal. An entity called watermark key is used during the process of embedding and detecting watermarks. The watermark key has a one-to-one correspondence with watermark signal (i.e., a unique watermark key exists for every watermark signal).The watermark key is private and known to only authorized parties and it ensures that only authorized parties can detect the watermark. Further, the communication channel can be noisy and hostile (i.e., prone to security attacks) and hence the digital watermarking techniques should be resilient to both noise and security attacks. [3]

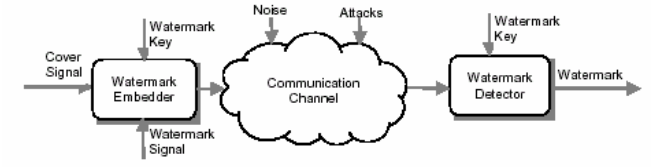

Fig1: Digital Watermarking Systems

D. TECNIQUES OF DIGITAL WATERMARKING

1) Spatial Domain Method

Spatial-domain method is been used for embedding the watermarks into a particular text, image by directly changing the pixel values of original host images. Some common spatial-domain algorithms include Least Significant Bit (LSB) Modification, Patchwork, Texture Block Coding, etc. The most serious drawback of spatial-domain technologies is that it tends to provides limited robustness. It is complex for spatial-domain watermarks to subsist under attacks such as lossy compression and low-pass filtering. Also the amount of information that can be embedded in spatial domain is also very limited. [5]

2) Frequency-Domain Technologies

In comparison to spatial-domain watermark, watermarks in frequency domain are more robust and much more compatible to popular image compression standards. Thus frequency-domain watermarking technique is more widely used and obtains more attention in comparision to spatial domain method. To embed a watermark, a frequency transformation needs to be applied to the host data. Then, modifications are made to the transform coefficients. Possible frequency image transformations include the Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) and others. In recent years they are becoming generally desolated. [5]

E. APPLICATION OF WATERMARKING

1) Copyright Protection

This is one of the most prominent application of watermarks. Due to huge exchange of images over insecure networks, copyright protection becomes a very important issue. Watermarking an image will prevent its redistribution.

2) Authentication

In some cases there arises the need to identify the ownership of the contents. All this can be done by embedding a watermark and providing the owner with a private key that gives him an access to the message. ID cards, ATM cards, credit cards are all examples of documents that require authentication.

3) Broadcast Monitoring

From the name it is clear that broadcast monitoring is been used to verify the programs that are broadcasted on TV or radio. It especially helps the advertising companies to see if their advertisements appeared for the right duration or not.

4) Content Labeling

Watermarks can be used for providing more information about the cover object. This process is named content labeling.

5) Tamper Detection

Fragile watermarks can be used to detect tampering in an image. If the fragile watermark is degraded in any way then we can say that the image or document in question has been tampered.

6) Digital Fingerprinting

This is a process that is been used for detecting the owner of the content. This is so because every fingerprint is the unique characteristics of the owner.

7) Content protection

In this process the content stamped with a visible watermark that is very difficult to remove so that it can be publicly and freely distributed. [1]

II. IMAGE COMPRESSION

It is defined as the reduction of amount of data used to represent an image by reducing redundant data, so that the image can be stored or transferred more efficiently. Image compression is minimizing the size in bytes of a graphics file without degrading the quality of the image to an unacceptable level. The reduction in file size allows more images to be stored in a given amount of disk or memory space. It also reduces the time required for images to be sent over the Internet or downloaded from Web pages. [9]

Compression is achieved by the removal of one or more of the three basic data redundancies:

1. Coding Redundancy
2. Interpixel Redundancy
3. Psychovisual Redundancy

Coding redundancy is present when less than optimal code words are used. Interpixel redundancy is due to correlations between the pixels of an image. Psychovisual redundancy occurs when the data is neglected by the human visual system (i.e. visually non essential information). Image compression techniques reduce the number of bits required to represent an image by taking advantage of these redundancies. An inverse process called decompression (decoding) is applied to the compressed data to get the reconstructed image. The

objective of compression is to reduce the number of bits as much as possible, while keeping the resolution and the visual quality of the reconstructed image as close to the original image as possible. [10]

A. Benefits of Compression

1) Storage Space

Storage space, such as that provided by computer hard drives, comes at a price. Compression of the data files allows to store more files in the storage space that is been available. Lossless compression, used in zip file technology, will typically reduce a file to 50 percent of its original size. However, no difference is seen in the file size if zip files are already in a compressed format, such as MP3 audio files or PDF (Portable Document Format) text-only files. [6]

2) Bandwidth and Transfer Speed

The download process uses network bandwidth whenever we download a file, such as an MP3 audio file, from a server on to the Internet. Bandwidth is the speed at which the network transfers data and is measured in terms of Mbps (megabits per second). Compressed files contain fewer "bits" of data than uncompressed files, and, as a consequence, use less bandwidth when we download them. This means that the transfer speed, that is to say the time taken by the file to be downloaded, is faster. It will take 10 seconds to download a file if we have bandwidth of 1Mbps available, and we are downloading a file that is 10Mb (megabits) in size. It will only take 5 seconds to download the file if the file is compressed to 5Mb. [6]

3) Cost

The costs of storing the large amount of data are reduced by compressing our files for storage because by doing so we can store more files in the available storage space when they are compressed. We need to buy a second 250MB drive if we have 500MB (megabytes) of uncompressed data and a 250MB hard drive on which to store it. We will not need to buy the extra hard drive if we compress the data files to 50 percent of their uncompressed size. This saving can be applied to the costs of maintaining an Internet connection. Many contracts with Internet Service Providers (ISP) include charges for the amount of data that you download. Download compressed files, and by doing so, one will be downloading much less data than one would do so when uncompressed files are been downloaded. The Internet download charges will be less as a consequence of this. [6]

4) Accuracy

It also reduces the chance of occurrence of transmission errors since fewer bits are transferred. [10]

5) Security

It also provides a level of security against illegal monitoring. [10]

B. IMAGE COMPRESSION TECHNIQUES

The image compression techniques are broadly classified into two categories depending whether or not an exact repro of the original image could be reconstructed using the compressed image.

These are:

1. Lossless technique
2. Lossy technique

1) Lossless compression

It is a compression technique that does not lose any data in the compression process. Lossless compression "packs data" into a smaller file size by using a kind of internal shorthand to signify redundant data. If an original file is 1.5MB (megabytes), lossless compression can reduce it to about half of that size, depending on the type of file that is being compressed. This makes lossless compression convenient for transferring files across the Internet, as smaller files transfer faster. Lossless compression is also handy for storing files as they take up less room. The zip convention, used in programs like WinZip, uses lossless compression. For this reason zip software is popular for compressing program and data files. That's because when these files are decompressed, all bytes must be present to ensure their integrity. If bytes are missing from a program, it won't run. If bytes are missing from a data file, it will be incomplete and falsified. GIF image files also use lossless compression.

Lossless compression has advantages as well as disadvantages. The advantage is that the compressed file will decompress to an exact duplicate of the original file, mirroring its quality. The disadvantage is that the compression ratio is not all that high, precisely because no data is lost. [7]

Following techniques are included in lossless compression:

1. Run length encoding
2. Huffman encoding
3. LZW coding
4. Area coding

2) Lossy Compression

It is a compression technique that does not decompress data back to 100% of the original. Lossy methods provide high degrees of compression and result in very small compressed files, but there is a certain amount of loss when they are restored.

Audio, video and some imaging applications can tolerate loss, and in many cases, it may not be noticeable to the human ear or eye. In other cases, it may be noticeable, but not that critical to the application. The more tolerance for loss, the smaller the file can be compressed, and the faster the file can be transmitted over a network. Examples of lossy file formats are MP3, AAC, MPEG and JPEG. Lossy compression is never used for business data and text, which demand a perfect "lossless" restoration. [8]

Lossy schemes tend to provide much higher compression ratios than lossless schemes. Lossy schemes are widely used since the quality of the reconstructed images is adequate for most applications. By this scheme, the decompressed image is not identical to the original image, but reasonably close to it. [10]

Lossy compression techniques includes following schemes:

1. Transformation coding

2. Vector quantization
3. Fractal coding
4. Block Truncation Coding
5. Sub band coding

III. PROPOSED SYSTEM

Neither watermarking nor compression can alone make the data transmission completely secure. Hence the new technique is been proposed in order to achieve secure transmission of data by making combine use of image watermarking using DCT technique and then applying image compression technique using improved adaptive Huffman algorithm to it.

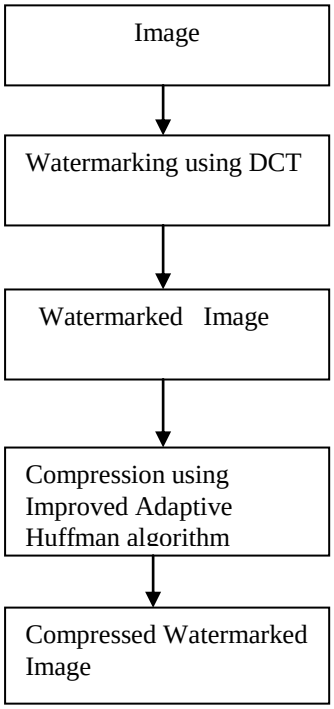

Fig. 2 "Flowchart showing the proposed technique"

A. *Image Watermarking Using DCT*

The DCT allows an image to be split up into the different frequency bands, making it convient for embedding the watermarking information into the middle frequency bands of an image. The middle frequency bands are chosen such that they have minimized they avoid the most visual important parts of the image (low frequency) without over-exposing themselves to removal through compression and noise attacks. [2]

1) Steps of DCT watermarking

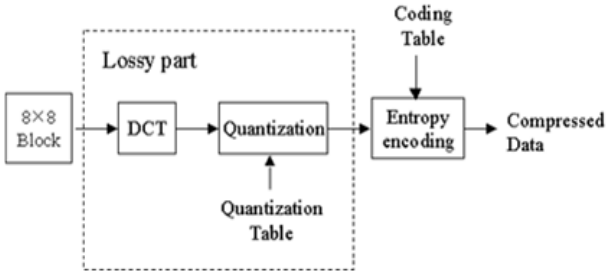

Fig. 3 "DCT Watermarking"

1.Transform the RGB color of the original image into the formation of Gray color.

2.Then the image is divided into 8 × 8 blocks by applying JPEG standard as below.

3.Transform the original 8 x 8 block into a cosine-frequency domain

- C(h) = if (h == 0) then 1/sqrt(2) else 1.0--C(h) is a auxiliary function been used in main function F(u,v)
- F(u,v) = ¼ x C(u) x C(v) Σx=0..7 Σy=0..7 Dxy x cos(π(2u + 1)u/16) x cos(π(2y + 1)v/16)

-Gives encoded pixel at row u, column v
-Dxy is original pixel value at row x, column y
-F(u,v) is new matrix value after DCT apply. [2]

2) Extracting Watermarked Image

Perform DCT transformation on watermarked image and the original host image. After doing so, subtract original host image from watermarked image. And finally do multiplication of extracted watermark by scaling factor to display. [2]

3) Advantages

1. DCT domain watermarking is comparatively much beneficial than the spatial domain encoding since DCT domain watermarking is able to survive against the attacks such as noising, compression, sharpening, and filtering.
2. It uses JPEG compression method to apply DCT watermarking as a parameter. One is able to use different kinds of parameters related to image processing and these parameters might provide equal or even stronger validity against various attacks based on image processing.
3. In Discrete cosine transform (DCT), pseudorandom sequences, such as M sequences, are been added at the middle frequencies of the DCT as signatures**. [2]**

After image has been watermarked it is been compressed using a compression algorithm that is based on Huffman coding. Huffman coding is one of the lossless compression techniques. Huffman algorithms have two ranges static as well as adaptive. Static Huffman algorithm is a technique that encodes the data in two passes. In first pass the frequency of each symbol is been calculated and in second pass a Huffman tree needs to be constructed. Adaptive Huffman algorithm is extended on the basis of Huffman algorithm that constructs the Huffman tree in one pass but take more space in comparison to Static Huffman algorithm. The algorithm that is used not only reduces the number of pass but also reduce the storage space needed in comparison to adaptive Huffman algorithm and comparable to static.

4) Static Huffman algorithm

The Static Huffman algorithm was developed by David Huffman in (1952) is used to generate the encoded data in two passes that are as follows:

1. The frequency of each of the different symbol present in the source data is been calculated. After calculating frequencies the table of all frequencies is been constructed in decreasing order by sorting each of the different symbol in decreasing order.

2. Then the Huffman tree is generated by combining the least two symbols into one composite symbol.

But in this method some time the source data available is so lengthy that it takes so much time to construct a table which in turns waste a lot of time as well as space required to store the table. [4]

5) Adaptive Huffman Algorithm

Expanding the static Huffman algorithm, Faller and Gallagher [Faller 1973; Gallagher 1978], and later Knuth [Knuth 1985] and Vitter [Vitter 1987], developed a way to perform Static Huffman algorithm by using one pass method that is as follows:

Initially Adaptive Huffman algorithm generates a Huffman tree with all different symbols with frequency count to one and then it takes the code for first symbol in the source data. For the second symbol it generates the second Huffman tree and takes the code for second symbol and so on till last bit (byte) of source data. [4]

Concept

The basic concept behind an adaptive compression algorithm is very simple:

1. At the beginning the model is to be initialized
2. Repetition is done for each character
3. {
4. Then encoding is done for the character
5. Finally the model is updated
6. }

Decompression also works in the same way. As long as both sides have the same initialize and update model algorithms, they will have the same information. [4]

In Adaptive Huffman algorithm, at the time of encoding of symbols, we need to update the tree after each symbol is encoded. Same thing is also been performed in decoding the code. That means there is processing overburden that is involved in the process. The encoded data by Adaptive Huffman algorithm requires more space than Static Huffman encoded data. The other major drawback of the Adaptive Huffman algorithm for encoding the data is that it requires in advance the amount of different symbols that are present in the source data. So it first needs to scan all the source data to determine that how many different symbols are present in the source data. Besides all this some other major drawbacks of adaptive Huffman algorithm are as follows:

- It is very time consuming, as it first constructs the tree and then take the code for the symbol, for the next symbol it adopt the same procedure and same is repeated (up to the last symbol).
- In adaptive Huffman algorithm many different symbols have same code in the encoded data that creates a lot of confusion while decompressing the data.
- In adaptive Huffman same symbol that occurs frequently has different code which can create a lot of confusion while decompressing the data.
- Finally while decompressing the data we need all trees, for smaller data it is ok but for large data it demands a huge storage space.

Thus to overcome all these shortcomings we have new compression algorithm named Improved adaptive Huffman. Improved Adaptive Huffman algorithm is based on existing Huffman algorithm. [4]

B. Improved Adaptive Huffman

It have one pass in comparison to the existing static Huffman algorithm and at the same times requires less space for storing the encoded data as compared to adaptive Huffman algorithm. The proposed method with this algorithm is as follows:

At initial step Improved adaptive Huffman algorithm will generate a strictly binary tree by reading first symbol from the source data, then for the next symbol it generates a tree and so on up to last symbol of source data. On reading the last symbol it makes the final Huffman tree. [4]

1) Advantages

Advantages of Improved Adaptive Huffman over Adaptive Huffman are:

- Improved adaptive Huffman utilizes less space to store the compressed data.
- It saves the time because here, there is no need to scan the whole string for constructing the first tree. It also saves the time while constructing trees e.g. it needs only one symbol for constructing the first tree unlike in adaptive Huffman that requires all different symbols to construct the tree.
- In Improved adaptive Huffman even if one symbol occurs frequently will tend to have same code.
- In improved adaptive Huffman, while constructing the next tree there is no need to remember the previous tree.
- Finally during the process of decompressing the only final tree is needed.

2) Algorithm

1. Scan the first Symbol and initialize its frequency to 1
2. Then next symbol is been scanned from the source data

If any previous symbol = next symbol **then**
the frequency of that previous symbol needs to be incremented
If any previous symbol frequency < recently incremented symbol frequency
then
Both nodes needs to be interchanged
Else
Initialize their frequency to 1

3. Create strictly binary tree with left and right node (Left or Right node can be NULL). The root is the composite Symbols of left and right nodes. Assign value 0 to Right node and 1to Left node.
4. Step 2 to 4 needs to be repeated till End of Source data is been reached. [4]

By the use of this algorithm the storage space will be reduced and time is also been saved. Thus the combination of the two processes watermarking and compression will results in providing high security level to the data to be transmitted.

IV. IMPLEMENTATION AND RESULT

The above described work is implemented in JAVA. Firstly visible watermarking using DCT technique is used to watermark an image. After that improved adaptive Huffman algorithm is used to compress the watermarked image to further enchance security needs of the system.

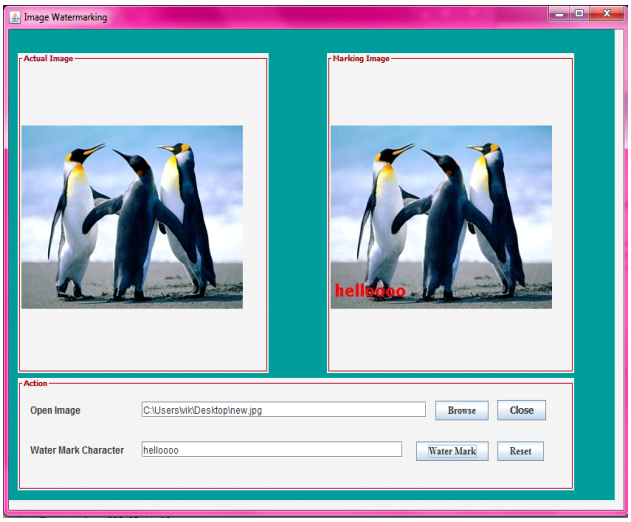

Fig. 4 Visible Watermarking using DCT

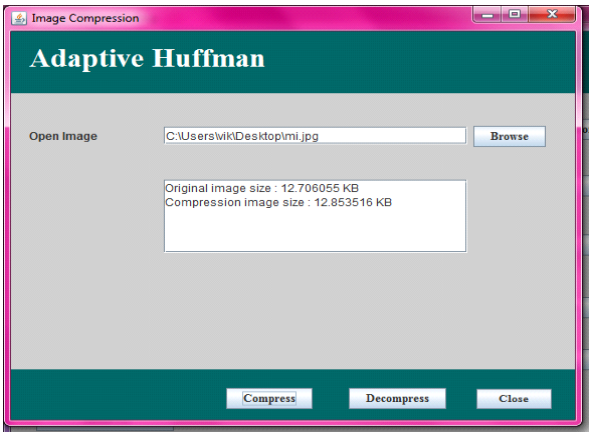

Fig. 5 Compression using adaptive Huffman algorithm

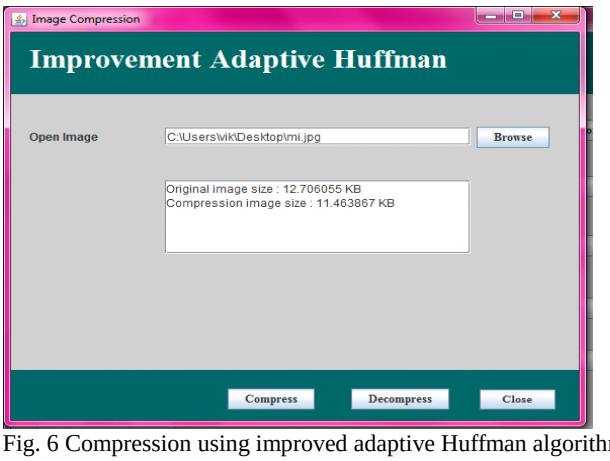

Fig. 6 Compression using improved adaptive Huffman algorithm

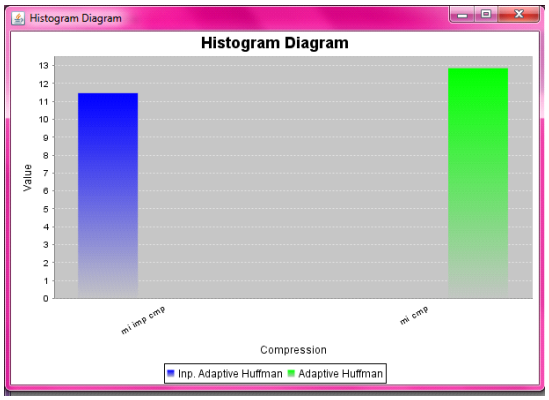

Fig. 7 Comparison graph for adaptive Huffman and improved adaptive Huffman

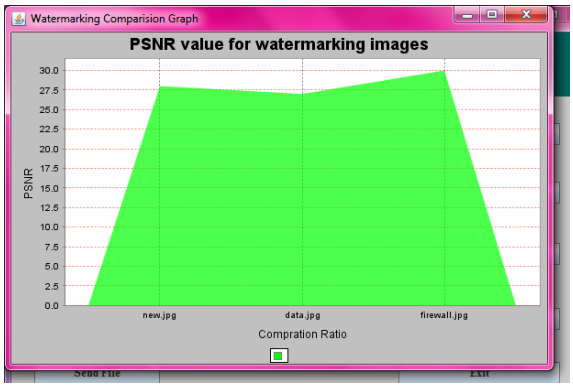

Fig. 8 PSNR value of watermarked image

V. CONCLUSION

Firstly watermarking is been discussed, its types, requirements, its various applications, techniques. Secondly the image compression is been discussed with its techniques. And finally the new technique watermarking using DCT combined with image compression using improved adaptive Huffman algorithm is been presented to enhance the level of security for the data to be transmitted.

ACKNOWLEDGMENT

I would like to give my sincere gratitude to my guide Dr. Ajit Singh who guided me throughout, to complete this topic.

REFERENCES

[1] http://www.google.com/#hl=en&sclient=psy-ab&q=Abrar+Ahmed+Syed_Digital+Watermarking.pdf&oq=Abrar+Ahmed+Syed_Digital+Watermarking.pdf&gs_l=hp.3...15306.18741.1.19731.5.5.0.0.0.0.1.2100.4803.2-1j6-1j0j1j1.4.0...0.0...1c.1.12.psy-ab.412g3Zw_g5g&pbx=1&bav=on.2,or.r_qf.&bvm=bv.46226182,d.bmk&fp=4fb3eea26210152c&biw=1366&bih=634J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," *IEEE Trans. Electron Devices*, vol. ED-11, pp. 34-39, Jan. 1959.

[2] Darshana Mistry / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, 2905-2909C. Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, and M. Miller, "Rotation, scale, and translation resilient public watermarking for images," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 767-782, May 2001

[3] Watermark Attacks And Applications in Watermarking in National Workshop-Cum-Conference on Recent Trends in Mathematics and

Computing (RTMC) 2011 Proceedings published in International Journal of Computer Applications® (IJCA)
[4] Improved Adaptive Huffman Compression Algorithm in International Journal of Computers & Technology Volume 1 No.1 Dec. 2011
[5] A Survey of Digital Watermarking Technologies www.ee.**sunysb.edu**/~cvl/ese558/.../**Lin**%20**Liu**/ese558report_**LinLiu**.pdf
[6] www.ehow.comInterne
[7] twww.wisegeek.com/what-is-**lossless-compression**.htm
[8] www.pcmag.com/encyclopedia/term/46335/**lossy-compression**
[9] searchcio-midmarket.techtarget.com/**definition**/**image-compressio**
[10] *www.rimt**engg**.com/coit2007/proceedings/pdfs/43.pdf* "A STUDY OF VARIOUS IMAGE COMPRESSION TECHNIQUES

Dr. Ajit Singh is presently working as Chairperson of School of Engineering & Sciences in BPSMV, Khanpur Kalan (Sonepat). He is also having the additional charge as a Director of University Computer Center (UGC). He posses qualifications of B.Tech, M.Tech, Ph.D. He is a member of BOG (Board of Governors) of Haryana State Counselling Society, Panchkula and also member of academic council in the University. He published approximate 20 papers in National/ International journals and conferences and holds a teaching experience of approximate 10 years. He holds the membership of Internal Quality Assurance cell, UG-BOS &PG-BOS and the NSS advisory committee. He is also an associate member of CSI & IETE. His research interests are in Network Security, Computer Architecture and Data Structure.

Ms. Meenakshi Gahlawat has completed her B.Tech degree in Computer Science from Maharishi Dayanand University, Rohtak in year 2011. She is pursuing M.Tech in Computer Science from BPSMV, Khanpur Kalan from July 2011. Her research interests are in Network Security and Computer Networks.