Scientific Research and Essays Vol. 6(10), pp. 2119-2128, 18 May, 2011
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE10.886
ISSN 1992-2248 ©2011 Academic Journals

Full Length Research Paper

A reinforcement learning method for decision making process of watermark strength in still images

Alimohammad Latif[1]*, Ahmad Reza Naghsh-Nilchi[1] and Vali Derhami[2]

[1]Department of Computer Engineering, Faculty of Engineering, The University of Isfahan, Isfahan, 81746, Iran.
[2]Department of Electrical and Computer Engineering, Yazd University, Yazd, Iran.

Accepted 10 December, 2010

Digital image watermarking is one of the most important techniques for copyright protection. The robustness and imperceptibility are the basic requirements of digital image watermarking that are contradictory. The key factor that affects both the robustness and imperceptibility is the watermark strength. This paper presents a new method to determine the watermark strength using Reinforcement Learning (RL) in Discrete Cosine Transform (DCT) domain. Thus, finding the watermark strength was formulated as an RL problem. In our study, the defined reinforcement function has two contradictory aspects, the one with positive aspect is with respect to the similarities between the host and watermarked image and the other with negative aspect is with respect to the robustness of the watermark. Therefore, a novel adaptive methodology is introduced to estimate watermark strength to ameliorate both imperceptibility and robustness at the same time. The experimental results show that the proposed RL algorithm for watermark strength estimation improves simultaneously the robustness and imperceptibility of the watermarking scheme.

Key words: Digital image watermarking, reinforcement learning, watermark strength, imperceptibility, robustness.

INTRODUCTION

The increasingly easy access to digital images through Internet and the ease of copying digital images identical to the original ones have made the image authentication an important issue. One way to solve this problem is digital image watermarking. Digital image watermarking is defined as a technique of embedding additional information called watermark into digital images by preserving perceptual quality of watermarked images. The watermark can be detected or extracted for purpose of owner identification or integrity verification of tested images (Cox et al., 1997).

In most digital image watermarking applications, the watermarked image is likely to be under some processing operations before it reaches the receiver. The processing operations include lossy compression, additive noise, enhancement and image filtering. An embedded watermark may intentionally or unintentionally be damaged by such processing operations. An attack is similar to the processing operations but it is deliberate and focused on impairing the detection of the watermark. The earliest malice attacks uses a trial and error procedure to estimate a combination of pixel values that has the largest influence on the detector for the least disturbance of the image, and then uses this estimate in order to eliminate the watermark (Voloshynovskiy et al., 2001). In general, all processed watermarked image, whether under a well meaning processing operation or a malice attack, is called an attacked watermarked image.

Depending on the domain in which the watermark information is embedded, digital watermarking techniques are classified as spatial and spectral domain techniques. The spatial approaches modify directly the intensity of image pixels to embed a watermark. One of the earliest spatial methods is based on the division of an image into two sub-images and adds a constant value to one group of the image (Pitas, 1996). Another method modifies the pixel information located in the specific position within an

*Corresponding author. E-mail: amlatif@gmail.com. Tel: +98-351-8241080.

image (Kutter et al., 1998).

Spectral domain approaches transform the original image into the frequency domain and modulate frequency coefficients to embed the watermark. In general, spectral domain methods are more robust than spatial domain against many common attacks. A fundamental advantage of the spectral techniques is that the transformed image has a good energy compactness properties and most of the image energy can be captured within a relatively small number of coefficients (Latif et al., 2010).

Discrete Fourier Transform (DFT) followed by the so-called Fourier-Mellin Transform (FMT), which are the earliest transforms that were used for image watermarking, can be proven to be against rotation, scale and translation invariant (Joseph et al., 1998). In another scheme introduced by Premaratne, the watermark is set as a spread spectrum signal that is embedded in the transform domain. For watermark detection in this scheme, the original image is subtracted from the watermarked image and the residual is transformed to the frequency domain where it is highly correlated with the watermark signal (Premaratne and Ko, 1999).

In addition, the DCT domain as other spectral based methods has been used extensively for embedding a watermark in images and videos. Using the DCT, an image is divided into frequency bands and the watermark is embedded in low and middle frequency bands. Sensitivities of the human visual system to changes in the DCT bands have been extensively studied in context of the JPEG compression. The results of these studies can be used to minimize the visual impact of the watermark embedding distortion. Note that, the JPEG and MPEG coding are based on the DCT decomposition, and embedding a watermark in the DCT domain makes it possible to integrate watermarking with image and video coding and produce real-time watermarking applications (Suhail and Obaidat, 2003). Thus, we focused on digital image watermarking scheme in the DCT domain.

Digital image watermarking algorithms have some requirements such as imperceptibility and robustness. The imperceptibility presents that the distortion between the host and watermarked image should remain imperceptible to a human observer. The robustness means the ability of the receiver to detect or extract the watermark from attacked watermarked images (Lee et al., 2008). It is important to note that the contradictions between the requirements of watermarking can cause a great deal of difficulty. Increasing the watermark strength in embedding procedure would increase the robustness, at one hand, and lower the imperceptibility, on the other hand, and vice versa. Many researchers attempted to provide different solutions in order to balance these two conflicting requirements. Traditional watermarking algorithms solved this problem by choosing the watermark strength via trial and error, which will always be inefficient (Cox et al., 1997).

Furthermore, some adaptive solutions based on human visual system are addressed. Wolfgang et al. (2002) presented an adaptive watermarking method using human visual model. Their model provided the maximum strength that can be inserted without visual distortion (Wolfgang et al., 2002). Mei et al. (2003) pointed out that the experiments on embedding watermark using visual mask into the DCT coefficients does not visually cause degradation of the images when watermark strength is larger than just noticed difference. They used an artificial neural network to model human visual system to decide watermark strength of the DCT coefficients (Mei et al., 2003). Jin and Wong (2007) introduced a neural network technique to estimate the watermark strength. They used different textural features and luminance to decide adaptively the watermark strength (Jin and Wang, 2007). In methods based on learning techniques such as neural network, many training samples are required. In addition, there are some masking effects that could be incorporated into the visual models and these effects are not considered in most studies in this category.

To tackle this problem in some recent researches, the watermark strength has been treated as an optimization or search problem. For example, Kumsawat et al. (2005) proposed a watermarking scheme based on the Genetic Algorithm (GA) in discrete wavelet transform domain. In their method, the GA was used to search the optimal watermark strength in order to improve the requirements of watermarking algorithm. The major advantage of using the GA for optimizing the strength is achieving high imperceptibility and good robustness simultaneously; however, the GA complexity is high (Aslantas et al., 2009; Kumsawat et al., 2009).

Thus, we introduce Reinforcement Learning (RL) technique to address the complexity problem of the GA method. Reinforcement learning is a powerful learning methodology that is based on interaction with the environment. This method uses only one scalar performance index called reinforcement signal to train agents in complex, nondeterministic and stochastic environments without need of a supervisor. The superiority of the RL algorithm over other alternatives, including GA, is its low cost computation and high explorations (Derhami et al., 2010). In this study, the scalar performance index in RL algorithm is defined to include the robustness and imperceptibility properties. Furthermore, to obtain the best value for the watermark strength to satisfy both required robustness and imperceptibility, simultaneously, a novel estimation methodology is offered. In experimental section, it is shown that the RL algorithm has higher performance and faster speed than the GA scheme.

METHODOLOGY

For the first time, we introduce a novel RL scheme to determine the watermark strength of DCT based digital image watermarking. This scheme guarantees the imperceptibility of embedded watermark and maximizes the watermark strength to achieve better robustness against most attacks. In this section, we first briefly introduce the

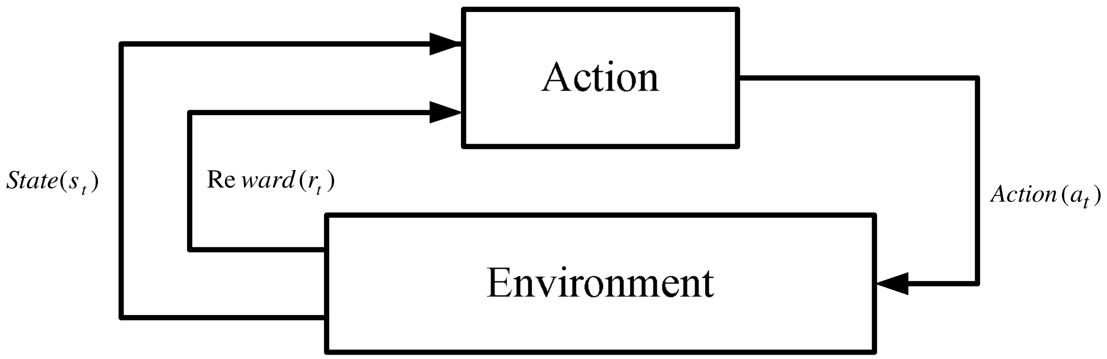

Figure 1. Reinforcement learning framework.

general RL scheme. Then, the embedding and detecting procedure of watermarking algorithm is discussed. Next, an adaptive estimation scheme to find the best watermark strength to achieve both the robustness and imperceptibility, simultaneously, is presented.

Reinforcement learning

Reinforcement learning is a learning method of how to behave in order to maximize a numerical reward signal. In the reinforcement learning, the learner or the decision-maker is called an agent. It is assumed that anything outside the agent in the system comprises the environment (Sutton and Barto, 1998).

Figure 1 shows the relation between the action and environment stated in the RL algorithm. At each time step, the agent observes the current state and selects an action from the set of actions under the decided policy and applies it to the environment. The environment states with transition probability $p(s_t, a_t, s_{t+1})$ goes to the next state $s_{(t+1)}$, and the agent receives the reinforcement signal $r_{t+1} = r(s_t, a_t)$. Policy is a rule that an agent uses in each state to select the corresponding action. It is denoted as π, where $\pi(s,a)$ is the probability of selecting action a in state s.

The key idea of the RL is the use of value functions to organize the search for good policies. The agent tries to take actions to reach the states with greater values. The value of a state is the sum of the discounted reinforcement signals that the agent can expect to receive after starting from that state. The value function of state s under policy π is denoted by $V^{\pi}(s)$ (*Sutton and Barto*, 1998):

$$V^{\pi}(s) = E_{\pi}\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s\}\} \quad 0 \le \gamma \le 1 \tag{1}$$

In this formula, γ is the discount factor, and $E_{\pi}\{.\}$ is the expected value. Similarly, the value of action a in state s under policy π is called the action value function that is denoted by $Q^{\pi}(s,a)$:

$$Q^{\pi}(s,a) = E_{\pi}\{\sum_{k=0}^{\infty} \gamma^k_{t+k+1} \mid s_t = s, a_t = a\} \tag{2}$$

There are some algorithms to estimate the action value function. The action value is a measure of suitability of the action. Q-learning algorithm is one of the most well known algorithms for this purpose. The Q-learning method estimates the maximum value of action a in state s and is denoted by $Q(s,a)$ on all possible policies according to the following update formula (Guo et al., 2004):

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha_t[r_{t+1} + \gamma \max_{b \in A} Q(s_{t+1}, b) - Q(s_t, a_t)] \tag{3}$$

Where α_t is the learning rate, and γ_{t+1} is the immediate reward received from the environment states after applying action a_t in state s_t.

There are several policies for action selection. Here, we use ε-greedy method. In this method, with probability $1-\varepsilon$ the action with highest estimated action value is selected or with small probability, ε, one of the actions is selected uniformly. The following formula shows selection probability of each action in ε-greedy method:

$$P(a) = \begin{cases} 1-\varepsilon+\dfrac{\varepsilon}{N} & a = \arg\max\limits_{b \in A}(Q(s,b)) \\ \dfrac{\varepsilon}{N} & otherwise \end{cases} \tag{4}$$

Where N is the total number of actions in state s.

Watermarking procedure

The watermarking schemes consist of two procedures. The former is embedding the watermark in the host image and the latter is detecting the watermark in the watermarked image. In this study, a pseudo-random sequence $W = \{w_1, w_2, \cdots, w_M\}$ of length M, which has a normal distribution with zero mean and unit variance, is generated as the watermark. The $N \times N$ DCT for an $N \times N$ gray scale image I is computed and the DCT coefficients $X = \{x_1, x_2, \cdots x_{N \times N}\}$ are reordered into a zig-zag scan (Wallace, 2002). The coefficients from $(K+1)th$ to $(M+K)th$ are taken where, the first K coefficients are skipped for embedding the middle band to achieve the perceptual invisibility of the watermark

Table 1. Summary of the embedding and extracting procedures.

Embedding procedure
Compute the DCT of the whole image;
Select the middle frequency coefficients;
Embed the watermark according to Equation 5;
Compute the inverse DCT of the result to obtain the watermarked image.
Detection procedure
Compute the DCT of watermarked image;
Select the middle frequency coefficients;
Compute NCC between the coefficients and original watermark;
Compare the NCC with the predefined threshold;
Decide the image is watermarked or not.

without loss of the robustness against signal processing operations (Liang, 2008).

Then, the watermark is scaled according to the watermark strength of the particular frequency component. The vector $X' = \{x'_{K+1}, x'_{K+2}, \cdots, x'_{K+M}\}$ with the marked DCT coefficients is computed according to the following rule:

$$x' = x_{K+1} + \lambda |x_{K+i}| w_i \quad (5)$$

Where $i = 1, 2, \cdots, M$ and λ is the watermark strength. Finally, X' is reinserted in the zig-zag scan and the inverse DCT is performed, thus obtaining the watermarked image I' (Barni et al., 1998).

In the detection procedure, since our study focuses on blind watermarking where the original host image is not available at the receiver, given a possibly corrupted image I^*, the $N \times N$ DCT is applied, then the DCT coefficients of I^* are reordered into the zig-zag scan and the coefficients from the $(K+1)th$ to the $(M+K)th$ are selected to generate a vector $X^* = \{x^*_{K+1}, x^*_{K+2}, \cdots, x'^*_{K+M}\}$. Note that it is impossible to get an estimate of the watermark by subtracting the non-watermarked DCT coefficients from X^*. Therefore, the Normalized Cross Correlation (NCC) between the attacked watermarked coefficients and the original watermark is taken into account as the measure of the watermark presence (Zheng et al., 2007). Finally, The NCC is compared with a predefined threshold and the presence of the watermark is determined. The embedding and detection procedure of watermarking algorithm are summarized in Table 1.

Watermark strength estimation

In this section, a system based on the RL is designed and incorporate into the watermarking scheme to satisfy the user's watermark strength preference. The main challenges of incorporating the RL in watermarking shown in Table 1 are as follows.

Define action space

The action space is defined as the range of the watermark strength. For example, the watermark strength value is a real number in the range of (0, 10), thus taking this range as the RL action space. The weights for the selected action values are picked equally at the beginning of the watermarking scheme. Then, the embedding and detecting procedure are performed as the implementing of the action. As it was explained earlier, selecting the proper value for watermark strength is a challenge, itself. Later in this paper, an adaptive estimation scheme is introduced to reach the best action with the desired robustness and imperceptibility accuracies.

Define reinforcement signal

To evaluate imperceptibility of each action the Peak Signal to Noise Ratio (PSNR) between the host image and watermarked image is applied (Petitcolas, 2000). This amount is considered as a positive reward signal. The more reward for each action, the higher action value for the next step. Then, to evaluate the robustness of watermarking scheme, some attacks are applied to the watermarked image and then, the NCC between the original watermark and the coefficients of attacked watermarked image is evaluated. The difference of the NCC between the coefficients of watermarked image and the attacked watermarked image with the original watermark is considered as the punishment. The bigger punishment in each action results into the lower action value for the next step.

With respect to the definition of positive reward and punishment, we define the reinforcement signal as follows:

$$r = \tanh(\rho_1 PSNR - \sum_{i=1}^{n} \rho_{i+1} (NCC_i - NCC_i^*)) \quad (6)$$

In this formula, r is the reinforcement signal, $PSNR$ is the peak signal to noise ratio between the host image and the watermarked image, NCC is the normalized cross correlation of the watermarked image, NCC^* is the normalized cross correlation of the attacked watermarked image, ρ_i is the weight factor for reward and punishment, n is the number of attacks and $\tanh(\cdot)$ is the hyperbolic tangent. Each weight factor represents the importance of each index during the search process. It should be noticed that the relationship $\sum_{i=1}^{n+1} \rho_i = 1$ must always be hold. It should also be noted that the hyperbolic tangent function in the formula is used for normalizing the reinforcement signal in the range of (-1, 1).

The watermarking procedure is repeated several times until the watermark with desired watermark strength is achieved. Every time the procedure is repeated, the agent selects an action (watermark strength) among the action set based on the action value of each action and the action selection rule, which is introduced in reinforcement learning. Then, the reinforcement signal is computed and, as a result, the action value of the selected action is updated according to Equation 3. The watermarking procedure is repeated with the updated action values from previous run and the updating procedure are redone to update the action value of the selected action, again. This learning phase proceeds until the termination condition is met or time is expired.

Adaptive estimation of watermark strength

As mentioned earlier, the standard Q-learning is a method that is used for problems with discrete states and action spaces, but the best watermark strength value may not be found in discrete space. Note that to reach a solution with desired accuracy, a fine-grained discretization method is required. Fine-grained discretization methods quickly result in a large number of actions and consequently increases the learning time and computation cost significantly (Tsitsiklis and Van Roy, 2002; Wiering, 2004). Thus, a new adaptive algorithm to estimate the watermark strength in the standard RL algorithm is introduced to overcome this weakness.

This algorithm begins with dividing the predefine interval of watermark strength into arbitrary equal subintervals. An arbitrary value in each interval, for example the midpoint, is selected as the RL actions. Then, the watermarking scheme is implemented separately for each action and the RL agent performs the learning process and updates the action value for each interval. The updated action values of these intervals are compared and if the difference between action values is higher than a threshold, the associated interval will be picked. Next, the picked interval would be divided into new equally spaced subintervals. The same process as before is repeated to find the interval with the largest action value among them. This divide and conquer process is repeated again and again until the terminal conditions are met. A value with high action value within final interval is picked as the required watermark strength.

To illustrate the decision making process, let us assume that the range of (0, 10) is divided into 10 subintervals, (0, 1), (1, 2) ..., (9, 10). The midpoint values of each interval, namely, 0.5, 1.5, ..., 9.5, are selected as the watermark strength and the watermarking scheme are performed 10 times. Let us assume after some iterations of the RL algorithm, the difference between two action values leads to higher value than predefined threshold. Then, the interval of the action with the highest action value and next action and the interval of the action with the highest action and previous action are divided to new subinterval. For example, assume the action value of 3 is the highest action value. Thus, the interval (2, 3) and (3, 4) are subdivided into 10 new ones, (2, 2.2), (2.2, 2.4) ..., (3.8, 4), and the whole learning scheme is repeated again for the midpoint of each these new intervals. This time, let's say again the difference between the action values is reached to threshold; the whole process would be repeated with two of these intervals and a new subinterval would be located. We repeat this process until the terminal conditions are reached. The midpoint of subinterval associated with the largest action value would be considered the desired watermark strength value.

Another example is given in Figure 2. In this figure, circles point out the actions and the numbers in the circles are watermark strength, also ellipses point out the value of each action. In this example, the mentioned threshold is set to 0.05. The difference between 0.20 and 0.26, which are the action value of 1 and 3, is 0.06. The maximum action value belongs to 3 and thus, the adaptive process has to be performed between (2, 3) and (3, 4). The interval (2, 4) is divided into equal subinterval and the action values are initialized by zeros. Then, new search around the best solution needs to be continued. As shown in Figure 2, in some next steps, the new subdivision takes place and the interval (2.29, 2.87) is divided to subinterval.

EXPERIMENTAL RESULTS

The RL algorithm and the watermarking scheme as well as the decision making process to pick the best value for the watermark strength is simulated using an Intel Pentium IV processor of 3 GHz and 2 GB RAM with Microsoft Windows XP operating system and the MATLAB 7.3 software package.

The choice of attacks, which is used for evaluating the robustness of watermarking scheme, depends on the application of watermarking scheme. In this study, the selected attacks are the JPEG compression (quality factor %50), median filtering (window 3×3) and averaging filtering (window 3×3). The JPEG compression is applied as the attacking function because of the popularity of transmitting JPEG images through the Internet. Also, median filter is a non-linear spatial filter which is normally used to remove noise spike from an image. The average filter smoothes image data to eliminate noises. This filter performs spatial filtering on each individual pixel in image using the gray level values in a square window of size 3×3 surrounding each pixel (Petitcolas, 2000).

In order to test the proposed watermarking algorithm, 500 watermarks are randomly generated (the 100th watermark is the correct one). Some gray level standard images are watermarked, and the selected signal processing attack techniques are applied to evaluate whether the detector can reveal the presence of the watermark; thus measuring the algorithm robustness.

Four host images are selected and they are gray-scale pictures of size 512×512 with 256 intensity levels. The images are identified by their names: Baboon, Cameraman, Lena, and Pepper. Baboon represents images with large areas of complex texture and includes homogeneous areas; Cameraman is chosen for its flat and high contrast regions; Lena has a mixture of characteristics (e.g. smooth background, big carves and the hat in the image includes complex textures); and Pepper provides luminosity changed (that is light reflection surfaces). An illustration of the host images are shown in Figure 3. The selected watermark is a pseudo-random sequence of length 1000, which has a normal distribution with zero mean and unit variance.

The aim of the proposed method is to find the best watermark strength value (λ_{opt}) between 0 and 10 to satisfy the requirements of watermarking scheme, simultaneously. For comparison, the results from another

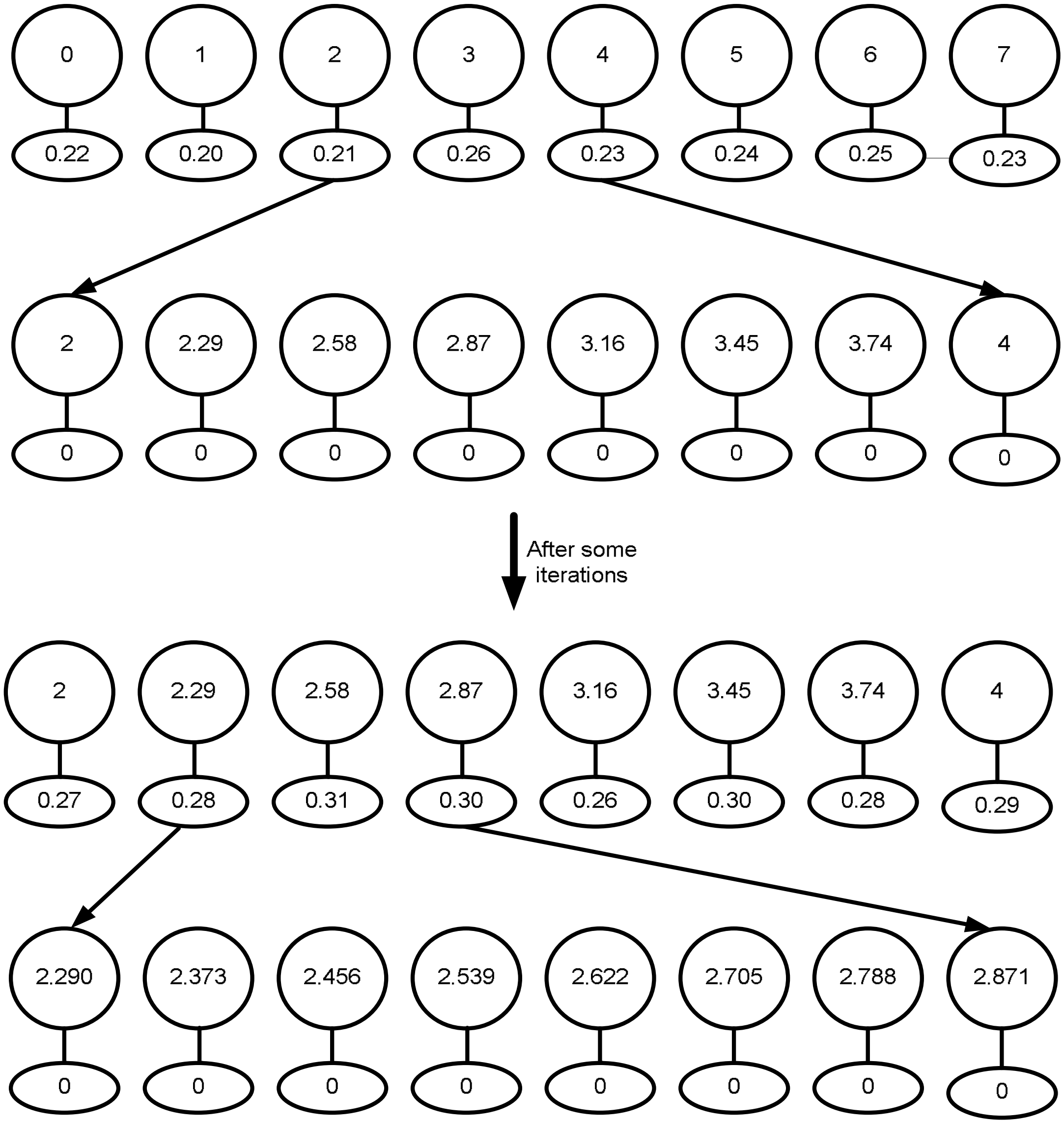

Figure 2. An adaptive discretization example.

work using genetic algorithm (GA) to find the watermark strength value in a watermarking scheme is also implemented (Shieh et al., 2004). The GA is a conventional search technique which is capable of optimizing non-linear function. It begins by defining the optimization parameters to find the fitness function; and it ends by testing for convergence. According to the GA methodology, designer needs to carefully define the necessary genetic elements for training.

In our method, the interval is equally divided into 10 sub-intervals. The weight for each action is initialized to zero. In the estimating the watermark strength, a threshold must be assigned. For our simulations, we set the threshold to 0.05. In addition, the required parameters K and M in the embedding part of the watermarking scheme are set to 16000 and 1000, respectively.

The watermarked images resulted from the proposed method and PSNR of them are illustrated in Figure 4 and the responses of watermark detector are illustrated in Figure 5. In these experiments, the watermark is embedded with the watermark strength which is evolved by the RL methodology. It is clear that the quality of the watermarked image using the proposed scheme is high and satisfactory. In addition, the computed NCC of the algorithm is large, thus the robustness of the proposed algorithm against many attacks is guaranteed.

DISCUSSION

The simulation results based on both the new reinforcement learning (RL) technique and the genetic algorithm (GA)

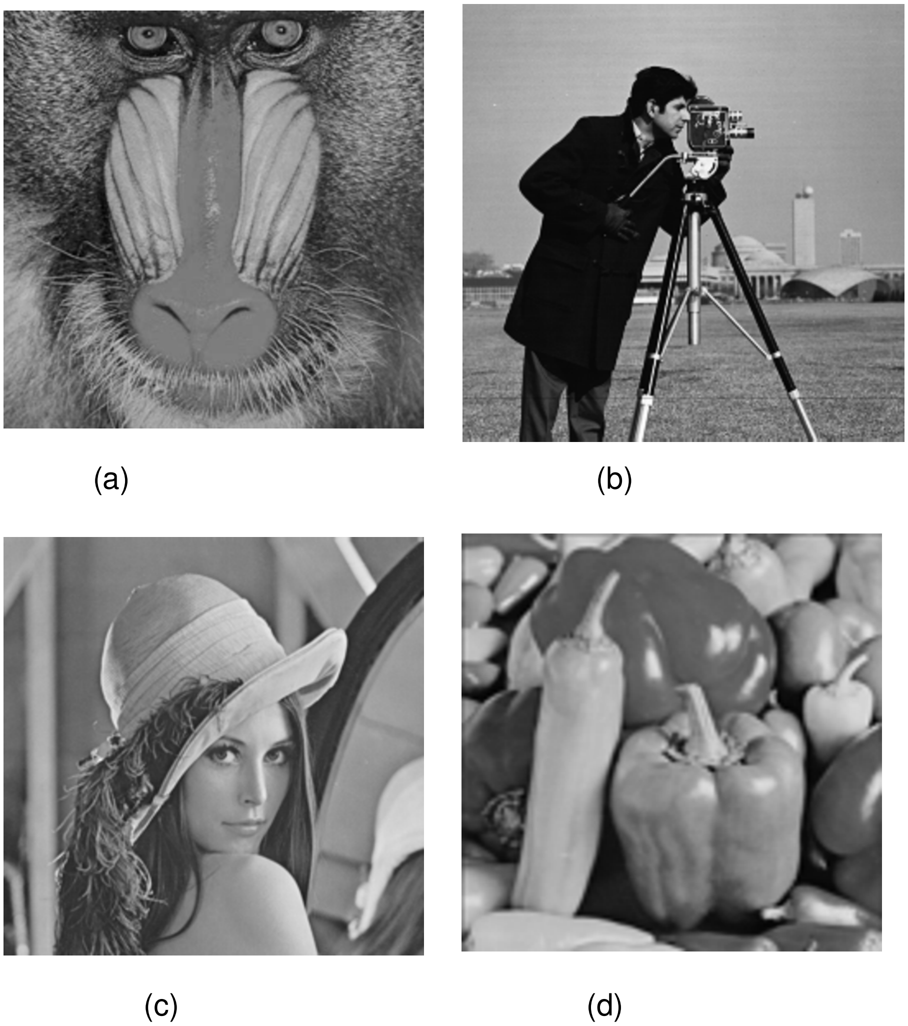

Figure 3. Host images (a) Baboon (b) Cameraman (c) Lena (d) Pepper.

are presented in Table 2. The table reports the average on the obtained results from 20 independent runs of both algorithms. As indicated earlier, the PSNR values correspond to the level of imperceptibility of the algorithm. The results show that the PSNR values obtained from both methods are sufficiently high and thus, the imperceptibility level may consider being fine enough for the human visual system. In addition, the normalized cross correlation between the original watermark and the transformed coefficients of the watermarked image against the JPEG compression, median and average filtering is computed and reported in this table. The results show that the robustness of the watermarking scheme using the RL algorithm is higher than the GA.

The sixth row of the table shows the execution time for both methods. It can be seen that the watermarking scheme using the RL algorithm is about three times faster than the GA based one. This is due to the fact that our proposed RL method searches in a set of limited candidate actions which is determined by the adaptive estimation methodology explained earlier. Moreover, in each step of the RL algorithm, only one action is selected and the reinforcement signal is computed for the selected action; whereas the GA searches all of the search space by beam search. That is, RL algorithm helps to overcome the time-consuming disadvantage of the GA watermarking.

The last row of the table shows the standard deviation of the found watermark strength in all independent runs. The smaller amount of the standard deviation for watermark strength in the RL algorithm shows that the search around solution is smoother and the RL converges to the desired solution faster.

Conclusion

In this paper, a new novel adaptive estimating method for decision making process and identifying the watermark

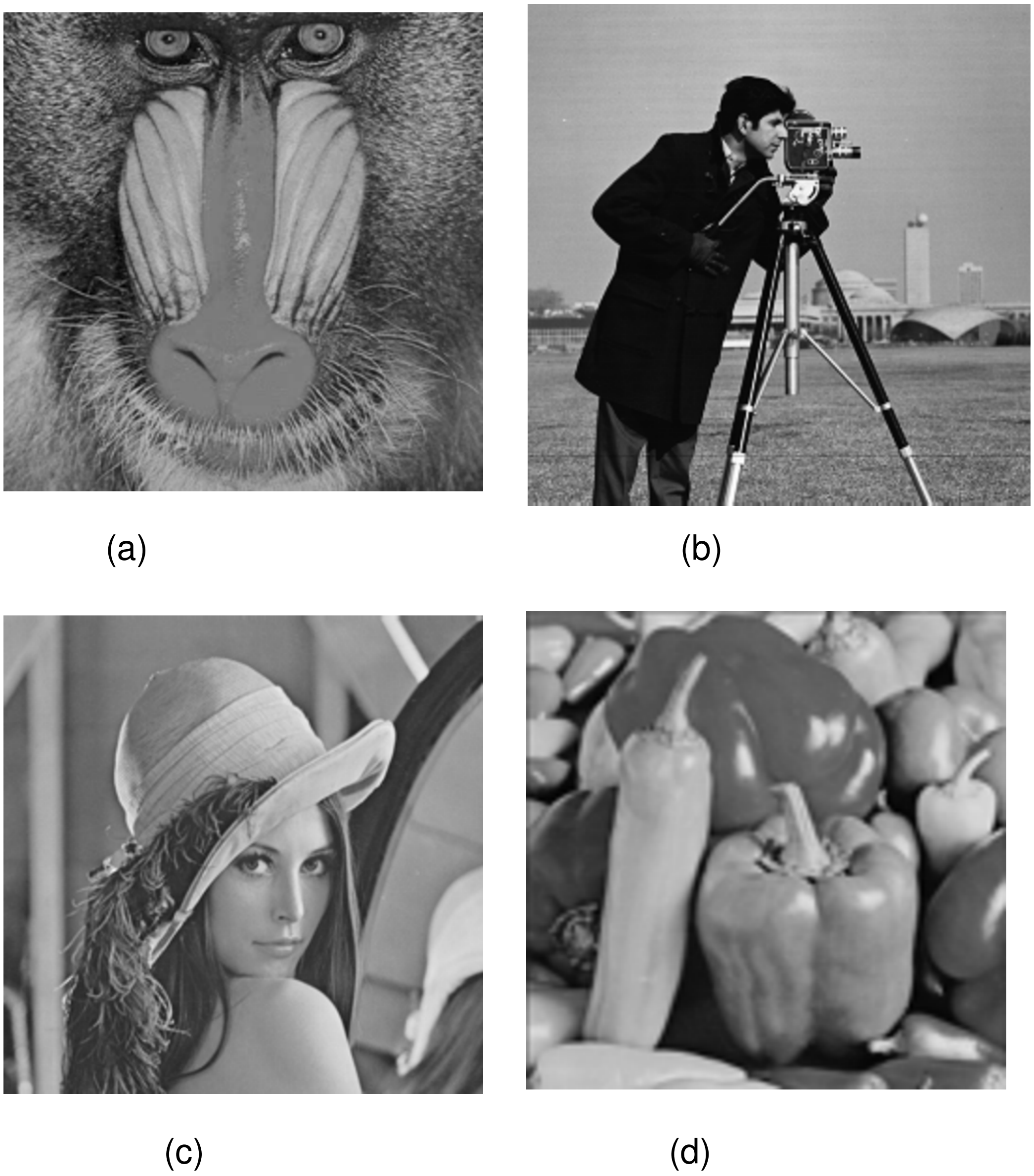

Figure 4. Watermarked images (a) Baboon; PSNR = 61.65 (b) Cameraman; PSNR = 62.43 (c) Lena; PSNR = 61.89 (d) Pepper; PSNR = 62.65.

Table 2. Numerical results for the GA and the RL algorithm.

Parameters	**Baboon**		**Cameraman**		**Lena**		**Pepper**	
	GA	**RL**	**GA**	**RL**	**GA**	**RL**	**GA**	**RL**
λ_{opt}	1.25	1.41	1.27	1.62	1.03	1.13	1.08	1.12
$PSNR$	61.6	62.9	61.9	63.3	61.1	62.4	62.3	63.6
NCC_1 [1]	0.601	0.679	0.679	0.692	0.627	0.677	0.635	0.644
NCC_2 [2]	0.615	0.682	0.682	0.685	0.623	0.681	0.606	0.637
$NCC3$ [3]	0.654	0.672	0.672	0.675	0.663	0.678	0.663	0.683
Time	196	63	163	54	191	73	227	74
S.D.[4]	1.08	0.61	0.61	0.43	1.2	0.72	0.94	0.52

1 for NCC of JPEG attack; 2 for NCC of Median attack; 3 for NCC of Average attack; 4 for Standard deviation.

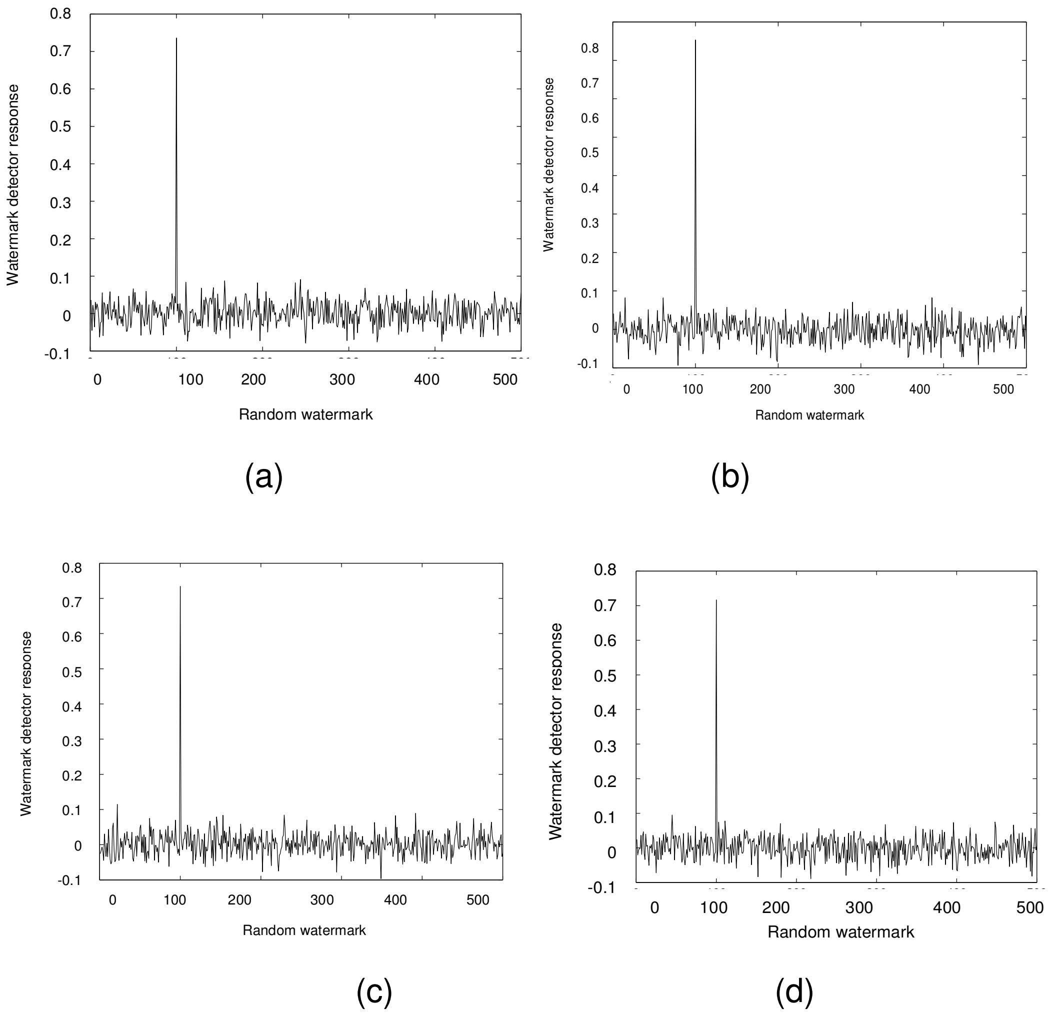

Figure 5. Detector response different images (a) Baboon (b) Cameraman (c) Lena (d) Pepper.

strength value using Reinforcement Learning (RL) in the DCT domain is presented. The motivation behind the proposed algorithm is to obtain the best watermark strength in order to have a required imperceptibility and robustness, simultaneously. Note that the determination of watermark strength is formulated as an RL problem. To count both robustness and imperceptibility in the watermarking scheme, a parameter called reward signal is defined in the RL scheme that relates both robustness and imperceptibility. This scheme guarantees the imperceptibility of embedded watermark and, at the same time, maximizes the watermark strength to achieve better robustness against most attacks. The main advantages of the proposed method over the Genetic Algorithm (GA) based method are low computational cost and converge to better robustness. The experimental results show that while the robustness of the watermarking scheme using the RL algorithm is higher than the GA, the same image imperceptibility quality is attained for both. Furthermore, the RL algorithm converges faster than the GA algorithm to the desired watermark strength value which results into lower computation cost. In our runs, we reached an average of three times faster speed using the RL algorithm in compare with the GA based method. The experimental results comparing the two RL based and GA based methods confirm the superiority of our methodology.

ACKNOWLEDGEMENTS

The office of graduate studies at the University of Isfahan is appreciated for their support. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

Aslantas V, Ozer S, Ozturk S (2009). A novel image watermarking method based on discrete cosine transform using genetic algorithm, Signal Processing and Communications Applications Conf., pp. 285-288.

Barni M, Bartolini F, Cappellini V, Piva A (1998). Copyright protection of digital images by embedded unperceivable marks, Image Vision Comput., 16(12-13): 897-906.

Cox, IJ, Kilian J, Leighton FT, Shamoon T (1997). Secure spread spectrum watermarking for multimedia, IEEE Trans. Image Process., 6(12): 1673-1687.

Derhami V, Majd VJ, Nili Ahmadabadi M (2010). Exploration and exploitation alance management in fuzzy reinforcement learning, Fuzzy Sets Syst., 161(4): 578-595.

Guo M, Liu Y, Malec J (2004), A new Q-learning algorithm based on the metropolis criterion, IEEE Trans. Syst, Man, Cybernetics, 34(5): 2140-2143.

Jin C, Wang S (2007). Applications of a neural network to estimate watermark embedding strength, IEEE Computer Society.

Joseph JK, Ruanaidh O, Pun T (1998). Rotation, scale and translation invariant digital image watermarking, Signal Process., 66(3). 303-317.

Kumsawat P, Attakitmongcol K, Srikaew A (2005). A new approach for optimization in image watermarking by using genetic algorithms, IEEE Trans. Signal Process., 53(12): 4707-4719.

Kumsawat P, Attakitmongcol K, Srikaew A (2009). Robust image watermarking based on genetic algorithm in multiwavelet domain, Int. conf. on Artificial intelligence, knowledge engineering and data bases, pp. 390-395.

Kutter M, Jordan F, Bossen F (1998), Digital signature of color images using amplitude modulation, J. of Electronic Imaging, 7(2): 326-332.

Latif A, Naghsh-Nilchi AR, Monadjemi SA (2010). A parametric slant-Hadamard system for robust image watermarking, J. Circuits, Syst, Comput., 19(2): 451-477.

Lee ZJ, Lin SW, Su SF, Lin CY (2008), A hybrid watermarking technique applied to digital images, Appl. Soft Comput., 8(1): 798-808.

Liang T, Zhi-jun F (2008). An adaptive middle frequency embedded digital watermark algorithm based on the DCT domain, Int. conf. on Management of e-Commerce and e-Government, pp. 382-385.

Mei S, Li R, Dang H, Wang Y (2003). Decision of image watermarking strength based on artificial neural-networks, Proc. Int. Conf. Neural Information Process., 5: 2430-2434.

Petitcolas F (2000). Watermarking schemes evaluation, IEEE Signal Process. Mag., 17(5): 58-64.

Pitas I (1996). A method for signature casting on digital images, Int. Conf. Image Process., 3: 215-218.

Premaratne P, Ko CC (1999). A novel watermark embedding and detection scheme for images in DFT domain, Int. Conf. Image Process. Applications, 2: 780-783.

Shieh CS, Huang HC, Wang FH, Pan JS (2004). Genetic watermarking based on transform-domain techniques, Pattern Recognition, 37(3), 555-565.

Suhail M A, Obaidat MS (2003), Digital watermarking-based DCT and JPEG model, IEEE Trans. Instrumentation Measure., 52(5): 1640-1647.

Sutton RS, Barto AG (1998). Reinforcement learning: An introduction, The MIT press.

Tsitsiklis JN, Van Roy B (2002). An analysis of temporal-difference learning with function approximation, IEEE Trans. Automatic Control, 42(5): 674-690.

Voloshynovskiy S, Pereira S, Iquise V, Pun T (2001). Attack modelling: Towards a second generation watermarking benchmark, Signal Process., 81(6): 1177-1214.

Wallace GK (2002). The JPEG still picture compression standard, IEEE Trans. Consumer Electronics, 38(1): 30-44.

Wiering MA (2004). Convergence and divergence in standard and averaging reinforcement learning, Machine Learning, pp. 477-488.

Wolfgang RB, Podilchuk CI, Delp EJ (2002). Perceptual watermarks for digital images and video, Proc. IEEE, 87(7): 1108-1126.

Zheng D, Liu Y, Zhao J, Saddik AE (2007). A survey of RST invariant image watermarking algorithms, ACM Comput. Surveys, 39(2): 5.