Asymmetric watermarking using nonlinear adaptive system trained on frequency domain

RUO ANDO, YOSHIYASU TAKEFUJI
Graduate school of media and governance, Keio University,
5322 Endo Fujisawa, Kanagawa, 252 Japan

Abstract: In this paper we propose a new asymmetric techniques using nonlinear adaptive system trained on frequency domain. Our system uses two public keys. One key is for encyption of discrete data. Another key is for extraction of continumous data. In embedding, location informaion of frequency domain, where adaptive system is trained, is binalized, expressed in hexadecimal number, and encrypted in assymetric cryptosystem. And encrypted location infomration is embedded in several parts of digital host contents. In generating key, supervised neural networks learn to assign the array of coefficients to teacher signal corresponding to the message to insert. This is one of the trasform-based methods to generate public key from private key. In extracting, location infomation is encrypted to hexadecimal expressio and then secrete information is extracted by feed-forward computation of nonliear adaptive system such as neural networks. Proposal method is tested in still image. Experimental results show that bit loss rate is roughly high to the extent that location information cannot be extracted or restored, while keeping the torelance for lossy compression with JPEG Qtable division.

Key-Words: - asymmetric watermark, fragile watermark, HVS watermark, trasform-based watermark, nonlinear adaptive system, location information embedding, RSA, backpropagtion, recognition prosess as watermark extraction

1 Introduction

With the rapid advance in digital network, digital libraries, and particularly WWW (world wide web) services, digital watermarking is becoming major research areas because the serious concern is raised about copyright protection and authority identification in digital media. Nowadays peer-to-peer network based contents distribution is discussed in the view of copyright protection. Digital multimedia contents on internet is easy to be distributed, reproduced and manipulated compared with conventional analog data.

A digital watermark is the suquence or code embedded into multimedia contents irremovably, imperceptibly in order to protect intellectual propriety rights. This paper presents a new model of asymmetric watermarking using both public key cryptosystem for discrete information embedding and supervised neural network for continuous information extratction. In our model, neural network is trained to assign coefficients to predefined secret code. Location information of training is encrypted and embedded into frequency domain.

2 Related Work

Digital watermarks can be divided into the following two points. At first, it could be categorized according to whether it uses original image or not. Private (nonblind) watermarking needs the original data. This technique depends on the pre-watermarked images to detect hidden code. Public (blind) watermarking requires performance without original image and watermark. Blind watermarking tend to be less robust and the most challenging method. Secondly, in embedding the hidden message, we should select the domain to process, i.e. spatial domain and frequency domain. In spatial watermarking, the secret code is added to the spatial domain. Spatial watermark is easy to implement in the sense that it is embedded directly into an image's pixel data. On the other hand, frequency watermarking is based on some transform method such as FFT, DCT and DWT [1]. Techniques in frequency domain are robust because they rely on the perceptual model of human vision. Recently, statistical models for coefficients in frequency domain are proposed, such as Gaussian, Laplacian and generalized Gaussian.

2.1 Asymmetric watermarking

The digital watermarking is classified into two categories according to the information disclosed for

detectors : symmetric and asymmetric watermarking. In simple meaning, asymmetric watermarking is the technique using different types of keys.

Traditional watermarking employs symmetric scheme that requires the same key in embedding and detecting. The key we need for message hiding and extraction are identical. So these conventional systems require the key for embedding watermark is disclosed completely in verification process[2]. This techniques result in security problem if attacker have similar signals of the embedding keys. It can be easily for them to remove watermark from subtraction attack[3].

In asymmetric key cryptosystem, it must be almost impossible even with current high performance computation to deduce the private key from the public key. Therefore an advantages of asymmetric watermarking is that the information about decoding key is not enough to remove watermark and that estimation of the secret message is not sufficient to eliminate watermark[4]. As long as the some relation properties of embedded signal are satisfied, different watermarks can be used for many kinds of content. Consequently, the scheme is particularly secure to averaging attack[5].

Furthermore, asymmetric watermarking are classified into two methods watermark characteristics based method and transform based method[6]. In transform based method, a detection key is generated from predefined key by some one way transform such as affine transform. To render asymmetry, many kinds of algorithms are proposed nowadays, i.e. periodic network, eigen vector network and neural network system etc. Furon and Duhamel pointed out that a one-way signal processing is necessary to render asymmetric, comparing to public-key cryptography.

2.2 Fragile watermarking

Digital watermarking are classified into two categories according to the purposes : robust and fragile watermarking. In copyright protection, the robust watermark is applied to prove the origin even after some manipulation or alteration.

Robust digital watermaks is used to provides a mechinism for copyright protection of digital contents by embedding authors and customer infomartion. Robust watermark is also called copyright or fingerprint watermark.

This watermark is for the contents that could be transferred, filtered and compressed owing to the user's purposes. Robust watermarks must be robust to many kind of attacks such as resizing, cropping and filtering. To survive these attacks, current watermark schemes usually employ a spread spectrum approach[7].

In authentication applications of multimedia contents, the main purpose is to find the modification o of data. The fragile watermark is designed for the objectives for authentication and integrity for host contents. This is authentication technique to insert a signature designed to be undetectable after the even slight manipulation of the content. So fragile watermark is best used for protection and verification of original documents.

As we discussed, in this paper the authentication of images is focused.Cryptography is probably the most common method used to authenticate the integrity of digital data. Cryptography is traditional authentication method to conceal the message to unauthorized person. In cryptography, digital signature preverse confidentiality and integrity. The signature is added to the encrypted message. Fragile watermark is surely similar to cryptography but can protect content by placing hidden information in the parts of content where it is not elliminated in normal usage. fragile watermarking method takes advantages in inserting secret messages directly with no addtional information to authenticate, while digital signature needs extra data transportation.

Although fragile watermarking is propoer in tamper proofing and authentication of host contents, most waternarked multimedia content is stored with compression due to constraint of disk space or transmission. As long as fragile watermark utilize hash function, this causes a problem when it comes to several manipilation such as lossy compression[8].

To address these problem, a watermarking techniques called semi-fragile is suggested. semi-fragile watermarking has the both charateristics of rosut and fragile watermarking. That is this technique is robust so that it could survive some manupilation frequently used for multimedia digital contents such as JPEG compression[9][10]. And also it can serve as a measure of tampering concerning other alterrations such as feture replacemet.

2.3 HVS watermarking

A good watermark is supposed to be detectable by human sense and can only be perceived by computing the correlation or periodicity. Recent watermarking method applies human visual system (HVS) features to insert signatures into host images. The actual human visual system is so complicated and can process a vast information.

In simple meaning, it is constructed of a receiver, with the eye and the retina, a transmission chennnel and processing engine such as visual cortex in the stage of preprocessing.HVS are less sensitive in textured, edge and rapid changing regions. This characteristics makes these zones appropriate for embedding watermark.

According to psychophysical experieces, the response of virtual cortex is turned to the limited band parts of the frequency domain. On the other hand,.it was verified that the brain decomposes the sensed spectrum in perceptual channels that are bands in spatial frequency. Currently many people apply HVS as a reasonable standard in measuring whether host contents are malformed or not since humen eyes are more sentitive to modifications than perception.

As we discussed before, a good watermark is supposed to be invisible for human eyes and undetectable without information to detect. And also it should be remained in spite of spatial / temporal modification. To satisfy these requirements, a watermark using HVS (human visual system) has been proposed [11]. According to HVS, human eyes are less accurate in the regions where the change is rapid, which make preferable zones to watermark. On the other hand, human visual system are sensitive in the regions where the value is changed smoothly, which in turn is not proper to embed the secret message.

Compared with previous watermarking techniques using global information, the adaptive watermarking based on HVS select the appropriate parts of frequency domain where the message will be imperceptible and robust for some changes [12].

These content adaptive watermarking has been researched mainly using a stochastic model based on frequency transform [13]. They utilize the local properties of contents to select the optimal embedding blocks with following subband quantization and perceptual model.

3 Algorithms

3.1 RSA

RSA is a public-key cryptosystem implemented for both encryption and authentication on internet. It was invented in 1977 by Ron Rivest, Adi Shamir and Leonard Adleman.

The steps of RSA are as follows :

1) The process begins to select P and Q, two large prime numbers (hundreds of digits).

2) Choose E such that E and (P-1)(Q-1) are relatively prime, which means they have no prime factors in common. E does not have to be prime but it must be odd. (P-1)(Q-1) can't be prime because it's an even number.

3) Find out D, its inverse, and mod (P-1)(Q-1) so that (DE-1) is evemtly divisible by (P-1)(Q-1).
This could be written as :
DE=1 mod (P-1)(Q-1), and D is called as the multiplicative inverse of E.

4) Once these steps is done, messsage can be encrypted in blocks, and manupilated on the following :

$$encrypt(T) = T^E \bmod PQ$$

where T is the plaintext (a positive number)

5) Similarly, C is decrypteed through follwoing equation.

$$decrypt(C) = C^D \bmod PQ$$

where C is the cipher text (a positive number)

it is impossible to deduce the private key from the public key. The cipher text can be decrypted only by corresponding private key. The public key is the pair(PQ,E). The primitive number D computed in step 3 must not be revealed to anyone. The prodct PQ is called the modulus. E selected in step 2 is the public exponent. We also call D as the decryption or secret exponent

3.2 Back propagation network

Backpropagation is the supervised learning algorithm of feed-forward neural network. This algorithm is processed in four steps: initialization, presentations of training examples, forward computation and backward

computation. In backward computation, we use the delta rule to modify connection weight of each neuron.

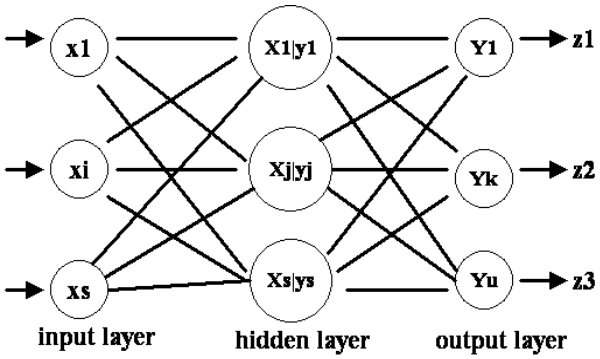

Fig. 1. Backpropagation network

4 Proposal watermarking scheme

In proposal method, we embed and detect location value. Location value is processed as input signal with nonlinear adaptive system such as back propagation network. In inserting watermark, key is generated. Compared with conventional model where signature is equivalent to the message to hide, we embed location code and train nonlinear adaptive system such as neural networks with teacher signal corresponding to secret message. In detecting secret message, output signal is converted to bit code according to threshold (about more than 0.95). we can extract secret message by processing input signal of the blocks where location value is indicating.

Our model is asymmetric in two meanings. First, location code is encrypted with private key and decrypted with public key. Second, the key to extract hidden bit code is transformed by one way signal processing of neural network. To achieve more secure watermarking, location value is encrypted. In symmetric watermarks, attackers can remove watermarking. We apply asymmetric watermarking and encrypt location code with private key. In detection we can decrypt location code with public key. Consequently we have two keys to extract watermarks.The features of our model is as follows:

(1) Compared with using ASCII code, location value embedding renders shorter code to insert. And all character is assigned to 3 bit constantly.

(2) The code to insert is encrypted with private key and we obtain more secure processing of asymmetric watermarking.

In proposal method, the range of location value is from 0 to 7 because we use discrete cosign transform with blocks 8*8.

(3) Location value is expressed in hexadecimal number. In hexadecimal number, one bit change may cause bigger changes in decimal number.

The schematic diagram of embedding process is depicted in figure.

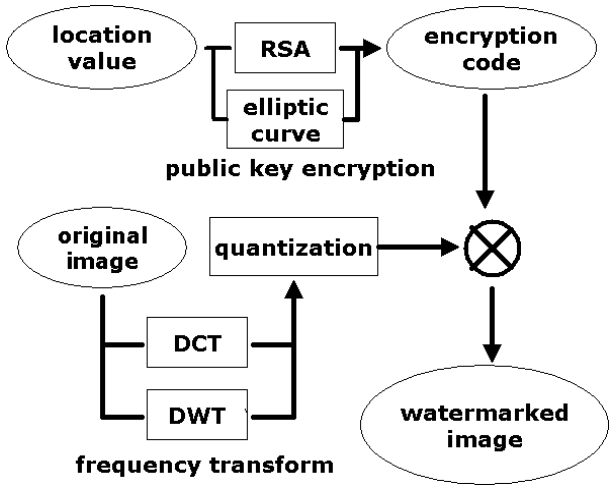

Fig. 2. Embedding watarmark

Watermark embedding process is divided into two steps : (1)embedding and (2)key generation.

As we mentioned before, transform-based method to make a detection key from a given embedding key by proper transform is applied. In encryption of location value, we use private key of RSA. But in generating key, we generate key from location value by training nonlinear adaptive system. Consequently, we use two keys to extract hidden information.

The schematic diagram of extraction process is illustrated in figure.

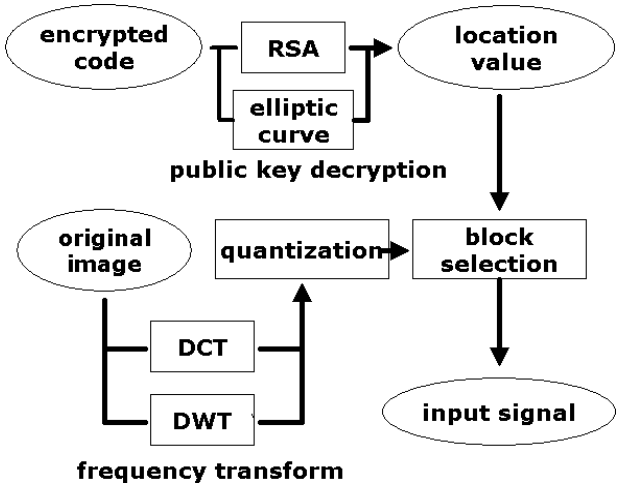

Fig. 3. Detecting watermark

In proposal model, extraction process is Feed-forward computing with input signal of the coefficients of the DCT / DWT blocks. We applied recognition process on frequency domain and convert analog output signal to bit code.

Extraction process is divided into twp process. First, we decrypt the location value with public key. Second, We set the input signal as the coefficients of the DCT block the decrypted value indicated. We use asymmetric cryptosystem because complete information disclosure of location value causes the possibility of removal attacks such as averaging attack. The details of each step is described in following section.

4.1 location code encryption

The proposal model aims for asymmetric and fragile watermarking. This step is concerned with asymmetric watermarking. The avoid removal attack, we use asymmetric cryptography for location value. In this paper we applied RSA algorithms discussed in section 3.1 to encrypt location value.

$$encrypt(T) = T^{E} \bmod PQ$$

4.2 adaptive watermarking and key generation using neural network

The adaptive watermarking based on HVS select the appropriate parts of frequency domain where the message will be imperceptible and robust for some changes

t

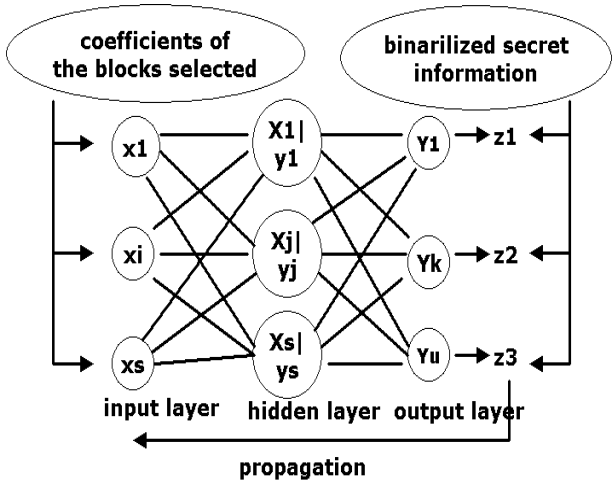

Fig. 4. Key generation. Asymmetric rendering

Watermark should be undetectable without key. In this case, the key is equivalent to connection weight matrix of back propagation As shown in figure, the supervised learning system process DCT coefficient as input signal. And in training, the hidden inserted is binalized as fixed output to neural net.

Once the location value are embed, the connection weight matrix is utilized as key to extract hidden code. In other words, learning process of neural network are transform process to generate key, we should connection weight matrix should be saved as key.

[STEP1]Calculate the DCT coefficients for each 8*8 block.
[STEP2]Quantize the DCTcoefficients by standard JPEG qauntization table.
[STEP3]Select the DCT block.
[STEP4] Embed the encrypted value in selected block.
[STEP5]Train the supervised neural network with the teacher signal corresponding to secret bit code. Input signal are the coefficients of DCT block.

In processing on frequency domain, neural network take advantages in calculating mainly in two aspects. (1) Neural network is nonlinear system that is able to process the nonlinear behavior well. (2) Neural network has fault tolerance, that is, the network can continue to perform acceptably in spite of the failure of some elements in the network.

4.3 location code decryption

Extraction step is divided into step steps : location code decryption and nonlinear signal processing.

In this steps, Location value is decrypted with public key. In embedding, location value is expressed in hexadecimal code. So if some manipulation is operated on host images, decrypted location value is changed.

$$decrypt(C) = C^D \bmod PQ$$

4.4 watermark extraction

In extraction of watermarks, we need

(1) information about the location of blocks where the neural network was trained
(2) public key
(3) connection weight matrix as key. Steps to detect watermark is as follows:

[STEP1]Calculate the DCT coefficients for each 8*8 block.
[STEP2]Quantize the DCT coefficients by standard JPEG qauntization table.
[STEP3]Process input signal. Input signal are the coefficients of the block selected.

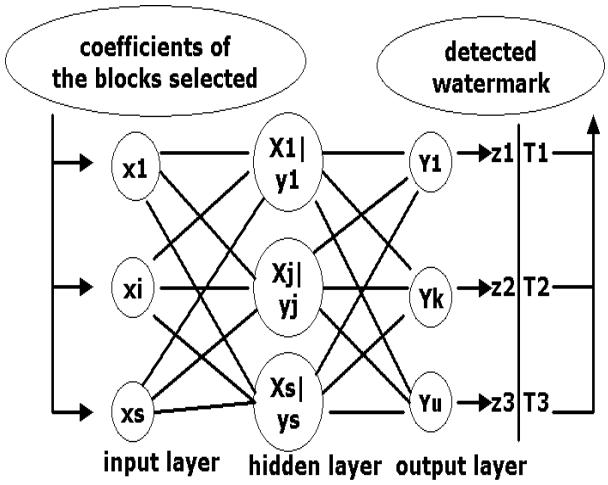

Fig. 5. Watermark extraction

In this phase, recognition is processed by forward phase in backpropagation computing. In forwarding computation, the input signal is propagated from layer to layer in network with the parameters fixed. This recognition phase is finished when each signals of the output layer are figured out.

5 Experimental results

To test this algorithm, we implemented a system for embedding and restoring 8 bits in images of size 256*256. We computed the DCT coefficients for each 8*8 block and quaintize those values by standard JPEG quantization table. In embedding the location value and training network, we set 8 perceptrons of input and output layer. In training network, we set the teacher signal (1,0,1,1,0,1,0,1).

Fig.6 shows the result of processing coefficients of the block selected correctly. The output of processing coefficients in irrelevant block is shown in Fig.6. Extraction failed when we selected the arrays of coefficients that were not processed as input signal in supervised learning.

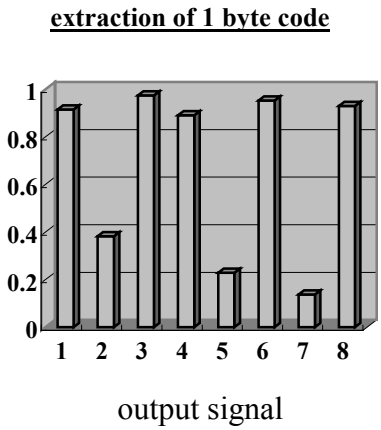

Fig. 6. Output signal by processing in the selected block

In learning process, we prepare the several insignificant pattern that all coefficients are extremely high or low and train network in order to assign these patterns to teacher signal such as (0,0,0,0,0,0,0,0) or (1,1,1,1,1,1,1,1). Experimental results show that our method is functional in recognizing the block that is not selected as meaningless array and in avoiding unexpected extraction from the incorrect blocks.

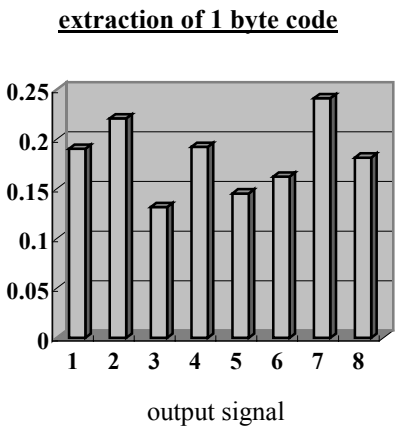

Fig. 7. Output signal by processing in the block where the network was not trained

As we know, in proposal model the location value on which nerural netwoek is trained is encrypted and embedded. Fig.8. shows the bit loss rate. The coefficients of high frequency domain is modifeid to binary number 0/1. After fileters listed in Fig,8, almost all location information expressed by binarized and hexaecimal nuber have become unable to be restored. We have obtained the results that compared with correlation based extraction, our method are more fragile in the sense that location infomation are break down for nonlinear adaptive systems output the signals near (0,0,0,0,0,0,0,0) as shown in Fig.7.

6 Conclusion

In this paper we have presented a new asymmetric watermarking techniques that is enhanced by fragile and adaptive watermarking method. Location value is binalized, emcrypted and embedded in frequecny domain after quaintization. Consequently, we achieved more secure and fragile wateramark compared with conventional asymmetric and fragile method. Experimental results show that bit loss rate is roughly high to the extent that location information cannot be extracted or restored, while keeping the torelance for lossy compression with JPEG Qtable division.

filter / attack	bit loss rate
median filter 2*2	0.92
median filter 3*3	0.65
median filter 4*4	0.69
FMLR	0.44
sharpening 0*1	0.84
sharpening 0*5	0.87
sharpening 1*0	0.83
sharpening 5*0	0.86
sharpening 5*5	0.88
Gaussian	1.00
Geomtric transform1	0.88
Geomtric transform2	0.89
Geomtric transform2	0.86

Fig. 8. Bit loss rate after filters / attacks. Location value cannot be restored.

Fig. 9 Host images in experient

References:

[1]M. Barni, F. Bartolini, V. Cappellini, A. Piva, "A DCT-domain System for robust image watermarking", Signal Processing, Vol. 66, No. 3,pp. 357-372,1998.

[2] S. Craver, S. Katzenbeisser, "Copyright Protection Protocols Based on Asymmetric Watermarking", in R. Steinmetz, J. Dittmann, M. Steinebach, Communications and Multimedia Security Issues of the New Century,

Proceedings of the Fifth Conference on Communication and Multimedia Security(CMS'01),Kluwer Academic Publishers, 2001, pp. 159—170,2001
[3] J.J. Eggers, J.K. Su and B. Girod, "Asymmetric Watermarking Schemes," *Tagungsband des GI Workshops "Sicherheit in Mediendaten,"*, Berlin, Germany, 19. September, Springer Reihe: Informatik Aktuell.Invited paper,2000
[4]Gaël Hachez and Jean-Jacques Quisquater. "Which Directions for Asymmetric Watermarking ? ". Proceedings of the *XI European Signal Processing Conference (EUSIPCO 2002)*. Toulouse, France. September 2002
[5] N.J. Hurley and G.C.M. Silvestre (2002). Nth Order Audio Watermarking. in Security and Watermarking of Multimedia Contents IV, Eward J. Delp III, Ping Wah Wong, Editors, Proc. of SPIE, Vol.4675, pages 102-110,2002.
[6]Tae Young Kim, Taejeong Kim and Hyuk Choi, "Correlation-Based Asymmetric Watermark Detector," International Conference on Information Technology: Coding and Computing (ITCC 2003),2003
The Orleans, Las Vegas, Nevada, US. April 28-30.
[7]I.J. Cox, J. Kilian, T. Leighton and T. Shamoon,"Secure spread spectrum watermarking for multimedia, IEEE Trans. on Image Processing,, pp.1673-1687,1997.
[8] L. M. Marvel, G. Hartwig, C. G. Boncelet Jr., 'Compression Compatible Fragile and Semi-Fragile Tamper Detection," SPIE International Conf. On Security and Watermarking of Multimedia Contents II,vol. 3971, No. 12, EI '00, San Jose, USA, Jan 2000.
[9] Junquan Hu, Jiwu Huang, Daren Huang, Yun Q. Shi: A DWT-Based Fragile Watermarking Tolerant of JPEG Compression. IWDW 2002: 179-188,2002
[10]Ching-Yung Lin and Shih-Fu Chang, "Semi-Fragile Watermarking for Authenticating JPEG Visual Content " SPIE Security and Watermarking of Multimedia Content II, EI '00, San Jose, CA, Jan.,2000
[11]C. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," IEEE Journal on Selected Areas in Communications. vol. 10, no. 4,pp. 525-540.,1998
[12]Ki-Ryong Kwon, Seong Geun Kwon, Jeho Nam, Ahmed H. Tewfik, "Content adaptive watermark embedding in the multiwavelet transform using a stochastic image model", IWDW2002, pp.249-263,2002
[13]D. Kundur and D. Hatzinakos, "A robust digital image watermarking scheme using wavelet-based fusion", Proc. ICIP-97, Vol. 1, 1997,pp. 544-547,2000