International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:5, 2007

Attack Detection through Image Adaptive Self Embedding Watermarking

S. Shefali, S. M. Deshpande, and S. G. Tamhankar

***Abstract*—**Now a days, a significant part of commercial and governmental organisations like museums, cultural organizations, libraries, commercial enterprises, etc. invest intensively in new technologies for image digitization, digital libraries, image archiving and retrieval. Hence image authorization, authentication and security has become prime need. In this paper, we present a semi-fragile watermarking scheme for color images. The method converts the host image into YIQ color space followed by application of orthogonal dual domains of DCT and DWT transforms. The DCT helps to separate relevant from irrelevant image content to generate silent image features. DWT has excellent spatial localisation to help aid in spatial tamper characterisation. Thus image adaptive watermark is generated based of image features which allows the sharp detection of microscopic changes to locate modifications in the image. Further, the scheme utilises the multipurpose watermark consisting of soft authenticator watermark and chrominance watermark. Which has been proved fragile to some predefined processing like intentinal fabrication of the image or forgery and robust to other incidental attacks caused in the communication channel.

***Keywords*—**Cryptography, Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) ,Watermarking.

I. Introduction

THE cost of computers, printers and digital transmission has made digital media increasingly popular over the conventional analog media. However, digital media also causes extensive opportunities for mass piracy of copyrighted material. It is therefore very important to have ways and means to detect copyright violations and control access to digital media. *Data-Hiding* or *Steganography*, is a rapidly growing field with potential applications for copyright protection, hiding executables for access control of digital multimedia data, embedded captioning, secret communications, tamper detection etc.

These novel techniques use digital watermark which is a sequence of binary information, containing the owner's rightful information for the protected multimedia data [1].

Manuscript received June 15 2007. This paper is the part of the project supported by Department of Science and Technology, Delhi, India [SR/FTP/[ETA]-35/2005] in the area of Digital Image Watermarking.

S. Shefali is a Ph.D student working as a lecturer at Department of C.S.E. at Walchand College of Engineering, Sangli, India (Corresponding author to provide phone:+91 09850574868; fax:+91 233 2300831; e-mail: k_shefali@yahoo.co.in).

S. M. Deshpande is a Professor at Department of C.S.E. at Walchand College of Engg., Sangli, India (e-mail: s_m_deshpande@yahoo.com).

S. G. Tamhankar is with Department of Electronics at Walchand College of Engg., Sangli, India (e-mail:sgtamhankar@yahoo.com).

It is a visible or an invisible mark inserted into digital multimedia data so that it can be detected in the later stage for the evidence of rightful ownership protection. A great deal of research efforts has been focused on digital watermarking in recent years. Fig. 1 gives idea of invisible digital image watermarking.

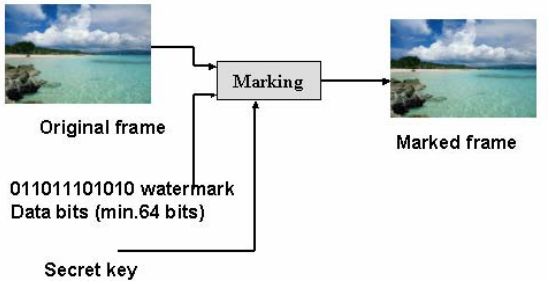


Fig. 1 Invisible Watermarking in an Image

Following this direction, the paper focuses on authentication and security of digital color image. Unlike, normal watermarking method, instead of some external watermark, here we embed an approximation of the image as a wateramrk into itself for unique authentication.

II. Soft Authentication

"Soft" image authentication is different from "Hard" image authentication. "Soft" authentication is more forgiving of small non-content distortion than "Hard" authentication is. The content of image is the logical relations among pixels. For example, the image after lossy processing such as JPEG could be found to be authentic by "Soft" image authentication, but it would fail "Hard" image authentication. For a "Hard" image authentication, one bit error in the message leads to a totally different authenticator, however, for a "Soft" image authentication, such an error does not necessarily alter the authenticator. "Hard" image authentication is highly sensitive and dependent on the exact value of image pixels, whereas "Soft" image authentication is sensitive just to content modification and serious image quality tampering. "Soft" image authentication is ideally dependent on the logical content-based, non-variant relation among image pixels.Authentication is the service of ensuring whether a given block of data has integrity (i.e. the associated content has not been modified) and is from the legitimate sender [2][3].

A. Self Embedding Technique - SET

In self embedding technique the image features are embedded into itself as an authentication stamp. The color and chrominance information based features of the image are extracted for the generation of the watermark. The proposed approach generates an image-dependent watermark using SET. Generated watermark will be unique for a set of image and cryptographic key used for security. The use of image-dependent watermark provides better security against fraud as compared to traditional schemes in which a random sequence or logo is used as a watermark for all images [4].

B. Contribution and Scope

The aim of this paper is to demonstrate the application of semi-fragile watermarking through SET for authentication and security of digital color images.

Section III formulates the joint authentication and watermarking problem using digital watermarking with data hiding framework. Our proposed solution is presented and analyzed under ideal conditions in Sections IV and Section V, respectively. Simulation and test results of the algorithmic behaviour are presented in Section VI followed by final remarks and conclusions.

III. Problem Formulation and Preliminaries

A. Formulation and Framework

Digital watermarking research has been proposed for a diverse set of applications including copy protection, image authentication, video error correction, and color image compression. Our system for authentication and security of digital color images consists of the following components:

1. Generating Function

It produces the watermark signal W to embed as follows:

$$W = f_g(l,k,Y)$$

Where k is the secret generation key known only to sender and receiver, Y is luminance of the host image, and l is watermark "payload" comprised of a bit sequence independent of k and Y. In our application, W has two parts: an authenticator watermark component W_a employed for security and a chrominance watermark component W_c to help with compression. We represent this relationship by: $W = [W_a \,||\, W_c]$ Where $||$ is the concatenation operator.

2. Embedding Function

It inserts the watermark signal W into the luminance host data Y with the help of a secret embedding key K known only to the sender and receiver, yielding the watermarked data Y^w.

$$Y^w = f_m(Y,W,K)$$

3. Extracting Function

It recovers the watermark information W from the received watermarked data Y^r, using the secret key K as follows:

$$W = f_x(Y^r,K)$$

4. Recovery Function

It employs W for authentication and color recovery of the image as follows:

$$[R_a, X_w] = f_r(Y^r,W,k')$$

Where k' is the key available to receiver, R_a is the statistic that allows the application-dependent authentication and tamper assesment of the received luminance image Y^r and X_w is overall color recovered version of Y^r.

B. Design Principles

Based on the analysis of strengths and limitations of semi-fragile watermarking, the following are helpful principles for system function design:

1. Authenticator Watermark: The authenticator watermark W_a should be a secure content-based adaptive authenticator. It should be a function of image features that are invariant to predefined content preserving image processing operations given by Ω_R and fragile to specified content modification attacks denoted by Ω_F.
2. Uniqueness of Authenticator Watermark Generation: Different values of k should produce distinct authenticator watermarks for the same host image X, to guarantee key-based security of W_a.
3. Chrominance Watermark: The component of the payload l corresponding to the chrominance watermark W_a should contain a version of the color information so that it can be later combined with the watermarked luminance image for color recovery.
4. Non-invertibility of Embedding: The keys K and k must not be easily identifiable even if the embedding method f_m and W are both known to attackers. Thus, authentication and embedding security lies in the secrecy of the key.
5. Chrominance Embedding and Lossy Compression: To achieve overall compression gains, chrominance embedding and lossy compression must work together. The chrominance information is no longer necessary to store separately, since it can be extracted from the luminance image. Thus overall volume of information is reduced.
6. Authenticator Generation and Embedding: For authentication, the watermark embedding should not affect the watermark generation else it can be shown that even under ideal situations, authentication is impossible because the changes imposed on host to embed the authenticator will render the image unauthentic.
7. Robustness and Fragility: The embedding and extracting functions f_m and f_x should together be robust to the image processing operations, specified by Ω_R and possibly fragile to malicious content changing attacks defined in Ω_F. This provides the necessary "soft-authentication" capability.

C. Orthogonality and Dual Domains

We use linear orthogonal separable transforms that work in orthogonal domains of the image for watermark generation and embedding [5]. This approach allows independent design

and analysis of the various system functions like watermark generation and watermark embedding. This method is used to overcome the drawbacks of some traditional digital watermarking techniques like:

- Employing a single domain host dependent watermark: In this, an image-dependent watermark generation and embedding are "mixed" in the same domain, it suffers from high sensitivity or inability to appropriately localize the degradations on the signal.
- Employing a host independent watermark: In this, watermark is a random sequence or logo independent of the image and embedded in a given domain. It requires the transmission of the watermark itself along with the image, which makes it susceptible to evesdropping and sophisticated attempts of fraud.

IV. Algorithm

A. Design Guidelines

Soft authentication should forgive high quality compression (e.g. higher than 0.5 bpp for JPEG), very small proportion (say below 1%) of random bit errors from transmission error, mild additive noise (such as salt and pepper below 1% and white Gaussian above 30 dB SNR) and mild linear filtering; these distortions collectively form Ω_R.

In contrast, the method should recognize severe compression (e.g. lower than 0.5 bpp in JPEG) that impedes image quality, forgery of the entire image, addition, removal or changes in spatially localized visual features; these attacks collectively form Ω_F.

B. Watermark Generation

To generate both components of the watermark, we first transform the host color image X into the YIQ color space to obtain the luminance Y and the chrominance images I and Q jointly representing saturation and hue.

Fig. 2 explains the watermark generation process. The chrominance watermark is created by taking the lowest resolution bands resulting from the second level Haar DWT [6]-[8] of both I and Q because subsampling chrominance has little visual effect on the overall color image; these bands are denoted by I_{2LL} and Q_{2LL}, respectively.

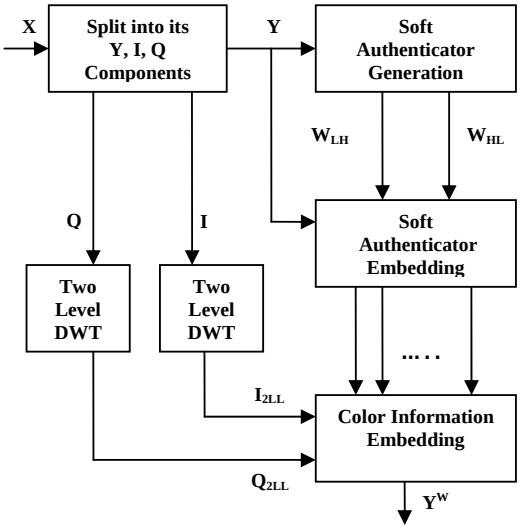


Fig. 2 Framework

Thus, the overall chrominance watermark is given by $W_c = [I_{2LL} \parallel Q_{2LL}]$. Generation of the authenticator watermark requires the use of a one-time only secret session key K_S known to both the sender and receiver.

The repeated use of the session key is employed for protection against block analysis, traffic analysis and replay attacks. In addition, another key (either private or secret) is used to perform asymmetric or symmetric encryption.

To provide "soft-authenticator" capabilities, the algorithm we propose, detailed in Fig. 3, consists of following stages:

- DCT and Feature Extraction: It identifies perceptually significant components of the image. The proposed measure is the dc coefficient of the 8x8 DCT blocks of the image.
- Binary Transform: It order-pairs dc coefficients so that their relative magnitudes are guaranteed to be maintained under content-preserving operations such as JPEG or SPIHT compression. The binary output of this stage is one component of the authenticator watermark denoted by W_{LH}.
- Permutation: It provides better security against fraud or forgery.
- Majority Function: It reduces the size of the output of the permuted binary transform, so as to make it more robust to content preserving operations.
- Map Function: It converts the output of the previous step to an appropriate size for encryption and subsequent watermarking.
- Encryption: it creates a watermark component denoted by W_{HL}, for sender authentication.

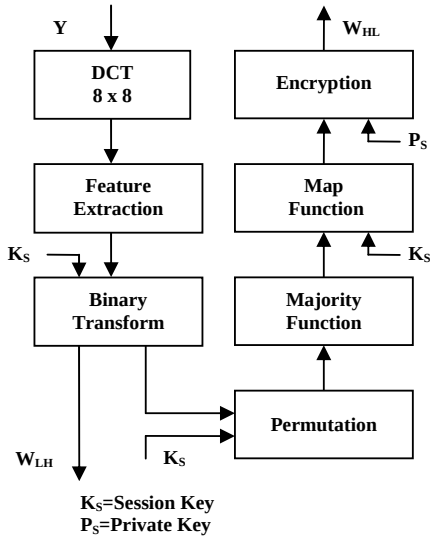


Fig. 3 Watermark Generation

To summarize, W_{HL} provides crucial cryptographic security and W_{LH} provides attack characterization capability to balance the requirements of tamper assessment. Therefore, $W_a = [W_{LH} \,||\, W_{HL}]$ and $W_c = [I_{2LL} \,||\, Q_{2LL}]$. Here W_{HL}, W_{LH}, I_{2LL} and Q_{2LL} are embedded in orthogonal bands of the luminance Haar DWT. Thus, the technique can be efficiently used for both i.e. better security as well as authentication of digital color images [9].

C. Watermark Embedding

The embedding process takes place in the Haar DWT domain [5] which we consider a "dual" to the DCT domain used for watermark generation.

Whole procedure is summerised in Fig. 4, in this two level Haar DWT of chrominance (i.e. Y) is taken and the resulting Y_{2LH} and Y_{2HL} bands are embedded with W_{LH} and W_{HL} respectively.

The subsampled chrominance components, I_{2LL} and Q_{2LL}, are embedded by simply replacing the specific Haar domain LL bands of Y_{LH} and Y_{HL}, respectively, thus obtaining $Y^e{}_{LH}$ and $Y^e{}_{HL}$.

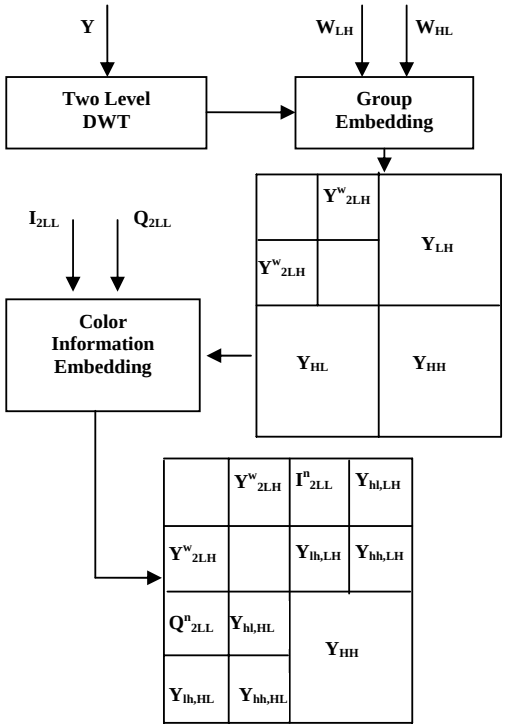


Fig. 4 Watermark and Color Information Embedding

D. Watermark Extraction, Authentication and Color Recovery

At the receiver side, image authentication and color recovery are performed. It is assumed that the receiver knows k' which includes the associated session key K_S and decryption key K_R.

Fig. 5 gives the procedure at the receiving side. The authentication watermarks are extracted from $Y^r{}_{2LH}$ and $Y^r{}_{2HL}$ bands of Y^r and are denoted by $W^e{}_{LH}$ and $W^e{}_{HL}$. These watermarks must be effectively compared to a corresponding set generated from Y^r for authentication and tamper assessment. Watermarks denoted $W^r{}_{LH}$ and $W^r{}_{HL}$ are generated from Y^r exactly as W_{LH} and W_{HL} were generated from Y initially. This is possible because the secret key K_R is known at receiver. The overall characterization process is conducted by computing the following authentication matrices A_{LH} and A_{HL} given by:

$$A_{LH}(i,j) = W^r{}_{LH}(i,j) \oplus W^e{}_{LH}(i,j)$$
$$A_{HL}(i,j) = W^r{}_{HL}(i,j) \oplus W^e{}_{HL}(i,j)$$

$$R_{LH} = \frac{\sum_{i=1}^{\lceil M_x/8 \rceil} \sum_{j=1}^{\lceil M_y/8 \rceil} A_{LH}(i,j)}{\lceil M_x/8 \rceil \bullet \lceil M_y/8 \rceil}$$

$$R_{HL} = \frac{\sum_{i=1}^{\lceil M_x/8 \rceil} \sum_{j=1}^{\lceil M_y/8 \rceil} A_{HL}(i,j)}{\lceil M_x/8 \rceil \bullet \lceil M_y/8 \rceil}$$

The authentication statistic $R_a = [R_{LH} \| R_{HL}]$ can be used to classify the received image as follows:

Level 1: $R_{LH} = R_{HL} = 0$ image content is credible and no modifications have been made; authentication of the sender is verified.

Level 2: $R_{LH}, R_{HL} < \tau$ image content is credible, but the image has been processed.

Level 3:

a) $R_{LH} < \tau$ and $R_{HL} > \tau$ image is not credible. R_{LH} can be used to characterize tampering; the sender is not legitimate;

b) $R_{LH}, R_{HL} > \tau$ image content is not credible and moreover the image is entirely fabricated.

V. Dual domains for Orthogonality

Watermark embedding (Haar DWT domain) does not interfere with watermark generation (DCT domain), so the authenticator generated from the original host image is the same as the authenticator generated from the watermarked image.

The dc coefficients of 8x8 DCT blocks collectively form a subspace of the Y_{2LL} band. Since Y_{2LL} band is orthogonal to the Y_{2HL}, Y_{2LH}, Y_{llHL} and Y_{llLH} bands of the Haar DWT, any change to these bands is guaranteed not to affect Y_{2LL} and its subspaces including the DCT dc coefficient.

Also embedding authenticator and chrominance information in Y_{2HL}, Y_{2LH}, Y_{llHL} and Y_{llLH} bands of DWT domain does not affect the 8x8 DCT dc domain. Watermark generation is based on these image features. So, we conclude that authenticator embedding will not affect authenticator generation.

VI. Simulations and Results

This section presents the simulation of the proposed technique and the results obtained.

A color image (to be protected) is taken as a sample image (Refer Fig. 6). Watermark is embedded in it using SET to give a watermarked image, shown in Fig. (a). Fig. (b) shows watermark component extracted from (a). Then the watermarked image is modified (fabricated by inserting a bird in it), as shown in (c). Fig. (d) shows watermark component extracted from (c), along with the location of tampering where the modifications have been made. This clearly indicates that the sample image has been tampered and it not the original image.

(c) $R_{LH} > \tau$ And $R_{HL} < \tau$ image content is not credible and more the image is entirely fabricated. Finally, the chrominance information from $Y^r{}_{llLH}$ and $Y^r{}_{llHL}$ bands of Y^r is used to reconstruct color image. Color recovery involves renormalizing the chrominance watermark and combining them using *YIQ* color space.

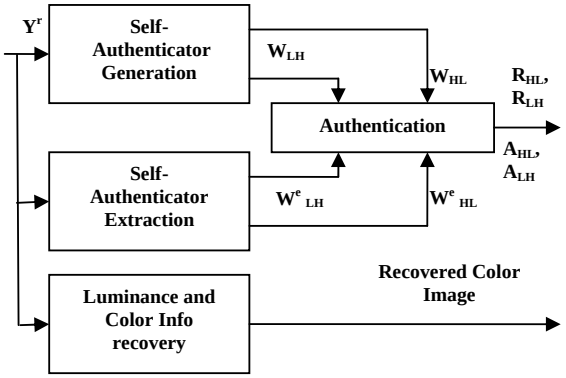


Fig. 5 Watermark Extraction

According to the authentication statistics given in section *IV.D*, Fig. 6(c) is categorised in Level 3(a) since the values obtained are: $R_{LH} = 0.479$ and $R_{HL} = 0.662$ The threshold selected is 0.6. ($R_{LH} < \tau$ and $R_{HL} > \tau$). This implies that the image is not credible and the sender is not legitimate (i.e. the image is not authentic).

Authentication performance can be assessed using two parameters: the missed detection rate P_m, and the false alarm rate P_f. P_m is the likelihood that a malicious attack is not detected by the given scheme (i.e. a tampered image is falsley classified as Level 1 or Level 2). P_f is the likelihood that the scheme gives an incorrect indication of malicious tampering in the absence of a malicious attack (i.e. an untampered image is falsely classified as Level 3).

The error rates are computed for different test images, each watermarked over multiple times using different session keys K_S. The attacks for which the error rates are computed include those from Ω_R and Ω_F. All tests were conducted using Matlab 7.0 in our laboratory [10]. We have also tested the robustness performance of this technique by applying different types of attacks on the watermarked images.

The average test results are shown in Table I with associated PSNR dB (Peak Signal to Noise Ratio) and MSE (Mean Square Error) values. It is observed that, average PSNR range for attacked watermark images is 58 to 70 dB . The theory of watermarking observes that [7], attacks, specially cropping or rotation will cause maximum perceptual degradation. We further examined the effect of rotation on composite image through 2~20 degrees and insertion of gaussian noise as shown in Chart I & II. We found that performance is within acceptable range. Hence, we can claim that this method is robust even against these attacks.

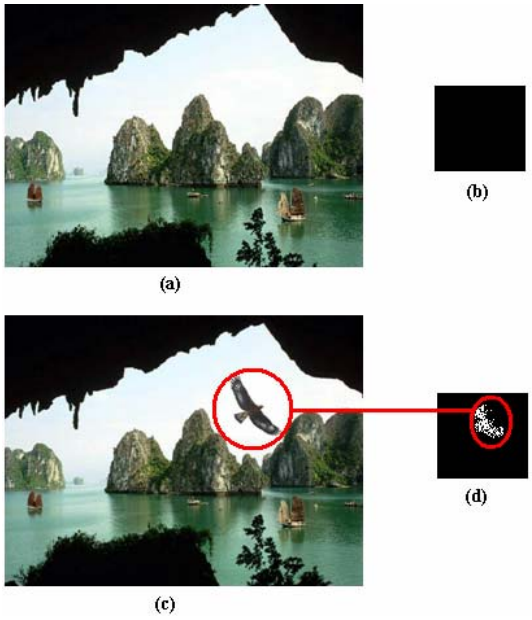


Fig. 6 Tamper Detection: a)Watermarked image b)Watermark component for (a) c)Watermarked image modified by inserting a bird. d)Watermark component for (c) indicating tampering location in fabricated image

VII. Conclusion

This paper proposes an approach for authentication and security of digital color images. It makes use of Self Embedding Technique, Orthogonal Dual Domains and Data Hiding Framework for an Integrated Algorithm. It is observed that the proposed method gives better results as it uses the properties of the image itself to create the watermark.

In this approach we have generated image-adaptive watermark instead of using a fixed watermark for all test images. We observed that better results are obtained for highly textured images as compared to the images containing smooth regions.

The technique is proved robust against various attacks performed. The use of semi-fragile property helps to detect the location of fraud in the image. The use of different cryptographic keys helps to maintain better security of the digital color images along with their authentication.

We find that, this technique is having a great scope of opportunities; especially in the field of cyber frauds, court evidences, certificate or identity forgery and even in the preservation and transmission of cultural heritage images.

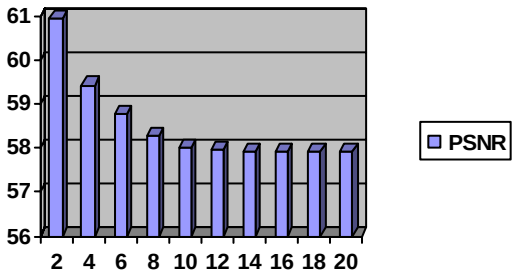
VIII. Limitations

As, this method works with two domains; DCT and DWT, the complexity of the algorithm is higher than any other algorithm which works in a single domain. But this can be compromised as the use of complementary orthogonal mismatching domains increases the image hiding capacity and hence the robustness of the technique [11].

TABLE I
Attacks on the Watermarked Image

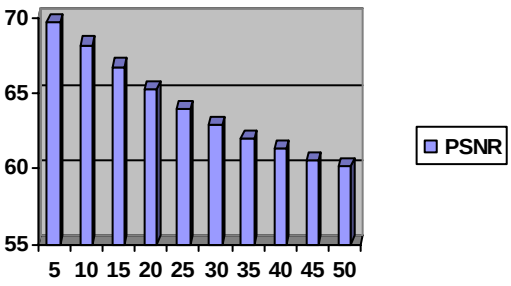
Type of Attack	R_{LH}	R_{HL}	MSE	PSNR(dB)
Rotation (10^o)	0.5801	0.5068	0.1028	58.0097
Cropping	0.4844	0.4775	0.1245	57.1787
Gaussian Noise (15 %)	0.4971	0.4951	0.0140	66.6588
Emboss	0.5078	0.5195	0.1021	58.0399
High Pass Filter	0.5059	0.4795	0.0156	66.1998
Brush Strokes	0.5137	0.4834	0.0287	63.5581
3D Transform (Circular)	0.5127	0.5264	0.0060	70.3154
Diffusion	0.5078	0.4961	0.0153	66.2719

Chart I



PSNR (dB) Vs Rotation Degrees

Chart II



PSNR (dB) % of Gaussian noise

REFERENCES

[1] M. Wu and B. Liu, "Watermarking for image authentication," in Proc. ICIP, Chicago, IL, Oct. 1998.
[2] W. Stallings, 'Network Security Essentials: Applications and Standards,' Prentice Hall, 2000.
[3] William Stallings, "Cryptography and network security," Third Edition, Pearson Publications, 2003.
[4] M. Goljan and J. Fridrich, "Protection of Digital Image Using Self Embedding," Symposium *on Content Security and Data Hiding in Digital Media*, May 1999.
[5] V. Cappellini, F. Bartolini, R. Caldelli, A. De Rosa, A. Piva, M. Barni, and M. Wada, "Copyright protection for CH multimedia data through digital watermarking techniques," in Proc. 11th IEEE Int. Workshop on Database and Expert Systems Applications, September 2000, pp. 935-939.
[6] C.-T. Hsu and J.-L. Wu, "Multiresolution Watermarking for digital images," *IEEE Trans. Consumer Electron.*, vol.45, pp.1097–101, Aug. 1998.
[7] Y. Zhao, Dual Domain "Semi-Fragile Watermarking for Image Authentication," M.A.Sc. Thesis, Univ. Toronto, Toronto, On, Canada, 2003.
[8] R. Mehul and R. Priti, "Discrete Wavelet Transform Based Multiple Watermarking Scheme," *Proceedings of IEEE Region 10 Technical Conference on Convergent Tech. for the Asia-Pacific*, Bangalore,India, October 14-17, 2003.
[9] V. Licks and R. Jordan, "On Digital Image Watermarking Robust to Geometric Transformations," *Proceedings of 2000 International Conference Image Proc (ICIP 2000)*, Vol. 3, Vancouver, BC, Canada, September 10-13, 2000, pp. 690-693.
[10] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, 'Digital Image Processing using Matlab,' Pearson Edition 2004.
[11] Prof. Ramkumar & Ali N. Akansu, "Capacity estimates for data hiding in compressed images," IEEE Tr. on Image proc. Vol. 10 no.8, 8Aug, 2001.