A Multi-Purpose Semi-Fragile Watermarking Scheme for Digital Images

Chetan K.R
Department of Computer Science, JNN College of Engineering, Shimoga-577201. India.
Email: krc_555@yahoo.co.in

ABSTRACT

Digital images can be easily shared via Internet and conveniently processed for queries in databases. Internet carries a large variety of images, which can tolerate minor changes. In general, minor data alterations may be acceptable if they still maintain the perceptual quality of the signal. For all-out forgeries, substantial modification of the content and other malicious attacks can be identified and rejected. So a "soft" image authenticator is desired for the Internet. Data hiding adds perceptually irrelevant information in order to embed data, while compression removes this irrelevancy and redundancy to reduce storage requirements. There exists a duality between data hiding and compression. A semi-fragile watermarking scheme has been proposed in this paper for the compression and authentication of digital images. The watermark consists of an authenticator watermark for authentication and tamper assessment for a given image, and chrominance watermark for "piggy-backing" colour components into the luminance component. The multipurpose watermark is designed by exploiting the orthogonality of various domains (DCT and DWT) used for authentication, colour decomposition and watermark insertion.

Keywords – DCT, DWT, Dual Domain, Soft-Authentication, Semi-fragile Watermarking

Date of Submission: July 03, 2010 Date of Acceptance: September 13, 2010

1. INTRODUCTION

The digital revolution has brought profound changes in communication and data processing. Digital images are easy to edit, modify and exploit. The Internet has become the most important information source and offers ubiquitous channels to deliver and exchange information. Consequently security of multimedia data on the Internet is a challenging topic.

The Internet carries a large variety of images, which can tolerate minor changes. The "loss tolerant" feature of images is exploited in lossy compression for the reduction in file size and is favoured in real time applications. In case of "lossy compression" or "low priority" bit losses during transmission, a conventional digital signature or Message Authentication Code (MAC) would fail the authentication protocol, since the received image data and the signed data are not same [1]. For all-out forgeries, substantial modification of content and other malicious attacks can be identified and rejected. So a "soft" image authenticator is desired for the Internet. Traditionally, data hiding and compression have had contradictory goals. The former problem adds perceptually irrelevant information in order to embed data, while the latter removes this irrelevancy and redundancy to reduce storage requirements.

Different semi-fragile watermarking schemes for practical image authentication have been proposed in the literature. Xie and Arce [2] proposed the embedding of signature AMAC or IMAC back to the image with a private key. Fridrich [3] looks at a robust hash function for watermark generation. Kundur [4] takes a key dependent random sequence as a watermark; the watermark is embedded in four level Haar DWT domain by quantizing the DWT coefficient to even or odd multiples of a step size. Liao and Lu [5] embed two complementary watermarks in the DWT domain by using cocktail watermarking - one Positive Modulation (PM) and the other Negative Modulation (NM) based on wavelet coefficients quantization. Delp and Lin [6] take a pseudo-random, zero-mean unit variance Gaussian noise sequence with a key controlled seed as a watermark.

Quelez [7] computes the rank order relationship of image projections on three secret directions to an image authenticator. Lin and Chang [8] use non-overlapping zones to generate and embed watermarking and the division method of zones is indicated by a secret mapping method using a seed. Fridrich and Goljan [9] take the concatenation of the compressed LSB of visited DCT coefficients and the hash of DCT coefficients as the watermark.

One of the main obstacles within the data hiding community has been developing a scheme which is robust to perceptual coding. Perceptual coding refers to the lossy compression of multimedia signals using human perceptual models. The compression mechanism is based on the premise that minor modifications of the signal representation will not be noticeable in the displayed signal content. These modifications are imposed on the signal in such a way as to reduce the number of information bits required for storage of the content. Human perceptual models are often theoretically and experimentally derived to determine the changes on a signal which remain imperceptible. A duality exists between the problems of perceptual coding and data hiding; the former problem attempts to remove irrelevant and redundant information from a signal, while the latter uses the irrelevant information to mask the presence of the hidden data. Thus, the goals of data hiding and perceptual coding can be viewed as being somewhat contradictory [10].

Several papers have dealt with integrating perceptual coding with data hiding [11, 12] and others have investigated the theoretical relationship between both processes [13]. Data hiding for media compression [14] is investigated. The method operates in the frequency domain and it is based on linear projection, quantization and perturbation. The central theme of all the works is that there must be an appropriate compromise between data hiding and compression to develop a method which performs both reasonably. It is assumed that each process hinders, not helps, the objective of the other.

A semi-fragile watermarking scheme is desired in which authenticator and chrominance watermarks are imperceptibly embedded in a digital image. This assures more practical image authentication desired in the Internet and also help improves the signal compression. The semi-fragile watermarking scheme is designed to tolerate occasional noise and common image processing such as lossy compression, but be fragile to any malicious tampering that modifies image content.

2. PROPOSED MODEL

In this section, a multipurpose semi-fragile watermarking algorithm is proposed for the authentication and compression of the digital images. The proposed model is robust to non-malicious content preserving operations but fragile to malicious content modification. Since the watermark generation domain is orthogonal to the embedding domain, the received image authentication needs only the public and session keys. The proposed algorithm can be easily incorporated with public key encryption systems to prevent active attacks, such as masquerade, replay, and modification of message and denial of service. It can also locate the modifications and differentiate images into three authentication levels. The algorithm exploits the use of data hiding for the coding of colour images [11].

2.1. Framework

Digital watermarking has been proposed for a diverse set of applications including copy protection, image authentication, video error correction and colour image compression. In each case, the existing inefficiencies in the host are exploited to provide value-added services and the design process involves reconciling fundamental compromises. This characteristic helps to believe that digital watermarking may present a useful paradigm for authentication and compression of digital images. It is a problem that also requires arbitration among competing objectives. Assuming such a framework, the proposed system consists of the following components:

1. The **generating function**, f_g, which produces the watermark signal W to embed is given by:
$$W = f_g(i, k, Y) \qquad (1)$$
where k is the secret *generation key* known only to the sender and receiver, Y is the luminance of the host image X and i is called the watermark “payload” which is comprised of a bit sequence independent of k and Y. W has two parts: an *authenticator watermark* component W_a employed for security and a c*hrominance watermark* component W_c to help with compression. This relationship can be expressed by:
$$W = [W_a \| W_c] \qquad (2)$$
where $\|$ is the concatenation operator.
2. The **embedding function**, f_m , which inserts W into Y with the help of secret embedding key K known only to the sender and receiver, yielding the watermarked data Y^w as given by:
$$Y^w = f_m(Y, W, K) \qquad (3)$$
such that Y^w is perceptually identical to Y.
3. The **extracting function**, f_x, which recovers the watermark information, $\hat{W}$, from the received watermarked data Y^r, using the secret key K given by:
$$\hat{W} = f_x(Y^r, K) \qquad (4)$$
4. The **recovery function**, f_r , employs $\hat{W}$ for authentication and colour recovery of the image and is represented by:
$$\lfloor R_a, \hat{X}_w \rfloor = f_r(Y^r, \hat{W}, k') \qquad (5)$$
where k' is a key available to the receiver. In the case of symmetric encryption scheme, k' is same as k and for the asymmetric encryption scheme, k' differs from k. R_a is a statistic that allows the application-dependent authentication and tamper assessment of the received luminance image Y^r and $\hat{X}_w$ is the overall colour-recovered version of Y^r .

It should be noted that there is no explicit payload detection stage in the proposed framework. This is because W_a , containing the authenticator information, needs only to be a function of Y and k. In contrast, W_c contains the chrominance information which is independent of Y and k as no security is required; thus, W_c is effectively the payload. The watermark generation step in the Equation (1) is a generalization of this process. As a result, the payload detection for W_a is unnecessary and trivial for W_c. Hence watermark extraction, authentication and colour recovery can be done without an explicit payload detection step.

2.2. Design Principles

Based on an empirical analysis of the strengths and limitations of semi-fragile watermarking and compressive data hiding, the following principles for system function design have been identified [11]:

• **Authenticator Watermark:** The authenticator watermark W_a should represent a secure content-based adaptive authenticator. Furthermore, the authenticator should be a function of image features that are invariant to predefined content-preserving image processing operations denoted by Ω_R, while fragile to specified content modification attacks denoted by Ω_F. Thus, designing f_g is equivalent to developing an effective adaptive authenticator that can distinguish Ω_R from Ω_F .

• **Uniqueness of Authenticator Watermark Generation:** Different values of k should produce distinct W_a for the same X and i ; different values of i should produce distinct W_a for the same X and k. This guarantees key-based security of W_a and unambiguous recoverability of i .

• **Chrominance Watermark:** The component of i corresponding to the W_c should contain a (possibly compressed) version of the colour information such that it can be later combined with the watermarked luminance image for colour recovery. No security or secrecy is required in the generation or embedding of W_c.

• **Non-invertibility of Embedding:** The keys k and K must not be identifiable even if both f_m and W are known to the attackers. Thus, authenticator and embedding security results from the secrecy of the key.

• **Watermark Embedding Structure:** The high resolution nature of digital images makes it practical to partition the host into distinct components, one to embed W_a and another to W_c and employ different embedding approaches for each. This facilitates more straightforward control over achieving both tasks of authentication and compression. Furthermore, embedding should not affect the authenticator watermark generation.

• **Chrominance Embedding:** The inefficiencies of compression should be exploited as the unused bandwidth available for W_c embedding. Hence it is no longer necessary to store chrominance and luminance separately thereby reducing the overall volume of information.

• **Authenticator Generation and Embedding:** For authentication applications, it is important that the watermark embedding does not affect the generation. If this requirement cannot be satisfied, then 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 inauthentic.

• **Blind Watermark Extraction:** The watermark extraction should naturally be blind for practicality. Otherwise there would be no necessity for watermarking nor image distribution as the authentic image would be available at the destination.

• **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 fragile to malicious content changing attacks defined in Ω_F. Together with proper authenticator watermark generation, this provides the necessary "soft-authentication" capability.

• **Computational Efficiency:** The watermarking components should be designed for effective hardware or software implementation for practical applicability. Therefore, only linear orthogonal separable transforms are used in the design of the different system functions.

2.3. Orthogonality and Dual domains

In the proposed system, linear orthogonal separable transforms are used. These transforms work in orthogonal domains of the image for watermark generation and embedding. This approach allows the independent design and analysis of the various system functions (e.g., f_g, f_m). The basic idea is to break an image into the following subspaces: V_c containing the chrominance information of the image to produce W_c and V_l containing the luminance component. Furthermore, V_l is partitioned into subspaces V_{gen} for authenticator watermark W_a generation, $V_{emb,a}$ for W_a embedding and $V_{emb.c}$ for W_c embedding. Ideally, all subspaces should be orthogonal, so that any signal processing involved in these domains will not interfere with one another. However it should be noted that V_{gen} , $V_{emb,a}$ and $V_{emb,c}$ do not necessarily span V_l . Moreover, based on the application to digital images, V_{gen} should allow access to "salient" image features that can be exploited by f_g to relate to the integrity of the image. Similarly, $V_{emb,a}$ should also contain features that are related to image credibility, but that can be used to characterize tampering.

2.4. Watermark Generation

The watermark generation mechanism is as shown in Fig. 1.

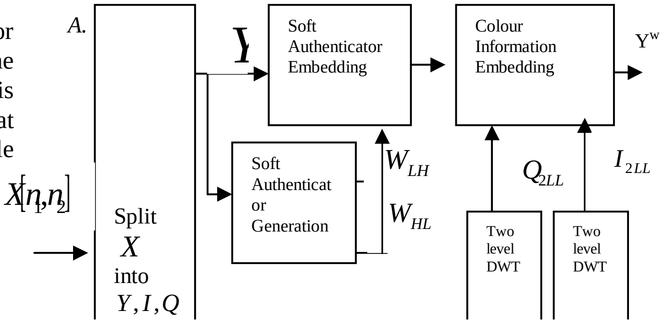


Fig. 1. watermark generation.

To generate both components of the watermark, the host colour image X is transformed into the YIQ color space to obtain the luminance Y and the chrominance images I and Q jointly representing saturation and hue. The chrominance watermark is created by taking the lowest resolution bands resulting from the second level Haar DWT of both I and Q, because subsampling chrominance has little visual affect on the overall colour image; these bands are denoted by $I_{2LL}$and Q_{2LL} , respectively.

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.

The first two steps of the authenticator watermark generation involves the 8 X 8 block DCT and Feature Extraction. These steps identify the components in the image that are of perceptual significance. The feature extracted is the *dc* coefficient of the 8 X 8 DCT blocks of the image. This low resolution representation provides a good metric to represent the raw spatial characteristics in the image. Next, the *Binary Transform* stage order-pairs *dc* coefficients so that their relative magnitudes are guaranteed to be maintained under content-preserving operations such as JPEG or SPIHT compression. For JPEG quality factors higher than 70% and moderate SPIHT compression [10], the sign of the difference between dc values in different 8 X 8 blocks is preserved as long as the magnitude of their difference is above 16. The sign of the differences between the ordered pairs is coded in a binary fashion. The binary output of this stage is one component of the authenticator watermark denoted by W_{LH}. W_{LH} should not change with high probability under content preserving modifications but should change with high probability for content manipulating attacks. Furthermore, the location of changes in W_{LH} point to possible 8 X 8 luminance image blocks that have been modified for tamper assessment capabilities. The other authenticator component W_{HL} is generated by continuing to process W_{LH}. For more security against fraud or forgery a *Permutation* is applied to W_{LH}. The *Majority Function stage* has the goal of reducing the size of the output of the permuted binary transform while coding it to make it more robust to content preserving operations. If the output of the majority function is zero (or one), then the corresponding input row or column contains, on the average, ordered *dc* coefficient pairs in which the first element of the pair is greater (or lower) than the second. Thus, the resulting sequence contains compressed information about the relative local luminance activity between 8 X 8 DCT blocks of the image. The *Map Function* stage converts the output of the previous step to an appropriate size for encryption and subsequent watermarking. The final *Encryption* stage creates a component W_{HL} denoted which allows for sender authentication. To summarize, W_{HL} provides crucial cryptographic security and W_{LH} provides attack characterization capability to balance the requirements of tamper assessment.

2.5. Watermark Embedding

The embedding process takes place in the Haar DWT domain which is considered as a "dual" to the DCT domain used for watermark generation. A two level Haar DWT is applied to Y and the resulting $Y_{2LH}$and Y_{2HL}bands are respectively embedded with W_{LH} and W_{HL} using group quantization method to produce Y^{W}_{2LH} andY^{W}_{2HL} . A quantization based strategy is one of the most popular methods for semi-fragile watermarking because it allows for the embedding of a reasonably long payload (in comparison to spread spectrum based methods), while having a convenient implementation structure [15]. Semi-fragile watermarking through this approach to selected image features provides robustness against perturbations of the features below a predefined threshold related to the quantization step size δ . Any modifications that exceed the threshold are detected.

The subsampled chrominance components, I_{2LL} and Q_{2LL}, are embedded by simply replacing Y_{LH} and Y_{HL} respectively, thus obtaining Y^{e}_{LH} and Y^{e}_{HL} . The rational behind this choice relies on the observation that, in order to obtain a good trade-off between robustness and transparency, many watermarking techniques [15] use "middle frequency" coefficients which makes subbands Y_{LH} and Y_{HL} intuitively suitable for embedding.

2.6. Watermark Extraction, Authentication and Colour Recovery

At the receiver, the image authentication and colour recovery are performed. The receiver is assumed to have appropriate session and decryption keys.

The authentication watermark is extracted from the Y^{r}_{2LH} and Y^{r}_{2HL} bands of Y^r. During extraction, the magnitude of the sum of the coefficients of each 2 X 2 block in Y^{r}_{2LH} andY^{r}_{2HL} is effectively placed in an appropriate "bin" to estimate the watermark bit embedded. Sums in even indexed bins decode to a zero and sums in odd numbered bins decode to a one. The extracted watermarks from Y^{r}_{2LH} andY^{r}_{2HL} are denoted by W^{e}_{LH} and W^{e}_{HL} respectively. These marks must be effectively compared to a corresponding set generated from Y^r for authentication and tamper assessment.

The authentication involves symmetric or asymmetric encryption, in which watermarks denoted by $\hat{W}_{LH}$ and $\hat{W}_{HL}$ are generated from Y^r . The overall characterization process is conducted by computing the authentication matrices $A_{LH}$and A_{HL} given by:

$$A_{LH}(i, j) = \hat{W}_{LH}(i, j) \oplus W^{e}_{LH}(i, j) \qquad (6)$$

$$A_{HL}(i,j)=\hat{W}_{HL}(i,j)\oplus W^{e}_{HL}(i,j) \tag{7}$$

where $\oplus$ is the exclusive OR binary operator.

The authentication statistic R_a 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 $R_{HL} > \tau$:some image content 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.

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

where, a user-defined threshold $0 < \tau < 0.5$ is used. Finally, the chrominance information from the Y^{r}_{llLH} and Y^{r}_{llHL} bands of Y^{r} is used to reconstruct the colour image. Colour recovery involves renormalizing the chrominance watermarks and combining them using the YIQ colour space.

3. THEORETCIAL ANALYSIS

The proposed algorithm is designed to be robust to non-malicious changes and fragile to malicious content tampering. This section analyzes the feasibility, and security analysis of the proposed system.

3.1. Feasibility Analysis

Embedding the watermark should not affect watermark generation. The watermark generated from the original image should be same as the watermark extracted from the watermarked image. To analyze the theoretical feasibility of the DCT-DWT combined domain, the difference between the watermarked and the original image is computed to characterize the embedding. Then, the effect of this embedding on the DCT domain watermark generation is evaluated.

Let $I(i,j)$ be a pixel of the image I at the ith row and jth column. The two dimensional first level Haar DWT coefficients of four bands $f_{LL}, f_{LH}, f_{HL}, f_{HH}$ are generated as :

$$\begin{vmatrix} f_{LL}(i,j) & f_{LH}(i,j) \\ f_{HL}(i,j) & f_{HH}(i,j) \end{vmatrix}=\frac{1}{2}\begin{pmatrix} \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} I(2i-1,2j-1)+\begin{vmatrix} 1 & 1 \\ -1 & -1 \end{vmatrix} I(2i-1,2j) \\ +\begin{vmatrix} 1 & -1 \\ 1 & -1 \end{vmatrix} I(2i,2j-1)+\begin{vmatrix} 1 & -1 \\ -1 & 1 \end{vmatrix} I(2i,2j) \end{pmatrix} \tag{8}$$

Every 2 X 2 block of image $I(i,j)$ is reconstructed as shown in (7):

$$\begin{vmatrix} I(2i-1,2j-1) & I(2i-1,2j) \\ I(2i,2j-1) & I(2i,2j) \end{vmatrix}= \tag{9}$$
$$\frac{1}{2}\begin{pmatrix} \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} f_{LL}(i,j)+\begin{vmatrix} 1 & 1 \\ -1 & -1 \end{vmatrix} f_{LH}(i,j)+ \\ \begin{vmatrix} 1 & -1 \\ 1 & -1 \end{vmatrix} f_{HL}(i,j)+\begin{vmatrix} 1 & -1 \\ -1 & 1 \end{vmatrix} f_{HH}(i,j) \end{pmatrix}$$

The following relations are obtained as shown in the Equations (10) – (17) :

$$f_{LL}(i,j)=\frac{1}{2}(I(2i-1,2j-1)+I(2i-1,2j)+I(2i,2j-1)+I(2i,2j)) \tag{10}$$

$$f_{LH}(i,j)=\frac{1}{2}(I(2i-1,2j-1)+I(2i-1,2j)-I(2i,2j-1)-I(2i,2j)) \tag{11}$$

$$f_{HL}(i,j)=\frac{1}{2}(I(2i-1,2j-1)-I(2i-1,2j)+I(2i,2j-1)-I(2i,2j)) \tag{12}$$

$$f_{HH}(i,j)=\frac{1}{2}(I(2i-1,2j-1)-I(2i-1,2j)-I(2i,2j-1)+I(2i,2j)) \tag{13}$$

$$I(2i-1,2j-1)=\frac{1}{2}(f_{LL}(i,j)+f_{LH}(i,j)+f_{HL}(i,j)+f_{HH}(i,j)) \tag{14}$$

$$I(2i-1,2j)=\frac{1}{2}(f_{LL}(i,j)+f_{LH}(i,j)-f_{HL}(i,j)-f_{HH}(i,j)) \tag{15}$$

$$I(2i,2j-1)=\frac{1}{2}(f_{LL}(i,j)-f_{LH}(i,j)+f_{HL}(i,j)-f_{HH}(i,j)) \tag{16}$$

$$I(2i,2j)=\frac{1}{2}(f_{LL}(i,j)-f_{LH}(i,j)-f_{HL}(i,j)+f_{HH}(i,j)) \tag{17}$$

According to the proposed algorithm, the watermark is only embedded in f_{LH} and f_{HL} to get $\hat{f}_{LH}$ and $\hat{f}_{HL}$. Let δ_{LH} and δ_{HL} be as shown in (18) and (19) respectively:

$$\delta_{LH}=\hat{f}_{LH}-f_{LH} \tag{18}$$

$$\delta_{HL}=\hat{f}_{HL}-f_{HL} \tag{19}$$

Based on (14) – (17), the watermarked image $\hat{I}$ is obtained as shown in (20) – (23):

$$\hat{I}(2i-1,2j-1)=I(2i-1,2j-1)+\frac{1}{2}(\delta_{LH}+\delta_{HL}) \tag{20}$$

$$\hat{I}(2i-1,2j)=I(2i-1,2j)+\frac{1}{2}(\delta_{LH}-\delta_{HL}) \tag{21}$$

$$\hat{I}(2i,2j-1)=I(2i,2j-1)+\frac{1}{2}(-\delta_{LH}+\delta_{HL}) \tag{22}$$

$$\hat{I}(2i,2j)=I(2i,2j)+\frac{1}{2}(-\delta_{LH}-\delta_{HL}) \tag{23}$$

The sum of every 2 X 2 block of pixel values in $\hat{I}$ is computed as in (24):

$$\hat{I}(2i-1,2j-1)+\hat{I}(2i-1,2j)+\hat{I}(2i,2j-1)+\hat{I}(2i,2j)$$

$$=I(2i-1,2j-1)+\frac{1}{2}(\delta_{LH}+\delta_{HL})+I(2i-1,2j)+\frac{1}{2}(\delta_{LH}-\delta_{HL})+$$
$$I(2i,2j-1)+\frac{1}{2}(-\delta_{LH}+\delta_{HL})+I(2i,2j)+\frac{1}{2}(-\delta_{LH}-\delta_{HL})$$

$$=I(2i-1,2j-1)+I(2i-1,2j)+I(2i,2j-1)+I(2i,2j) \tag{24}$$

According to (24), one can infer that sum of every 2 X 2 block pixel values of the original image is same as that of the watermarked image. Thus the watermark embedding has little effect on the sum of every 2 X 2 block pixel values.

The DCT coefficients $f_d(i,j), i,j=1,2\cdots 8$ of Image I in every 8 X 8 block are given by :

$$f_d(k,l) = \sum_{i=1}^{8}\sum_{j=1}^{8} I(i,j)a(k,i)a(l,j) \quad (25)$$

where $k,l = 1,2\cdots 8$

According to the DCT watermark generation of the proposed algorithm, the watermark is created from the *dc* coefficients $f_d(1,1)$ sub-domain as shown in the (26). It can be seen that:

$$f_d(1,1) = \frac{1}{8}\sum_{i=1}^{8}\sum_{j=1}^{8} I(i,j) \qquad (26)$$

The DC coefficients $\hat{f}_d(1,1)$ of the watermarked image $\hat{I}$ are given by:

$$\hat{f}_d(1,1) = \frac{1}{8}\sum_{i=1}^{8}\sum_{j=1}^{8} \hat{I}(i,j) \qquad (27)$$

The sum of every 2 X 2 block pixel values of the original image is same as that of the watermarked image. So the sums of every 8 X 8 should be same before and after watermark embedment as shown in (28) :

$$\begin{aligned}\hat{f}_d(1,1) &= \frac{1}{8}\sum_{i=1}^{8}\sum_{j=1}^{8} \hat{I}(i,j)\\ &= \frac{1}{8}\sum_{i=1}^{8}\sum_{j=1}^{8} I(i,j)\\ &= f_d(1,1) \qquad (28)\end{aligned}$$

So the DCT *dc* coefficient sub-domain for the watermark generation is the same before and after watermark embedding. Therefore, it can be concluded that DWT watermark embedding should not affect DCT watermark generation.

3.2. Attack Analysis

The proposed algorithm aims to prevent images from being modified or fabricated by estimating the distortion on the watermark inserted in the image. The authentication watermark should keep the image safe from a series of active attacks that involve some modification of data stream or the creation of a false stream. Since the watermark is a crucial measure for authentication and integrity verification, the watermark information should be disguised against passive attacks [1].

Defending Active Attacks

The proposed dual domain semi-fragile watermark can effectively detect all four categories of active attacks

- **Masquerade:** Assume that an image is sent by an unauthorized party by inserting a watermark that has been used by an authorized party. Since the watermark used in the proposed algorithm is an image content based watermark, different images have completely different watermarks. The fabrication will fail in both bands, since the reused watermark does not have the content information of the image it inserted and the session key is expired. If an impersonator uses the same method to create a watermark and insert it into an image, the decrypted watermark in the DWT HL band will completely fail the authentication, since the impersonator does not own the right private key for watermark encryption. Thus the proposed algorithm is effective against masquerade attacks.
- **Replay:** If an unauthorized party re-sends an image of an authorized party to pretend to be him/her, the replay attack will be detected since the session key used for watermark generation has expired and this causes the authentication fail completely.
- **Modification:** The content modification of an image will lead the authentication to fail in the DWT *hl* band, since the watermark generated from this image is different from the original one, which is embedded in the DWT domain. The watermark matrix embedded in *hl* is an encrypted watermark, so even a single bit difference between the watermark generated from this image and the original one leads to a complete authentication failure.
- **Denial of service:** Since the proposed watermark is incorporated with the public system, which has a comprehensive audit service to detect and prevent this kind of special attack, the proposed algorithm requires just the keys. If the public key and session key infrastructure can run normally, the proposed algorithm will effectively authenticate an image.

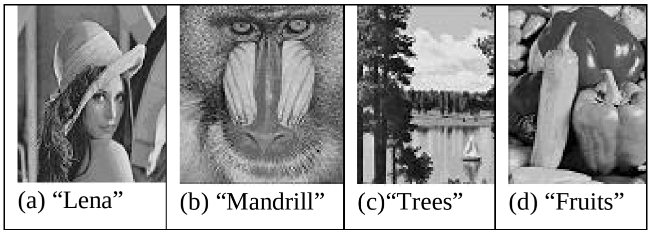
Defending Passive Attacks

The proposed dual domain watermark can resist passive attacks on the true watermark information.

- **Release of Content:** If the image transmitted is eavesdropped by a third party, the attacker can extract the watermark matrix from the DWT domain. However, since he/she does not own the session key, true watermark could not be extracted.
- **Traffic Analysis:** The session key is used to decide block permutation, DCT *dc* coefficients pair combination and the watermark embedding pattern for watermark generation. It makes the watermark a completely random sequence every time; the same image sent by the same sender would have a different watermark at a different time. Hence it is impossible for an intruder to get useful information for finding a rule to analyze the watermark, despite all the image transmissions on the Internet could be observed.

4. RESULTS AND ANALYSIS

The proposed algorithm has been simulated using MATLAB. The simulations are taken on the different images of size 512 X 512 as shown in Fig. 2. The watermark embedding method has been tested to be perceptually transparent by using both subjective evaluation criteria such as human visual perceptibility measures and objective metrics like PSNR.



(a) “Lena” (b) “Mandrill” (c)“Trees” (d) “Fruits”

Fig. 2. Test images.

4.1. PSNR Evaluation

The PSNR values obtained after watermarking on the *Y* component of the “Lena” image for different values of δ is shown graphically in Fig. 3. It can be observed that PSNR is a function of δ. As the value of δ increases, the PSNR of the watermarked image decreases and at $\delta = 12$, PSNR value falls below the acceptable distortion limit.

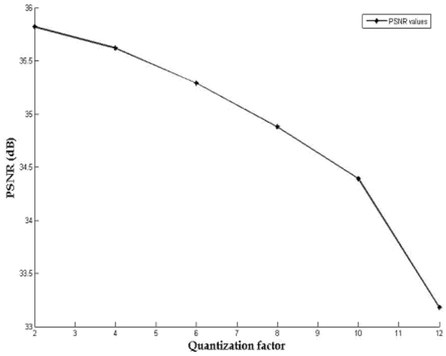


Fig. 3. PSNR versus Quantization factor for the Lena image.

PSNR is evaluated between the host and watermarked image containing only the “authenticator” information (i.e. chrominance information is not embedded) denoted by “auth”, only the “chrominance information” denoted by “IQ” and containing both the information denoted by “auth + IQ”. The results are presented in the Table 1. The obtained values for PSNR metric indicate that there is no perceptual change in the quality of the watermarked images for each test case.

Table 1. PSNR (DB) Values of the Luminance Component (Y)

Images	**PSNR with “auth”**	**PSNR with “IQ”**	**PSNR with “auth + IQ”**
Lena	40.03	35.87	34.46
Mandrill	40.13	29.67	29.30
Trees	40.10	31.85	31.24
Fruits	40.04	33.56	32.68

4.2. Robustness Analysis

The test images shown in Fig. 2 have been watermarked and authenticated using symmetric and asymmetric encrypion scheme. The watermarked image has been subjected to common image processing operations such as salt and pepper noise, histogram equalization and low-pass filtering. The corresponding values of R_{LH} and R_{HL} are shown in the Tables 2 and 3. It can be observed that the calculated block error rates $R_{LH}, R_{HL} < \tau$. Hence it can be concluded that the proposed scheme is robust to common image processing operations.

Table 2. Robustness To Common Image Processing Operations Under Symmetric Encryption

Images	**Salt & Pepper Noise**		**Histogram Equalization**		**Low-pass filtering**	
	R_{LH}	R_{HL}	R_{LH}	R_{HL}	R_{LH}	R_{HL}
Lena	0.26	0.48	0.46	0.49	0.41	0.48
Mandrill	0.27	0.49	0.49	0.49	0.49	0.49
Trees	0.28	0.48	0.45	0.48	0.45	0.49
Fruits	0.27	0.49	0.48	0.49	0.42	0.47

Table 3. Robustness To Common Image Processing Operations U Nder Asymmetric Encryption

Images	**Salt & Pepper Noise**		**Histogram Equalization**		**Low-pass filtering**	
	R_{LH}	R_{HL}	R_{LH}	R_{HL}	R_{LH}	R_{HL}
Lena	0.26	0.40	0.46	0.45	0.41	0.45
Mandrill	0.27	0.40	0.49	0.49	0.49	0.49
Trees	0.29	0.41	0.45	0.47	0.45	0.47
Fruits	0.26	0.47	0.49	0.48	0.41	0.47

Similarly, the watermarked images have been subjected to different degrees of JPEG compression and the corresponding R_{LH} and R_{HL} values are shown in the Tables 4 and 5 for symmetric and asymmetric encryption respectively. It can be observed that the calculated block error rates, $R_{LH}, R_{HL} < \tau$ upto 30%. Below that, R_{LH}, R_{HL} exceeds τ and thus classified as “inauthentic”. Thus the proposed scheme is robust to mild and moderate JPEG compression, but does not tolerate severe compressions.

Table 4. Robustness To JPEG Compression Under Symmetric Encryption

Images	**70**		**50**		**20**	
	R_{LH}	R_{HL}	R_{LH}	R_{HL}	R_{LH}	R_{HL}
Lena	0.04	0.10	0.11	0.35	0.44	0.59
Mandrill	0.11	0.34	0.25	0.46	0.49	0.50
Trees	0.10	0.22	0.21	0.41	0.47	0.51
Fruits	0.05	0.17	0.14	0.36	0.45	0.54

Table 5. Robustness To JPEG Compression Under Asymmetric Encryption

Images	**70**		**50**		**20**	
	R_{LH}	R_{HL}	R_{LH}	R_{HL}	R_{LH}	R_{HL}
Lena	0.22	0.45	0.35	0.48	0.40	0.54
Mandrill	0.20	0.45	0.35	0.49	0.44	0.50
Trees	0.23	0.41	0.36	0.47	0.43	0.52
Fruits	0.24	0.42	0.37	0.48	0.41	0.55

A graph has been plot as shown in the Fig. 4, that compares the robustness against JPEG and JPEG2000.

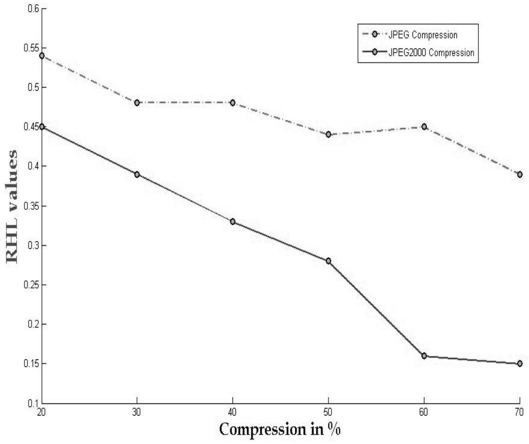


Fig. 4. Robustness to JPEG and JPEG200.

It is evident from the Fig. 4 that the compression curve under JPEG reaches the threshold τ at a lower compression rate than JPEG2000. For e.g., as the compression ratio decreases from 30% to 20%, curve under JPEG reaches τ, but in the case of JPEG2000 it can tolerate well beyond 20%. This emphasizes the fact that proposed scheme is more robust to JPEG2000.

4.3. Fragility Analysis

The watermarked images have also been subjected to content manipulation attacks such as adding an object, substitution and removal of the objects. The corresponding values of R_{LH} and R_{HL} are shown in the Table 6. It can be inferred from the Tables 6 and 7 that $R_{HL} > \tau$ for all the test cases and hence attacked image has been regarded as "inauthentic". Also the A_{LH} matrices for each test case gives the corresponding tampered regions. Thus the proposed scheme has been fragile to the slightest of intentional modifications.

Images	Adding an object		Substi-tution		Removal of an object	
	R_{LH}	R_{HL}	R_{LH}	R_{HL}	R_{LH}	R_{HL}
Lena	0.04	0.54	0.57	0.59	0.04	0.53
Mandrill	0.05	0.53	0.53	0.55	0.02	0.52
Trees	0.05	0.54	0.54	0.52	0.03	0.51
Fruits	0.03	0.51	0.55	0.56	0.03	0.55

Table 6. Fragility to Content Manipulation Operations Under Symmetric Encryption

Images	Adding an object		Substi-tution		Removal of an object	
	R_{LH}	R_{HL}	R_{LH}	R_{HL}	R_{LH}	R_{HL}
Lena	0.02	0.52	0.53	0.51	0.02	0.52
Mandrill	0.03	0.51	0.53	0.52	0.02	0.52
Trees	0.02	0.51	0.54	0.51	0.03	0.52
Fruits	0.02	0.50	0.55	0.56	0.04	0.50

Table 7. Fragility to Content Manipulation Operations Under Asymmetric Encryption

4.4. Colour Image Recovery

After successful authentication, the corresponding colour images are recovered. Normalized Colour Distance (NCD) [16] is caluclated between the host and recovered images and tabulated as in the Table 8. It can be inferred from the table that the proposed scheme has a good perceptual performance and colour image compression does not affect the visual quality of the recovered image.

Table 8. NCD of Recovered Image

Lena	Mandrill	Trees	Fruits
0.1909	0.4101	0.3239	0.2414

4.5. Comparison between Symmetric and Asymmetric Encryption schemes for Soft Authentication.

It can be inferred from the Tables 2 and 3 that, higher R_{HL} values are obtained for common image processing operations under asymmetric authentication. It is evident from the Tables 4 and 5, that higher R_{HL} are obtained for JPEG compression under asymmetric authentication. Hence it can be concluded that asymmetric authentication is more sensitive to content-preserving operations. Comparing the Tables 6 and 7, it can be noted that higher R_{HL} values under asymmetric authentication makes it more fragile to content-manipulation attacks. The graphs plotted in the Fig. 5-8 emphasize the sensitivity analysis of both the authentication schemes.

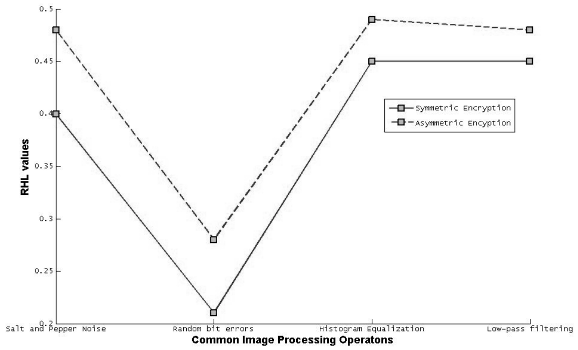


Fig. 5. Robustness to Common Image Processing operations

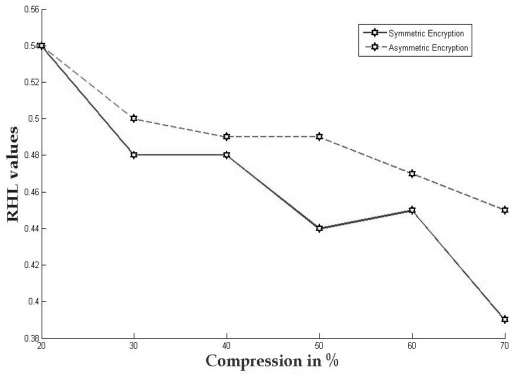


Fig. 6. Robustness to JPEG Compression

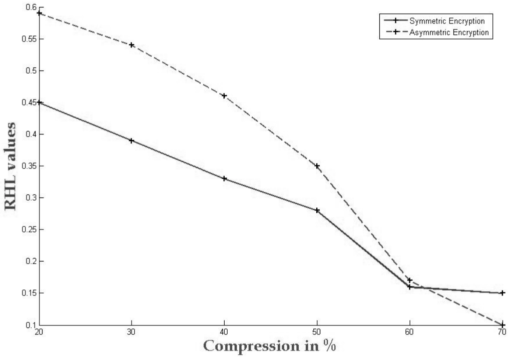


Fig. 7. Robustness to JPEG2000 Compression

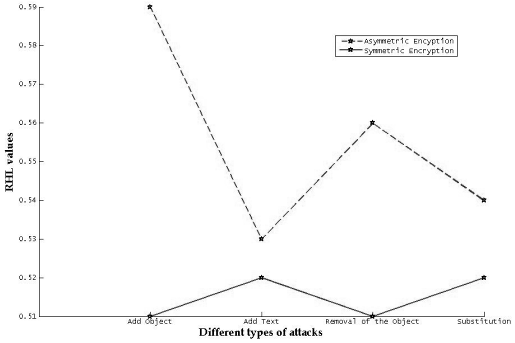


Fig. 8. Robustness to Content Manipulation operations

It can be observed from the Fig. 5 that, for each common image processing operation, the plot for the symmetric authentication is lower than the asymmetric case. Hence asymmetric authentication has the higher tendency to reach τ. It can be observed from the Fig. 6 and 7 that the compression curves for asymmetric authentication reaches τ quickly than the symmetric authentication. Also, for the same compression ratio, the asymmetric authentication scheme posses higher R_{HL} value than symmetric authentication. Thus, the asymmetric scheme is more robust to JPEG and JPEG2000 compression. It can be revealed from the Fig. 8 that, for each type of content-manipulation attack, asymmetric authentication produces relatively higher R_{HL} value, making it more fragile than symmetric authentication.

4.6. Assesing the Authentication Performance

The performance of the proposed approach, in terms of authentication capabilities, is tested in comparison to the following popular algorithms:

1. "Combined Watermarking for Image Authentication and Protection" (CWIMP) [17].
2. "Invertible Authentication Watermark for JPEG Images" (IAWJI) [9].
3. "A Robust Image Authentication Method Distinguishing JPEG Compression from Malicious Manipulation" (ARIAMDJCMM) [8].
4. "A Class of Authentication Digital Watermarks for Secure Multimedia Communication" (ACADWSMC) [18].

To assess the authentication performance, two figures of merit are used: the missed detection rate P_m and the false alarm rate P_f; these standard measures are used to assess baseline performance of authentication watermarking schemes [15]. The first measure, P_m is defined as the likelihood that a malicious attack (from Ω_F) is not detected by the given scheme. In the proposed approach this means that a tampered image is falsely classified as Level 1 or 2 (when it should really be Level 3). Similarly, P_f is the likelihood that a scheme gives the incorrect indication of malicious tampering in the absence of a malicious attack. In the proposed scheme, it refers to an erroneous Level 3 classification when there is no distortion or the modification is from Ω_R.

The error rates are computed over ten different test images each watermarked ten times using different session keys K_S (that affect Steps 3, 4 and 6 of the watermark generation algorithm). The quantization factor is set to $\delta = 10$. The comparisons of the authentication capabilities of the proposed scheme with the other popular semi-fragile watermarking algorithms is shown in the Table 9. The table reports the better overall performance of the proposed scheme. The other methods are each appropriate for different attacks, but do not exhibit the attractive global behaviour of the proposed scheme.

Algorithms	Substitution attack P_m %	Signal Processing attacks P_f %		
		Histogram Equalization	JPEG 70%	Low pass filtering
CWIMP	3.2	23	3.4	15
IAWJI	1.0	31	6.7	58
ARIAMDJCMM	0.0	45	7.2	58
ACADWSMC	0.8	47	45	53
Proposed	0.1	21	1.5	36

Table 9. Authentication Performance of the Proposed Scheme

5. CONCLUSION

An approach to combine image authentication with colour image compression within the digital watermarking paradigm has been proposed. The proposed scheme generates authentication watermark from the DCT domain of the luminance component and chrominance watermark from the chrominance component of the image. The authentication watermark provides "soft" image authentication and chrominance watermark "piggybacks" the colour information into the grayscale component of the image. These watermarks are embedded into the DWT domain of the image to realize "dual" domain watermarking authentication. It is an "oblivious" watermarking scheme as the users just need the public and session keys to authenticate the received image. The watermark is embedded in a group style, so that the watermark embedding can tolerate common image processing and noise. The watermark generation, which is based on the invariant features of the image, is fragile to content modification, but robust to common image processing. Therefore the proposed semi-fragile watermarking scheme is a practical scheme for image authentication on the Internet. Future work includes further cryptanalysis of the proposed scheme. Further improvement for locating the complex modification is also required.

REFERENCES

[1] William Stallings, *Cryptography and Network Security*, (Pearson Education,2003), Third Edition

[2] G.R.A.L. Xie (Oct 1998) et.al. , *Joint Wavelet Compression and Authentication Watermarking, Proc.* ICIP 98, Chicago, IL, USA, Vol. 2, pp. 427-431

[3] J. Fridrich, *Robust bit extraction from images*, Proceedings of the 1999 IEEE International Conference on Multimedia Computing and Systems, Centro Affari, Florence, Italy, Vol. 2, Jul. 1999, pp. 536-540.

[4] D.H.D. Kundur, *Digital Watermarking for Telltale Tamper-proofing and Authentication*, Proceedings of the IEEE Special Issue on Identification and Protection of Multimedia Information, Vol. 87, No. 7, Jul. 1999, pp. 1167 - 1180.

[5] C. S. C.S.Lu, H.Mark Liao, *Combined Watermarking for Image Authentication and Protection*, Proc. ICME 2000, New York, NY, USA, Vol. 3, Feb. 2000 , pp. 1415-1418.

[6] C.E.T. Lin and E.J. Delp, *Detection of Image Alterations Using Semi-fragile Watermarks*, Proc. of SPIE Int. Conf. on Security and Watermarking of Multimedia Contents II, San Jose, CA, Vol. 3971, Jan. 2000, pp. 23-28.

[7] P. L. M. P. Quelez, *Spatial Watermark for Image Verification*, Proc. of SPIE Security and Watermarking of Multimedia Contents II, San Jose, CA , Vol. 3971, Jan. 2000, pp. 120-127.

[8] C. Lin and S. F. Chang, *A Robust Image Authentication Method Distinguishing JPEG Compression from Malicious Manipulation*, IEEE Transactions on Circuits and Systems of Video Technology, Vol. 11, No. 2, Feb. 2001, pp. 153-168.

[9] R. D. J. Fridrich, M. Goljan, *Invertible Authentication Watermark for JPEG Images*, Proc. of International Conference on Information Technology: Coding and Computing, Las Vegas, NV, USA, Apr. 2001, pp. 223-227.

[10] P. Campisi, D. Kundur, D. Hatzinakos, and A. Neri, *Compressive data hiding: An unconventional approach for improved color image coding*, EURASIP J. Appl. Signal Process. Special Issue on Emerging Applications of Multimedia Data Hiding, Vol. 2002, No. 2, Feb. 2002, pp. 152–163.

[11] S. Bhattacharjee and M. Kutter, *Compression tolerant image authentication*, Proc. IEEE ICIP 98, Chicago, IL, USA, Vol. 1, Oct. 1998, pp. 435-449.

[12] J. Lacy, S. R. Quackenbush, A. R. Reibman, and J. H. Snyder, *Intellectual property protection systems and digital watermarking*, Optics Express, Vol. 3, No. 12, 1998, pp. 478–484.

[13] C. Fei,D. Kundur, and R. H. Kwong, *The choice of watermark domain in the presence of compression*, Proc. IEEE Int. Conf. on Information Technology: Coding and Computing, Las Vegas, NV, USA, Apr. 2001, pp.79–84.

[14] B. Zhu and A. H. Tew.k, *Media compression via data hiding*, Thirty-First Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, USA Vol. 1, Nov. 1997, pp. 647–651.

[15] Ozgur Ekici et.al., *Comparative Evaluation of Semi-fragile Watermarking algorithms, Journal of Electronic Imaging, 13*(1), Jan. 2004, pp. 209 – 216.

[16] K. N. Plataniotis and A. N. Venetsanopoulos, *Color Image Processing and Applications*, Springer-Verlag, Berlin, 1st Edition, 2000.

[17] C. S. C.S.Lu, H.Mark Liao, *Combined Watermarking for Image Authentication and Protection*, Proc. ICME 2000, New York, NY, USA, Vol. 3, Feb. 2000 , pp. 1415-1418.

[18] G. R. A. L.Xie, *A Class of Authentication Digital Watermarks for Secure Multimedia Communcation*, Proc. ICIP 2001, Vol. 10, No.11, Nov. 2001, pp. 1754-1764.

Author Biography

Mr. Chetan K.R. is a Sr. Lecturer in the Dept. of CS& E, JNN College of Engineering. He has done his undergraduate course in Information Science & Engineering and graduate course in Network & Internet Engineering. He has published several papers in the area of Information Hiding. His research interest is in the area of Digital Rights Management.