M. Karimi, M. Mohrekesh, S. Azizi and S. Samavi., "Transparent Watermarking Based on Psychovisual Properties Using Neural Networks," ***Machine Vision and Image Processing*****, September 2013.**

Transparent Watermarking Based on Psychovisual Properties Using Neural Networks

Maryam Karimi, Majid Mohrekesh, Shekoofeh Azizi, Shadrokh Samavi
Department of Electrical and Computer Engineering
Isfahan University of Technology, Iran
Maryam.karimi@ec.iut.ac.ir

Abstract**—The extreme growth of using digital media has created a need for techniques that can be used to protect the copyrights of digital contents. One approach for copyright protection is to embed an invisible signal, known as a digital watermark, in the image. One of the most important features of an effective watermarking scheme is transparency. A good watermarking method should be invisible such that human eye could not distinguish the dissimilarities between the watermarked image and the original one. On the other hand, a watermarked image should be robust against intentional and unintentional attacks. There is an inherent tradeoff between transparency and robustness. It is desired to keep both properties as high as possible In this paper we propose the use of artificial neural networks (ANN) to predict the most suitable areas of an image for embedding. This ANN is trained based on the human visual system (HVS) model. Only blocks which produce least amount of perceivable changes are selected by this method. This block selection method can aid many of the existing embedding techniques. We have implemented our block selection method in addition to a simple watermarking method. Our results show a noticeable improvement of imperceptibility in our approach compared to other methods.**

Keywords—watermarking, imperceptibility,psychovisual,HVS, neural network.

I. INTRODUCTION

Due to the extreme use of digital media through the internet, the necessity of copyright protection of multimedia products has emerged. Watermarking is one of the most popular techniques in this area. Watermarking is a kind of information hiding that encodes the publisher's copyright information into the media content [1]. Unlike encryption, watermarking does not limit access to the media and it is used for copyright protection. To be effective, the watermark should be perceptually invisible to human eye such that the distortion caused by embedding in the image would not be visible. Robustness to protocol, cryptographic, geometrical and removal distortions are other characteristic of any watermarking approach. Lossy compression, linear and non-linear filtering, cropping and scaling are some instances of these attacks. A good watermarking scheme also guaranties statistical invisibility to unauthorized removal and simple extraction by the owner[2]. These characteristics are usually inconsistent and finding a suitable tradeoff between them would be a main challenge in many watermarking methods. For example, a robust watermarking degrades invisibility and vice versa. Thus, to achieve a robust watermarking we lose imperceptibility therefore maintaining imperceptibility in a robust watermarking would be a challenge

Watermarking concept covers a wide range of methods and different categorizing forms have been introduced in this field. Embedding can be done in spatial and transform domains. The simplest form of spatial domain is to modify the least significant bits (LSB) of pixels [1].In transform domain techniques, watermarking is performed on Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT),Discrete Fourier Transform (DFT) or Contourlet Transform (CT) of a media [3-5]. Spatial domain methods are easier to implement and almost hard to percept, but very fragile to attacks while transform domain methods are more sophisticated and robust [5]. From another point of view the watermarking algorithm can be fixed or adaptive. In former methods, watermarking parameters are constant and the same set of parameters used for all intended images. Therefore fixed approaches are independent of input image while adaptive ones are dependent to characteristics of the input image and to image consequently. It is expected that adaptive methods result in better watermarking features like robustness, transparency or tradeoff between them.

To implement the adaptive algorithms many methods have been proposed. Due to need for enhancing the invisibility, we require a system to imitate human psychovisual model. The adaptiveness in Human Visual System (HVS) is produced by the nature of neural structure of the human eye. Therefore using Artificial Neural Network (ANN) maybe a good choice to achieve imperceptibility in the watermarking process. Embedding the ownership information will inevitably change some of pixels in the image. But our eyes are not sensitive to all parts of an image and all levels of luminance change. Thus the embedding process can be done in locations where eyes are least attracted to this areas. Some researches tried to model this biological system and invented some watermarking methods based on HVS [2,6]. Since the human eye sensitivity is relatively complex, ANN can learn the process and helps watermarking methods to develop their results on the basis of modeling human neural system.

Yu *et al.* [7] proposed a watermarking technique based on neural network for color images which can remind the relation between logo and watermarked image. Since it modifies the intensity values of luminance in spatial domain, the watermark can easily be lost by image compression. In [8], neural networks are used to strengthen watermarking robustness based on the frequency component features of the cover image. Zhang [9] proposed a blind watermarking algorithm based on Hopfield neural network to estimate the capacity of an image for watermarking. Later in [10] Wang proposed a blind watermarking algorithm that adds the watermark bits to the particular coefficients of wavelet transform of image which picked by an ANN. Exact recovering of the watermark is the main feature of this method. In [11]an RBF neural network is used to determine the stability of DCT coefficients in watermark embedding. It is noticeable that all studied methods that mentioned above, use random block selection and their talent is in embedding algorithm.

In this paper a novel use of ANN is introduced. We tried to minimize the degradation of watermarked images such that less obvious differences could be perceived between the original and watermarked one. This approach attempts to define an adaptive watermark scheme based on Multi-Layer Feed-forward (MLF) neural networks. We use MLF to pick original image blocks which are less detectable by human eye after embedding. These blocks will be introduced as the most appropriate places for embedding. Then, to compare our proposed method with similar works in this domain, random blocks have been used for embedding and have shown that the watermarked image quality degrades while preserving the robustness comparing to our approach.

The rest of this paper is organized as follows. After a brief review of the human visual system and its characteristics in section 2, our proposed method based on psychovisual properties and ANN is presented in section 3. Experimental results are described in section 4, followed by a conclusion in section 5.

II. MODEL OF HUMAN EYE

The human eye can not completely distinguish between minor differences of luminance and this weakness can be exploited to embed the watermark image bits[6]. However the study of human perception ability is not simple, a lot of watermarking methods attempted to identify some of HVS properties to improve imperceptibility of watermarked images. Contrast sensitivity, texture sensitivity and frequency sensitivity are some of these properties modeled in HVS [12].Also, the human eye sensibility depends on the average luminance of the background. Luminance sensitivity is a measure of this property in HVS and it obtains the average block intensity of an image as an estimation of the block saliency function of eye [2]. The last metric, contrast sensitivity, refers to the ability of detecting a signal in the presence of another signal. Itsmajor contribution occurs when both signals are of the same spatial frequency, orientation and location [3] [4].Human eye sensitivity to image distortion decreases from smooth blocks to texture and edge blocks. This is another property called texture sensitivity. Most of the works in perceptual watermarking use these characteristics of human visual system to determine Just Noticeable Difference (JND) thresholds on these criteria.

Although these methods try to formulize specific functionalities in human eye using JND, the operation of this biological system completely depends on the situation and the viewed scene. This adaptiveness makes these metrics inefficient to a biological system which takes advantage of dynamic thresholds. Therefore we would prefer more dynamic model which can be adaptable to the psychovisual model of the eye.

There has been much work to realize the vision system for media application systems. While statistical modeling techniques try to examine the image quality, it is ultimately the viewer who can decide how the modified image has maintained its previous quality. Perceptual models take advantage of characteristics of the human visual system in order to qualify image processing applications. As we said, the complexity of human visual system cannot exactly be formulized; therefore, using the visual system itself to monitor an artificial neural network after embedding and back propagating the errors to the network would have better approximation of this dynamic system.

A. Weber ratio

The image formation in the human eye is not a simple phenomenon and only some of the visual properties are measurable. Weber ratio is one of nonlinear characteristics of this system. Let us consider a spot of intensity $I + dI$ in a background having intensity I. dI is increasing from 0 until it becomes noticeable. The ratio dI/I, is called Weber Ratio [14]. It is obvious that brightness discrimination is low at low illumination levels and it improves at high levels of illumination. Fig. 1 demonstrates the Weber ratio diagram for a wide range of luminance values and states more sensibility of human eye to changes for bright regions of image. This fact will be helpful to introduce the amount of tolerable extensive changes in each block as the target value for training ANN.

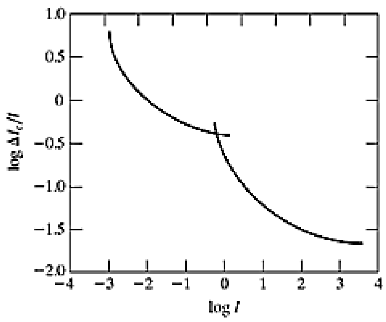

Fig. 1. Weber ratio for different levels of luminance

III. PROPOSED METHOD

To maintain invisibility while embedding, we have proposed a method to find proper blocks. As mentioned above, we have designed a Multi-Layer Perceptron (MLP) neural network which can predict blocks that will not make major changes in quality of image after watermarking. The input patterns of our neural network are the block features effective in perceptual quality of image. The targets are the number of pixels in each watermarked block that intolerably changed on the basis of Weber ratio. Output of this network is a set of corrected weights and biases. After this phase, our trained network can be used for targeting in other images. The block diagram of our proposed method for training neural network is shown in Fig. 2.

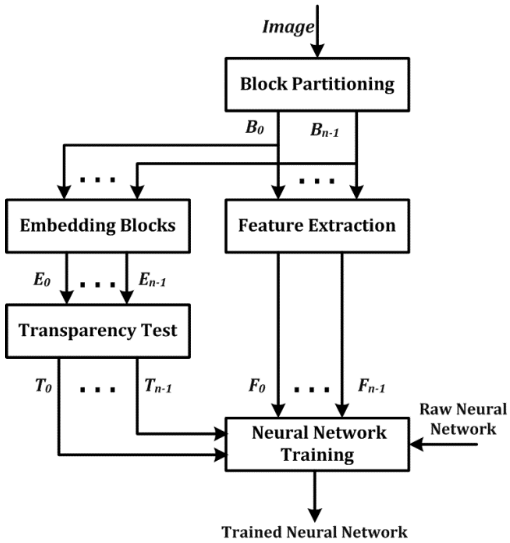

Fig. 2. Block diagram of ANN training

A. Feature Extraction

Based on the description of human visual system in section two, frequency sensitivity, contrast sensitivity and texture sensitivity are effective issues in an imperceptible watermarking.

To achieve this quality of watermarking we should define some features that can cover all of these sensitivities. These features should provide a suitable description of human visual system based on these issues. Simplicity of these descriptors is another necessity which causes to decrease the computational load in calculation of input values and also number of neurons in input layer of our network. Other HVS based algorithms use intricate formulas that impose high computational load and they are time consuming. In addition, these formulas are not so accurate according to such a dynamic system of human eye. Thus we devolve this complexity to the training phase of ANN and we select only the simple features that their combination will provide a thorough description of psychovisual system.

To support these sensitivities we did a study on different blocks of test images. According to the relationships we have selected these four special domain features as the input pattern vectors, which are simple but sufficient descriptors to meet the needs of imperceptible watermarking.

- Number of edges
- Entropy
- SIFT
- Variance

If we increase the contrast level in a block the edge probability enhances. Entropy is a criterion for amount of information in a block and thus, the capacity of new information which could be embedded with minor perceptible distortion. SIFT and variance values are also relatively large amounts in complex regions which have more capability of imperceptible embedding.

After the training phase this network is expected to find proper blocks for watermarking. Extensive HVS researches indicate that the sensitivity of information distortion to human eye diminishes from edge block to smooth block and texture. To extract the edges in blocks we tried some prevalent edge detectors and among them, the Canny edge detector got the best results. The entropy values of image blocks represent the amount of the information in a block. The entropy values of smoother blocks are smaller than those of edge and textured blocks [2]. The entropy is calculated by the following equation:

$$E_k = -\sum_{x,y=0}^{7} p_k(x,y).\log p_k(x,y)$$

Where:

$$p_k(x,y) = \frac{X_k(x,y)}{\sum_{x,y=0}^{7} X_k(x,y)}$$

In order to have another effective parameter related to contrast sensitivity we have selected Scale Invariant Feature Transformation (SIFT). SIFT provides a set of features of an object that are not affected by many of the complications experienced in other methods, such as object scaling and rotation. SIFT image features are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation. The SIFT features are also robust to changes in illumination and noise. The last feature, variance measures the value of distribution in image blocks.

B. Training ANN based on HVS

Perceptual transparency is one of the most important requirements of a proper watermarking method. HVS model results in scoring the blocks of cover images on the basis of their luminance changes after embedding. Human eye is tolerant to changes less than a specific threshold for each luminance level. So the best blocks for embedding logo pixels are the ones make less perceptible difference comparatively to others.

Artificial neural networks are well known in approximating adaptive nonlinear decision. Therefore theses networks are able to easily learn the intricacies of the biological neurosystem. The selected ANN in this paper is a two-layer feed-forward ANN including 5 neurons in the hidden layer. Fig.3 shows the network architecture. The four feature components of a feature vector $F = [f_1, f_2, f_3, f_4]$include variance, SIFT, entropy and number of edges in respective blocks for the explained reasons above. Different numbers for neurons in the hidden layer have been experimented and the best test results obtained for 5 neurons. The desired outputs are the number of pixels in each block which are changed tangibly by human eye on the basis of Weber ratio. The activation function in the hidden layer is a sigmoid function, but the output neuron uses the *purelin* activation function. We have selected *Levenberg-Marquardt* as the network training function which is often one of the fastest back propagation algorithms and it won't get into local minima

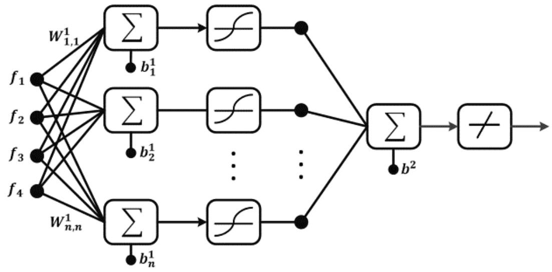

Fig. 3. The structure of used Artificial Neural Network

during training process. It is highly recommended as a first choice supervised algorithm and although it does require more memory than other algorithms. During training operation, information is propagated back to the network and used to update connection weights. It repeats learning experiment many times for every pattern vector in the training set until achieving acceptable values for output errors. This neural network provides an automatic system in order to locate the best host blocks for embedding.

C. Block identification byANN

Now we have a trained ANN which can get the same features of image blocks as utilized in the training phase. This network selects the most appropriate blocks for embedding. As the human visual system is too complicated to be completely formulized, training the neural network helps finding the host blocks which will be less modified after embedding. The block diagram of using trained ANN for embedding watermark in appropriate blocks of an image is shown in Fig. 4.At first step image is partitioned into 8×8 blocks and each block is sent to the trained ANN. In ANN, complex block is determined and used for embedding. After embedding step, complex blocks that contain logo and non-complex blocks retile in appropriate place and will form watermarked image. Embedding process is explained in section 4.

IV. EXPERIMENTAL RESULTS

After training the neural network by images of the training set, it can estimate the best blocks of a test set without initial watermarking. To evaluate our technique, we select the cover blocks using the trained ANN and embed each bit of the logo in one block. For evaluating the performance of our technique from the imperceptibility point of view, two image quality metrics, Peek Signal-to-Noise Ratio (PSNR) has been employed. The results of our selection method have been compared with the random selection of the blocks for the same watermarking method.

A. Embedding process

However the proposed block selection method in this paper is independent of watermarking algorithm, we have implemented a simple robust watermarking method to evaluate the performance. After applying the DCT on a selected block B to embed the k^{th} logo bitb_k, we swap $TB_{i,j}$ and $TB_{j,i}$ or not with a positive certainty margin α called strength factor as follows:

$$TB_{i,j} = \begin{cases} TB_{j,i} - \alpha b_k = 1, TB_{i,j} > TB_{j,i} \\ TB_{i,j} + \alpha b_k = 0, TB_{i,j} > TB_{j,i} \end{cases}$$

$$TB_{i,j} = \begin{cases} TB_{j,i} + \alpha b_k = 0, TB_{i,j} < TB_{j,i} \\ TB_{i,j} - \alpha b_k = 1, TB_{i,j} < TB_{j,i} \end{cases}$$

$TB_{i,j}$ is the (i,j) index of the transformed block. Because of the fragility of the embedding in high frequency and the non-transparency of embedding in low frequency DCT coefficients, the medium frequency coefficients $TB_{i,j}$ and $TB_{j,i}$of the transformed block are chosen. We used$(i,j) = (4,5)$ in our experiments. The constant added (α) is also for refining the extraction phase. The bigger the α, the more accurate the extracted logo and the less transparent the output block will be. The extraction phase will be simply done by a comparison of the two coefficients.

B. Extraction process

The binary matrix of block locations is in the owner's possession. Therefore he can extract the clear logo using a simple comparison.

$$b_k = \begin{cases} 0 & TB_{i,j} > TB_{j,i} \\ 1 & TB_{i,j} < TB_{j,i} \end{cases}$$

Given an original image I and a watermarked imageI^*, both of size $M \times N$, the PSNR between them is:

$$PSNR = 10log_{10}\left(\frac{X_{max}^2}{MSE}\right) = 10log_{10}\left(\frac{255^2}{MSE}\right)$$

Where MSE value is calculated by the following equation:

$$MSE = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(I_{ij} - I_{ij}^*)^2}{M \times N}$$

In these equations, $X_{max}$is the maximum value of luminance and MSE is the Mean-Square Error between the original image and the watermarked one. A larger PSNR indicates that the watermarked image has more resemblance to the original image. In other words, large PSNR means that the watermarking method is more transparent and imperceptibility is preserved [15].

We have measured PSNR values for some images after watermarking in more than half of the blocks using the two techniques for $\alpha = 10$and$\alpha = 20$. The values in TABLE I indicates that PSNRs are improved using our technique for block selection. These results suggest that our watermarked images are visually improved comparing to the other technique. The main reason for these results is using the visual properties of human eye to train an ANN. Thus the ANN locates the distorted blocks in complex areas far from the eye attention.

TABLE I. PSNR AND SSIM VALUES FOR $\boldsymbol{\alpha = 20}$

Lena		**Baboon**		**Goldhill**	
ANN	Random	ANN	Random	ANN	Random
48.9929	47.5281	44.0209	43.6599	48.7280	46.9086
Ship		**Peppers**		**Plane**	
ANN	Random	ANN	Random	ANN	Random
48.5842	47.7019	47.9306	47.0698	47.7163	46.8482

In Fig.5 we have raised theα value to 100 just to highlight the embedded blocks. As it is shown, the destination blocks are spread all over the image in Fig.5 (a). Many blocks of the face which can attract the attention of eye, have impaired. Fig.5 (b) represents another watermarked version of the same image that the embedded blocks have been selected using the estimation of our trained ANN. The selection of candidate blocks is much better than the image in Fig.5 (b). Employing the trained ANN using psychovisual properties of human eye has resulted in concentration of damaged blocks in the imperceptible areas of image.

a. Water marking using random selection of blocks

b. Selection of blocks for water marking using ANN

Fig. 1. Comparison of our method to other methods

V. CONCLUSION

Watermarking is a popular method in order to protect the copyright while permitting to access the media content. To save both the transparency and robustness, much work has been done. In this paper we intended to propose a method to find places of an image which are more suitable for watermarking. This was done by exploiting the psychovisual properties of human eye to train an ANN. This multi-layer network gets some related features and predicts the proper blocks for embedding watermark. Using the trained ANN, the publisher can chose non perceptual blocks with minor preprocessing. It is possible to use this proposed method with different embedding routines. The experimental results demonstrated that the proposed approach significantly improve the perceptual quality of the watermarked image independent of the watermarking method.

REFERENCES

[1] V. M. Potdar, S. Han, and E. Chang, "A survey of digital image watermarking techniques," IEEE International Conference on Industrial Informatics, pp. 709-716, 2005 .

[2] S.Oueslati, A. Cherif, "Adaptive Image Watermarking Scheme based on Neural Network", International Journal of Engineering Science and Technology (IJEST), vol. 3, pp. 748-756, Jan 2011.

[3] P. Meerwald and A. Uhl, "A survey of wavelet domain watermarking algorithms," in Electronic Imaging, Security and Watermarking of Multimedia Contents, vol. 4314 in Proc. 2001 SPIE.

[4] M. Jayalakshmi, S. N. Merchant and U. B. Desai, "Digital Watermarking in Contourlet Domain," International Conference on Pattern Recognition, pp. 861-864, 2006.H. Song, S. Yu, X. Yang, L. Song, and C. Wang, "Contourlet-based image adaptive watermarking," Signal Processing: Image Communication, vol. 23, no. 3, pp. 162-178, 2008.

[5] S. Oueslati, A. Cherif, and B. Solaiman, "A Fuzzy Watermarking Approach Based on the Human Visual System," International Journal Of Image Processing (IJIP), vol. 4, pp. 218-231, 2011.

[6] P. T. Yu, H. H. Tsai, and J. S. Lin, "Digital watermarking based on neural networks for color images," Signal Processing, vol. 81, pp. 663-671, 2001.

[7] Z. M. Zhang, R. Y. Li, and L. Wang, "Adaptive Watermark Scheme with RBF Neural Networks," in Proc. 2003. International Conf Neural Networks and Signal Processing, vol. 2, pp. 1517-1520.

[8] F.Zhang, H. Zhang, "Applications of neural network towatermarking capacity," in Proc. 2004. IEEE International Symposium on Communications and Information Technology, vol. 1, pp. 340-343.

[9] Z. Wang, N. Wang, " A Novel Blind Watermarking Scheme Based on Neural Network in Wavelet Domain," in Proc. 2006 Sixth World Congress on Intelligent Control and Automation, WCICA, vol. 1, pp. 3024-3027.

[10] Q. Liu, J. Xuemei, "Design and Realization of a Meaningful Digital Watermarking Algorithm Based on RBF Neural Network," in Proc. 2006 World Congress on Intelligent Control and Automation (WCICA 2006). vol. 1, pp. 2878-2881.

[11] N.S. Jayant, J.D. Johnson, and R. Safranek, "Signal compression based on models of human perception," in Proc.1993 IEEE, vol. 81, No. 10, pp. 1385-1422, 1993.

[12] N. Sakr et al, "Copyright protection of learning using wavelet objects using wavelet-based watermarking and fuzzy logic," in Proc. 2006. annual e-learning conference on Intelligent Interactive learning.

[13] R. C. Gonzalez, R. E. Woods, Digital Image Processing, 2nd edition, Prentice Hall, 2002, p 51.

[14] K. H.Thung, P. Raveendran, "A Survay of Image Quality Measures," in Proc. 2009,Technical Postgraduates(TECHPOS), pp. 1-4.