MLP Neural Network to improve Digital Watermark detection in gray scale images

KHALIL ABOUTAMMAM[1], M'HAMED AIT KBIR[2], AHMED TAMTAOUI[1], DRISS ABOUTAJDINE[1]

[1]LRIT unit associated to CNRST, Faculty of Sciences, UMVA University, Rabat, Morocco
[2] LIST, Faculty of Sciences and Techniques, UAE University, Tangier, Morocco
[1]k_aboutammam@hotmail.com, [2]m.aitkbir@fstt.ac.ma

Abstract. Image watermarking has today a growing success in the community of image processing. Many methods were already proposed making it possible to obtain increasingly more powerful algorithms in spatial and frequency domain. Most of the spatial watermarking schemes are based on the image decomposition into a grid of blocks, in order to insert a sequence of bits (message). To reach good performances, content-based watermarking schemes aim to use feature points to link the mark with the content of the image [8]. The detection step of all these methods perform a thresholding operation on the correlation function, computed with respect to the block to be processed and the mark or the estimated mark using Wiener filtering method [14]. It's a hard task to fix the threshold value to be used in this step and any improvement here can enhance performances of the global scheme.

In this paper we look for a suitable alternative to perform this task easily and to improve the detection step by using artificial neural networks. In fact, a training phase is performed using a MLP neural network that can be feed by an image block and gives a float value as an output that we can use to take a decision about the presence of the mark. Although, the training phase is time consuming, it's performed separately. This method gives good results even when mark estimation using Wiener filtering isn't used.

Keywords: Watermarking, Neural Networks, Image processing.

(Received June 30th, 2012 / Accepted October 25th, 2012)

1 Introduction

The watermarking was introduced at the beginning of the years 90, as a complementary safety mechanism to data encoding. This second defense line has gained increasingly the interest of content distribution systems, like control and follow-up of copies [18] [10]. The inserted mark is only known by the owner (identification problem) or by the diffuser (follow-up problem), its characteristics are unique and depend on keys supplied by the owner. In this paper, we aim to explore the huge potential of neural networks by integrating them on the detection scheme of the watermarking algorithms. This method is applied using an uniform decomposition of the image. In the next section we present image watermarking techniques in the spatial domain, give principles behind the embedding and the detection schemes and present the problem to be solved. The MLP neural network model used in the detection step is discussed in the third section. Implementation and results are presented in he last section.

2 The proposed approach and related works

Digital image watermarking consists in embedding imperceptibly and indelibly a signature into an image. In fact, a good watermarking algorithm should have two main qualities: the first one is the invisibility of the inserted signature, while the second consists in the robustness of the algorithm against attacks [19]. Thus, the researcher's objective in this domain is to design new watermarking approaches that could achieve a good compromise between the imperceptibility and robustness. To reach this objective multiple watermarking methods have been proposed both in spatial and frequency domain, where each one has some advantages and disadvantages [21] [17] [25] [24] [16] [12]. The next subsections present the standard embedding and the detection schemes and dresses the problem to be solved by our approach.

2.1 Standard embedding and detection in spatial domain

Most of the insertion schemes in spatial domain are based on a decomposition of the image to a grid of similar and square sized blocks, in order to allow the insertion of several information bits (One information bit by block) [26]. Firstly, the mark R_w is generated using a random sequence generator. The mark is zero mean and can take values in $\{-1, 1\}$. Then, we decompose the image to have an uniform tessellation, a grid of disjoint squares: $R = \{R_i, 0 \leq i < N\}$. These squares are processed sequentially to insert bits of the message, let M be the number of bits of the message. Message embedding is done by inserting the mark (bit 1) or the inverse of the mark (bit 0) inside each square of the tessellation. Lets R_p be the result of the multiplication of R_w by a psychovisual ponderation [27], which depends on the processed block R_i, see figure 1. The marked block R_s is obtained by the following equation $R_s = R_i + R_p$. The marked image is finally obtained by replacing the first $min(N, M)$ blocks of the grid by there corresponding marked ones.

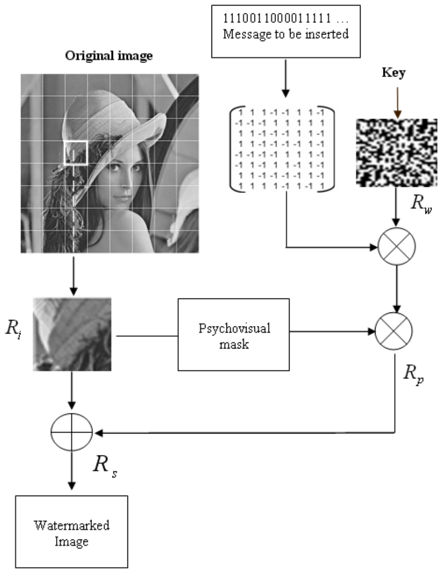

Figure 1: Message embedding scheme

In the detection step R_w is generated from a random sequence depending on a secret key (the same key as in the embedding scheme). A grid of squared blocks: $R = R_i, 0 \leq i < N$, similar to the embedding step, is generated. The presence or the absence of the mark must be decided by processing blocks of the grid sequentially, see figure 2 .

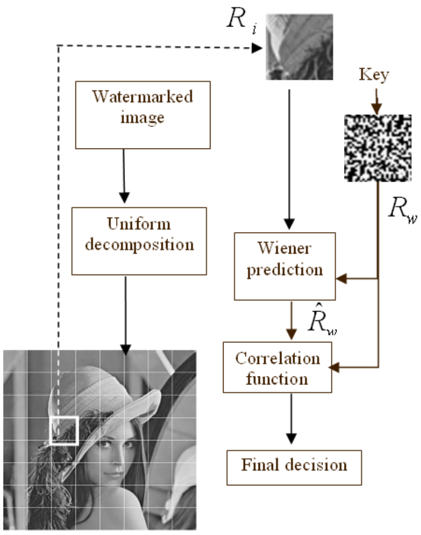

Figure 2: Message detection scheme

R_i and R_w are used to perform statistical prediction using Wiener method to eliminate components provided by the original image to compute the estimated mark $\hat{R_w}$. In fact, Wiener prediction can be considered as a denoising operation that allows separation of the image components from the mark ones [14]. A global decision is obtained by computing the correlation function:

$$\langle R_w, \hat{R_w} \rangle = \sum_{i,j} R_w(i,j) \times \hat{R_w}(i,j) \qquad (1)$$

if the absolute value of the result is greater than an experimentally fixed threshold, then the mark is present in the processed block. It's a hard task to fix experimentally this threshold value that can be different, depending on image blocks homogeneity.

2.2 Content based Watermarking

Other works use decomposition schemes based on a Delaunay triangulation combined with image content. For mark embedding, we must first extract feature points from the image and perform Delaunay tessellation, based on these points, to decompose the image into a set of disjoint triangles. The set of triangles is then processed to insert the mark using a classical additive scheme. This operation requires applying affine transformations to the original mark in order to have a mark that is identical geometrically to each processed triangle. The mark detection is based on the same image content analysis, discussed above and uses the correlation function of different triangles with the transformed mark. As the reference mark triangle must be transformed to fit with each triangle of the decomposition. Firstly, this operation doesn't preserve angles. Secondly, the mark will not have zero mean, which is a very interesting property in the detection step that is based on correlation computing. We proposed in a recent work a new hierarchical method using rectangular decomposition to fix the problem by avoiding the use of triangular tessellation. In fact, we used Harris feature points detector, in order to get a content based image description and generate a set of adjacent rectangles. In fact, feature points are used to guide the decomposition. Each rectangle is combined with the rectangular mark, computed by transforming a square reference mark, in order to compute the marked rectangle [2].

We have also proposed in recent works a decomposition based on QuadTree technique [3] [1]. In fact, the image is decomposed in a set of blocks of various sizes with a big degree of similarity. Indeed, the QuadTree structure is relatively adapted to image content representation, where an image having numerous small blocks of identical pixels will be represented by a relatively minimalist tree. As block sizes are not the same, the number of bits inserted in each block depends on bloc dimensions. In fact, largest blocks can contain more than one bit, one bit by layer, as proposed by CDMA approach [26]. Our approach uses a scalar coefficient that depends on blocks homogeneity, to perform the psycho-visual watermark ponderation and avoid computing this coefficient locally for each pixel, which is very time consuming.

2.3 The problem to be solved

All schemes discussed above use a thresholding operation based on the correlation function to determine the presence of the mark in the detection step. In fact, the correlation computing is performed with respect to the block to be processed and the mark or the estimated mark using Wiener filtering method [14]. A statistical model which aims digital watermark detection and blind threshold computing is proposed in [13], this model uses pre-processing tools such as pre-detection filtering and geometrical distortions compensation. It's a hard task to fix the threshold value to be used in the detection step. This paper aims to use MLP neural networks to perform supervised blocks classification in order to improve performances in the detection step. In fact, a supervised learning function based on MLP (Multi-Layer Perceptron) neural network is introduced. The training data consist of a column version of image blocks. As this kind of networks involves supervised learning, we must associate a label, target class, to each block. To fix the target output, a real value in our case, for each block, a label is associated to each block as follow : 1 if bit 1 was inserted; -1 if bit 0 is inserted and 0 if no bit is inserted.

The computing world has a lot to gain from neural networks. Their ability to learn by example makes them very flexible and powerful. Furthermore we don't need an algorithm to perform a specific task; i.e. there is no need to understand the internal mechanisms of the process under study. In fact, we must just build data samples using vector version of image blocks to perform the training phase. Neural networks are also very well suited for real time systems because of their fast response and computational time which are due to their parallel architecture.

3 Mark detection using neural networks

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the new structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurones) working in union to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between neurones. This is true of ANNs as well.

3.1 Multi-layer perceptron neural network (MLP)

Feed-forward multilayer perceptron (MLP) is applied in many fields due to its powerful and stable learning algorithm [22].

A MLP model contains one or more hidden layers, neurons in the hidden layers must arbitrate between the input and the output of neural network. The input feature vector is feed into the source nodes in the input layer of the neural network at first. The neurons of the input layer constitute the input signals and apply them to neurons of the first hidden layer. The output signals of the hidden first layer can be used as inputs to the next hidden layer or the output layer, see figure 3. Finally, the output layer products the output results and terminates the neural computing decision.

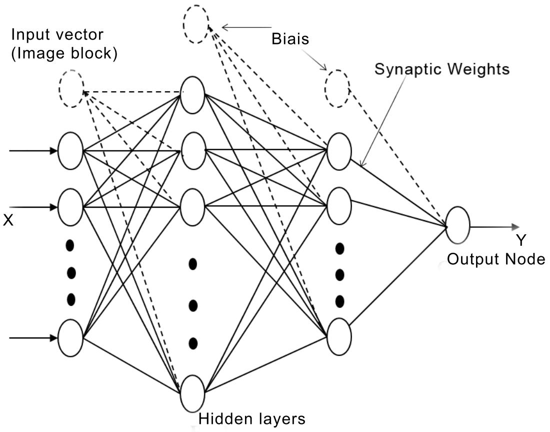

Figure 3: MLP Neural Network architecture

The backpropagation algorithm is used to fix the MLP network parameters, synaptic weight vectors, where the training process is based on two different steps: the forward step and the backward step. The power of the back-propagation algorithm lies in two main aspects: local for updating the synaptic weights and biases and efficient for computing all the partial derivatives of the cost function with respect to these free parameters [7] [15] [11]. The weight-update rule in back-propagation algorithm is defined as follows:

$$\Delta w_{ji}(n) = \alpha \delta_j x_i + \eta \Delta w_{ji}(n-1) \tag{2}$$

where $\Delta w_{ji}(n)$ is the weight update performed during the n^{th} iteration to the connection that join node i and node j, α is a positive constant called learning rate, δ_j is the error term associated with node j, x_i is the state of the input node i and $0 \leq \eta < 1$ is a constant called the momentum. To be conform to the Kolmogorov theorem [20], an MLP neural network must have two hidden layers of neurons, the first layer perform linear separation in the parameter space, the second layer performs convex regions and the output layer gives a decision related to regions with an arbitrary shape. The network was tested in many applications [4] [6] [5].

The training set must be large enough and suitably chosen with the same number of examples of each class to reach good performances. To optimize our network training procedure and reach a best generalization, we use training data and testing data that are drawn randomly from the same data set, same image. A sub-set of the training data that we do not train the network on, the validation data set, is used to estimate what the performance after each training epoch, when all the training data samples are presented to the network. This approach is called the hold out method [23] and is best to provide the best generalization to the testing set.

3.2 How to learn the mark presence using MLP neural network

In our approach, we aim to use MLP neural network to detect the mark in an image block. To reach this goal we must first work out the training data samples. As, this network uses supervised learning the target values: -1, 1 or 0 correspond for each sample vector, column version of a block, to a block that contain respectively bit 0, bit 1 or no bit. In fact, after decomposing the test image, for each block we add the mark to insert the bit 1 or the opposite of the mark to insert the bit 0. Samples of these three classes are presented alternatively to the network to perform the training phase. To do this, the back-propagation learning algorithm is executed with respect to the set of sample blocks, randomly extracted from the image. As the output layer of the network contains one neurone that take real values in the interval $]-1,1[$ (output of tangent sigmoïd function) [9]. The neural network learns the training data structures by adjusting the synaptic weight of neurons according to the error occurred on the output layer. After a successful training phase that is guaranteed by the convergence of the algorithm, optimal synaptic weights are fixed.

The resulted MLP neural network can be used to predict the presence of the mark with blocks that doesn't belong to the training data. In fact, bit 1 is detected when the network output is higher than a threshold $1 > T > 0$, bit 0 is detected if this output is lower than $-T$ and not bit of information is detected when the output belong to the interval $]-T,T[$.

4 Simulations and results

To prove our approach capabilities, we make a comparison against correlation computing method, by using the rate of the correct detection as a criteria. The training phase is performed with $\alpha = 0.01$ and $\eta = 0.9$. 300 samples, 16×16 blocks randomly extracted from a 512×512 gray scale image (Lena), are used to from the training data set. First and second hidden layers of the MLP network are used with different number of neurons. Experiments are performed with three architectures : 256-9-6-1, 256-6-4-1 and 256-3-2-1. Neurons of all architectures have tangent sigmoïd as activation function. Figure 4 shows the convergence of the algorithm after 100 iterations.

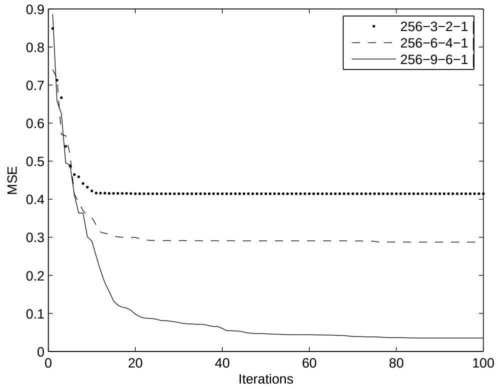

Figure 4: Mean square error versus iterations

We can see that the convergence is reached for the three architectures; we can also observe good performances for the networks that have more hidden neurons. The following table gives performances of the obtained networks against correlation computing method.

Table 1: Performances of the obtained networks

Correlation		Neural Networks		
Block based method	Estimated mark based method	256-9-6-1	256-6-4-1	256-3-2-1
50,48%	62,89%	87,50%	79,29%	56,05%

Note that the threshold used for all the experimented methods is fixed so that the central 20% range of possible values given by the measurement criteria or the network output correspond to the absence of the mark. Lets T be this threshold, if the value of the used criteria is superior to T the detected bit is 1, if this value is lower than $-T$ the bit detected is 0 in the other case no bit is detected. When using block based correlation criteria T is equal to $S_b \times S_b \times 0.2$. S_b is the block width. T is fixed to 0.2 for all experiments with MLP network.

Table 2: Percentage of correct detection

Mark height	Correlation based method(%)	MLP network(%)
$H = 1$	50,3438	86,6211
$H = 2$	54,9805	91,3086
$H = 5$	70,8008	95,8008

The mark take values in the set $\{-H, H\}$. we can see that when H is higher it becomes easy to detect the mark by the MLP neural network. Even if the network training is time consuming, this step is performed outline and doesn't belong to the decision computing time. Finally, we can judge that our method based on MLP neural network is largely better than a simple correlation computing.

5 Conclusion

Neural networks have a remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. To make the training in supervised mode, it is necessary to generate a training data set in order to drive the network by adjusting the synaptic weights its architecture. In our case input vectors are constituted from vector version of image blocks. Blocks are categorized to three classes : ones that contain the mark, others that contain the negative of the mark and those that are not marked. The network can learn from image blocks the structure of the mark and can be used to take a decision with not seen blocks usually used to detect the presence of the mark in image watermarking schemes in spatial domain.
Neural networks are very useful when computational time is required, especially in video processing. We aim, in future works, to use other neural network models for the same purposes and to extend and adapt the proposed approach to video watermarking.

References

[1] Aboutammam, K., Aït Kbir, M., and Tamtaoui, A. Content based image watermarking using multi-level decomposition and cdma technique. *The 5th IEEE International Symposium on Image Video Communications and Mobile Networks (ISIVC'10)*, September 2010.

[2] Aboutammam, K., Aït Kbir, M., Tamtaoui, A., and Aboutajdine, D. A new spatial decomposition scheme for image content-based watermarking. *The 7th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-2009)*, May 2009.

[3] Aboutammam, K., Aït Kbir, M., Tamtaoui, A., and Aboutajdine, D. Tatouage des images numériques par décomposition non-uniforme à base de la texture. *TELECOM'2009|6th JFMMA*, March 2009.

[4] Aït Kbir, M., Benkirane, H., and Benslimane, R. An adaptive median filter based on a neural network. *17emes Journées Tunisiennes d'Electronique et d'Automatique à Nabeul en Tunisie*, November 1997.

[5] Aït kbir, M., El Ouaazizi, A., and Benslimane, R. A genetic learning algorithm for improving mlp

neural network ability. *Quality control by artificial vision, Trois Rivières, Canada*, pages 191–196, May 1999.

[6] Aït kbir, M., Gadi, T., Ouremchi, R., and Benslimane, R. Fingerprint identification for security access using mlp network classifier. *International Conference on Quality Control by Artificial Vision, Takamatsu, Kagawa, Japan*, pages 6–10, November 1998.

[7] B. Demuth, H., H. Beale, M., and T. Hagan, M. Neural network design. *PWS Publish Company*, 11:1–47, January 1996.

[8] Bas, P. Méthodes de tatouage d'images fondées sur le contenu. *Thèse de Doctorat de l'INPG (Grenoble) soutenue le*, 5 octobre 2000.

[9] Bourlard, H. and Morgan, N. Connectionist speech recognition. *Kluwer Academic Publishers*, pages 191–196, June 1994.

[10] CCA. "dvd copy control association, site web : ". *"http://www.dvdcca.org"*, July 2012.

[11] Comon, P. Classification supervisée par réseaux multi-couches. *Revue Traitement du Signal*, 8(6):387–407, March 1992.

[12] Ganesan, K. and Guptha, T. K. Multiple binary images watermarking in spatial and frequency domains. *Signal and Image Processing*, Vol.1, No.2:148–159, December 2010.

[13] Gonzalez-Lee, M., Nakano-Miyatake, M., and Perez-Meana, H. M. Optimal detection system of digital watermarks in spatial domain. *Telecommunications and Radio Engineering*, Volume 65 (Issue 6-10):739–751, 2006.

[14] Hernandez, J. R. and Perez-Gonzalez, F. Satistical analysis of watermarking schemes for copyrights protection of images. *Proceedings of IEEE*, pages 1142–1143, July 1999.

[15] Hornik, K. Approximation capabilities of mutltilayer feed-forward networks. *Neural Networks*, 4(2):251–258, 1991.

[16] Huang, W. T., Tan, S. Y., Chang, Y. J., and Chen, C. H. A robust watermarking technique for copyright protection using discrete wavelet transform. *WSEAS transactions on computers*, 9(5):485–495, November 2010.

[17] J. Cox, I., Doerr, G., and Miller, M. Applying informed coding and embedding to design a robust, high capacity watermark. *IEEE Transactions on Image Processing*, 13(6):792–807, June 2004.

[18] J. Cox, I., Miller, M., and Bloom, J. Digital watermarking. *Morgan Kaufmann Publishers*, 2001.

[19] Kerckhoffs, A. La cryptographie militaire. *Journal des sciences militaires, IX:5-83*, January 1883.

[20] Kolmogorov, A. N. On the representation of continuous functions of many variables by superposition of continious functions of one variable and addition. *American Math. Soc. Series Trans*, pages 55–59, 1963.

[21] Le Guelvouit, G. Tatouage robuste par étalement de spectre avec prise en compte de l'information adjacente. *PhD thesis, INSA Rennes*, November 2003.

[22] Lippman, R. An introduction to computing with neural nets. *ASSP Magazine IEEE*, 4(2):4–22, 1987.

[23] M. Bishop, C. Neural networks for pattern recognition. *Oxford University Press*, 1995.

[24] Pérez-Freire, L., Perez-Gonzalez, F., Furon, T., and Comesana, P. Security of lattice-based data hiding against the known message attack. *IEEE Transactions on Information Forensics and Security*, 1(4):421–439, December 2006.

[25] Seddik, H., Sayadi, M., and Fnaiech, F. Nouveau schéma de tatouage par substitution s'appliquant aux techniques spatiales robuste aux attaques asynchrones. *3rd International Conference: Sciences of Electronic,Technologies of Information and Telecommunications "SETIT"*, March 2005.

[26] Vassaux, B. Technique multicouches pour le tatouage d'images et adaptation aux flux vidéo mpeg-2 et mpeg-4. *Thèse de Doctorat de l'INPG (Grenoble) soutenue le 7 Novembre*, 2003.

[27] Voloshynovsky, S., Herrigel, A., Baumgaertner, N., and Pun, T. A stochastic approch to content adaptive digital image watermarking. *Processing of Internationnal workshop on Information hiding, Dresden, Deutschland*, September 1999.