

## PAPER

# Adaptive Spread-Transform Dither Modulation Using a New Perceptual Model for Color Image Watermarking\*

Lihong MA<sup>†a)</sup>, Member, Dong YU<sup>††</sup>, Gang WEI<sup>†</sup>, Jing TIAN<sup>†††</sup>, and Hanqing LU<sup>††††</sup>, Nonmembers

**SUMMARY** Major challenges of the conventional *spread-transform dither modulation* (STDM) watermarking approach are two-fold: (i) it exploits a *fixed* watermarking strength (more particularly, the quantization index step size) to the whole cover image; and (ii) it is fairly vulnerable to the amplitude changes. To tackle the above challenges, an *adaptive spread-transform dither modulation* (ASTDM) approach is proposed in this paper for conducting robust color image watermarking by incorporating a new perceptual model into the conventional STDM framework. The proposed approach exploits a new perceptual model to adjust the quantization index step sizes according to the local perceptual characteristics of a cover image. Furthermore, in contrast to the conventional Watson's model is vulnerable to the amplitude changes, our proposed new perceptual model makes the luminance masking thresholds be consistent with any amplitude change, while keeping the consistence to the properties of the human visual system. In addition, certain color artifacts could be incurred during the watermark embedding procedure, since some intensity values are perceptibly changed to label the watermark. For that, a color artifact suppression algorithm is proposed by mathematically deriving an upper bound for the intensity values according to the inherent relationship between the saturation and the intensity components. Extensive experiments are conducted using 500 images selected from Corel database to demonstrate the superior performance of the proposed ASTDM approach.

**key words:** Color image watermarking, perceptual model, spread-transform dither modulation, amplitude scaling, requantization.

## 1. Introduction

Digital watermarking of multimedia content has become a very active research area over the last decade, due to the fact that it is potential for the protection of ownership rights and policing information piracy of multimedia elements [1], [2]. With the rapid development of digital technology and the popularization of Internet, the storage and copy of the digital products have appeared as convenient and easy jobs. How to effectively protect the copyright of digital products has become a very significant issue. For that, the watermark-

ing technique has played an important role in the copyright protection, to identify the content owner, and in the authentication, to determine whether the data has been altered in any manner from its original form.

Over the last decade, a variety of watermarking approaches have been developed. They can be classified into the *spatial*-based approach and the *frequency*-based approach. The first approach embeds the watermark into the pixels of a cover image via directly changing the intensity values of the pixels, while the second one embeds the watermark into a certain band of frequency. Recently, various quantization-based watermarking approaches [3], such as *quantization index modulation* (QIM) and *spread-transform dither modulation* (STDM), have been proposed to use a structured lattice code to provide a high coefficient watermarking performance. However, this method exploits a fixed watermarking strength to the whole image, which could cause poor transparency performance in certain regions of the watermarked image. Furthermore, it is extreme sensitive to the amplitude changes, which are commonly incurred in image processing (e.g., image enhancement). In view of this, an *adaptive spread-transform dither modulation* (ASTDM) approach is proposed in this paper by incorporating a new perceptual model into the conventional STDM framework to tackle the following three challenges.

The first challenge is to overcome the drawback of the conventional STDM approach that a *fixed* watermarking strength is applied to the whole cover image. Motivated by various perceptually-optimized watermarking schemes that exploit the properties of the *human visual system* (HVS) [4], [5], a new perceptual model is proposed in this paper and is further incorporated into the conventional STDM framework. The basic idea behind is to adjust the watermarking strength (i.e., the adaptation of the quantization step sizes) according to the local characteristics (i.e., the distortion visibility threshold determined by the perceptual model) of a cover image. More watermark information could be embedded into the regions that have higher distortion visibility thresholds.

The second challenge is to improve the robustness of the conventional STDM approach to resist the amplitude changes. For that, various solutions have been developed [6]–[12]. Eggers *et al.* [6] estimated the amplitude scaling by “securely embedding SCS pilot watermark”. However, this method requires to perform a signal calibration; thus it may lead to the security weaknesses. Lee *et al.* [7] proposed to exploit the *Expectation-Maximization*

Manuscript received February 20, 2009.

Manuscript revised October 7, 2009.

<sup>†</sup>The authors are with Guangdong Key Lab of Wireless Network and Terminal, School of Electronic and Information Engineering, South China University of Technology, Guangzhou, 510641, China.

<sup>††</sup>The author is with Ericsson Mobile Data Applications Technology R&D (Guangzhou) Co., Ltd. 6F, No. 44, Jian Zhong Road, Tianhe Software Park, Tianhe District Guangzhou, 510665, China.

<sup>†††</sup>The author is with School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore.

<sup>††††</sup>The author is with National Lab of Patt. Recog. Institute Automation, China Academy of Science, Beijing, 100080, China.

\*A preliminary version of this paper was presented in the *IEEE GLOBECOM 2008*, New Orleans, CA, Dec. 2008.

a) E-mail: eelhma@scut.edu.cn

DOI: 10.1587/transinf.E93.D.843

(EM) algorithm [8] to compute a global scaling factor. However, this method yields a fairly high computational load. Furthermore, the global optimum of the EM algorithm is not guaranteed. Lagendijk and Shterev [9] presented another method based on the characteristic signal functions; however, this method requires the prior knowledge of both the host signal and the noise. Perez-Gonzalez *et al.* developed a gain-invariant adaptive quantization step size at both the watermark embedder and the watermark decoder [10]. Zhu and Tang [11] proposed a gain-invariant quantization via dividing the host signal by a statistical feature extracted from the target content. Li and Cox proposed an amplitude scaling invariant method by using a perceptual model [12]. However, this method sacrifices the transparency performance, since it changes the perceptual model's luminance masking threshold within the watermark embedding procedure; thus it is not consistent with perceptual characteristics of the HVS. In view of this, a new perceptual model is proposed which scales linearly with the amplitude while keeping consistent with the properties of the HVS.

The third challenge is how to exploit the conventional STDM approach for handling the color image data. Most of the approaches mentioned above address the gray image only, without the consideration of the color fidelity issue. In this paper, the watermark embedding is proposed to be performed on the intensity component of the color cover image in the *hue, saturation and intensity* (HSI) color space, instead of the *red, green and blue* (RGB) color space. Since the intensity values are changed to embed the watermark, certain color artifacts could be incurred in the watermarked image. To deal with that, an upper bound for the intensity values is mathematically derived to remove the above color artifacts, according to the inherent relationship between the saturation and the intensity components.

The rest of this paper is organized as follows. Section 2 provides a brief introduction to the conventional QIM and STDM watermarking approaches. A new perceptual model is proposed in Sect. 3, and then incorporated into the conventional STDM framework to develop an ASTDM approach in Sect. 4, followed by the development of the proposed color artifact suppression algorithm. Extensive experimental results are provided in Sect. 5 to demonstrate the superior performance of the proposed ASTDM approach. Finally, Sect. 6 concludes this paper.

## 2. Quantization Index Modulation and Spread-Transform Dithering Modulation

### 2.1 Quantization Index Modulation

Suppose that a *host image*  $\mathbf{X}$ , which is with a size of  $MN \times 1$  representing an  $M \times N$  image in the lexicographic-ordered vector form, is to be embedded with one bit *message* information  $m$ , which is either 1 or 0 (i.e.,  $m \in \{0, 1\}$ ). The conventional QIM method [3] quantizes the host image by applying the chosen dithered quantizers according to the message bit  $m$  and produces a *watermarked image*  $\mathbf{Y}$  via [3]

$$\mathbf{Y} = Q_{\Delta}(\mathbf{X} - \mathbf{d}_m) + \mathbf{d}_m, \quad (1)$$

where  $Q_{\Delta}(\cdot)$  is a quantizer defined as [3]

$$Q_{\Delta}(s) = \Delta \cdot \text{round}\left(\frac{s}{\Delta}\right), \quad (2)$$

in which the function  $\text{round}(\cdot)$  performs rounding a value to its nearest integer;  $\mathbf{d}_m$  is a pseudo-randomly generated dither vector with a same size as  $\mathbf{X}$ , each entry of which is determined by [3]

$$d_m = \begin{cases} -\frac{\Delta}{4}, & m = 0; \\ \frac{\Delta}{4}, & m = 1. \end{cases} \quad (3)$$

On the other hand, the watermark detector decodes a message  $\hat{m}$  according to [3]

$$\hat{m} = \arg \min_{m \in \{0,1\}} \text{dist}(\mathbf{Y}, \Lambda_m), \quad (4)$$

where  $\Lambda_m$  is the lattice set determined by the quantizers, the function  $\text{dist}(\cdot)$  represents the  $L_2$ -norm distance.

### 2.2 Spread-Transform Dithering Modulation

Instead of directly performing quantization to the host signal, *spread-transform dither modulation* (STDM) [3] first projects the host signal to a dithering vector that is derived from the original host signal, and then performs the dithering modulation on this dithering vector. Due to its vector quantization mechanism, STDM has superiority in anti-requantization attacks carried out in a specific direction. Given a host vector  $\mathbf{X}$ , the message information to be embedded  $m$ , and a projection axis  $\mathbf{p}$ , the watermarked signal is [3]

$$\mathbf{Y} = \begin{cases} \mathbf{X} + (Q_0(\mathbf{X}^T \mathbf{p}) - \mathbf{X}^T \mathbf{p}) \mathbf{p} & m = 0; \\ \mathbf{X} + (Q_1(\mathbf{X}^T \mathbf{p}) - \mathbf{X}^T \mathbf{p}) \mathbf{p} & m = 1. \end{cases} \quad (5)$$

where the superscript  $T$  denotes vector transpose. On the other hand, the watermark decoder projects the received  $\mathbf{Y}$  onto the direction  $\mathbf{p}$  and decides whether quantizer  $Q_0$  or  $Q_1$  was used [3]

$$\hat{m} = \arg \min_{m \in \{0,1\}} \text{dist}(\mathbf{Y}^T \mathbf{p}, \Lambda_m), \quad (6)$$

where  $\Lambda_m$  is the lattice set formed by the quantizers. Note that the distortion, which is incurred due to the watermark embedding, takes place in the direction  $\mathbf{p}$  only; no other component of  $\mathbf{X}$  is modified. Therefore, the watermark embedder can allocate the entire distortion budget in the direction  $\mathbf{p}$  only.

The projection axis (i.e.,  $\mathbf{p}$  in (6)) used in the conventional STDM approach could be determined by a private key [13]. Denote an  $MN \times 1$  host vector as  $\mathbf{X}$ , and the  $N \times 1$  watermark sequence  $\mathbf{m} = \{m_1, m_2, \dots, m_N\}$ , where  $m_i \in \{0, 1\}$ . Then the host vector is divided into  $N$  sub-vectors  $\mathbf{X}_i$  with a length of  $M \times 1$ , each of which is embedded with 1-bit message  $m_i$ . The  $MN \times 1$  private vector  $\mathbf{K}$  is

also divided into  $M$  sub-vectors. Now  $Q_0$  and  $Q_1$  quantize the projected host vector onto the axis  $\mathbf{K}_i$ , and compute the  $i$ -th watermarked vector  $\mathbf{Y}_i$  by

$$\mathbf{Y}_i = \mathbf{X}_i + \frac{(Q_k(\mathbf{X}_i^T \mathbf{K}_i) - \mathbf{X}_i^T \mathbf{K}_i) \mathbf{K}_i}{\|\mathbf{K}_i\|^2}, \quad (7)$$

where the function  $\|\cdot\|$  is the  $L_2$ -norm distance. When a watermark bit  $m = 1$  is embedded, the quantizer  $Q_1$  is used to quantize the projection of  $\mathbf{X}_i$  in the direction  $\mathbf{K}_i$ . The other quantizer  $Q_0$  is used while the message  $m$  is 0. Therefore the quantizer  $Q_k(\cdot)$ , where  $k = 0, 1$ , is defined as

$$Q_k(s) = \Delta \cdot \text{round}\left(s - \frac{\Delta}{d_i} + d_i\right), \quad k = 0, 1, \quad (8)$$

where the dither factor  $d_i$  is selected as

$$d_i = \pm \frac{\Delta}{4}, \quad \text{for } m_i = 0, 1. \quad (9)$$

On the other hand, the watermark detector calculates the nearest quantized point to  $\mathbf{Y}_i$  and decodes the message information as

$$\hat{m}_i = \arg \min_{m_i \in \{0,1\}} \text{dist}(\mathbf{Y}_i^T \mathbf{K}_i, \Lambda_{m_i}), \quad (10)$$

where  $\Lambda_m$  is the lattice set determined by the quantizers, the function  $\text{dist}(\cdot)$  represents the  $L_2$ -norm distance.

### 3. Proposed New Perceptual Model

The motivation of using a perceptual model to perform watermarking is to embed more information at the regions where the changes are less perceptible to the HVS [5]. Therefore, the perceptual model serves as the foundation for adaptively adjusting the quantization step sizes used in the proposed ASTDM approach. In this section, the conventional Watson's model and Li and Cox's perceptual model [12], plus our proposal of a new perceptual model will be presented.

#### 3.1 Conventional Watson's Model

The conventional Watson's model [14] exploits the properties of the HVS to determine the distortion visibility thresholds (called *slack*) for different DCT coefficients, by computing acceptable data distortion that is considered the minimum perceptible distortion (called *just noticeable distortion* (JND)). The slack is defined by three components: *frequency sensitivity*, *luminance masking* and *contrast masking* [14]; they are briefly described as follows.

- *Frequency sensitivity* (FS) reflects the smallest magnitudes of the noticeable distortion of the block DCT coefficients. A smaller value indicates that the HVS is more sensitive to the change of this frequency (i.e., the DCT coefficient). Suppose that the image is divided into disjoint  $8 \times 8$  blocks, then the block DCT transform

is performed for each block, the block DCT coefficients of the image can be obtained. Denote each coefficient of the  $k$ -th block as  $C(i, j, k)$ ,  $0 \leq i, j \leq 7$ ,  $C(0, 0, k)$  is the DC coefficient of the  $k$ -th block. The FS value of each coefficient (denoted as  $F(i, j, k)$ ) could be determined by a table defined in [14]. Note that different block DCT coefficients have the same FS values; that is,  $F(i, j, k) = F(i, j, l)$ , for any  $k$  and  $l$ .

- *Luminance masking* refers to the fact that a DCT coefficient could be changed by a larger amount before becoming perceptible, if the average intensity of its block is brighter. The luminance masking threshold (denoted as  $L(i, j, k)$ ) is defined as [14]

$$L(i, j, k) = F(i, j, k) \left( \frac{C_o(0, 0, k)}{C_{0,0}} \right)^{0.7}, \quad (11)$$

where  $F(i, j, k)$  represents its FS value,  $C_o(0, 0, k)$  is the DC coefficient of the  $k$ -th block in the original image,  $C_{0,0}$  represents the mean intensity of the whole image.

- *Contrast masking* refers to the reduction in the visibility of a change in one frequency due to the energy present in that frequency; it is defined as [14]

$$M(i, j, k) = \max\left(1, |C_o(i, j, k)|^{0.7} \times (L(i, j, k))^{-0.7}\right). \quad (12)$$

The overall distortion visibility threshold for each DCT coefficient (denoted as  $S(i, j, k)$ ) is defined as [14]

$$\begin{aligned} S(i, j, k) &= L(i, j, k)M(i, j, k) \\ &= \max\left(L(i, j, k), |C_o(i, j, k)|^{0.7} \times (L(i, j, k))^{0.3}\right), \end{aligned} \quad (13)$$

where the function  $|\cdot|$  denotes the absolute value. The above slack represents a threshold for each individual DCT coefficient that could be changed before its alternation becoming perceptible.

The key challenge of conventional Watson's model (i.e., (13)) is that it is sensitive to the amplitude scaling of the image. To be more specific, suppose the amplitude of the image is scaled by a factor  $\beta$  (i.e., an amplitude scaling attack), the resulting luminance masking threshold (denoted as  $\hat{L}(i, j, k)$ ) is

$$\begin{aligned} \hat{L}(i, j, k) &= F(i, j, k) \left( \frac{\beta C_o(0, 0, k)}{\beta C_{0,0}} \right)^{0.7} \\ &= F(i, j, k) \left( \frac{C_o(0, 0, k)}{C_{0,0}} \right)^{0.7}. \end{aligned} \quad (14)$$

Comparing (11) and (14), one can see that the luminance masking threshold under the attack (i.e.,  $\hat{L}(i, j, k)$ ) remains the same as the original threshold (i.e.,  $L(i, j, k)$ ); that is

$$\hat{L}(i, j, k) = L(i, j, k) \neq \beta \times L(i, j, k). \quad (15)$$

On the other hand, the resulting slack (denoted as  $\hat{S}(i, j, k)$ ) under the above amplitude scaling attack is

$$\hat{S}(i, j, k)$$

$$= \max\left(\widehat{L}(i, j, k), |\beta C_0(i, j, k)|^{0.7} \times \left(\widehat{L}(i, j, k)\right)^{0.3}\right). \quad (16)$$

Substituting (15) into (16) we have

$$\begin{aligned} \widehat{S}(i, j, k) &= \max\left(L(i, j, k), |\beta C_0(i, j, k)|^{0.7} \times (L(i, j, k))^{0.3}\right) \\ &= \max\left(L(i, j, k), \beta^{0.7} |C_0(i, j, k)|^{0.7} \times (L(i, j, k))^{0.3}\right). \end{aligned} \quad (17)$$

Comparing (13) and (17), one can see that

$$\widehat{S}(i, j, k) \neq \beta \times S(i, j, k). \quad (18)$$

Therefore the conventional Watson's model can not resist the amplitude scaling attack, since it remains unchanged even if the image's amplitude is scaled.

### 3.2 Li and Cox's Perceptual Model

To overcome the above drawbacks of the conventional Watson's model, Li and Cox [12] proposed a new luminance-masked threshold model by multiplying (11) with the mean DC components of the image; that is,

$$L^M(i, j, k) = F(i, j, k) \left(\frac{C_O(0, 0, k)}{C_{0,0}}\right)^{0.7} \left(\frac{C_{0,0}}{128}\right). \quad (19)$$

Furthermore, the new slack function (denoted as  $S^M(i, j, k)$ ) is

$$\begin{aligned} S^M(i, j, k) &= \max\left(L^M(i, j, k), |C_0(i, j, k)|^{0.7} \times (L^M(i, j, k))^{0.3}\right). \end{aligned} \quad (20)$$

Suppose the amplitude of the image is scaled with a factor  $\beta$ , then the luminance masking threshold under the amplitude scaling attack with a scale  $\beta$  is

$$\begin{aligned} \widehat{L}^M(i, j, k) &= F(i, j, k) \left(\frac{\beta C_O(0, 0, k)}{\beta C_{0,0}}\right)^{0.7} \left(\frac{\beta C_{0,0}}{128}\right) \\ &= \beta F(i, j, k) \left(\frac{C_O(0, 0, k)}{C_{0,0}}\right)^{0.7} \left(\frac{C_{0,0}}{128}\right). \end{aligned} \quad (21)$$

Comparing (21) and (19), one can see that the luminance-masked threshold under the attack (i.e.,  $\widehat{L}^M(i, j, k)$ ) remains as same as the original threshold (i.e.,  $L^M(i, j, k)$ ); that is

$$\widehat{L}^M(i, j, k) = \beta \times L^M(i, j, k). \quad (22)$$

Therefore, Li and Cox's model [12] can resist the amplitude scaling attack, since it remains unchanged even if the amplitude is scaled.

However, Li and Cox's model [12] changes the slack values so that it is not consistent with the properties of HVS; consequently, it could result in inferior transparency performance. More specifically, the slack value defined in (20) is different with that of the conventional Watson's model (consistent with the properties of the HVS) within the watermark embedding procedure; that is, Li and Cox's model [12] is *not* consistent with the properties of the HVS.

### 3.3 Proposed New Perceptual Model

To tackle the above challenges of both the conventional Watson's model and Li and Cox's model [12], our idea is to make the slack scale linearly with the amplitude, and keep the slack consistent with the properties of the HVS. For that, a new perceptual model is proposed in this paper by modifying (11) to be a new luminance masking thresholding function (denoted as  $L^P(i, j, k)$ )

$$L^P(i, j, k) = F(i, j, k) \left(\frac{C_O(0, 0, k)}{C_{0,0}}\right)^{0.7} \frac{C_{0,0}}{C_{0,0}^O}, \quad (23)$$

where  $C_{0,0}$  and  $C_{0,0}^O$  denote the mean DC coefficient of the watermarked and the cover image, respectively. Furthermore, the new slack function (denoted as  $S^P(i, j, k)$ ) is

$$\begin{aligned} S^P(i, j, k) &= \max\left(L^P(i, j, k), |C_0(i, j, k)|^{0.7} \times (L^P(i, j, k))^{0.3}\right). \end{aligned} \quad (24)$$

The proposed new perceptual model has the following two advantages, compared with the conventional Watson's model.

First, the proposed perceptual model can resist the amplitude scaling attack, since it will be scaled accordingly if the image's amplitude is scaled. Suppose the amplitude of the image is scaled with a factor  $\beta$ , then the luminance masking threshold under the amplitude scaling attack with a scale  $\beta$  is

$$\begin{aligned} \widehat{L}^P(i, j, k) &= F(i, j, k) \left(\frac{\beta C_O(0, 0, k)}{\beta C_{0,0}}\right)^{0.7} \left(\frac{\beta C_{0,0}}{C_{0,0}^O}\right) \\ &= F(i, j, k) \left(\frac{C_O(0, 0, k)}{C_{0,0}}\right)^{0.7} \left(\frac{\beta C_{0,0}}{C_{0,0}^O}\right). \end{aligned} \quad (25)$$

Comparing (23) and (25), one can see that the luminance masking threshold under the attack (i.e.,  $\widehat{L}^P(i, j, k)$ ) will scale according to the original threshold (i.e.,  $L^P(i, j, k)$ ); that is

$$\widehat{L}^P(i, j, k) = \beta \times L^P(i, j, k). \quad (26)$$

On the other hand, the resulting slack (denoted as  $\widehat{S}^P(i, j, k)$ ) under the above amplitude scaling attack is

$$\begin{aligned} \widehat{S}^P(i, j, k) &= \max\left(\widehat{L}^P(i, j, k), |\beta C_0(i, j, k)|^{0.7} (\widehat{L}^P(i, j, k))^{0.3}\right) \\ &= \max\left(\beta L^P(i, j, k), |\beta C_0(i, j, k)|^{0.7} (\beta L^P(i, j, k))^{0.3}\right) \\ &= \max\left(\beta L^P(i, j, k), \right. \\ &\quad \left. \beta^{0.7} |C_0(i, j, k)|^{0.7} \beta^{0.3} (L^P(i, j, k))^{0.3}\right) \\ &= \max\left(\beta L^P(i, j, k), \beta |C_0(i, j, k)|^{0.7} (L^P(i, j, k))^{0.3}\right) \end{aligned}$$

$$= \beta \max \left( L^P(i, j, k), |C_0(i, j, k)|^{0.7} \left( L^P(i, j, k) \right)^{0.3} \right). \quad (27)$$

Comparing (24) and (27), one can see that

$$\widehat{S}^P(i, j, k) = \beta \times S^P(i, j, k). \quad (28)$$

Therefore the proposed perceptual model can resist the amplitude scaling attack, since both the luminance masking threshold (see (26)) and the slack threshold (see (28)) will be changed according to the amplitude changes of the image.

The second advantage of the proposed perceptual model is its consistence to the properties of the HVS within the watermark embedding procedure. To be more specific, the slack value of the proposed model remains the same as that of the conventional Watson's model, since  $C_{0,0} = C_{0,0}^O$  within the watermark embedding procedure; that is, they are consistent with the properties of the HVS.

It is worth pointing out that the fundamental difference between our proposed model and Li and Cox's model [12] mentioned in the previous section, where a new luminance masking threshold model is proposed by multiplying (11) with the mean DC components of the image. Li and Cox' model increases the slack values of the original Watson's model, which is precisely calculated and should not be changed easily [14]. This change may result in the inconsistency with the properties of HVS, especially when the mean DC value is high (the deviation increases linear with mean DC value). Consequently, it could result in inferior transparency performance in the watermarked image. To justify this, extensive experiments are conducted in Sect. 5 to demonstrate that our proposed model outperforms Li and Cox's model [12].

#### 4. Proposed Adaptive Spread-Transform Dither Modulation Approach

In this section, an ASTDM approach is proposed by incorporating our new perceptual model into the conventional STDM framework to adaptively adjust the quantization step sizes. Consequently, the proposed ASTDM approach can embed more information at the regions where the intensity changes are less perceptible. This is in contrast to that the conventional STDM approach applies a fixed quantization step size for the whole cover image.

##### 4.1 Adaptive Quantization Step Size Selection

Denote a  $MN \times 1$  host vector as  $\mathbf{X}$ , and the  $N \times 1$  watermark sequence  $\mathbf{m} = \{m_1, m_2, \dots, m_N\}$ , where  $m_i \in \{0, 1\}$ . The objective is to divide the host vector into  $N$  sub-vectors, each of which is denoted as  $\mathbf{X}_i$  with a length of  $M \times 1$ , followed by embedding each sub-vector  $\mathbf{X}_i$  with 1-bit message  $m_i$ . In addition, the  $MN \times 1$  private vector  $\mathbf{K}$  is also divided into  $N$  sub-vectors. According to (7), for  $i = 1, 2, \dots, N; k = 0, 1$ , the watermarked vector  $\mathbf{Y}_i$  is

$$\mathbf{Y}_i = \mathbf{X}_i + \frac{\left( Q_k(\mathbf{X}_i^T \mathbf{K}_i) - \mathbf{X}_i^T \mathbf{K}_i \right) \mathbf{K}_i}{\|\mathbf{K}_i\|^2}, \quad (29)$$

where the function  $\|\cdot\|$  is  $L_2$ -norm,  $Q_k(\cdot)$  is the  $k$ -th quantizer. After embedding the message  $m_i$  into the host vector  $\mathbf{X}_i$ , the error introduced by the quantization should not exceed the slack  $\mathbf{S}_i^P$ ; otherwise, the embedded watermark will become perceptible. More specifically, the difference between each component of the host vector (denoted as  $\mathbf{X}_i(j)$ ) and that of the watermarked vector (denoted as  $\mathbf{Y}_i(j)$ ) should be smaller than its corresponding component of the slack (denoted as  $\mathbf{S}_i^P(j)$ ); that is

$$|\mathbf{Y}_i(j) - \mathbf{X}_i(j)| \leq \mathbf{S}_i^P(j). \quad (30)$$

The objective is to derive the quantization step  $\Delta_i$  used for the host vector  $\mathbf{X}_i$  according to (30). First, according to (29), the left term of (30) is

$$\begin{aligned} |\mathbf{Y}_i(j) - \mathbf{X}_i(j)| &= \left| \mathbf{X}_i(j) + \frac{\left( Q_k(\mathbf{X}_i^T \mathbf{K}_i) - \mathbf{X}_i^T \mathbf{K}_i \right) \mathbf{K}_i(j)}{\|\mathbf{K}_i\|^2} - \mathbf{X}_i(j) \right| \\ &= \left| \frac{\left( Q_k(\mathbf{X}_i^T \mathbf{K}_i) - \mathbf{X}_i^T \mathbf{K}_i \right) \mathbf{K}_i(j)}{\|\mathbf{K}_i\|^2} \right|. \end{aligned} \quad (31)$$

Substituting (31) into (30), we get

$$\left| \frac{\left( Q_k(\mathbf{X}_i^T \mathbf{K}_i) - \mathbf{X}_i^T \mathbf{K}_i \right) \mathbf{K}_i(j)}{\|\mathbf{K}_i\|^2} \right| \leq \mathbf{S}_i^P(j). \quad (32)$$

Next, since the maximum quantization error of the quantizer for each  $\mathbf{Y}_i$  is  $\frac{\Delta_i}{2}$  [3], substituting  $\left| Q_k(\mathbf{X}_i^T \mathbf{K}_i) - \mathbf{X}_i^T \mathbf{K}_i \right| = \frac{\Delta_i}{2}$  into (32), we have

$$\left| \frac{\frac{\Delta_i}{2} \mathbf{K}_i(j)}{\|\mathbf{K}_i\|^2} \right| \leq \mathbf{S}_i^P(j). \quad (33)$$

Then, (33) can be rewritten as

$$\Delta_i \leq \frac{2|\mathbf{S}_i^P(j)|\|\mathbf{K}_i\|^2}{|\mathbf{K}_i(j)|}. \quad (34)$$

Finally, denote a modification factor  $\lambda \in (0, 1)$ ,  $\overline{|\mathbf{K}_i|}$  and  $|\overline{\mathbf{S}_i^P}|$  are the average absolute values of  $\mathbf{K}_i$  and  $\mathbf{S}_i^P$ , respectively. To satisfy (34), the quantization step size for the host vector  $\mathbf{Y}_i$  yields the form as

$$\Delta_i = \frac{2\lambda|\overline{\mathbf{S}_i^P}|\|\mathbf{K}_i\|^2}{\overline{|\mathbf{K}_i|}}. \quad (35)$$

Owing to (35), the quantization step size  $\Delta_i$ , which is performed to the host vector  $\mathbf{X}_i$ , is adaptively adjusted according to the JND slack ( $\mathbf{S}_i^P$ ), which could be flexibly controlled by the proposed new perceptual model, and the modification factor  $\lambda$ , which is determined by the users in practice.

## 4.2 Representation, Embedding and Extraction of $C_{0,0}^O$

Referring to Sect. 3, the main contribution of our proposed new perceptual model is to incorporate the mean intensity value of the cover image  $C_{0,0}^O$  (see (23)) into the luminance masking function. Therefore, there are certain significant issues needed to be investigated.

The first issue is the representation of the mean intensity information  $C_{0,0}^O$ . For that, an 11-bit binary representation is used in this paper. More specifically, the first 7 bits are copied from that of the binary representation of  $C_{0,0}^O$ ; the next 3 bits represent the summation of the “1” bits among the above-mentioned 7 bits (the maximum possibility is that there are seven 1s, while the minimum possibility is that there is no 1; therefore, three bits are sufficient); and the last 1 bit represents the parity check bit.

The second issue is how to embed the mean intensity information  $C_{0,0}^O$  into the cover image. For that, the above 11-bit representation of  $C_{0,0}^O$  is embedded into the cover image, after the embedding of the watermark. Furthermore,  $C_{0,0}^O$  could be embedded for more than one time, as long as there is any host signal available, so as to resist any transmission error. Suppose there is a  $512 \times 512$  cover image and a  $60 \times 60$  watermark image, since only a single message information will be embedded into each  $8 \times 8$  block DCT coefficients, the available space to embed the mean intensity information  $C_{0,0}^O$  is  $(\frac{512 \times 512}{8 \times 8} - 60 \times 60) / (11) \approx 45$ . That is, the mean intensity information  $C_{0,0}^O$  can be embedded 45 times in the cover image.

The third issue is how to extract  $C_{0,0}^O$  from the received watermarked image. Suppose the mean intensity information has been embedded for  $N$  (e.g.,  $N = 45$  in the above example) times during the watermark embedding procedure, then the watermark detector should extract  $N$  mean intensity values, each of which is denoted as  $x_i$ , where  $i = 1 \sim N$ . First, the majority voting rule is applied to determine the value of each bit of the 11-bit representation of the extracted mean intensity information (denoted as  $\widehat{C}_{0,0}^O(j)$ ), according to that of the above  $N$  extracted mean intensity values  $x_i(j)$ . That is, for each  $j$ ,  $1 \leq j \leq 11$ , we have

$$\widehat{C}_{0,0}^O(j) = \begin{cases} 1, & \sum_{i=1}^N x_i(j) \geq \frac{N}{2}; \\ 0, & \sum_{i=1}^N x_i(j) < \frac{N}{2}. \end{cases} \quad (36)$$

Next, a parity check is performed to determine whether the summation of first 7 bits of  $\widehat{C}_{0,0}^O$  equals to its next 3 bits of  $\widehat{C}_{0,0}^O$ . If it is correct, then, the extracted mean intensity value  $\widehat{C}_{0,0}^O$  is used to compute the slack for detecting the watermark; otherwise, the extracted  $\widehat{C}_{0,0}^O$  is neglected, and the conventional Watson’s model is used to calculate the slacks for detecting the watermark.

## 4.3 Proposed Artifacts Suppression Algorithm

To perform color watermarking, the proposed ASTDM ap-

proach embeds the watermark into the *intensity* component of the cover image, rather than perform watermark embedding in the red, green or blue component directly. This is motivated by the fact that the correlations among the hue, saturation and intensity components are much less than that among the red, green and blue components [18]. However, a slightly modification to the intensity component may cause great color distortion in the watermarked image. For that, an artifacts suppression algorithm is proposed as a post-processing procedure to correct the possible color distortion caused by watermark embedding.

The idea is to correct the intensity values, which have been changed due to the watermark embedding operation, according to the inherent constraint among the hue, saturation and intensity components. This is motivated by that both the hue and saturation components remain unchanged during the watermark embedding procedure; thus, they are considered as reliable.

Denote the normalized red, green and blue components of the pixel (at the position  $(i, j)$ ) are  $R(i, j)$ ,  $G(i, j)$  and  $B(i, j)$ , respectively. Further denote its corresponding hue, saturation and intensity values are  $H(i, j)$ ,  $S(i, j)$  and  $I(i, j)$ , respectively. The upper bound of the intensity value  $I(i, j)$  (denoted as  $I^U(i, j)$ ) is

$$I^U(i, j) = \begin{cases} \frac{2}{S(i,j)+2}, & \text{if } \frac{2}{3} < I(i, j) \leq 1; \\ 1, & \text{otherwise.} \end{cases} \quad (37)$$

The detailed derivation of (37) is presented in Appendix. Any intensity value, which exceeds the above upper bound should be corrected. Furthermore, even after the correction of the intensity values, the transformation from the HSI color space to the RGB space could also cause the resulting RGB values exceeding the bounds. Therefore, a normalization operation is performed to clip those outstanding RGB values as follows. If any value of the pixel’s red, green or blue components (e.g., the red component at the position  $(i, j)$ , that is  $R(i, j)$ ), is larger than 255, then it is set to be 255; followed by shrinking its corresponding green and blue components’ values as  $\hat{G}(i, j) = \frac{255}{R(i,j)}G(i, j)$  and  $\hat{B}(i, j) = \frac{255}{R(i,j)}B(i, j)$ , respectively.

The computational complexity of the proposed artifact suppression method is discussed as follows in terms of the numbers of comparisons, multiplications and additions. Suppose that the image has a size of  $n \times n$ , of which  $K$  pixels exceed the upper bounds (i.e., (37)) of intensity values, then  $O(n^2)$  comparison operations need to be performed for comparing the intensity of each pixel with its respective upper bound. Furthermore,  $O(K)$  multiplications and  $O(K)$  additions will be applied on  $K$  pixels that exceed the upper intensity bound. In addition, if any of these  $K$  pixels has outstanding RGB values, three extra multiplications are needed to correct its RGB values (i.e., one multiplication for each value of the RGB components).

#### 4.4 Summary of the Proposed ASTDM Watermarking Approach

Figure 1 illustrates an overview of the proposed ASTDM approach, which has two modules: the *embedding module* and the *extracting module*. Both of these two modules are described in detail as below:

- The *embedding module*.
  - Convert the cover image into the HSI color space to obtain its intensity component for watermark embedding.
  - Divide the intensity component into disjoint  $8 \times 8$  blocks, followed by performing the block DCT. Next, the proposed new perceptual model is exploited to compute the slacks via (24), and the DC coefficients are used to compute the mean intensity information that will be embedded into the cover image.
  - Use the ASTDM embedder to embed the watermark, together with the mean intensity information, into the DCT coefficients of the intensity component of the cover image. Next, apply the inverse DCT to the watermarked DCT coefficients.
  - Apply the proposed color artifact suppression algorithm to remove any artifacts incurred due to the watermark embedding operation and then convert the image back to the original format.
- The watermarked image may receive certain attacks during the transmission before reaching the receiver.
- The *extracting module*.
  - Convert the received watermarked image into the HSI color space to obtain its intensity component for watermark extracting.
  - Divide the intensity component into disjoint  $8 \times 8$  blocks, followed by performing block DCT. Next, extract the mean intensity information, and the proposed perceptual model is exploited to compute the slacks via (24).
  - Use the ASTDM detector to extract the watermark

from the DCT coefficients of the intensity component of the received watermarked image.

### 5. Experimental Results

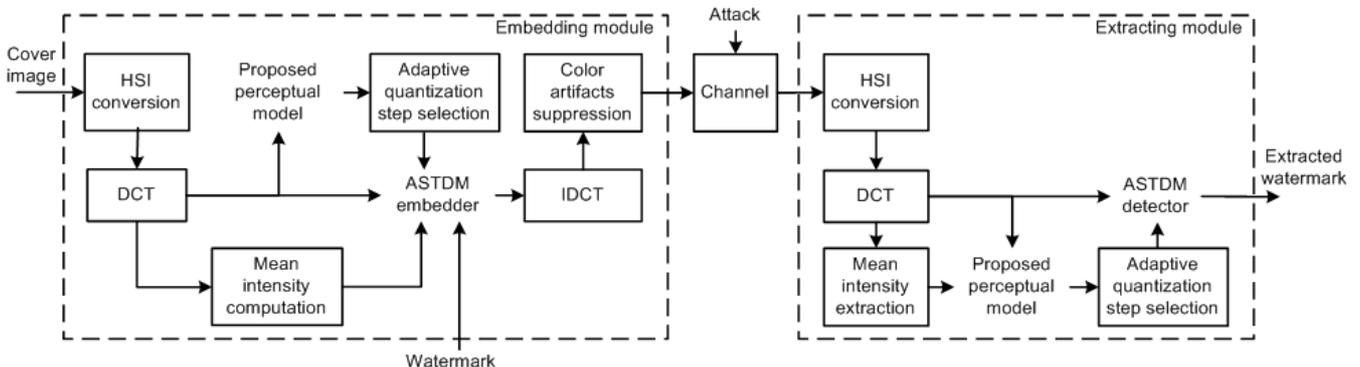
Experiments are conducted to compare the proposed ASTDM approach with the conventional STDM approach [3], plus Li and Cox’s approach [12]. 500 images that are selected from Corel database [17] (including Buildings, Flowers, Elephants, Food, Mountains and Portraits images) are used as the cover images (as shown in Fig. 2), and a binary image *Runner* with a size of  $60 \times 60$  is used as the watermark to be embedded into the cover image. Each watermark algorithm performs an  $8 \times 8$  block DCT on the cover image, then selects 20 DCT coefficients (from the 2-nd to the 21-st zig-zag-scanned DCT coefficients) from each block to form a host vector, followed by embedding one bit watermark into this host vector. The embedding positions are crucial to the transparency performance of the watermarking algorithm, the above selection (i.e., from the 2-nd to the 21-st zig-zag-scanned DCT coefficients) is experimentally determined in this paper and has been proved to yield fairly good transparency performance. The private key is generated from a Gaussian distribution with zero mean and variance of 16.

#### 5.1 Transparency Performance

Two image quality measures—PSNR and *Structural Similarity* (SSIM) [19] are used to evaluate the transparency performance of various watermarking approaches. The SSIM measurement is fairly consistent with the perceptual assessment. Note that a larger SSIM value indicates a better im-



**Fig. 2** Selected images from Corel database used as cover images in the experiments (This is a color figure): From the left to the right: *Baboon*, *Lena*, *Pills* and *Watch*.



**Fig. 1** Overview of the proposed ASTDM approach.

age quality. To demonstrate how the proposed ASTDM approach can adaptively adjust the embedding strength according to our proposed perceptual model, we let our approach produce a watermarked image that yields a same (or nearly same) SSIM performance as that of the conventional STDM approach, plus Li and Cox's approach [12]. More specifically, the modification factor  $\lambda$  (see (34) in the proposed ASTDM approach) and the quantization step size  $\Delta$  (see (8) in the conventional STDM approach) are carefully adjusted to make their resulting watermarked image approximately yield equal SSIM performance as 0.982.

The first experiment compares the averaged quantization steps between the conventional STDM approach and the proposed ASTDM approach, as presented in Table 1. The proposed approach yields larger quantization step sizes than that of the conventional STDM approach; consequently, the proposed approach yields a higher robustness performance

against certain attacks. This is due to the fact that a larger quantization step size indicates a higher robustness performance against certain attacks [3].

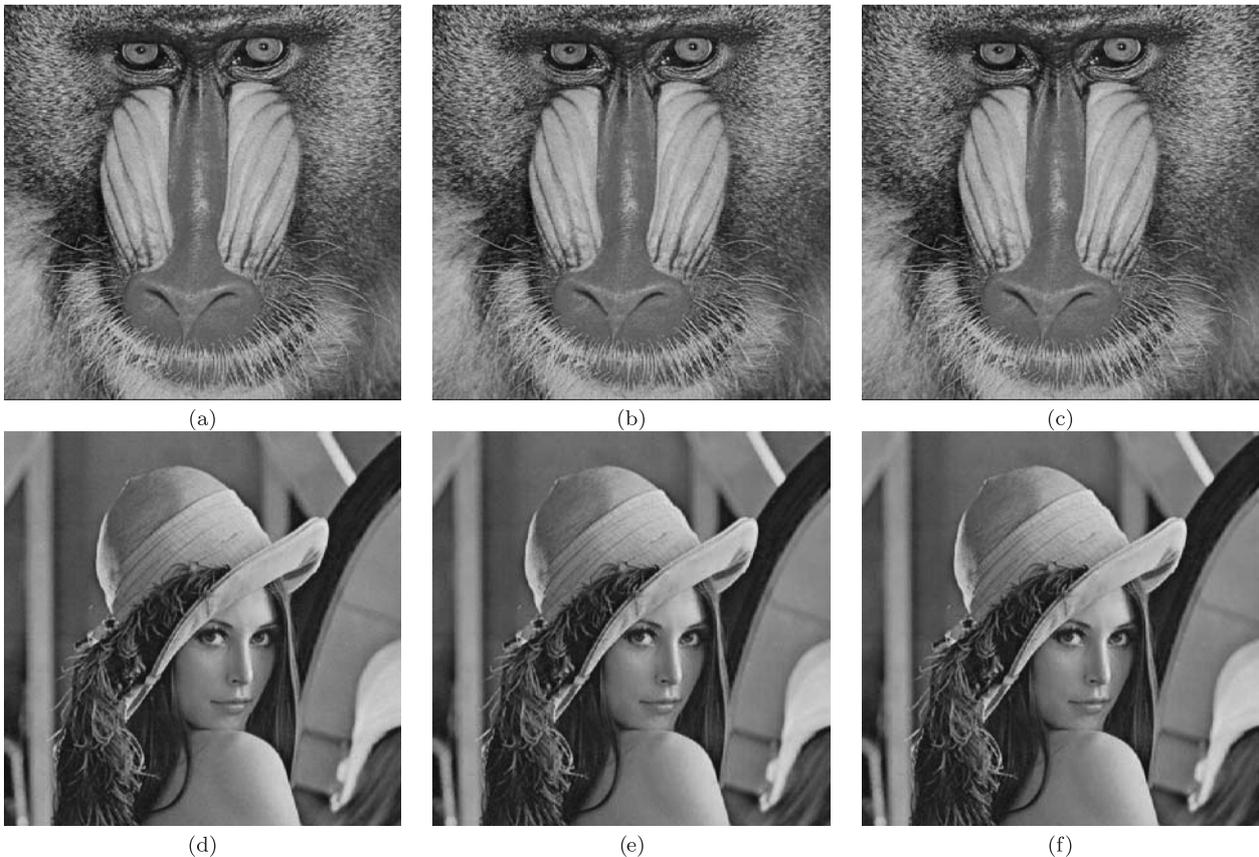
The second experiment compares the PSNR and SSIM performance (see Table 2) and the watermarked images (see

**Table 1** The comparison of the averaged quantization steps, which are carefully tuned so that their resulting watermarked images approximately yield equal SSIM performance as 0.982.

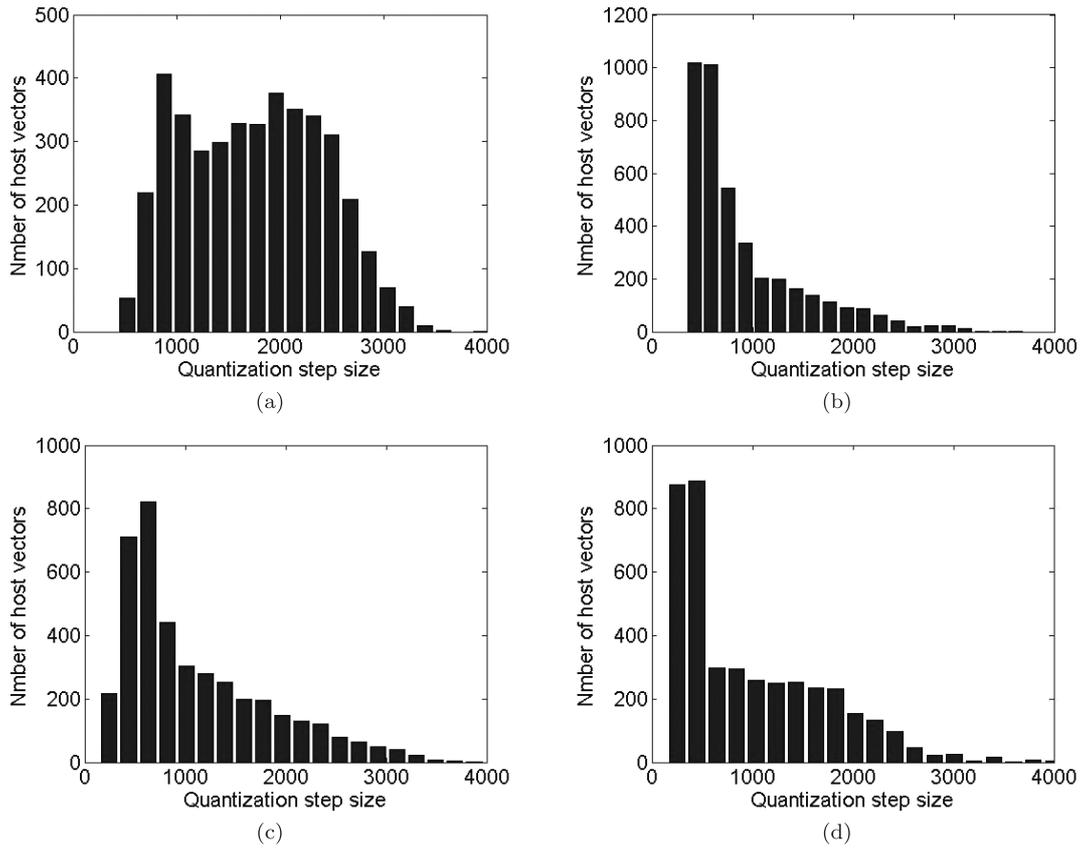
Test image	Conventional STDM approach [3]	Proposed ASTDM approach
<i>Baboon</i>	1450	1746
<i>Lena</i>	800	913
<i>Pills</i>	800	1100
<i>Watch</i>	700	984
Averaged over 500 test images	938	1186

**Table 2** Transparency performance comparison.

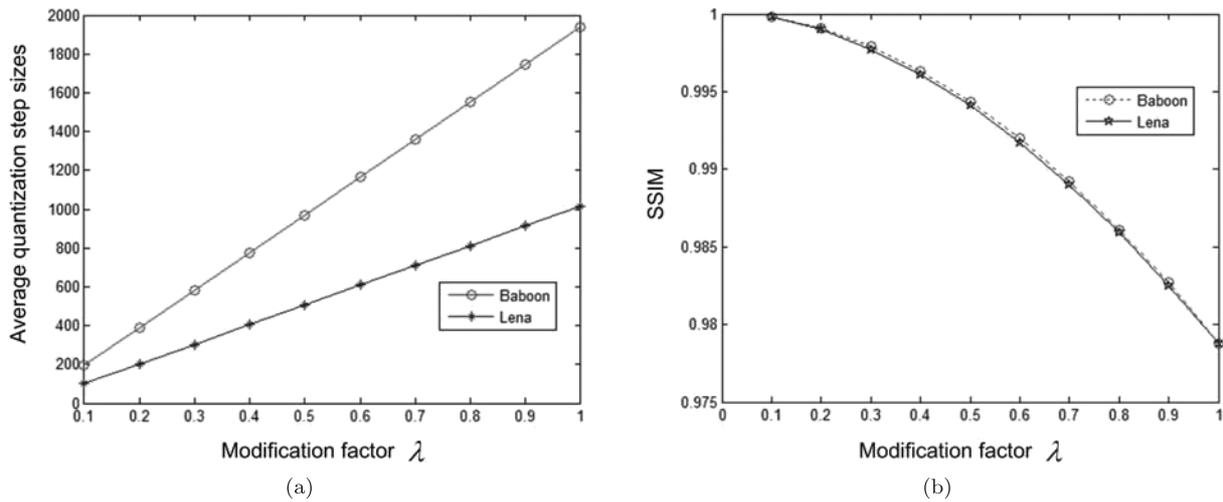
Image	Conventional STDM approach [3]			Li and Cox's approach [12]			Proposed ASTDM approach		
	$\Delta$	SSIM	PSNR	$\lambda$	SSIM	PSNR	$\lambda$	SSIM	PSNR
<i>Baboon</i>	1450	0.9829	38.21 dB	0.86	0.9827	37.84 dB	0.9	0.9826	37.53 dB
<i>Lena</i>	800	0.9828	42.17 dB	0.85	0.9825	42.23 dB	0.9	0.9826	41.60 dB



**Fig. 3** Various watermarked images (This is a color figure): (a) the conventional STDM approach [3] ( $\Delta = 1450$ ); (b) Li and Cox's approach [12] ( $\lambda = 0.86$ ); (c) the proposed ASTDM approach ( $\lambda = 0.9$ ); (d) the conventional STDM approach [3] ( $\Delta = 800$ ); (e) Li and Cox's approach [12] ( $\lambda = 0.85$ ); (f) the proposed ASTDM approach ( $\lambda = 0.9$ ).



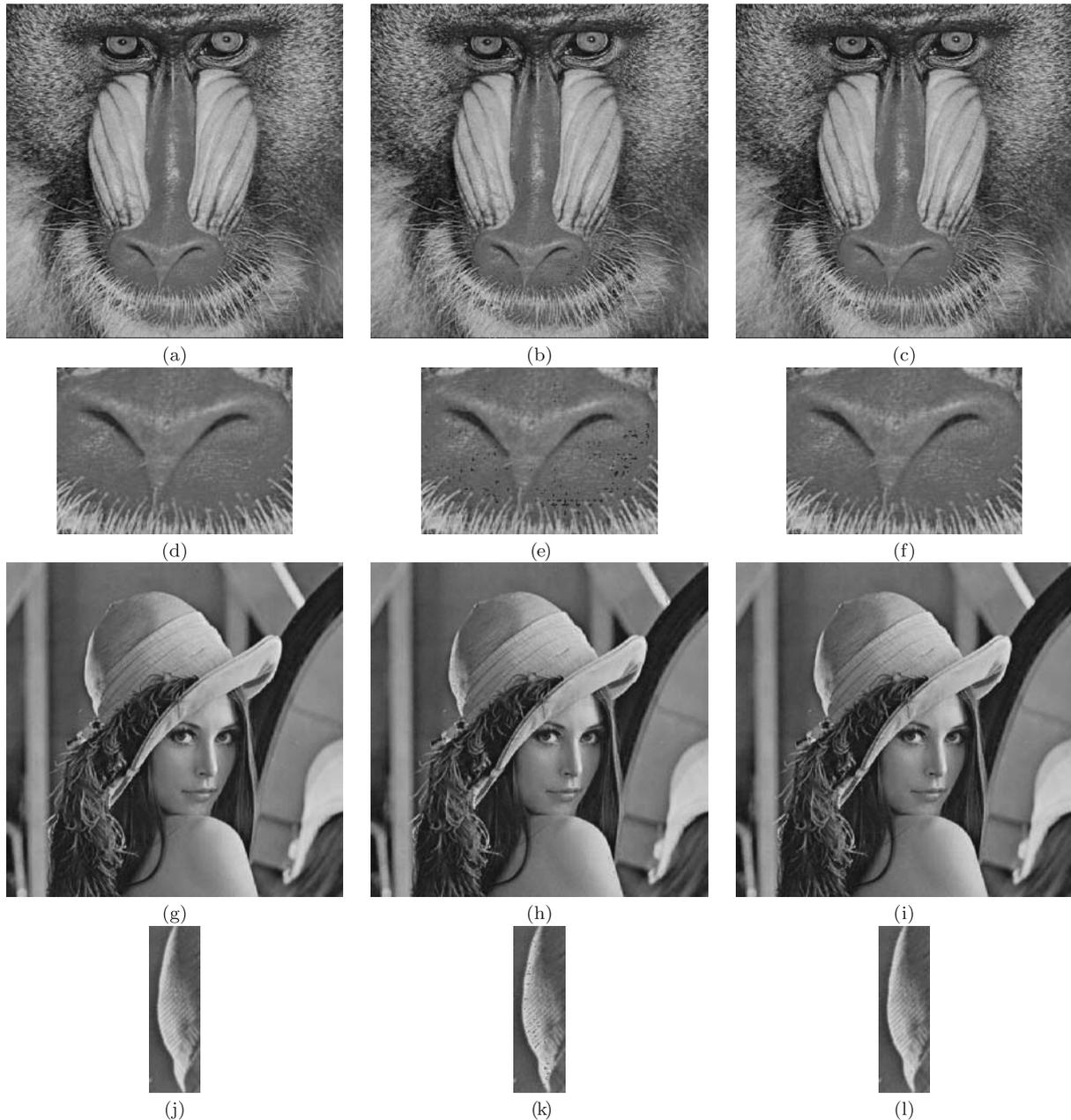
**Fig. 4** The histogram of the quantization step sizes of the proposed ASTDM approach ( $\lambda = 0.9$ ): (a) *Baboon*; (b) *Lena*; (c) *Pills*; and (d) *Watch*.



**Fig. 5** The proposed ASTDM approach can adaptively adjust the quantization step sizes according to the image’s content to achieve that the watermarked image has the same transparency performance (in terms of the SSIM performance [19]): (a) the average quantization step sizes of the proposed ASTDM approach using various modification factors  $\lambda$ ; (b) the SSIM [19] performance of the proposed ASTDM approach using various modification factors  $\lambda$ .

Fig. 3) of the above three approaches using *Baboon* and *Lena*. The proposed ASTDM approach yields an inferior PSNR performance than that of the conventional STD M approach. This is due to the fact that the proposed ASTDM

approach is able to automatically enhance the embedding strength in the areas that could not be easily distinguished by human eyes, so that the PSNR of these areas is lower than that in the conventional STD M algorithm. Consequently,



**Fig. 6** Various watermarked images (This is a color figure): (a) original *Baboon* image; (b) the watermarked image using the proposed ASTDM approach; (c) applying the proposed color artifact suppression method on the image (b); (d)-(f) are the zoomed parts of images (a)-(c), respectively; (g) original *Lena* image; (h) the watermarked image using the proposed ASTDM approach; (i) applying the proposed color artifact suppression method on the image (h); (j)-(l) are the zoomed parts of images (g)-(i), respectively.

the proposed ASTDM approach is able to provide a higher embedding strength than the conventional STDM algorithm, while achieving same SSIM performance. One can also see that  $\lambda$  of Li and Cox's approach [12] is lower than that of our approach when they yield nearly equal SSIM performance. Since lower  $\lambda$  indicates lower embedding strength, it is proved that Li and Cox's approach [12] increases the precisely calculated slack values of original Waston's model in embedding procedure which will result in an inferior water-

mark transparency.

The third experiment aims to demonstrate that the proposed ASTDM approach is able to adjust the quantization step sizes according to the cover image's content. This is in contrast to that the conventional STDM approach applies a *fixed* quantization step size for the whole image. Figure 4 presents the histograms of step sizes for *Baboon*, *Lena*, *Pills* and *Watch* images. Furthermore, the proposed ASTDM approach can automatically adjust the quantization step sizes

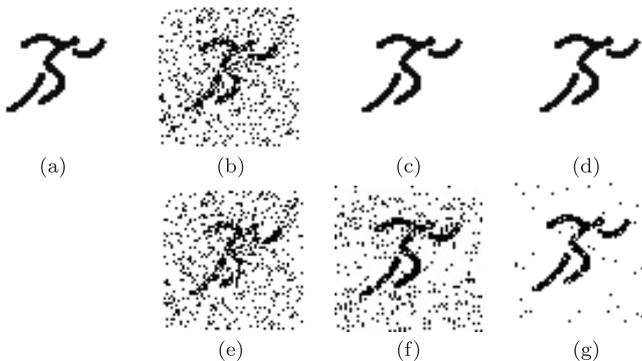
(see Fig. 5 (a)) according to the image content, while maintaining an equal transparency (see Fig. 5 (b)).

The fourth experiment is to demonstrate the performance of the proposed color artifact suppression algorithm, by comparing the watermarked images of the proposed ASTDM approach *with* the incorporation of the proposed color artifact suppression algorithm and that of the proposed ASTDM approach *without* the incorporation of the proposed color artifact suppression algorithm, as presented in Fig. 6. In this experiment, the watermarked image is obtained as follows. The proposed ASTDM approach selects 20 DCT coefficients (specifically, from the 2-nd to the 21-st zig-zag-scanned DCT coefficients) in each  $8 \times 8$  DCT block to form a host vector and embeds one bit into this host vector, with  $\lambda = 0.92$ . As seen from Fig. 6, there are some color distortions (black spots) in the nose area of *Baboon* in Fig. 6 (b), while these artifacts can be removed by the proposed color artifact suppression algorithm as shown in Fig. 6 (h). Likewise, the above conclusion is also applicable to the hair area of *Lena* (see Fig. 6 (c) and (i)).

### 5.2 Robustness Performance

Experiments are conducted to demonstrate the robustness performance of the three watermarking algorithms against various attacks in 500 cover images: (i) the amplitude scaling attack; (ii) the JPEG compression attack; (iii) the additive white Gaussian noise attack and (iv) the Salt and Pepper noise attack.

- *Amplitude scaling attack*: The amplitudes of the watermarked image are scaled by a scaling factor, which varies from 0.1 to 2.0 with a step 0.1. In addition, to provide the subjective performance comparison, Fig. 7 demonstrates the detected watermarks from *Lena* for the cases that the amplitude scaling factors are  $\beta = 0.6$  and  $\beta = 1.4$ , respectively. As seen from this figure, the proposed ASTDM approach can provide a better detected watermark (in terms of visual quality) than the



**Fig. 7** Various extracted watermark images that are extracted from the amplitude scaling attacked image: (a) original watermark; (b) conventional STD approach [3] ( $\beta = 0.6$ ); (c) Li and Cox's approach [12] ( $\beta = 0.6$ ); (d) proposed ASTDM approach ( $\beta = 0.6$ ); (e) conventional STD approach [3] ( $\beta = 1.4$ ); (f) Li and Cox's approach [12] ( $\beta = 1.4$ ); (g) proposed ASTDM approach ( $\beta = 1.4$ ).

conventional STD approach (in all range), plus Li and Cox's approach [12] (when scaling factor is larger than 1.0). This justifies that our proposed ASTDM approach can better resist amplitude scaling attack while maintaining the consistency to the properties of the HVS.

- *JPEG compression attack*: The watermarked image is compressed using the JPEG compression algorithm with different *quality factors* (QFs), which varies from 40 to 100 with a step 5. Note that a large QF value indicates a low compression ratio. For example, a 100 QF value indicates that a lossless JPEG compression is performed.
- *Additive white Gaussian noise attack*: The watermarked image is added with an additive white Gaussian noise with a zero mean and a variance that varies from 0.5 to 6 with a step 0.5.
- *Salt and Pepper noise*: The watermarked image's pixels are randomly corrupted by two fixed extremal values, 0 and 255 that are generated with the same probability  $\rho/2$ , where  $\rho$  represents the noise ratio and it varies from 0.5% to 6% with a step 0.5%.

In each experiment, the *normalized cross-correlation* (NC) between the original watermark (denoted as  $m$ ) and the detected watermark (denoted as  $\hat{m}$ ) is computed to provide the objective performance comparison. The NC performance is defined as

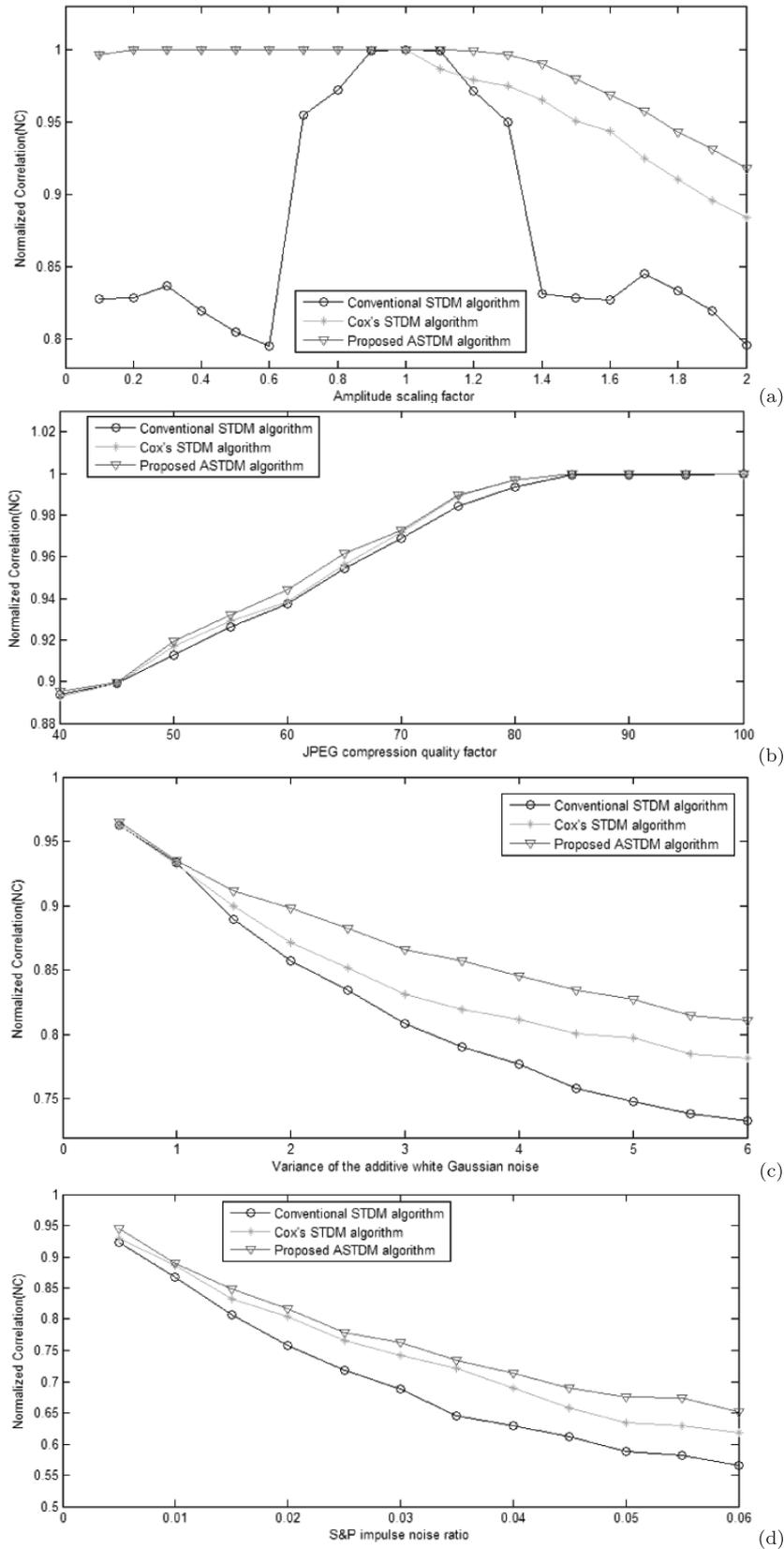
$$NC = \frac{\sum_{i=1}^N m_i \times \hat{m}_i}{\sqrt{\sum_{i=1}^N m_i^2} \times \sqrt{\sum_{i=1}^N \hat{m}_i^2}}, \quad (38)$$

where  $N$  represents the length of the watermark,  $m_i$  and  $\hat{m}_i$  represent the  $i$ -th components of the watermark and the extracted watermark, respectively. Figure 8 compares the NC performance of three algorithms for the above four attack experiments, using the mean value from the experiment results of 500 images. As seen from Fig. 8, the proposed ASTDM approach always outperforms the conventional STD approach, plus Li and Cox's approach [12] by achieving higher NC performance when the watermarked image is attacked with various attacks.

### 5.3 Computational Complexity

A discussion is provided in this section to compare the computational complexity of the proposed approach with that of the conventional STD approach and Li and Cox's approach [12].

- Compared to the conventional STD algorithm, Li and Cox's approach [12] needs to perform an extra multiplication for computing the luminance mask value (see (19)); that is, the total additional computational burden is  $O(M)$  multiplications. On the other hand, the proposed approach also needs to perform an extra multiplication for computing the luminance mask value (see



**Fig. 8** The average NC performance comparisons of the conventional STD approach [3], Cox's approach [12] and the proposed ASTD approach over 500 images from Corel database using various attacks: (a) amplitude scaling attack; (b) JPEG compression attack; (c) additive white Gaussian noise attack; (d) Salt and Pepper noise attack.

(24)); that is, the total additional computational burden is  $O(M)$  multiplications. To summarize, both Li and Cox’s approach and the proposed approach yield an extra  $O(M)$  multiplications, compared with the conventional STDM approach.

- Compared with Li and Cox’s approach [12], the additional computations incurred in the proposed approach are:

- The proposed approach needs to calculate the mean intensity  $C_{0,0}^O$  that needs  $O(n)$  additions, since the mean intensity values of each block is pre-calculated as  $C_{0,0}$ .
- The proposed approach needs to perform the color artifacts suppression operation. Its computational load has already been discussed in the last paragraph in Sect. 4.3.

## 6. Conclusions

In this paper, a new perceptual model is proposed and incorporated into the conventional STDM framework to develop an ASTDM approach for conducting robust color image watermarking. Our proposed new perceptual model improves the conventional Watson’s model by developing a new luminance masking function to resist the amplitude changes, while keeping the consistence to the properties of the HVS. Our proposed new perceptual model is further exploited to adaptively adjust the quantization step sizes for performing watermark embedding, which is contrast to that a fixed quantization step size is utilized in the conventional STDM approach. In addition, a color artifact suppression algorithm is proposed to remove the artifacts that could be presented in the watermarked image. Extensive experiments are conducted to show that the proposed ASTDM approach yields superior transparency performance to that of several state-of-the-art approaches, and provides more robust performance against various attacks, as verified in our extensive experiments using 500 images selected from the Corel database.

## Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 60972133) and Guangdong Science Fund (Grant No. 9351064101000003).

## References

[1] F. Hartung and M. Kutter, “Multimedia watermarking techniques,” Proc. IEEE, vol.87, no.7, pp.1079–1107, July 1999.  
 [2] C.I. Podilchuk and E.J. Delp, “Digital watermarking: algorithms and applications,” IEEE Signal Process. Mag., vol.18, no.4, pp.33–46, July 2001.  
 [3] B. Chen and G.W. Wornell, “Quantization index modulation: A class of provably good methods for digital watermarking and information embedding,” IEEE Trans. Inf. Theory, vol.47, no.4, pp.1423–1443, May 2001.

[4] I.J. Cox and M.L. Miller, “A review of watermarking and the importance of perceptual modeling,” Proc. SPIE Conf. on Electronic Imaging, pp.92–99, Feb. 1997.  
 [5] I.J. Cox, M.L. Miller, and J.A. Bloom, Digital Watermarking: Principles and Practice, Morgan Kaufmann, 2002.  
 [6] J.J. Eggers, R. Bauml, and B. Girod, “Estimation of amplitude modifications before SCS watermark detection,” Proc. SPIE Conf. on Multimedia Systems and Applications, pp.387–398, San Jose, CA, June 2002.  
 [7] K. Lee, D.S. Kim, T. Kim, and K.A. Moon, “EM estimation of scale factor for quantization-based audio watermarking,” Proc. International Workshop on Digital Watermarking, pp.312–367, Seoul, Korea, Oct. 2003.  
 [8] P. Dempster, N.M. Laird, and D.B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” J. Royal Statist. Soc. B, vol.39, no.1, pp.1–38, 1977.  
 [9] R.L. Lagendijk and I.D. Shterev, “Estimation of attacker’s scale and noise variance for QIM-DC watermark embedding,” Proc. IEEE Int. Conf. on Image Processing, pp.55–58, Singapore, Oct. 2004.  
 [10] F. Perez-Gonzalez, C. Mosquera, M. Barni, and A. Abrardo, “Rational dither modulation: A high-rate data-hiding method invariant to gain attacks,” IEEE Trans. Signal Process., vol.53, no.10, pp.3960–3975, Oct. 2005.  
 [11] X. Zhu and Z. Tang, “Improved quantization index modulation watermarking robust against amplitude scaling distortions,” Proc. IEEE Int. Conf. on Multimedia and Expo, pp.237–240, Hannover, Germany, June 2008.  
 [12] Q. Li and I.J. Cox, “Using perceptual models to improve fidelity and provide resistance to volumetric scaling for quantization index modulation watermarking,” IEEE Trans. Information Forensics and Security, vol.2, no.2, pp.127–139, June 2007.  
 [13] P. Moulin and R. Koetter, “Data-hiding codes,” Proc. IEEE, vol.93, no.12, pp.2083–2126, Dec. 2005.  
 [14] A.B. Watson, Digital Image and Human Vision, MITPress, Cambridge, 1993.  
 [15] Q. Li and I.J. Cox, “Improved spread transform dither modulation using a perceptual model: Robustness to amplitude scaling and JPEG compression,” Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp.185–188, Honolulu, HI, April 2007.  
 [16] Q. Li and I.J. Cox, “Rational dither modulation watermarking using a perceptual model,” Proc. IEEE Int. Workshop on Multimedia Signal Processing, pp.1–4, Shanghai, China, Oct. 2005.  
 [17] Corel Database, <http://wang.ist.psu.edu/docs/related/>  
 [18] R.C. Gonzalez and R.E. Woods, Digital Image Processing, Prentice Hall, 2007.  
 [19] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol.13, no.4, pp.600–612, April 2004.

## Appendix: Derivation of (37)

Denote the normalized red, green and blue components of the pixel (at the position  $(i, j)$ ) of the color image are  $R(i, j)$ ,  $G(i, j)$  and  $B(i, j)$ , respectively. Further denote its corresponding hue, saturation and intensity values are  $H(i, j)$ ,  $S(i, j)$  and  $I(i, j)$ , respectively. The relationship among the above red, green, blue components and their corresponding hue, intensity and saturation values can be mathematically expressed as [18]

$$S(i, j) = 1 - \frac{3 \times \min(R(i, j), G(i, j), B(i, j))}{R(i, j) + G(i, j) + B(i, j)}, \quad (\text{A} \cdot 2)$$

$$I(i, j) = \frac{R(i, j) + G(i, j) + B(i, j)}{3}. \quad (\text{A} \cdot 3)$$

$$H(i, j) = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R(i, j) - G(i, j)) + (R(i, j) - B(i, j))]}{\sqrt{(R(i, j) - G(i, j))^2 + (R(i, j) - B(i, j))(G(i, j) - B(i, j))}} \right\}. \quad (\text{A} \cdot 1)$$

Substituting (A·3) into (A·2), we have

$$\begin{aligned} S(i, j) &= 1 - \frac{3 \times \min(R(i, j), G(i, j), B(i, j))}{R(i, j) + G(i, j) + B(i, j)} \\ &= 1 - \frac{\min(R(i, j), G(i, j), B(i, j))}{I(i, j)}. \end{aligned} \quad (\text{A} \cdot 4)$$

The objective is to derive the upper bound (i.e., (37)) for each pixel's intensity values of the color image, according to its inherent relationship with its corresponding hue and saturation values (i.e., (A·4)). Now we analyze the bound for the following two cases  $I(i, j) \leq \frac{2}{3}$  and  $\frac{2}{3} < I(i, j) < 1$ , respectively.

First, for the case  $I(i, j) \leq \frac{2}{3}$ , we have  $R(i, j) + G(i, j) + B(i, j) \leq 2$ . Therefore, the pixel's any red, green or blue component could yield a value of 0; that is

$$\min(R(i, j), G(i, j), B(i, j)) \geq 0, \quad (\text{A} \cdot 5)$$

where the equality is achieved for the case that anyone of  $R(i, j), G(i, j), B(i, j)$  is zero. Substituting (A·5) into (A·4), we have,

$$\begin{aligned} S(i, j) &= 1 - \frac{3 \times \min(R(i, j), G(i, j), B(i, j))}{R(i, j) + G(i, j) + B(i, j)} \\ &\leq 1 - \frac{3}{R(i, j) + G(i, j) + B(i, j)} \times 0 \\ &= 1. \end{aligned} \quad (\text{A} \cdot 6)$$

On the other hand, according to (A·3), we have

$$\begin{aligned} \min(R(i, j), G(i, j), B(i, j)) &\leq \frac{R(i, j) + G(i, j) + B(i, j)}{3} \\ &= I(i, j), \end{aligned} \quad (\text{A} \cdot 7)$$

where the equality is achieved for the case that  $R(i, j) = G(i, j) = B(i, j)$ . Substituting (A·7) into (A·4), we have

$$\begin{aligned} S(i, j) &= 1 - \frac{3 \times \min(R(i, j), G(i, j), B(i, j))}{R(i, j) + G(i, j) + B(i, j)} \\ &\geq 1 - \frac{3}{R(i, j) + G(i, j) + B(i, j)} \times I(i, j) \\ &= 1 - \frac{3 \times (R(i, j) + G(i, j) + B(i, j))}{(R(i, j) + G(i, j) + B(i, j)) \times 3} \\ &= 0. \end{aligned} \quad (\text{A} \cdot 8)$$

Referring to (A·6) and (A·8), the saturation value  $S(i, j)$  always belongs to the range  $[0, 1]$ . Therefore, there is no need to correct the intensity value in this case.

Second, for the case  $\frac{2}{3} < I(i, j) < 1$ , we have  $R(i, j) + G(i, j) + B(i, j) > 2$ . Therefore, the pixel's all red, green or blue components must yield a non-zero value; that is

$$\min(R(i, j), G(i, j), B(i, j)) \geq 3I(i, j) - 2, \quad (\text{A} \cdot 9)$$

where the equality is achieved for the case that any two of  $R(i, j), G(i, j), B(i, j)$  equal to one. Substituting (A·9) into (A·2), we have,

$$\begin{aligned} S(i, j) &= 1 - \frac{3 \times \min(R(i, j), G(i, j), B(i, j))}{R(i, j) + G(i, j) + B(i, j)} \\ &\leq 1 - \frac{3}{R(i, j) + G(i, j) + B(i, j)} \times (3I(i, j) - 2) \\ &= 2 \left( \frac{1}{I(i, j)} - 1 \right). \end{aligned} \quad (\text{A} \cdot 10)$$

Furthermore (A·10) can be rewritten as

$$I(i, j) \leq \frac{2}{S(i, j) + 2}. \quad (\text{A} \cdot 11)$$

On the other hand, according to (A·3), we have

$$\begin{aligned} \min(R(i, j), G(i, j), B(i, j)) &\leq \frac{R(i, j) + G(i, j) + B(i, j)}{3} \\ &= I(i, j), \end{aligned} \quad (\text{A} \cdot 12)$$

where the equality is achieved for the case that  $R(i, j) = G(i, j) = B(i, j)$ . Substituting (A·12) into (A·2), we have

$$\begin{aligned} S(i, j) &= 1 - \frac{3 \times \min(R(i, j), G(i, j), B(i, j))}{R(i, j) + G(i, j) + B(i, j)} \\ &\geq 1 - \frac{3}{R(i, j) + G(i, j) + B(i, j)} \times I(i, j) \\ &= 1 - \frac{3 \times (R(i, j) + G(i, j) + B(i, j))}{(R(i, j) + G(i, j) + B(i, j)) \times 3} \\ &= 0. \end{aligned} \quad (\text{A} \cdot 13)$$

Finally, to summarize the analysis in the above two cases, the only constraint imposed on the intensity values is (A·11) for the second case. Therefore, the upper bound of the intensity value (denoted as  $I_{max}^U(i, j)$ ) as

$$I^U(i, j) = \begin{cases} \frac{2}{S(i, j) + 2}, & \text{if } \frac{2}{3} < I(i, j) \leq 1; \\ 1, & \text{otherwise.} \end{cases} \quad (\text{A} \cdot 14)$$



**Lihong Ma** received the B.Sc. in Radio Engineering, M.Sc. and Ph.D. degrees both in Communication and Circuit Systems from South China University of Technology, Guangzhou, China, in 1987, 1990 and 1999, respectively. She was a Research Assistant at The Hong Kong Polytechnic University in 2000. She is currently an Associate Professor of the School of Electronic and Information Engineering, South China University of Technology. Her research interests include digital signal transformation, image/Video analysis and coding. She has published about 30 research papers appeared in international journals and conferences. She has been a Technical Program Co-Chair of IEEE ICCS 2008. She has received awards such as the Nature Science Award of Guangdong Province, the Distinguished Researcher of Guangzhou and the Best Paper Award from *CIS 2006*.



**Dong Yu** received the B.Sc. degree in information engineering from South China University of Technology in 2002, Guang Zhou, China, and is currently pursuing his M.Sc. degree in the Department of Electrical and Electronic Engineering at South China University of Technology, and has published his research results in IEEE ICIP 2008 and IEEE GLOBECOM 2008. His research interests include color image processing, information security, multimedia and communications.



**Gang Wei** received the B.Sc. and M.Sc. degrees in Radio Engineering from Tsinghua University, Beijing, China, and the Ph.D. degree in Communication and Circuit Systems from South China University of Technology, Guangzhou, China, in 1984, 1987, and 1990, respectively. He was a Visiting Scholar at University of Southern California, Los Angeles, from June 1997 to June 1998. He is currently a Chair Professor and Head of the School of Electronic and Information Engineering, South China University of Technology. His research interests are in signal processing and communications. He has published more than 100 research papers appeared in international journals, such as IEEE Trans. on Circuits and Systems, IEEE Trans. on Multimedia, Signal Processing, etc., and the top 3 journals in China. He has been a Keynote/Invited speaker and Chairperson of some international conference. He has received awards such as the honorable Distinguished Young Researcher of China. He is a Committee Member of the National Natural Science Foundation of China.



**Jing Tian** received his B.Eng. degree (Electronic and Information Engineering), from School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China, M.Eng. degree (Electronic and Information Engineering), from School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China, Ph.D. degree (Electrical and Electronic Engineering), from School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore. His research interests include image processing, pattern recognition and computer vision.



**Hanqing Lu** received the B.S. and M.S. degrees in computer science from the Harbin Institute of Technology, Harbin, China in 1982 and 1985, and the Ph.D. degree in Huazhong University of Sciences and Technology, Wuhan, China in 1992. Currently, he is a Professor of Institute of Automation, Chinese Academy of Sciences. His research interests include image similarity measure, video content analysis, object recognition and tracking. He published more than 200 papers in those areas, appeared in IEEE Trans. on Multimedia, etc., and the top 3 journals in China. He has received many awards such as the China Nature Science Award, etc.