EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

A game-theoretic architecture for visible watermarking system of ACOCOA (adaptive content and contrast aware) technique

Min-Jen Tsai* and Jung Liu

Abstract

Digital watermarking techniques have been developed to protect the intellectual property. A digital watermarking system is basically judged based on two characteristics: security robustness and image quality. In order to obtain a robust visible watermarking in practice, we present a novel watermarking algorithm named adaptive content and contrast aware (ACOCOA), which considers the host image content and watermark texture. In addition, we propose a powerful security architecture against attacks for visible watermarking system which is based on game-theoretic approach that provides an equilibrium condition solution for the decision maker by studying the effects of transmission power on intensity and perceptual efficiency. The experimental results demonstrate that the feasibility of the proposed approach not only provides effectiveness and robustness for the watermarked images, but also allows the watermark encoder to obtain the best adaptive watermarking strategy under attacks.

Keywords: copyright protection, visible watermarking, watermarking game, Nash equilibrium, wavelet

1. Introduction

Due to the advancement of digital technologies and rapid communication network deployment, a wide variety of multimedia contents have been digitalized which makes their duplication or circulation easy through both authorized and unauthorized distribution channels. With the advantages of effortless editing and digital data reproduction, the protection of the intellectual rights and the authentication of digital multimedia have become issues of great importance in recent years [1-3]. Among different techniques, visible watermarking schemes protect copyrights in a more active way since the logo watermark are generally embedded in the host image (Figure 1a). Such approach not only allows the observers to easily recognize the property owner of multimedia but also discourage the action of pirates.

In this study, we have explored the inter-relationship between the image fidelity and robust requirement of visible watermarking and propose a powerful secure watermarking architecture which is based on game-theoretic methodology. The system provides an equilibrium condition solution for the copyright manager to make a decision by studying the effect of transmission power on intensity and perceptual efficiency. In addition, we have formulated the watermark embedding problem as a dynamic non-cooperative game with complete information [4]. Complete information requires that every player knows the strategies of the other players but not necessarily the actions. Under the complete information, we present a game-theoretic architecture as a watermarking game to analyze the different situation and get the best strategy between the embedding watermark energy and the perceptual translucence for visible watermark where the best strategy is defined by the Nash equilibrium of the game [4]. Tsai and Liu's research [5] has preliminary study for visible watermarking which only applies peak signal noise ratio (PSNR) and correlation for the payoff functions. However, visual image quality measure is very critical for visible watermarking and such an issue should be included and weighted during the algorithm design. Therefore, we here leverage the previous research of [5] not only to consider the above discussion but also improve the visible watermarking technique for a novel payoff function under the game-theoretic architecture.

* Correspondence: mjtsai@cc.nctu.edu.tw
Institute of Information Management, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsin-Chu, 300, Taiwan

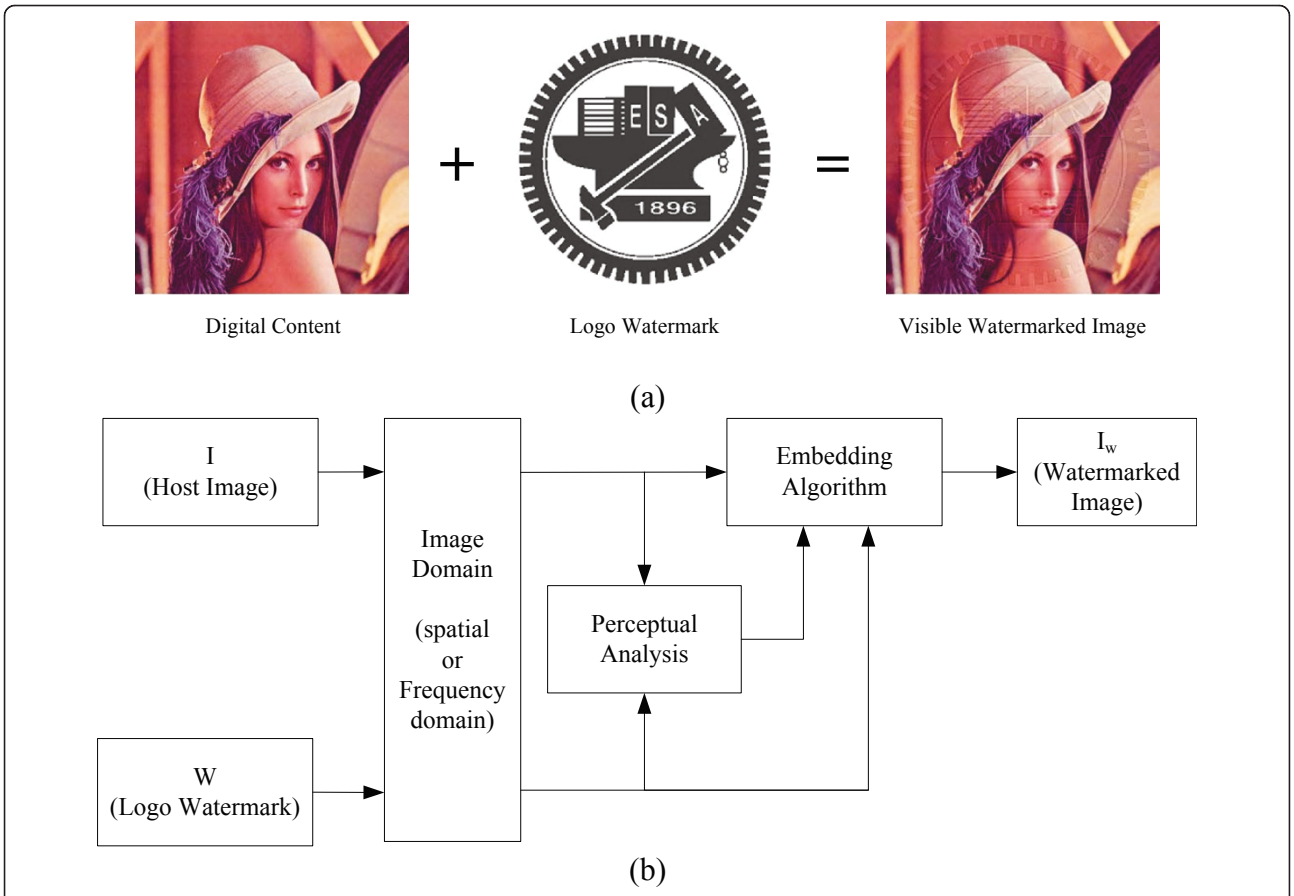

Figure 1 The visible watermark embedding procedures. **(a)** An example of visible watermark embedding. **(b)** A generic visible watermark embedding diagram.

The rest of this article is organized as follows. In section 2, related works about visible watermarking and game-theoretic architecture will be introduced briefly. In section 3, we will give the detailed description of the proposed watermarking algorithm called ACOCOA (adaptive content and contrast aware) and a power security watermarking architecture design. In section 4, numerical results with discussion will be presented. Finally, the conclusions and future works are in section 5.

2. Related works

2.1. Digital watermarking

Digital watermarking techniques are the process of possibly irreversibly embedding information into a digital signal and they are used to protect copyright of digital multimedia like sound, music, audio, images, or video files that have to be delivered for certain purpose, such as digital multimedia used in exhibition, digital library, advertisement, or distant learning web, while illegal duplication is forbidden.

A review of the literature indicates that the visible watermarking studies have captured significant attention since their applications meet the requirements of many media industries [2,3].

Through the survey, Braudaway et al. [6] proposed one of the early approaches for visible watermarking by formulating the non-linear equation to divide the linear brightness scale into two regions and accomplish the brightness alteration in spatial domain. Meng and Chang [7] proposed an efficient compressed-domain content-based algorithm which applied the stochastic approximation model for Braudaway's method in the discrete cosine transform (DCT) domain by adding visible watermarks in video sequences. Kankanhalli et al. [8] proposed a coefficient modulation in the DCT domain where the scaling factors are calculated by exploiting the human visual system (HVS), to ensure that the perceptual quality of the host image is preserved. Mohanty et al. proposed a watermarking technique called dual watermark, which is a combination of a visible watermark and an invisible watermark in the spatial domain. The visible watermark is adopted to establish the owner's right to the image and invisible watermark is used to check the intentional and unintentional tampering of

image [9]. Due to the watermark insertion is done in the spatial domain, the image fidelity and robustness under attacks is pretty low. Tsai and Lin have developed more advanced approach in [10] by considering the global and local characteristics of the host and watermark images in the discrete wavelet transform (DWT) domain. Consequently, Mohanty et al. [11] also proposed a mathematical modification model for exploiting the texture sensitivity of the HVS in DCT domain. The weakness of this approach is the necessity to keep the watermark secret which is very unrealistic for visible watermarking. Better design is achieved in [12] and the approach is leveraged in this research. Chen [13] has proposed a visible watermarking mechanism to embed a gray level watermark into the host image where the strength of the embedded watermark locally depends on the standard deviation of luminance.

Vehel and Manoury [14] proposed a method for digital image watermarking which is based on the modification of certain subsets of the wavelet packet decomposition (WPD) and the WPD is a generalization of the dyadic wavelet transform with low-pass subbands. Hu and Kwang implemented an adaptive visible watermarking in the wavelet domain by using the truncated Gaussian function to approximate the effect of luminance masking for the image fusion. Based on image features, they first classify the host and watermark image pixels into different perceptual classes. Secondly, they use the classification information to guide pixel-wise watermark embedding. In high-pass subbands, they focus on image features, while in the low-pass subbands, they use truncated Gaussian function to approximate the effect of luminance masking [15,16]. Yong et al. [17] also proposed a translucent digital watermark in the DWT domain and use error-correct code to improve the ability of anti-attack.

Each of above mentioned schemes was not devoted to better feature-based classification and the use of sophisticated visual masking models. Huang and Tang [18] later presented a contrast sensitive visible watermarking scheme with the assistance of HVS. They utilized the contrast sensitive function (CSF) mask of the DWT domain with square function to determine the mask weights and at last they adjusted the scaling and embedding factors based on the block classification with the texture sensitivity of the HVS for watermark embedding. Tsai [12] improved their approach and further proposed a novel visible watermarking algorithm based on the content and contrast aware (COCOA) technique. He utilized the global and local characteristics of the host and watermark images and considered HVS model in the DWT domain by using the CSF, noise visibility function (NVF), and DWT basis amplitude modulation for the best quality of perceptual translucence and noise reduction.

In summary, Figure 1 describes the generic structure for visible watermark embedding processes. First, a host image (original image) directly embeds watermark in spatial domain or is transformed into frequency domain through the well-known spread spectrum approach [19], i.e., Discrete Fourier Transform (DFT), DCT, or DWT. However, the algorithms using transform domain approach develop more robust watermarking techniques than directly embedding watermark into the spatial domain [3,18]. Consequently, coefficients are passed through a perceptual analysis block that determines how strong the watermark in embedding algorithm can be, so that the resulting watermarked image is acceptable. The watermark is embedded through using a well-designed algorithm based on mathematical or statistical model. If the host image is employed in frequency domain, the inverse spread spectrum approach is then adopted to obtain a watermarked image [2,3]. The watermark extraction applies to the similar operations in embedding processes with reverse procedures.

Digital contents embedded with visible watermarks will overlay recognizable but unobtrusive copyright patterns to identify its ownership. Therefore, a visible watermarking technique should retain details of contents and ensure embedded patterns difficult or even hard to be removed, and no one could use watermarked data illegally. How to solve the conflict problem and to determine the best tradeoff between the intensity of embedded watermark and the perceptual translucence for adaptive visible watermark under intentional attacks is becoming a subject of importance [5,12,18]. In this article, we present a game-theoretic architecture to solve this gap by proposing the ACOCOA (adaptive content and contrast aware) algorithm that provides more flexible design for encoder to set the energy of embedding watermark. We will introduce the ACOCOA technique and a game-theoretic architecture for visible watermarking system in details.

2.2. Game theory

Game theory is the formal study of the conflict and cooperation. The concepts of a game-theoretic approach help to formulate structure, analyze and understand strategic scenarios, and make a decision whenever the actions of the several agents are interdependent [4]. Game theory aims to help us to understand the situations in which decision-makers interact. Therefore, decision-makers can better estimate the potential effects of their actions and then make the ideal decisions to avoid the conflict.

There are two types of game theory. One is non-cooperative game, which focuses on analyzing each game player to maximize their own profit. The other is the cooperative game, which concentrates on groups of

players and may enforce cooperative behaviors. Game theory has applications in several fields, such as economics, auctions, bargaining, politics, law, biology, social network, and voting systems. Some games have been proposed and we will briefly address different game techniques here.

Cohen and Lapidoth [20] computed the coding capacity of the watermarking game for Gaussian covertext and squared-error distortions. Both the public version of the game (covertext known to neither attacker nor decoder) and the private version of the game (covertext unknown to attacker but known to decoder) are treated. Moulin et al. [21] proposed an information-theoretic analysis of information hiding. They describe the fundamental limits of information-hiding system, formulate the information-hiding problem as a communication problem, and seek the maximum rate of reliable communication through the communication system.

Among the various theories of game, Nash equilibrium is one of the most important and widespread equilibrium concepts in the twentieth century. Nash equilibrium is a solution concept of a game involving two or more players, in which each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing only his or her own strategy unilaterally. If each player has chosen a strategy and no player can benefit by changing his or her strategy while the other players keep theirs unchanged, then the current set of strategy choices and the corresponding payoffs constitute Nash equilibrium [4]. Under such scenario, the situation of visible watermark embedding strategies against attacks can be formulated as a competition game based on the actions of encoder and attackers. Therefore, we proposed a secure watermarking system based on game-theoretic methodology to achieve the objective of watermarking management. The idea of Nash equilibrium is adopted to develop the solution for the non-cooperative problem. Section 3 will describe how we can apply such a concept to make the game design for making decision of the visible watermark embedding procedures.

2.3. Image quality measure

Image quality measure has become crucial for the most image processing application. It can evaluate the numerical error between the original image and the tested image. Several image quality measure metrics have been developed for incorporating the texture sensitivity of the HVS [22]. However, in the real world, there is yet no universal standard for an objective assessment of image quality. From the image visual quality study of [23], Ponomarenko et al. exploited the color image database TID2008 using a wide variety of known image quality metrics by the rank correlations of Spearman and Kendall. TID2008 database contains 1700 distorted images and 17 different types of distortions. They evaluated both full set of distorted test images in TID 2008 and for particular subsets of TID2008 that include distortions most important for digital image processing applications. Under their investigation, MSSIM, PSNR-HVS, and PSNR-HVS-M perform better correlation correspondence of HVS where PSNR-HVS and PSNR-HVS-M produce similar numerical results. In addition, VIF and WSNR show consistent presentation behavior under our study. Therefore, we will briefly explain several used metrics in this article including peak signal-to-Noise Ratios (PSNR), visual information fidelity (VIF), structural similarity (SSIM), mean structural similarity (MSSIM), the PSNR human visual system masking metric (PSNR-HVS-M), and weighted signal-to-noise ratio (WSNR) since several image quality measures will be adopted in the payoff function under the game-theoretic architecture. The formulas of VIF, SSIM, MSSIM, PSNR-HVS-M, and WSNR are explained in Appendix for details.

(1) PSNR is the most commonly used quality measure for reconstruction of lossy compression codecs such as image compression, image distortion, and so on. The definition of PSNR is as following:

$$\mathrm{PSNR} = 10\log_{10}(255^2/\mathrm{MSE}) \tag{1}$$

where MSE is the mean square error between original and tested images. In general, typical values for the PSNR in lossy image are between 30 and 50dB [24] and a higher PSNR means that the tested image is less degraded and provides a higher image quality.
(2) VIF is based on local mutual information which measures how much information could flow from the reference image through the image distortion process to the human observer [22]. It uses natural scene statistics modeling in conjunction with an image-degradation model and the HVS model. The VIF measure can have values in the range [1], with VIF equal to 1 when the two compared images are identical.
(3) SSIM is a method for measuring the similarity between original and tested images [25]. Typically, it is computed from three measurement comparisons: luminance, contrast and structure with the window sizes of 8 × 8. The window can be displaced pixel-by-pixel on the image but the authors propose to use only a subgroup of the possible windows to reduce the complexity of the calculation. In practice, one usually requires a single overall quality measure of the entire image; thus, the mean SSIM index is

computed to evaluate the overall image quality. The SSIM can be viewed as a quality measure of one of the images being compared, while the other image is regarded as perfect quality. Similar to SSIM, the MSSIM [25] method is a convenient way to incorporate image details at different resolutions. The results of SSIM and MSSIM can be between 0 and 1, where 1 means excellent quality and 0 means poor quality.

(4) PSNR-HVS-M is peak signal to noise ratio taking into account of CSF and between-coefficient contrast masking of DCT basis functions [26,27]. Similar to PSNR, a higher PSNR-HVS-M value means that the tested image is less degraded.

(5) WSNR [28] is a method, which uses the CSF as the weighting function by defining WSNR as the ratio of the average weighted signal power to the average weighted noise power. As HVS is not equally sensitive to all spatial frequencies, CSF is taken into account where CSF is simulated by a low-pass or band-pass frequency filter. Similar to PSNR, a higher WSNR value means that the tested image is less degraded.

3. The proposed approach

For visible watermarking techniques, robustness and translucence are generally required; but they are unfortunately conflicted with each other. If encoder increases the energy of watermark to improve its robustness against attack, the watermarked image will be more degraded under such a scenario. Therefore, it is necessary to find a balance position in order to keep the image quality acceptable. To figure out the ideal strategies in various situations by applying visible watermarking between encoder and attacker, an example is shown in Figure 2 where the amount of watermark embedding intensity increases, the quality of watermark logo also increases as well as the robustness against attacks. On the other hand, the attacker degradation intensity is decreased simultaneously. Accordingly, an equilibrium condition exists when the ideal strategies are encountered for both sides.

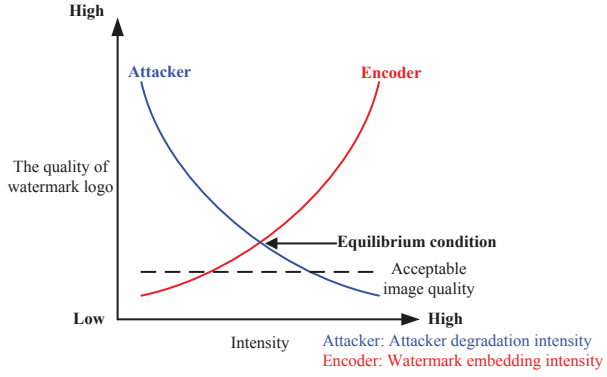

Figure 2 The illustration of equilibrium condition for the strategies between encoder and attacker.

In practice, the receiver will request the sender to send the watermarked image again if the received image quality is below an acceptable criterion. Such a condition forms a constraint for the application of visible watermarking since the image feasibility is essential to convince the receiver to take what is offered. In Figure 2, a horizontal dash line represents the acceptable image quality requirement where the equilibrium condition for both encoder and attack must above it. Otherwise, the attacked watermarked image will be rejected by the receiver. To fulfill our design methodology, we leverage the study of COCOA [12] to adaptive COCOA (ACOCOA) approach and develop a dynamic game-theoretic architecture for the watermark embedding problem which is described as a dynamic non-cooperative game with complete information [4]. The ideal strategy developed in Section. 3.2 is defined by the Nash equilibrium of the game [4]. The detailed information about ACOCOA will be explained in the following.

3.1. The ACOCOA (adaptive content and contrast aware) technique

HVS researches offer the mathematical models about how human sees the world. Psychovisual studies have shown that human vision has different sensitivity from various spatial frequencies. Tsai [12] proposed the COCOA algorithm with the consideration of HVS model by using the CSF and NVF for the best quality of perceptual translucence and noise reduction. However, the scaling factor $\alpha_{\lambda,\theta}$ and embedding factor $\beta_{\lambda,\theta}$ of COCOA algorithm are based on the CSF perceptual importance and wavelet basis function amplitudes. They both need further flexibility to fit the dynamic adjustment under game-theoretic architecture where encoder can make different decisions. Therefore, we propose an ACOCOA technique which modifies the perceptual weighting as following:

$$\alpha_{\lambda,\theta} = 1 - 0.7\beta_{\lambda,\theta} \tag{2}$$

$$\beta_{\lambda,\theta} = \begin{cases} 1 - \mathrm{NVF}_{x,y}, & \text{if } 1 - \mathrm{NVF}_{x,y} < P \times T_{\lambda,\theta}, \\ P \times T_{\lambda,\theta} & \text{otherwise} \end{cases} \tag{3}$$

$$T_{\lambda,\theta} = \begin{cases} A_{\lambda,\theta}, & \text{if } A_{\lambda,\theta} < G_{\lambda,\theta}, \\ G_{\lambda,\theta} & \text{otherwise} \end{cases} \tag{4}$$

$$G_{\lambda,\theta} = 0.01 + \frac{(7.20 - r_{\lambda,\theta})^2}{7.2^2} \tag{5}$$

Here, $T_{\lambda,\theta}$ is the perceptual weight which is determined by basis function amplitudes and CSF masking in order to avoid adding too much energy in the low frequency subbands. $r_{\lambda,\theta}$ is the perceptual weight in [18]. λ is the DWT level and θ is the orientation, and NVF is the characteristic of the local image properties. P is the watermark weighting factor in the range of [1] where a higher P value means that host image has stronger watermark embedded. Table 1 shows $A_{\lambda,\theta}$ for a 5-level 9/7 DWT from [12]. Table 2 shows $G_{\lambda,\theta}$ values after a 5-level wavelet pyramidal decomposition, which are calculated by Equation 5. Figures 3 and 4 illustrate $r_{\lambda,\theta}$ and $T_{\lambda,\theta}$ values in different DWT level and orientation, respectively.

In order to further improve the application of block classification by simply categorizing three type blocks in [18], the local and global characteristics in DWT domain is considered. In ACOCOA scheme, a stochastic image model for watermark embedding is adopted by using the NVF which characterizes the local image properties.

$$\mathrm{NVF}_{x,y} = \frac{w(x,y)}{w(x,y) + \sigma_I^2} \tag{6}$$

$w(x,y) = \gamma[\eta(\gamma)]^{\gamma} \frac{1}{\|r(i,j)\|^{2-\gamma}}$ and σ_I^2 are the global variance of the cover image I, $\eta(\gamma) = \sqrt{\Gamma(3/\gamma)/\Gamma(1/\gamma)}$, $\Gamma(t) = \int_0^{\infty} \mathrm{e}^{-u} u^{t-1} \mathrm{d}u$ (gamma function) and $r(x,y) = (I(x,y) - \bar{I}(x,y))/\sigma_I$, γ is the shape parameter, and $r(x, y)$ is determined by the local mean and the local variance. For most of real images, the shape parameter is in the range $0.3 \le \gamma \le 1$.

In our scheme, we use the stationary GG model in the embedding stage, and the estimate shape parameter for $\gamma = 0.65$, and width of window is 1. Regarding the visible watermarking algorithm, the algorithm in [12] is modified based on the consideration of the image quality where the controlling parameters of watermark embedding are selected. The watermarking procedures are briefly described as following and the flow chart is shown in Figure 5.

Table 1 Basis function amplitudes for a 5-level 9/7 DWT [12]

Orientation	Level				
	1	2	3	4	5
LL	0.62171	0.34537	0.18004	0.09140	0.045943
HL	0.67234	0.41317	0.22726	0.11792	0.059758
LH	0.67234	0.41317	0.22726	0.11792	0.059758
HH	0.72709	0.49428	0.28688	0.15214	0.077727

Table 2 CSF masking with 11 unique weights for a 5-level wavelet pyramidal decomposition

Orientation	Level				
	1	2	3	4	5
LL					0.23563
HL	0.46750	0.12674	0.07963	0.26699	0.27694
LH	0.46750	0.12674	0.07963	0.26699	0.27694
HH	0.75151	0.23960	0.01000	0.27694	0.31710

(1) The host color image is converted in the color space domain from RGB to YCrCb.
(2) By using Bi9/7 filter, compute the 5-level 2-D wavelet coefficients of Y component from host color image and grayscale watermark image.
(3) Modify the DWT coefficients of the host image by using the following equation:

$$I_{x,y}^{w} = \alpha_{\lambda,\theta} \times I_{x,y} + (\beta_{\lambda,\theta} + \mathrm{NVF}_{x,y} \times K) \times w_{x,y} \tag{7}$$

Note: (x,y) indicates the spatial location. I and w are the decomposed wavelet coefficients of the host image and the watermark image. $NVF_{x,y}$ is defined in Equation 6 and the relationship of $\alpha_{\lambda,\theta}$ and $\beta_{\lambda,\theta}$ are defined in Equations 2 and 3. The constant K value is 0.08.
(4) Inverse transform the DWT coefficients of the host image to obtain a watermarked image.

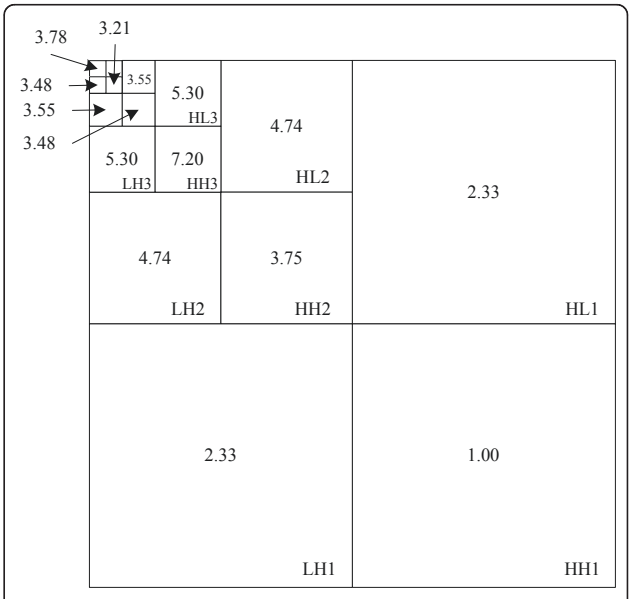

Figure 3 DWT CSF mask with 11 unique weights in different DWT level and orientation [18].

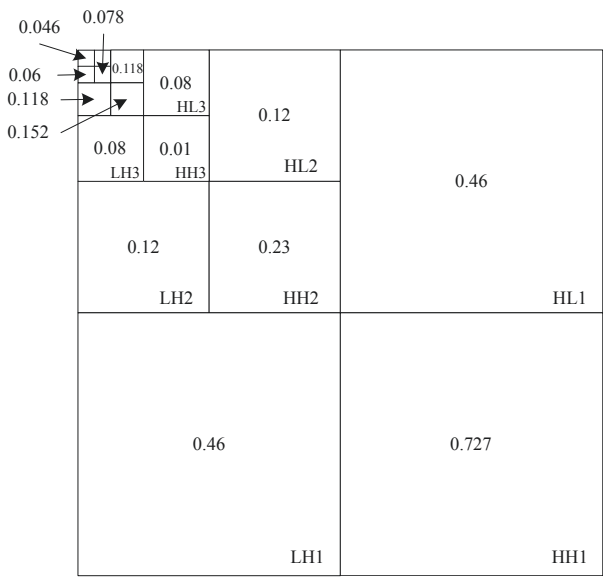

Figure 4 $T_{\lambda,\theta}$ **in different DWT level and orientation**.

3.2. A game-theoretic architecture design for visible watermarking system

Take the ACOCOA algorithm as an example and the formula from Equation 7 where $I_{x,y}$, $I^{w}_{x,y}$, and $w_{x,y}$ are the (x,y)th pixels of the host image, the watermarked image, and the visible logo image, respectively. $\alpha_{\lambda,\theta}$ in Equation 2 and $\beta_{\lambda,\theta}$ in Equation 3 are the two weighting factors that contain the adjustable parameter value of P for host image and watermark intensity. While the image quality of $I^{w}_{x,y}$ is a constraint during the watermark embedding, the selection of $\alpha_{\lambda,\theta}$ and $\beta_{\lambda,\theta}$ will be critical points since they will determine the expected image quality of $I^{w}_{x,y}$. After the watermark embedding stage, encoder will send the watermarked image to the receiver via Internet or other communication channels, while the attackers would try various ways to remove or destroy the watermark if they can intercept the transmission. Under such scenario, the robustness of the watermarking technique is essential to protect the intellectual property. Therefore, the visible watermark embedding action can be stated as a non-cooperative game where individual player decides the strategy to cope with the different situations.

We adopt the definition of Nash equilibrium in [29]. Suppose that there are N players in a game. Let X_i denote the set of possible strategies for player i. $V_i(s_1, \ldots s_N)$ denotes player i's payoff function where $s_1, \ldots s_N$ are the strategies chosen by players 1, ..., N,respectively. An Nash equilibrium is a strategy profile $\{s^*_1, \ldots, s^*_N\}$ where $s^*_1 \in X_i$ is the equilibrium strategy of player i and the function $f_i(x) = V_i(S^*_i, \ldots, S^*_{i-1}, x, S^*_{i+1}, \ldots, S^*_N)$ is optimized, for all $x \in X_i$. That is, in Nash equilibrium, a player's equilibrium strategy is the best response to the belief where the other players will also adopt their Nash equilibrium strategies.

There are two stages in Nash equilibrium. First, each player's optimal strategy is identified in response to what the other players might do. This is done for every combination of strategies by the other players. Second, Nash equilibrium is identified when all players are playing their optimal strategies simultaneously, and every player's strategy is ideal given under the other players use their equilibrium strategy. If both the set of players and set of strategies are not infinite, at least one such equilibrium exists in any time.

This study proposes a security architecture of watermarking system, which is based on the game theory and extended from Figure 1 as the generic structure for visible watermark embedding processes. A game-theoretic architecture consists of four main parts where the roles and functions are defined below:

(1) a set of players;
(2) for each player, each has a set of strategies/actions;
(3) for each player, there is existing a payoff function to evaluate the gain/profit associated with the adopted strategy/action;
(4) for each player, there are a set of constraints.

Figure 6 demonstrates the complete flow diagram of the game-theoretic architecture design for two players–encoder vs. attacker for the visible watermarking technique. The encoder and attacker player will design a payoff function to estimate the gain/profit in order to select the best strategies/actions in the watermarking game. In the mean time, the acceptable image quality is the constraint for both players. That is, the system will request to recreate a watermarked image if the image quality is below the acceptable level. The detailed description of each parts of the game-theoretic architecture for visible watermarking is as following:

(1) Players
In this case, there are two players in the game security system. One player is the encoder player and the other one is the attacker player.
(2) Strategies/actions
Due to the dynamic property during the watermark embedding stage, there are certain strategies/actions for each player to determine the best parameters based on its own interest. Let V_i and V_j denote the state of encoder and attacker players. The set of strategies for encoder player is $V_i(s_1, \ldots s_N)$ where $s_1, \ldots s_N$ are N different parameter/strategy selections for watermarking algorithm. On the other hand, we assume that

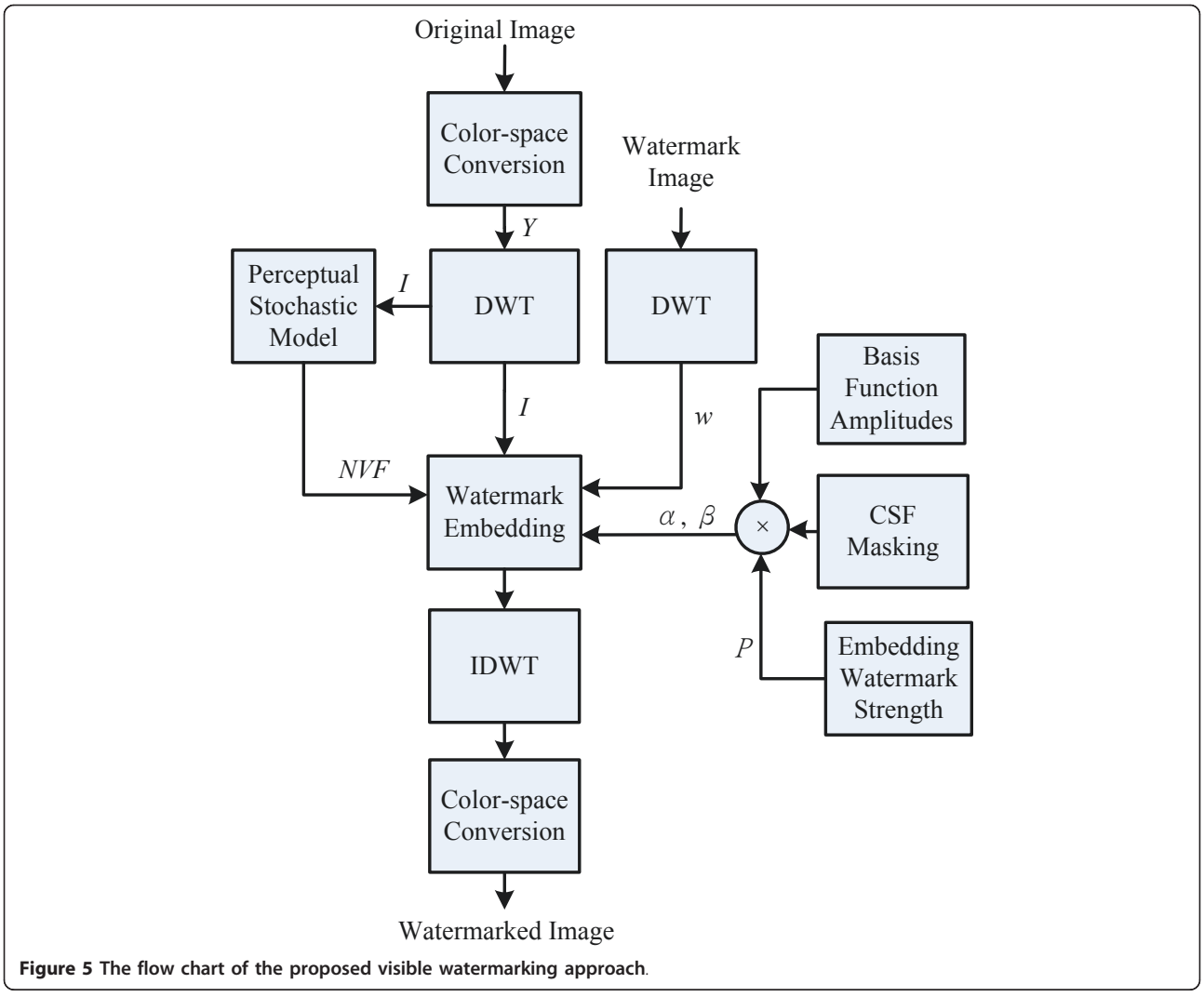

Figure 5 The flow chart of the proposed visible watermarking approach.

attacker adopts the technique to remove or destroy the watermark from the watermarked image. Here, the set of actions for attacker player is $V_j(s_1, ...s_M)$ where $s_1, ...s_M$ are equivalent to M different parameter/strategy selections for attacking algorithm.

(3) Payoffs

The payoffs represent the welfare of the players at the end of the game. They are on the basis of each player choosing his strategy and the payoff function of a player is defined as the total profit/gain. From encoder player point of view, the image quality between the host image and the watermarked image is critical since the encoder need to reserve the highest fidelity after watermark embedding. Based on the quality assessment metric study of Ponomarenko et al [23], we apply four quality assessment metrics that produce reasonably good results from [23], such as MSSIM, VIF, PSNR-HVS-M, and WSNR. In addition, the correlation between the logo watermark and the extracted watermark after attack is also important since the robustness of the watermark embedding technique is critical for the encoder player. Therefore, four image quality assessment metric and correlation functions will be adopted in the payoff function for encoder player.

The payoff function f_1 of encoder player is defined as a weighted sum of the strategy profiles e^M (quality assessment metric) where m is from 1 to 4 and e^5 (correlation). The complete formula of f_1 is shown in Equation 8

$$f_{1\,(N,M)} = W_1 \times \left(\frac{1}{4}\right) \times \sum_{m=1}^{4} \frac{e^m_{(N,M)} - \min(e^m_{(.,M)})}{\max(e^m_{(.,M)}) - \min(e^m_{(.,M)})} + W_2 \times \frac{e^5_{(N,M)} - \min(e^5_{(.,M)})}{\max(e^5_{(.,M)}) - \min(t^5_{(.,M)})} \quad (8)$$

where

$e^5_{(N,M)} = \text{correlation}((I_w - I), w)_{N,M}$,

$e^5_{(N,M)} = \text{correlation}((I_w - I), w)_{N,M}, 0 \le W_1 \le 1$, $0 \le W_2 \le 1$, and $W_1 + W_2 = 1$.

e^m represents image visual quality metric where e^1 is MSSIM, e^2 the VIF, e^3 the PSNR-HVS-M, and e^4 the WSNR. W_1 and W_2 are the weighting parameters for image quality and the robustness of watermark respectively in Equation 8.

The meaning of $e^m_{(.,M)}$ represents the payoff value of a certain M for whole set of N where N is from 1 to N_{Max}.

Note:

I is the original host image; w is the logo watermark; and I_w is watermarked image.

In order to achieve the objective of encoder player's evaluation, the payoff should get a balanced function value between the intensity of embedded watermark and the perceptual translucence for watermark. Therefore, the payoff function f_1 is defined as a normalized operation from four quality assessment metrics (MSSIM, VIF, PSNR-HVS-M, and WSNR) and correlation where the encoder's best strategy is $f_1^* = \arg\max f_{1\,(.,M)}$.

In the similar way, the same quality assessment metrics (MSSIM, VIF, PSNR-HVS-M, and WSNR) used for the payoff function of the encoder are evaluated here for the attacker player since the image quality between the watermarked image and the attacked watermarked image is decisive for the receiver. That is, the attacker expects that the receiver will not be conscious of the action of attacks. Therefore, the image quality plays an important role for the payoff function f_2 of attacker player and the formula is defined in Equation 9. Compared Equations 8 with 9, there is no correlation component in Equation 9 since the attacker does not have the original watermark logo for comparison.

$$f_{2\,(N,M)} = \left(\frac{1}{4}\right) \times \sum_{n=1}^{4} \frac{e^n_{(N,M)} - \min(e^n_{(N,.)})}{\max(e^n_{(N,.)}) - \min(e^n_{(N,.)})} \quad (9)$$

where

$$e^n_{(N,M)} = \text{quality assessment metric}(I_w, I'_w)^n_{N,M}.$$

Note: e^n represents image visual quality metric where e^1 is MSSIM, e^2 is VIF, e^3 is PSNR-HVS-M, and e^4 is WSNR.

The meaning of $e^n_{(N,.)}$ represents the payoff value of a certain N for whole set of M where M is from 1 to M_{max}.

Note: I_w is watermarked image and I'_w is the attacked watermarked image.

Accordingly, the payoff function f_2 is defined as a normalized operation from four quality assessment metrics where the attacker's best strategy is $f_2^* = \arg\min f_{2\,(N,.)}$.

(4) The constraints

From the receiver point of view, the received image must be above an acceptable image quality which is the horizontal line as shown in Figure 2. This becomes the same requirement of the watermarking game for encoder and attacker to make an acceptable watermarked image to receiver. Therefore, the encoder's payoff function should be higher than average value with no attack which can be described as $f_{1\,(N,I)} \ge 0.5$.. On the other hand, the attacker has various actions so we set a constraint μ value where μ defined in Equation 10 is the average value of attacker's payoff function in different strategies and actions.

$$\mu = \begin{cases} \frac{1}{N \times M} \times \sum_{n=1}^{N} \sum_{m=1}^{M} f_{2\,(n,m)}, & \text{if } \mu > 0.5 \\ 0.5, & \text{otherwise} \end{cases} \quad (10)$$

(5) Equilibrium condition

We adopt the concept of the Nash equilibrium and analyze the strategies/actions of the players in the watermarking system. If there has a solution profile $(f_1^*, f_2^*) = (\arg\max(f_{1\,(.,M)}), \arg\min(f_{2\,(N,.)}))$, we can say (f_1^*, f_2^*) is an equilibrium condition result of the game-theoretic architecture for visible watermarking.

4. Experimental results

The proposed ACOCOA visible watermarking algorithm and game-theoretic architecture have been implemented and intensively tested by using the commonly available color images from USC image database [30] with 512 × 512 images. The image quality metrics for the payoff function are available at the following website: MeTriX MuX Visual Quality Assessment Package [31]. The grayscale watermark of logo image adopted in the experiments is the school logo shown in Figure 1a. Different signal processing and geometric attacks have been thoroughly tested. Due to the limit of enough space to

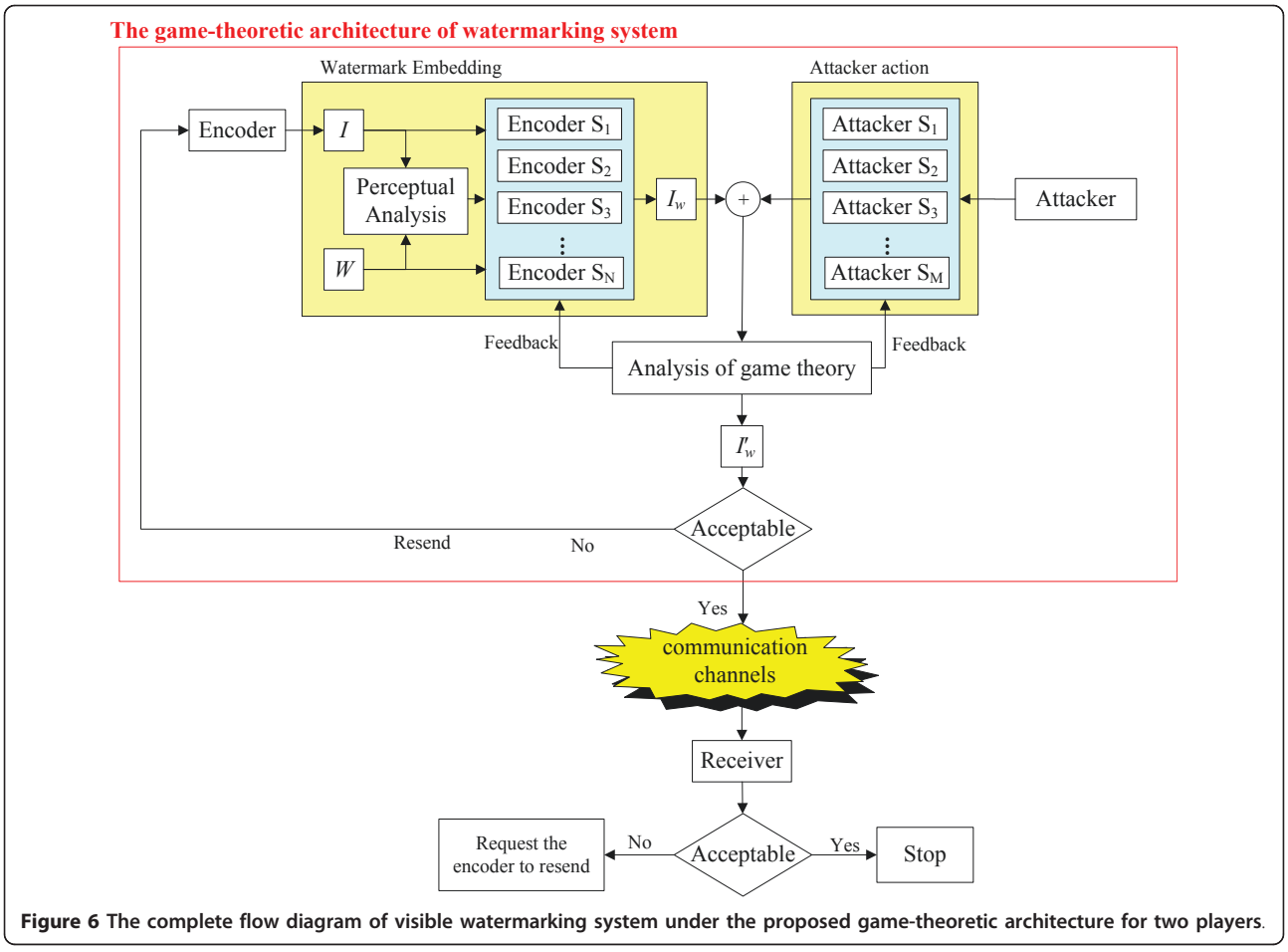

Figure 6 The complete flow diagram of visible watermarking system under the proposed game-theoretic architecture for two players.

tabulate all attacks, the experimental results show similar behavior which provides the best selection of Nash equilibrium condition under different attacks. The performance analysis can be categorized as follows.

4.1. JPEG2000 compression

Here, we tabulate all details of strategies/actions for encoder and attacker using JPEG2000 compression as attacker's action. Such procedures can be applied to any different attack. The actions for encoder player are $V_j(s_1, ... s_N)$ where $s_1, ...s_N$ are different watermark weightings of 0.0, 0.1, 0.2,..., 1.0 for $\beta_{\lambda,\theta}$. On the other hand, the actions for attacker player are $V_j(s_1, ...s_M)$ where $s_1, ...s_M$ are equivalent to compression ratio of no compression, 0.1, 0.09, ..., 0.01 for total 11 states. The meaning of compression ratio like 0.01 represents 100:1 between the uncompressed image and compressed image. Other settings from 0.1 to 0.02 are with the same operation.

It is the assumption here that the encoder knows the potential attack and it will apply the game theory to obtain the best strategy for watermark embedding. Through detailed examination, the watermark robustness plays an important role for the payoff function so we set the two weighting parameters W_2 = 0.6 and W_1 = 0.4 for Equation 8.

The performance summaries by different encoder's strategies and attacker's actions for Lena image of MSSIM, VIF, PSNR-HVS-M, WSNR, and Correlation are demonstrated in Figure 7. The results reveal that the values of the four image quality metrics and correlation are decreasing while the compression ratio is increasing. On the other hand, the correlation values are increasing while the embedded watermark is stronger for different encoder strategy. Table 3 illustrates the encoder's payoffs $f_1(N,M)$ where N and M are from 1 to 11, respectively, and the best selection for each attacker action occurs among different encoder strategy. In the mean time, the best selection characterizes the goal of the encoder for not only achieving the highest perceptual image quality but also enduring the watermark robustness against the attacker.

From the attacker's viewpoint, it is reasonable to assume that the watermarking algorithm is unknown to the attackers. Thus, we make the hypothesis that

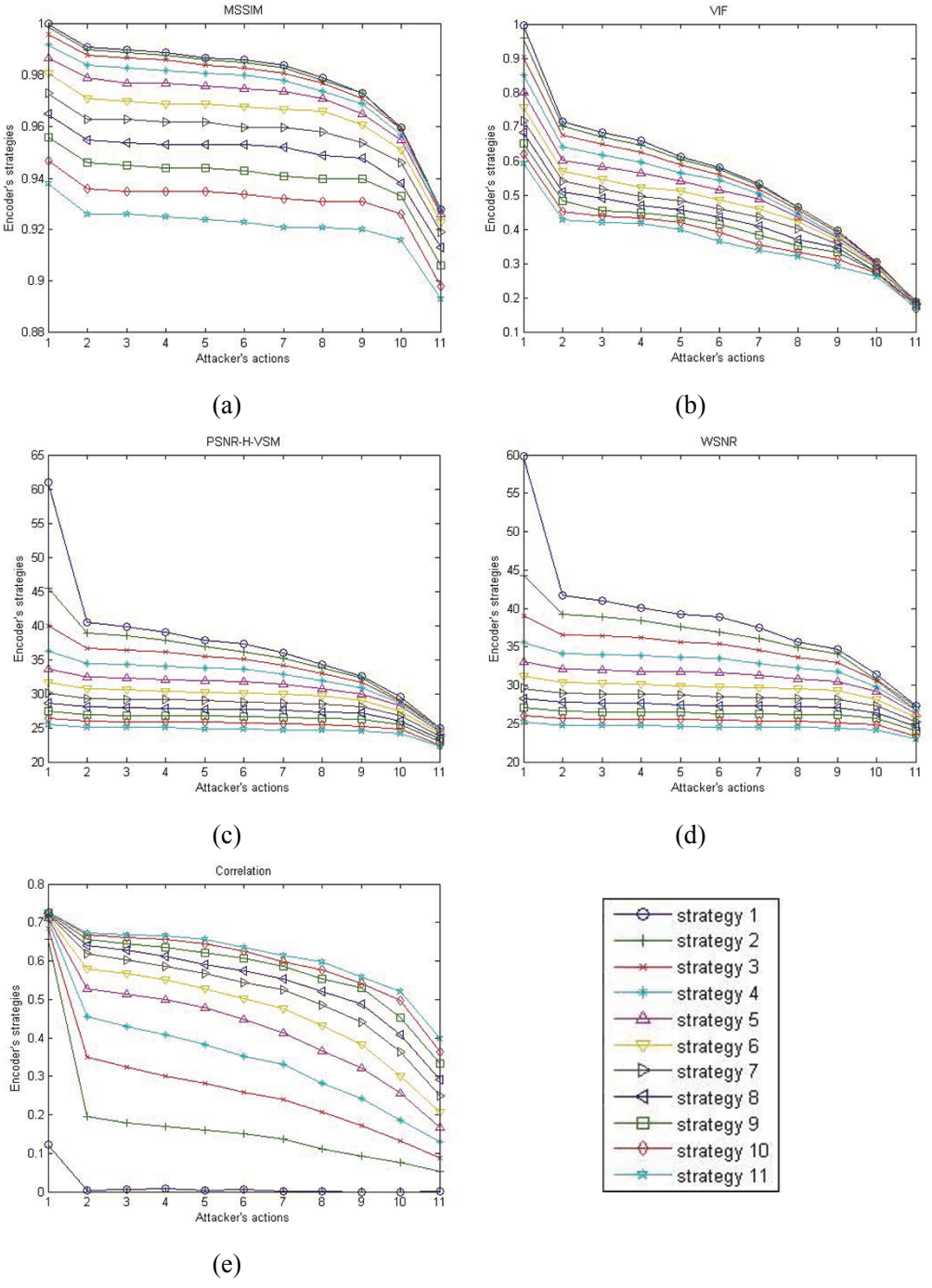

Figure 7 Performance summaries by different encoder's strategies and attacker's actions for Lena image of (a) MSSIM, (b) VIF, (c) PSNR-HVS-M, (d) WSNR, and (e) Correlation.

attacker wants to undermine the watermark but to maintain the attacked image with acceptable image quality. Table 4 illustrates the attacker's payoffs $f_2(N,M)$ where N and M are from 1 to 11, respectively, and the best selection for each encoder strategy occurs among different attacker's action.

Table 5 demonstrates the equilibrium condition from the encoder's payoffs and the attacker's payoffs under the game-theoretic system security design. With the constraint of attacked watermarked image, the equilibrium condition occurs at the state of $(N,M) = (7, 7)$ for Lena image which is equivalent to WSNR value at

Table 3 The encoder's payoffs and the best selection

Image: Lena											
	M										
	Attacker										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
N											
Encoders											
1	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.398	0.400	0.389
2	0.830*	0.541	0.529	0.523	0.521	0.513	0.509	0.499	0.491	0.455	0.473
3	0.809	0.634	0.620	0.604	0.600	0.586	0.581	0.567	0.546	0.520	0.486
4	0.789	0.682	0.667	0.655	0.647	0.638	0.635	0.609	0.586	0.555	0.567
5	0.762	0.702	0.696	0.687*	0.690*	0.682*	0.675	0.652	0.635	0.616	0.590
6	0.734	0.704*	0.697*	0.683	0.685	0.681	0.684	0.676	0.663	0.618	0.611
7	0.706	0.700	0.688	0.676	0.678	0.675	0.689*	0.681*	0.681	0.644	0.621
8	0.679	0.680	0.673	0.658	0.655	0.666	0.671	0.656	0.684*	0.623	0.634*
9	0.652	0.656	0.644	0.639	0.645	0.656	0.656	0.641	0.684	0.644	0.623
10	0.626	0.629	0.626	0.625	0.633	0.635	0.623	0.620	0.636	0.647*	0.573
11	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.612

Note: * means the best selection from encoder's payoffs.

28.10dB image quality under JPEG2000 compression with the compression ratio of 100:5 attack with f_1 value of 0.69. Similar game-theoretic design for Tiffany image is also performed and tabulated in Table 6. With the constraint of attacked watermarked image, the equilibrium condition occurs at the state of $(N,M) = (6, 8)$ for Tiffany image which is equivalent to 35.03dB image quality under 100:5 JPEG2000 compression attack with f_1 value of 0.70.

In Table 7, we tabulate the visual quality performance of Lena and Tiffany images before and after JPEG 2000 compression at compression ratio 100:3. There are three rows for both images. The 'before' row means that the image quality measure values are compared between the original image and the watermarked image. The 'after' row means the values of image quality measure are compared between the original image and the attacked watermarked image. The 'after(wm)' row means the

Table 4 The attacker's payoffs and the best selection

Image: Lena											
	M										
	Attacker										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
N											
Encoders											
1	**1.000***	0.602	0.579	0.554	0.518	0.497	0.456	0.391	0.330	0.211	0.000
2	**1.000**	**0.795**	**0.759**	**0.721**	**0.674**	**0.639***	0.586	0.504	0.430	0.251	0.000
3	**0.946**	**0.914**	**0.869**	**0.825**	**0.775**	**0.739**	**0.678***	0.582	0.489	0.305	0.000
4	**0.829**	**0.940**	**0.874**	**0.841**	**0.790**	**0.760**	**0.691***	0.586	0.490	0.303	0.000
5	**0.741**	**0.960**	**0.908**	**0.862**	**0.827**	**0.786**	**0.713***	0.608	0.510	0.332	0.000
6	**0.653**	**0.977**	**0.913**	**0.869**	**0.828**	**0.779**	**0.711***	0.611	0.522	0.329	0.000
7	0.572	**0.996**	**0.934**	**0.894**	**0.857**	**0.783**	**0.722***	0.616	0.523	0.337	0.000
8	0.487	**1.000**	**0.945**	**0.908**	**0.850**	**0.788**	**0.699***	0.586	0.520	0.317	0.000
9	0.413	**1.000**	**0.941**	**0.925**	**0.877**	**0.793**	**0.674***	0.589	0.531	0.332	0.005
10	0.374	**1.000**	**0.968**	**0.948**	**0.908**	**0.802**	**0.665***	0.604	0.504	0.363	0.015
11	0.303	**1.000**	**0.975**	**0.953**	**0.884**	**0.764**	**0.664***	0.602	0.473	0.338	0.037

Note:

(1) The constraint for attacker of acceptable image quality $\mu = 0.626$.

(2) The bold numbers mean the values are large than the constraint.

(3) * means the best selection from attacker's payoffs.

Table 5 Payoff function value for Lena image under JPEG2000 attack and the best selection of (*N*, *M*) is (7, 7) under acceptable image quality constraint

Image: Lena																						
M																						
Attacker	1		2		3		4		5		6		7		8		9		10		11	
N																						
Encoders																						
1	0.40	**1.00***	0.40	0.60	0.40	0.58	0.40	0.55	0.40	0.52	0.40	0.50	0.40	0.46	0.40	0.39	0.40	0.33	0.40	0.21	0.39	0.00
2	**0.83***	1.00	0.54	0.80	0.53	0.76	0.52	0.72	0.52	0.67	0.51	**0.64***	0.51	0.59	0.50	0.50	0.49	0.43	0.46	0.25	0.47	0.00
3	0.81	0.95	0.63	0.91	0.62	0.87	0.60	0.83	0.60	0.78	0.59	0.74	0.58	**0.68***	0.57	0.58	0.55	0.49	0.52	0.31	0.49	0.00
4	0.79	0.83	0.68	0.94	0.67	0.87	0.66	0.84	0.65	0.79	0.64	0.76	0.64	**0.69***	0.61	0.59	0.59	0.49	0.56	0.30	0.57	0.00
5	0.76	0.74	0.70	0.96	0.70	0.91	**0.69***	0.86	**0.69***	0.83	**0.68***	0.79	0.68	**0.71***	0.65	0.61	0.64	0.51	0.62	0.33	0.59	0.00
6	0.73	0.65	**0.70***	0.98	**0.70***	0.91	0.68	0.87	0.69	0.83	0.68	0.78	0.68	**0.71***	0.68	0.61	0.66	0.52	0.62	0.33	0.61	0.00
7	0.71	0.57	0.70	1.00	0.69	0.93	0.68	0.89	0.68	0.86	0.68	0.78	**0.69***	**0.72***	**0.68***	0.62	0.68	0.52	0.64	0.34	0.62	0.00
8	0.68	0.49	0.68	1.00	0.67	0.95	0.66	0.91	0.66	0.85	0.67	0.79	0.67	**0.70***	0.66	0.59	**0.68***	0.52	0.62	0.32	**0.63***	0.00
9	0.65	0.41	0.66	1.00	0.64	0.94	0.64	0.93	0.65	0.88	0.66	0.79	0.66	**0.67***	0.64	0.59	0.68	0.53	0.64	0.33	0.62	0.01
10	0.63	0.37	0.63	1.00	0.63	0.97	0.63	0.95	0.63	0.91	0.64	0.80	0.62	**0.67***	0.62	0.60	0.64	0.50	**0.65***	0.36	0.57	0.02
11	0.60	0.30	0.60	1.00	0.60	0.98	0.60	0.95	0.60	0.88	0.60	0.76	0.60	**0.66***	0.60	0.60	0.60	0.47	0.60	0.34	0.61	0.04

Note:

(1) The constraint for attacker of acceptable image quality $\mu = 0.626$.

(2) * means the best selection from encoder's or attacker's payoffs.

(3) The underlined numbers (0.69*, 0.72*) represent the best selection of Nash Equilibrium after encoder's and attacker's payoff evaluation.

Table 6 Payoff function value for Tiffany image under JPEG2000 attack and the best selection of (*N*, *M*) is (6, 8) under acceptable image quality constraint

Image: Tiffany																						
	M																					
	Attacker																					
	1		**2**		**3**		**4**		**5**		**6**		**7**		**8**		**9**		**10**		**11**	
N																						
Encoders																						
1	0.40	**1.00***	0.40	0.59	0.40	0.57	0.40	0.53	0.40	0.50	0.40	0.48	0.40	0.44	0.40	0.38	0.40	0.30	0.40	0.19	0.39	0.00
2	0.81	1.00	0.54	0.74	0.52	0.69	0.53	0.66	0.52	**0.62***	0.52	0.60	0.51	0.55	0.49	0.46	0.49	0.36	0.48	0.23	0.45	0.00
3	**0.84***	0.97	0.65	0.88	0.63	0.84	0.62	0.77	0.61	0.72	0.61	0.70	0.59	**0.65***	0.58	0.56	0.55	0.43	0.52	0.26	0.50	0.00
4	0.81	0.86	0.69	0.93	0.69	0.89	0.68	0.83	0.67	0.78	0.66	0.75	0.64	**0.68***	0.64	0.60	0.61	0.45	0.58	0.29	0.54	0.00
5	0.77	0.74	**0.71***	0.96	**0.70***	0.90	**0.69***	0.83	0.69	0.79	0.69	0.77	0.68	0.71	0.68	**0.61***	0.64	0.46	0.62	0.30	0.60	0.00
6	0.74	0.63	0.71	0.98	0.70	0.92	0.69	0.86	**0.70***	0.83	**0.69***	0.81	**0.70***	0.72	**0.70***	**0.63***	**0.67***	0.48	0.64	0.31	**0.63***	0.00
7	0.71	0.54	0.69	1.00	0.69	0.92	0.68	0.89	0.68	0.86	0.68	0.81	0.68	**0.71***	0.69	0.60	0.67	0.48	0.66	0.33	0.61	0.00
8	0.68	0.43	0.68	1.00	0.67	0.92	0.67	0.89	0.67	0.87	0.67	0.81	0.67	**0.70***	0.68	0.56	0.66	0.46	**0.67***	0.32	0.62	0.00
9	0.65	0.34	0.65	1.00	0.65	0.93	0.64	0.91	0.64	0.87	0.65	0.82	0.66	**0.71***	0.67	0.57	0.65	0.47	0.66	0.32	0.61	0.01
10	0.63	0.27	0.62	1.00	0.62	0.93	0.62	0.92	0.62	0.89	0.62	0.80	0.63	**0.67***	0.63	0.54	0.63	0.48	0.63	0.32	0.60	0.03
11	0.60	0.20	0.60	1.00	0.60	0.95	0.60	0.93	0.60	0.89	0.60	0.79	0.60	**0.65***	0.60	0.54	0.60	0.49	0.60	0.31	0.60	0.04

Note:

(1) The constraint for attacker of acceptable image quality $\mu = 0.609$.

(2) * means the best selection from encoder's or attacker's payoffs.

(3) The underlined numbers (0.70*, 0.63*) represent the best selection of Nash Equilibrium after encoder's and attacker's payoff evaluation.

Table 7 Performance summaries of watermarked color images before and after JPEG 2000 compression at compression ratio 100:3

Method	MSSIM			VIF			PSNR-HVS-M (dB)			WSNR (dB)		
	A(1)	A(2)	A(3)	B(1)	B(2)	B(3)	C(1)	C(2)	C(3)	D(1)	D(2)	D(3)
Lena												
Before	0.933	0.943	0.973	0.693	0.599	0.718	22.685	27.462	30.056	21.598	28.129	29.592
After	0.912	0.927	0.954	0.326	0.306	0.357	22.092	26.420	28.036	21.213	27.337	28.100
After (wm)	0.971	0.969	0.968	0.467	0.465	0.378	32.047	31.605	31.859	34.588	34.276	34.478
Tiffany												
Before	0.910	0.931	0.975	0.664	0.572	0.733	24.008	27.869	32.861	27.167	32.968	36.949
After	0.885	0.910	0.952	0.287	0.273	0.344	23.625	27.186	30.286	27.039	32.453	35.031
After (wm)	0.966	0.964	0.966	0.453	0.451	0.364	32.364	32.016	32.712	38.579	38.572	39.148

Note: A, B, C, and D are image quality metric of MSSIM, VIF, PSNR-HVS-M, and WSNR, respectively.
(1) is Huang and Tang's method [18].
(2) is Tsai's method [12].
(3) is the proposed ACOCOA approach.

image quality measure values are compared between the watermarked image and the compressed watermarked image (attacked image). From Table 7, the visual image quality measures of MSSIM, VIF, PSNR-HVS-M, and WSNR for ACOCOA are better than those of method [18] and [12]. To further investigate the attack effect of compression, the visual difference can be illustrated in Figure 8 and by the close-up comparison in Figure 9. We observe that the watermark patterns for Figure 8d and 8h are still with sharp edges and the logo watermark can be clearly and easily identified by human eyes for Figure 9f and 9l. Therefore, from the experimental results, we demonstrate the ACOCOA technique that is with flexibility and robustness under game-theoretic architecture. Further studies for other images are also performed and we can see similar results for visual image quality measure values and visual comparison.

4.2. Median filter

Applying the same approaches under proposed game-theoretical architecture, the attacks in StirMark [32] have been thoroughly tested and we have found that the experimental results show similar behavior, which provides the best selection of Nash equilibrium under different attacks. Due to the limited space to tabulate all attacks, we only explain median filter attack here but the scheme can be applied for other attacks.

Here the actions for encoder player are $V_i(s_1, ...,s_N)$ where $s_1, ...s_N$ are different watermark weightings of 0.0, 0.1, 0.2, ..., 1.0 for $\beta_{\lambda,\theta}$. On the other hand, the actions

Figure 8 The visual quality comparison of original and watermarked images. **(a), (e)** are original Lena and Tiffany images, respectively. **(b), (f)** are watermarked images by method [18]. **(c), (g)** are watermarked images by method [12]. **(d), (h)** are watermarked images by the ACOCOA technique.

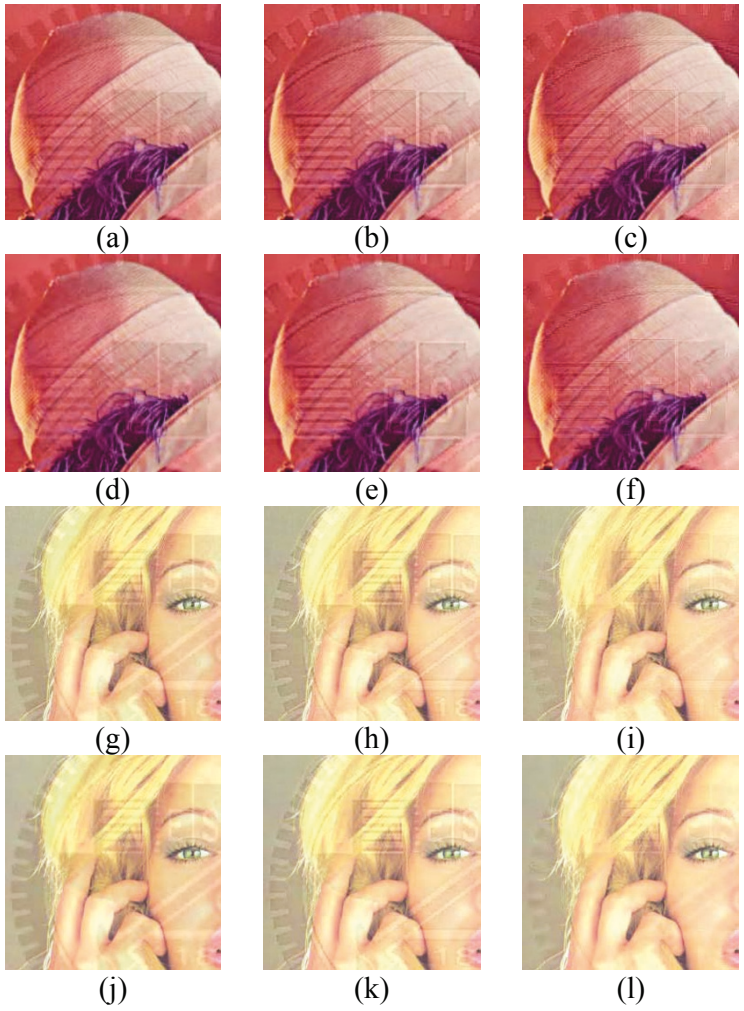

Figure 9 The visual quality comparison of close-ups for Lena and Tiffany images after JPEG 2000 compression. **(a), (g)** are watermarked images by method [18]. **(b), (h)** are watermarked images by method [12]. **(c), (i)** are watermarked images by the ACOCOA technique. **(d), (j)** are watermarked images by method [18] after JPEG 2000 compression. **(e), (k)** are watermarked images by method [12] after JPEG 2000 compression. **(f), (l)** are watermarked images by the ACOCOA technique after JPEG 2000 compression.

for attacker player are $V_i(s_1, ...,s_M)$ where s_1, ...,s_M are equivalent to filtering ratio of no filtering, 1 × 1, 3 × 3, 5 × 5, 7 × 7, 9 × 9, 11 × 11 for total seven states under game-theoretic architecture. In Table 8, it is clear that the image quality measure values using ACOCOA technique perform better than those using method [12] and [18] under median filtering. Therefore, the data support that the proposed method is with flexibility and robustness.

4.3. Image recovery and watermark removal attack

To further examine ACOCOA's robustness, we have implemented the method of watermark removal attack [33] for comparison. Figure 10 illustrates the results of the image recovery attack by method [12,18] and ACOCOA. In Figure 10, the logo pattern by method [18] is completely removed but the contours of logo pattern by method [12] and ACOCOA still exist. By using ACOCOA with game-theoretic architecture, we can easily find the best parameters for visible watermarking technique. In summary, the proposed technique can resolve the issue for watermark encoder to obtain the best watermarking strategy under attacks.

4.4. Discussions

There are several issues that the authors would like to address in this session on game-theoretic architecture for ACOCOA technique.

(1) Multiple equilibrium conditions
In theory, it is possible to have multiple equilibrium conditions under Nash equilibrium explanation. To

Table 8 Performance summaries of watermarked color images before and after 7 × 7 median filtering

Method	MSSIM			VIF			PSNR-HVS-M (dB)			WSNR (dB)		
	A(1)	A(2)	A(3)	B(1)	B(2)	B(3)	C(1)	C(2)	C(3)	D(1)	D(2)	D(3)
Lena												
Before	0.933	0.943	0.973	0.693	0.599	0.718	22.685	27.462	30.056	21.598	28.129	29.592
After	0.888	0.913	0.923	0.255	0.256	0.272	16.772	18.914	18.497	15.857	18.329	17.766
After (wm)	0.946	0.940	0.942	0.535	0.512	0.501	20.260	20.231	20.186	20.718	20.463	20.586
Tiffany												
Before	0.910	0.931	0.975	0.664	0.572	0.733	24.008	27.869	32.861	27.167	32.968	36.949
After	0.867	0.929	0.904	0.203	0.209	0.215	19.184	19.405	18.704	23.219	23.484	22.518
After (wm)	0.940	0.904	0.929	0.548	0.510	0.492	19.565	19.681	19.715	23.798	23.935	24.101

Note: A, B, C, and D are image quality metric of MSSIM, VIF, PSNR-HVS-M, and WSNR, respectively.
(1) is Huang and Tang's method [18].
(2) is Tsai's method [12].
(3) is the proposed ACOCOA approach.

clarify such assumption, experimental results of best selections for Peppers, Baboon, and Splash under JPEG2000 attack are listed in Table 9 and it is true that there exist multiple equilibrium conditions. Take Peppers as an example, (7, 7) means weighting factor is 0.6 and compression ratio is 100:5 since there are 11 different weighting actions of 0.0, 0.1, 0.2,..., 1.0 for encoder player and 11 states for compression attacks. Similarly, (6, 8) means weighting factor is 0.5 and compression ratio is 100:4. When the multiple solutions exist in this study, the security concern has higher priority than the image quality. Therefore, the selection of (7, 7) outweighs (6, 8) and the underlined selection is the final choice in Table 9. Consequently, same approach is applied for Baboon and Splash images.

(2) New types of attack actions

While the technology improves continuously, there are always new types of attacks for the visible watermarking. Under such situation, the

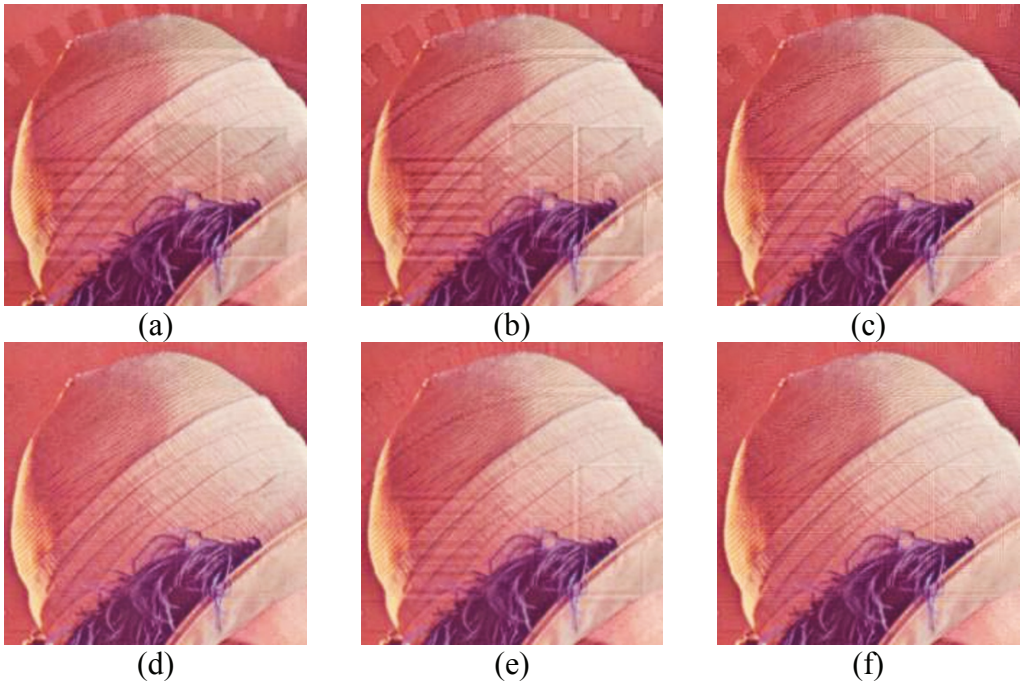

Figure 10 The visual quality comparison of close-ups between watermarked images and the watermarked images after image recovering. **(a)** is the watermarked image by method [18]. **(b)** is the watermarked image by method [12]. **(c)** is the watermarked image by ACOCOA technique. **(d)** is the watermarked image by method [18] after image recovering. **(e)** is the watermarked image by method [12] after image recovering. **(f)** is the watermarked image by the ACOCOA technique after image recovering.

Table 9 Summary of best selections under JPEG2000 attack for Nash equilibrium solution

Image	Attack
	JPEG 2000
Peppers	(7,7),(6,8)
Baboon	(8,7),(2,1)
Splash	(6,7),(5,8)

Note: The underline means the final selection under multiple equilibrium conditions.

proposed game theoretic architecture is universal and can be used ubiquitously. Therefore, the encoder can simulate the attacker's actions to get the best selection of Nash equilibrium condition.

(3) Multiple attack actions

It is possible that the watermarked image suffers multiple attacks through the communication channel and the encoder can still apply the game-theoretic architecture to simulate the multiple attacks in order to obtain the best selection among the attacks. For example, the image can be under JPEG2000 compression attack first and then median filtering attack later. Under such scenario, the joint attack will degrade the image further compared with single attack. Accordingly, the attackers should adjust the settings for multiple attacks in order to preserve the acceptable image quality. Otherwise, the image will be requested to resend if the image quality is below the threshold as shown in Figure 2.

(4) The weighting parameters

The weighting parameters W_1 and W_2 in Equation 8 are 0.4 and 0.6, respectively, in this study which are obtained empirically. Even the encoder can flexibly choose those parameters, more systematical analyses regarding the relationship between image quality and watermark robustness are suggested for future studies. For example, human objective evaluation can be collected for different parameter settings during the use of the game-theoretic security system. Therefore, the sensitivity of parameters between the original image, watermarked image and attacked image can be evaluated by the analysis of variance (ANOVA) technique in order to get the systematic influence values of the correlation coefficients.

(5) The equal weightings of quality assessment metrics for payoff functions in Equations 8 and 9.

Here, we assume that each quality assessment metric for payoff function has the equal weighting. However, such assumption is adaptable since the quality metrics may play unequal importance for Equations 8 and 9 under the game-theoretic architecture. Since there is no such discussion available during the literature survey, this topic could be further investigated as the further research.

(6) Selected best parameters for different attacks

Due to the constraints of image quality requirement as shown in Figure 2, the Nash equilibrium will be achieved under attack for best parameters selection based on the game-theoretic architecture. For example, the parameters selected under JPEG 2000 compression are still efficient for other attacks even they may not be the best selection under certain circumstances. If the decision maker wants to obtain a better watermarked image to against the specific attack, the proposed game-theoretic architecture is still the best approach to obtain the most efficient parameters under constraints.

(7) The computation time for using game-theoretic architecture for ACOCOA watermarking

The computation time for using game-theoretic architecture is determined by each player's strategies/actions and payoffs. The whole complexity should be examined by calculating each individual visual quality metric's computation.

For VIF, the fastest way of computing the determinant of a matrix is actually to use good old Gaussian elimination [34]. The determinant of a triangular matrix is simply the product of the diagonal elements. Every matrix can be reduced to a triangular matrix through elementary row operations, and all of these change the determinant in an easily predictable manner. The complexity of VIF is closely related with Equations A2 and A4 and the total amount of calculation approximately equals to the image size (we can use static array to store the results). Thus, the complexity of variance takes $O(n^2)$ computation and the natural logarithm operation also takes roughly $O(\log\ n)$. Hence, the complexity of mutual information between X and its perceptual image Y can be computed as $O(n^3 \log n)$ ($O(n \bullet n^2 \bullet \log n)) \approx O(n^2 \bullet \log n)$ for $n \times n$ image size.

Regarding the complexity of MSSIM, Equation A7 is determined by the global mean, the global variance, the local mean, the local variance, and global covariance. The complexity of global mean, global variance, and global covariance are $\approx O(n^2)$. The complexity of local mean and the local variance is $\approx O(l^2)$, $l = 2L + 1$ is the window size. In this study, the window size is 8×8. Thus, the total amount of calculation approximately equals to the image size and the overall time complexity for MSSIM is no more than $O(n^2)$ ($O(n^2 + l^2) \approx O(n^2)$) since image width n is

much larger than l). Consequently, WSNR and correlation have similar operations with the same complexity of $O(n^2)$. On the other hand, PSNR-HVS-M utilizes the weighted energy of DCT coefficients of an 8 × 8 image block. By using the fast cosine transform algorithms, the complexity of DCT could be as low as $O(n\log n)$ and the total complexity of PSNR-HVS-M will be $\approx O(n^2 \cdot n\log n) = O(n^3 \log n)$. In summary, Table 10 tabulates the complexity for each image visual quality metric adopted in this study.

From our simulation of JPEG 2000 compression, encoder player has $N = 11$ strategies and its payoff needs to calculate five different values of image quality (MSSIM, VIF, PSNR-HVS-M, WSNR, and correlation). In the mean time, attack player has $M = 11$ strategies and its payoff needs to calculate four different values of image quality (MSSIM, VIF, PSNR-HVS-M, and WSNR).

In Table 11, we tabulate the average computation time for each image visual quality metric by using a 512 × 512 testing image. The whole loop of game-theoretic architecture for ACOCOA by considering JPEG2000 attack will take about 1592.463 seconds (26.54 min) under Intel Core2 Quad CPU 2.66GHz, 2G RAM computer. The computation is performed by using Matlab software which can be further optimized by using low level language like C or C^{++} and parallel processing (cloud computing) to speed up the computation.

5. Conclusions

The researchers have been working hard to pursue the visible digital watermarking techniques for copyright protection. There are two essential characteristics: first, robustness for common signal processing operations and the second, perceptual translucence of the watermark with acceptable image quality. Since these two issues are correlated closely, how to find the best parameter settings has become a critical factor for the watermarking applications.

Table 10 The complexity of image visual quality metric for a $n \times n$ testing image

Image visual quality metric	Complexity
VIF	$\approx O(n^3 \log n)$
MSSIM	$\approx O(n^2)$
PSNR-HVS-M	$\approx O(n^3 \log n)$
WSNR	$\approx O(n^2)$
Correlation	$\approx O(n^2)$

Table 11 The average computation time for each image visual quality metric

Image visual quality metric	Computation time (s)
VIF	3.7828144
MSSIM	0.2330334
PSNR-HVS-M	2.3939420
WSNR	0.1506786
Correlation	0.0399120

In order to resolve these concerns, the ACOCOA technique and a security watermarking system, which is based on game-theoretic approach that provides the best selection for the decision maker, are proposed by studying the effect of transmission power on intensity and perceptual efficiency. The game-theoretic architecture helps us to analyze the watermarking competition game between the encoder and the attacker. In the mean time, it also provides the solution to acquire the best selection between watermark transparency and robustness for digital contents in different strategies/actions with complete information in the dynamic non-cooperative situations.

After thorough simulation and examination, the experimental results demonstrate that the proposed scheme can provide the useful information for the encoder to determine the best watermarking strategy. On the other hand, further investigations of research topics are suggested to get more precise inter-relationship among constituted components of payoff functions for the players. In summary, the proposed game-theoretic technique provides a useful decision methodology for encoder who can make the best selection among choices. Accordingly, our research could help each player to maximize its utility benefits under different situation and resolve the security issue of visual communication.

Appendix

Formulas of image quality measures

Here are the brief descriptions of the image quality measures (IQM) formulas used for payoff function in this study. Interested readers should refer the references for the detailed information.

A.1. VIF [22,35]

VIF is an image quality assessment approach based on information theory. In reference [22], Sheikh et al. defined the HVS as a typical additive noise channel. An image X is treated as a random signal and sent in at one end. The other end, the brain, receives the visual information Y, which is defined in Equation A1.

$$Y = X + V \tag{A1}$$

where V is the vision noise and obeys normal distribution with zero mean and σ_V^2.

The mutual information between X and its perceptual image Y can be computed as:

$$I(X;Y) = \log\left(1 + \frac{\sigma_X^2}{\sigma_V^2}\right) \tag{A2}$$

In contrast to the existing methods, Sheikh et al. built a relationship with the reference X and distorted image X_d with a distortion model.

$$X_d = kX + Z \tag{A3}$$

where k is a scalar and Z is the Gaussian noise with zero mean and σ_Z^2. Based on the distortion model, the common information between X and the perceptual image Y_d of the distorted image X_d can be computed as:

$$I(X;Y_d) = \log\left(1 + \frac{k^2\sigma_X^2}{\sigma_Z^2 + \sigma_V^2}\right) \tag{A4}$$

From Equation A4, it can be seen that as the scalar k decreases (blur effect) or noise Z increases (noise, compression effect, or quantization noise), $I(X;Y_d)$ is going to decrease. The visual information fidelity in each frequency band is defined as the ratio between the two mutual information $I(X;Y)$ and $I(X;Y)_d$. The overall visual information fidelity can be computed as the ratio between two mutual information in all channels.

$$\mathrm{VIF} = \frac{\sum_{j\in\text{channels}} I\left(X^j; Y_d{}^j\right)}{\sum_{j\in\text{channels}} I\left(X^j; Y^j\right)} \tag{A5}$$

A.2. MSSIM [25]

The definition of MSSIM is as following:

$$\mathrm{MSSIM}(X,Y) = \frac{1}{M}\sum_{j=1}^{M}\mathrm{SSIM}(x_j, y_j) \tag{A6}$$

where X and Y are the reference and the distorted images respectively; x_j, y_j are the image contents at the jth local window and M is the number of local windows in the image.

The SSIM metric is calculated on various windows of an image. The measure between two windows of the size $N \times N$, x and y are two nonnegative image signals. The definition of SSIM is as following:

$$\mathrm{SSIM}(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\ \left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\ \left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{A7}$$

with

μ_x the average of x; μ_y the average of y;

σ_x^2 the variance of x;σ_y^2 the variance of y;

σ_{xy} the covariance of x and y;

C_1 and C_2 are two variables to stabilize the division with weak denominator. Typically, it is calculated on window-sizes of 8 × 8.

A.3. PSNR-HVS-M [26,27]

In reference [27], authors denote a weighted energy of DCT coefficients of an image block 8 × 8 as E_w (X):

$$E_w(X) = \sum_{i=0}^{7}\sum_{j=0}^{7} X_{ij}{}^2 C_{ij} \tag{A8}$$

where X_{ij} is a DCT coefficient with indices i,j, C_{ij} is a correcting factor determined by the CSF.

The DCT coefficients X and Y are visually undistinguished if E_w $(X - Y) < \max$ $(E_w(X)/16, E_w$ $(Y)/16)$, where E_w $(X)/16$ is a masking effect E_m of DCT coefficients X (normalizing factor 16 has been selected experimentally).

Reducing of the masking effect due to an edge presence in the analyzed image block, they propose to reduce a masking effect for a block D proportionally to the local variances $V(.)$ in blocks D_1, D_2, D_3, D_4 in comparison to the entire block:

$$E_m(D) = E_w(D)\delta(D)/16$$

where

$\delta(D) = (V(D_1) + V(D_2) + V(D_3) + V(D_4))/4V(D)$, $V(D)$ is the variance of the pixel values in block D.

Below is a flowchart of PSNR-HVS-M calculation (see Figure 11).

Reduction by value of contrast masking in accordance to the proposed model is carried out in the following manner. First, the maximal masking effect E_{max} is calculated as $\max(E_m(X_e), E_m(X_d))$ where X_e and X_d are the DCT coefficients of an original image block and a distorted image block, respectively. Then, the visible difference between X_e and X_d is determined as:

$$\mathrm{PSNR\text{-}HVS\text{-}M} = X_{\Delta ij} = \begin{cases} X_{eij} - X_{dij}, & i = 0, j = 0 \\ 0, & \left|X_{eij} - X_{dij}\right| \le E_{norm}/C_{ij} \\ X_{eij} - X_{dij} - E_{norm}/C_{ij}, & X_{eij} - X_{dij} > E_{norm}/C_{ij} \\ X_{eij} - X_{dij} + E_{norm}/C_{ij}, & \text{otherwise} \end{cases} \tag{A9}$$

where E_{norm} is $\sqrt{E_{max}/64}$.

A.4. WSNR [28]

The CSF was used as a weighting function for noise measurement and the error measurement criterion is the WSNR (weighted SNR):

$$\mathrm{WSNR} = 10\log_{10}\frac{\sum_{n=1}^{N}\left[x_n * c(x_n)\right]^2}{\sum_{n=1}^{N}\left[\left((x_n) - (y_n)\right) * c(x_n)\right]^2} \tag{A10}$$

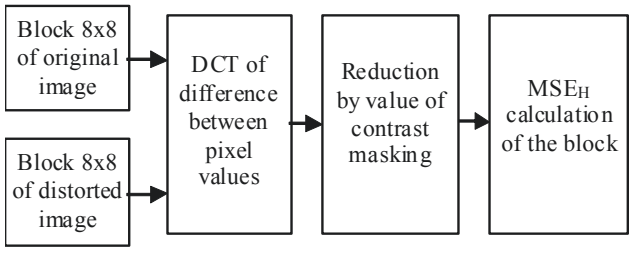

Figure 11 Flowchart of PSNR-HVS-M calculation.

where x_n and y_n denote the original image and the noisy image. * denotes linear convolution and $c(x_n)$ is CSF in the spatial domain.

List of abbreviations

ACOCOA: adaptive content and contrast aware; COCOA: content and contrast aware; CSF: contrast sensitive function; DCT: discrete cosine transform; DFT: discrete Fourier transform; DWT: discrete wavelet transform; HVS: human visual system; MSSIM: mean structural similarity; NVF: noise visibility function; PSNR: peak signal-to-noise ratios; PSNR-HVS-M: PSNR human visual system masking metric; SSIM: structural similarity; VIF: visual information fidelity; WPD: wavelet packet decomposition; WSNR: weighted signal-to-noise ratio.

Acknowledgements

This work was partially supported by the National Science Council in Taiwan, Republic of China, under Grant NSC99-2410-H-009-053-MY2.

Competing interests

The authors declare that they have no competing interests.

Received: 7 January 2011 Accepted: 30 August 2011
Published: 30 August 2011

References

1. World Intellectual Property Organization (WIPO), http://www.wipo.int/
2. IJ Cox, M Miller, J Bloom, J Fridrich, T Kalker, *Digital Watermarking and Steganography*, 2nd edn. (Morgan Kaufmann, Boston, 2007)
3. D Kundur, CY Lin, B Macq, H Yu, Special issue on enabling security technologies for digital rights management, in *Proceedings of the IEEE*, 879–882 (June 2004)
4. MJ Osborne, *An Introduction to Game Theory* (Oxford University Press, New York, 2003)
5. MJ Tsai, J Liu, A game-theoretic system security design for the visible watermarking, in *Proceedings of the ACM Multimedia 2010 Workshop, Firenze Italy*. 19–23 (October 2010)
6. GW Braudaway, KA Magerlein, F Mintzer, Protecting publicly-available images with a visible image watermark, in *Proceeding of SPIE Conference Optical Security and Counterfeit Deterrence Techniques*, **2659**, 126–133 (1996)
7. J Meng, S-F Chang, Embedding visible video watermarks in the compressed domain. in *International Conference on Image Processing (ICIP)*, **1**, 474–477 (Oct 1998)
8. MS Kankanhalli, KR Ramakrishnan, Adaptive visible watermarking of images, in *IEEE International Conference on Multimedia Computing and Systems*, **1**, 568–573 (1999)
9. SP Mohanty, KR Ramakrishnan, MS Kankanhalli, A dual watermarking technique for image, in *Proceedings of the 7th ACM International Multimedia Conference (ACMMM)*, **2**, 49–51 (Oct/Nov 1999)
10. MJ Tsai, CW Lin, Wavelet based multipurpose color image watermarking by using dual watermarks with human vision system models. IEICE Trans Fundam. **E91-A**(6), 1426–1437 (2008). doi:10.1093/ietfec/e91-a.6.1426
11. SP Mohanty, KR Ramakrishnan, MS Kankanhalli, A DCT domain visible watermarking technique for images, in *IEEE International Conference on Multimedia and Expo*. **2**, 1029–1032 (2000)
12. MJ Tsai, A visible watermarking algorithm based on the content and contrast aware (COCOA) technique. J Vis Commun Image Rep. **20**(5), 323–338 (2009). doi:10.1016/j.jvcir.2009.03.011
13. PM Chen, A visible watermarking mechanism using a statistic approach, in *Proceedings of the International Conference on Signal Processing (ICSP)*, **2**, 910–913 (Aug. 2000)
14. JL Vehel, A Manoury, Wavelet packet based digital watermarking, in *International Conference on Pattern Recognition (ICPR)*, **3**, 413–416 (2000)
15. Y Hu, S Kwong, Wavelet domain adaptive visible watermarking. Electron Lett. **37**(20), 1219–1220 (2001). doi:10.1049/el:20010838
16. Y Hu, S Kwong, An image fusion-based visible watermarking algorithm, in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*. **3**, 794–797 (May 2003)
17. L Yong, LZ Cheng, Y Wu, ZH Xu, Translucent digital watermark based on wavelets and error-correct code. Chin J Comput. **27**(11), 1533–1539 (2004)
18. BB Huang, SX Tang, A contrast-sensitive visible watermarking scheme. IEEE MultiMedia, **13**(2), 60–66 (2006). doi:10.1109/MMUL.2006.23
19. IJ Cox, J Kilian, FT Leighton, T Shamoon, Secure spread spectrum watermarking for multimedia. IEEE Trans Image Process. **6**(12), 1673–1687 (1997). doi:10.1109/83.650120
20. AS Cohen, A Lapidoth, The Gaussian watermarking game. IEEE Trans Inf Theory, **48**(6), 1639–1667 (2002). doi:10.1109/TIT.2002.1003844
21. P Moulin, JA O'sullivan, Information-theoretic analysis of information hiding. IEEE Trans Inf Theory, **49**(3), 563–593 (2003). doi:10.1109/TIT.2002.808134
22. HR Sheikh, AC Bovik, Image information and visual quality. IEEE Trans Image Process. **15**(2), 430–444 (2006)
23. N Ponomarenko, F Battisti, K Egiazarian, J Astola, V Lukin, Metrics performance comparison for color image database, in *4th International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, (Scottsdale, Arizona, USA, Jan 2009)
24. Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/PSNR
25. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error measurement to structural similarity. IEEE Trans Image Process. **13**(4), 600–612 (2004). doi:10.1109/TIP.2003.819861
26. PSNR-HVS-M page, http://ponomarenko.info/psnrhvsm.htm
27. N Ponomarenko, F Silvestri, K Egiazarian, M Carli, J Astola, V Lukin, On between-coefficient contrast masking of DCT basis functions, in *CD-ROM Proceedings of the Third International Workshop on Video Processing and Quality Metrics, USA* (2007)
28. N Damera-Venkata, TD Kite, WS Geisler, BL Evans, AC Bovik, Image quality assessment based on a degradation model. IEEE Trans Image Process. **9**(4), 636–650 (2000). doi:10.1109/83.841940
29. HS Bierman, L Fernandez, *Game Theory with Economic Application* (Addison-Wesley, New York, 1997)
30. USC SIPI - The USC-SIPI Image Database, http://sipi.usc.edu/database/
31. MeTriX MuX Visual Quality Assessment Package, http://foulard.ece.cornell.edu/gaubatz/metrix_mux/
32. StirMark, http://www.petitcolas.net/fabien/software/StirMarkBenchmark_4_0_129.zip
33. SC Pei, YC Zeng, A novel image recovery algorithm for visible watermarked image. IEEE Trans Inf Forens Security **1**, 543–550 (2006)

34. JR Bunch, JE Hopcroft, Triangular factorization and inversion by fast matrix multiplication. Math Comput, 231–236 (1974)
35. D Zhang, Information Theoretic Criteria for Image Quality Assessment Based on Natural Scene Statistics (PhD Thesis, University of Waterloo, 2009)

doi:10.1186/1687-6180-2011-48
Cite this article as: Tsai and Liu: A game-theoretic architecture for visible watermarking system of ACOCOA (adaptive content and contrast aware) technique. *EURASIP Journal on Advances in Signal Processing* 2011 **2011**:48.