

RESEARCH

Open Access



A new blind color watermarking based on a psychovisual model

Pascal Lefevre¹, David Alleysson² and Philippe Carre^{1*} 

*Correspondence:

philippe.carre@univ-poitiers.fr

¹Laboratoire XLIM CNRS UMR 7252, Université de Poitiers, Bât. H1 - SP2MI, 11 Bd Marie et Pierre Curie, TSA 41123, 86073 Poitiers Cedex 9, France

Full list of author information is available at the end of the article

Abstract

In this paper, we address the problem of the use of a human visual system (HVS) model to improve watermark invisibility. We propose a new color watermarking algorithm based on the minimization of the perception of color differences. This algorithm is based on a psychovisual model of the dynamics of cone photoreceptors. We used this model to determine the discrimination power of the human for a particular color and thus the best strategy to modify color pixels. Results were obtained on a color version of the lattice quantization index modulation (LQIM) method and showed improvements on psychovisual invisibility and robustness against several image distortions.

Keywords: Metrical space; Robust watermarking; Psychovisual model; Lattice QIM

1 Introduction

Nowadays, technologies related to multimedia are more accessible than ever, and it is easier for anyone to abuse unprotected contents. Watermarking has been one of the numerous multimedia security solutions for almost thirty years. A lot of interest has been dedicated to color image watermarking algorithms very early, which remains today a complex problem.

Literature contains numerous color watermarking algorithms. Earlier solutions proposed to apply classical grayscale strategy only on one of the three color channels. For example, in the past, [12, 20] proposed to select the blue channel in order to minimize perceptual changes in the watermarked image at the cost of a weaker robustness. Other contributions proposed to modify the intensity component [19] or the saturation component [10] in order to increase the robustness at the cost of a stronger embedding distortion.

Though there are some recent works on color image watermarking, the majority of the published works use grayscale images as host images for embedding the watermark information. Using colors, some works suggest that color images can be treated as grayscale information and each color component has to be processed in the same fashion separately (for example, [9]). Since color images contain more information than the grayscale counterparts, color image embedding has a different scale of impact; moreover, it is admitted that color watermarking method tradeoffs (robustness vs invisibility) can be improved by considering a vector approach. According to this strategy, some recent works such as [1]

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

described such approach using a PCA decomposition, or introduced particular adapted vectorized strategy such as [16] that proposed a vector approach and modified wavelet coefficients computed from the RGB color space or color transform. Others use specific color transform: in [13] a quaternionic discrete cosine transform is proposed. However, the psychovisual invisibility of the mark is not directly addressed.

One step further towards better psychovisual invisibility is to take into account how the HVS deals with color distortions. In [5], a two-dimensional mark is embedded in the chromatic plane xy . This approach uses a histogram manipulation technique proposed in [8]. To improve the invisibility, color distances corresponding to embedding distortions are computed in the $L^*a^*b^*$ colorspace. Then, this method was improved in [6] by adding the mark directly into the $L^*a^*b^*$ colorspace. In other words, the embedding is processed in the color domain and takes advantage of the low sensitivity of the HVS to perceive small color differences.

This paper aims to develop a watermark embedding scheme that handles color images more efficiently taking care of the local color discrimination power of the human eye. So far, no-one has fully investigated the perception of color differences in the context of color watermarking. In our work, we propose a 3D quantization algorithm (Sect. 2) minimizing the perception of color differences of the HVS. It is based on a psychovisual model of photoreceptor dynamics in the human eye (Sect. 3). After explaining how to adapt a vector quantization method for color quantization, we show performance improvements thanks to some experimental results on psychovisual invisibility and robustness against various image distortions.

Before starting the presentation of our method, we introduce some elements concerning watermarking. There are several ways in which we can model a watermarking process. In this article, we use a model which is based on a communication-based view of watermarking. Watermarking is a process of communicating a message from the watermarking embedder to the watermarking receiver, and the cover image is the channel. In a general secure communication model, on one side we would have the sender which would encode a message using some kind of encoding key in order to be resistant to channel-related perturbation (for example, correcting code). Then the message would be transmitted on a communication channel, which would add some noise to the encoded message (in the watermarking context, it is the attack aspect). The resulting noisy message would be received by the receiver which would try to decode it using the decoding key to get the original message back. In this article, we use the quantization process. Every watermarking system has some very important desirable properties which are often conflicting. The first important property is the image fidelity. Watermarking is a process that alters an original image to add a message to it. We want to keep the degradation of the image's quality to a minimum: it is the notion of distortion. This property can be influenced by a parameter of the watermarking process: the embedding rate which measures the ratio between the size of the watermark (bits) and the number of samples of the host (cover image). The second important property is the robustness which is crucial for most watermarking systems. The watermarked message is altered during its lifetime, for example, by transmission over a lossy channel (attacks). A robust watermark should be able to withstand many operations: additive Gaussian noise, compression, color operations, etc.

2 Color quantization

In this section, we describe a general three-dimensional color quantization method in a three-channel colorspace. As set out in the introduction, the color watermarking process is based on lattice QIM. We briefly review this classical strategy.

2.1 Lattice QIM (LQIM)

Quantization index modulation (QIM) was introduced by Chen et al. in [7] and later extended as the lattice QIM method. We propose to describe the version studied in [15]. To embed a binary information $m \in \{0, 1\}$, a quantizer Q_m is used to send a host sample x on a coset Λ_m of $\Delta\mathbb{Z}^L$ corresponding to bit m . x is a vector of dimension L whose values can be pixel values, DCT or DWT coefficients, etc. We have

$$y = Q_m(x, \Delta) = \left\lfloor \frac{x}{\Delta} \right\rfloor \Delta + (-1)^{m+1} \frac{\Delta}{4} \tag{1}$$

y is the result of the quantization. Cosets are defined as follows:

$$\begin{cases} \Lambda_0 = \Delta\mathbb{Z}^L - \frac{\Delta}{4}, \\ \Lambda_1 = \Delta\mathbb{Z}^L + \frac{\Delta}{4}. \end{cases}$$

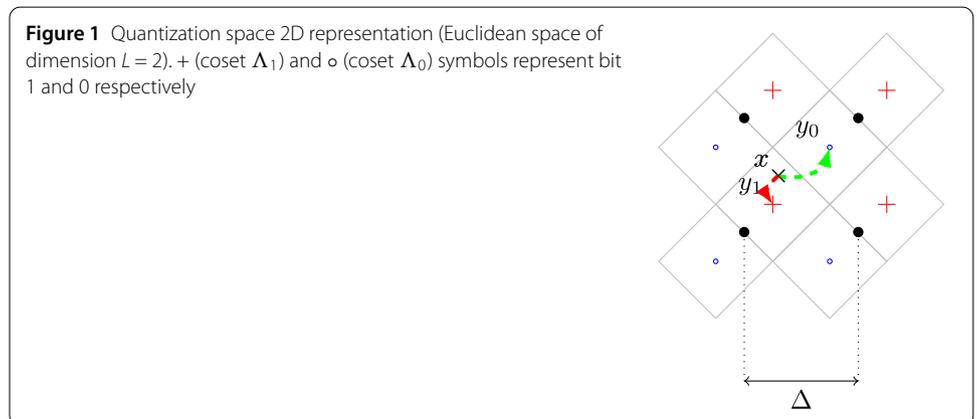
An example of quantization is depicted in Fig. 1 where a vector x is quantized to embed bit m . If $m = 0$, the result of the quantization is $y_0 = Q_0(x, \Delta)$ and $y_1 = Q_1(x, \Delta)$ otherwise.

The construction of the quantization space $\Lambda_0 \cup \Lambda_1$ relies on the idea of equally spreading binary information in the Euclidean space so that any $x \in \mathbb{R}^2$ can be transformed into y_m without important distortions. These distortions are controlled by the quantization step Δ . More formally, the distortion in dimension $L = 2$ is bounded so that

$$\|y_m - x\|_2 \leq \Delta \frac{\sqrt{2}}{2}. \tag{2}$$

Hence, one can note that a larger Δ leads to potentially more distortions in the image being watermarked.

At the detection step, one receives a vector $z \in \mathbb{R}^L$ that is modified from the original watermarked value y_m . In order to estimate bit m , one needs to find the closest coset Λ_m



(with respect to the Euclidean distance $\|\cdot\|_2$). In other words, we have

$$\hat{m} = \arg \min_{m \in \{0,1\}} \text{dist}(z, \Lambda_m)$$

with

$$\text{dist}(z, \Lambda) = \min_{y \in \Lambda} \|z - y\|_2.$$

Using Fig. 1 again, we plotted a diagonal grid (or cells centered with + and \circ) as a visual example in dimension $L = 2$. To illustrate the detection of z , it then suffices to identify the cell center y_m where z is located.

2.2 Color adaptation

For a given color pixel P , the embedding distortion is controlled by a direction vector u_P , i.e., the concept of scalar quantization is applied along the axis u_P .

The new value of the pixel after the watermarking by quantization (denoted by P_w) is obtained as follows:

$$P_w = P + (s_w - s) \cdot u_P \quad (3)$$

with u_P being a normalized direction vector depending on the color pixel P , $s = \langle P, u_P \rangle$ is the projection of the pixel vector P onto the direction vector u_P . s_w is the modified scalar product by the quantification procedure introduced above:

$$s_w = Q_{m(P)}(s, \Delta)$$

with $m(P)$ being the message bit associated with the pixel P .

Let us note that this approach has some consequences on the color perception. Given a pair of pixels P and P_w separated by a distance $d_2(P, P_w)$, the perceived color distortion may be different for the HVS when considering a different pair of pixels (P', P'_w) separated by the same distance $d_2(P, P_w)$.

At the detection step, s_w is the coefficient to estimate in order to extract the embedded information:

$$s_{w,a} = \langle P_{w,a}, u_{P_{w,a}} \rangle, \quad (4)$$

where $P_{w,a}$ is the watermarked and modified color pixel after some attacks and $u_{P_{w,a}}$ is its associated estimated direction vector. Consequently, to work properly, it is necessary that direction vectors u_P and $u_{P_{w,a}}$ are close enough to correctly estimate s_w . If, as we will show later, the direction u_P is given by a model of the direction of lower visibility, it is necessary for the psychovisual model to be stable. If the color difference $P_{w,a} - P$ is small (i.e., the color information of the cover image is not degraded by the watermarking process and the attacks), then u_P and $u_{P_{w,a}}$ have to be closed.

The following section defines the general process to embed and detect a message into a color image. The key of the method is the setting of parameter u_P .

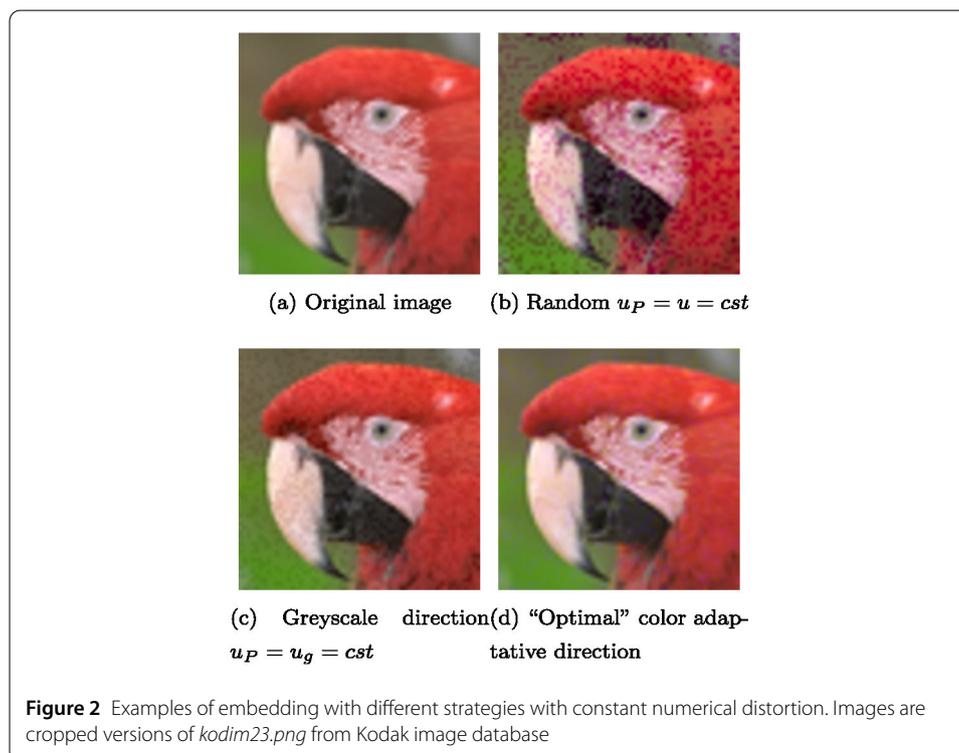
2.3 Color direction embedding

We can use different strategies to choose a color direction vector u_P for the embedding step. One of the aims is to find a direction that minimizes the color quantization noise for the HVS. An elementary approach consists in choosing a constant direction $u_P(P) = u$ for every color.

To illustrate the influence of this choice, we randomly choose $u_P = u$ constant, then it is easy to perceive color distortions (Fig. 2b) because this direction introduces colors that are not relevant to the image content.

It is possible to decrease the perceived quantization noise by using, for example, the constant direction $u_g = (1, 1, 1)/\sqrt{3}$ associated with the grayscale or luminance information. This strategy is used by many authors. As illustrated in Fig. 2c, we experimentally concluded that u_g produces a lower number of irrelevant colors. Compared to Fig. 2b, we can perceive fewer wrong colors for the same level of numerical distortion. However, this result is not totally satisfying for watermarking invisibility. Since the perception of quantization noise is different in function of the color information associated with the pixel, it is necessary to introduce an adaptive approach. For every color pixel P , a direction vector $u_P(P)$, depending on P , will be determined so that the color quantization noise is minimized for the HVS. A simple example using the *adaptive approach* is illustrated in Fig. 2d. We can see that the quantization noise is more difficult to see. This figure illustrates the main objective of the article: the use of the human visual system (HVS) model to improve the watermark invisibility.

In the next section, we propose a psychovisual model and a method to extract adaptive direction vectors required to set up the adaptive approach.



3 Psychovisual approach to the HVS

Understanding how humans perceive color differences can allow us to minimize embedding distortions perceptually, i.e., improve the invisibility of a watermarking system. To achieve this goal, we require a model of color discrimination. This model is a simplification of the sophisticated neural process along the whole visual chain and solely involved photoreception layer. The model is three-dimensional because of human trichromacy, and it mimics the representation of light by cone receptors for their spectral sensitivities as well as their dynamics of intensity coding.

The model is an oversimplification of what is called neurogeometry which was proposed in the context of form perception based on achromatic information. A detailed description of the construction of forms by the primary visual cortex is described in [17, 18]. Few others have considered neurogeometry in the context of color vision [4, 11, 21].

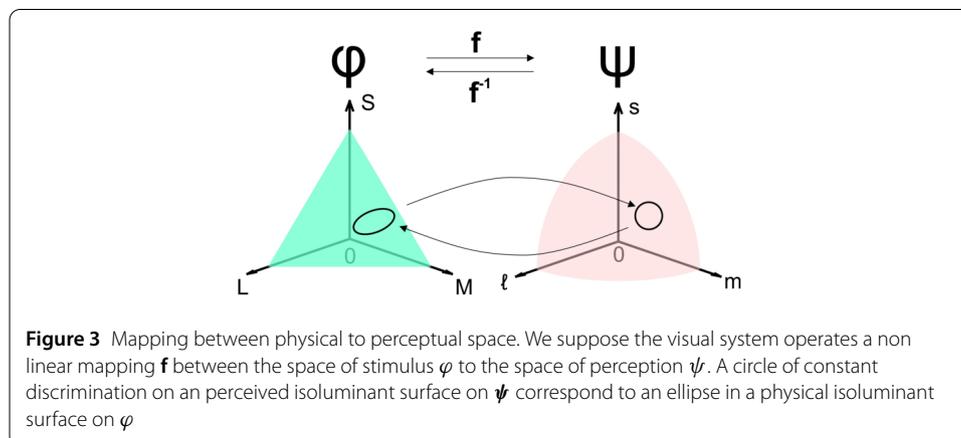
In the model we suppose that the visual system operates a nonlinear mapping between the space of physical stimuli ϕ and the perceptual space ψ . The map between those spaces is given by the neural function. This model is illustrated in Fig. 3. A circle of constant discrimination on a perceived isoluminant surface on ψ corresponds to an ellipse in a physical isoluminant surface on ϕ .

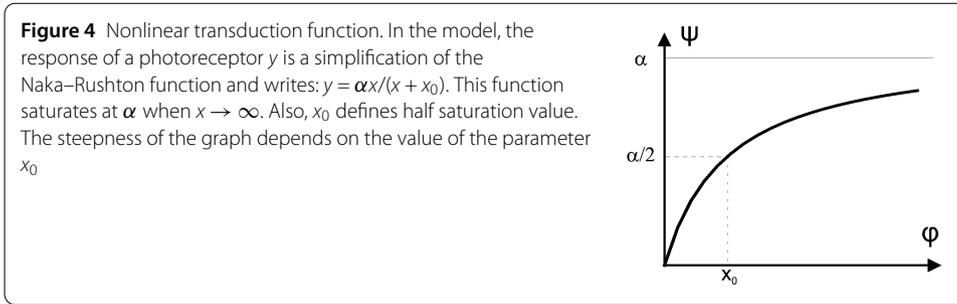
We will give an explicit nonlinear and parametric transform between ϕ and ψ . We made the hypothesis that the perceptual space ψ is an Euclidean space; consequently, contours of constant discrimination are spheres. Using classical tools of differential calculus one can endow the space ϕ with a metric obtained from the Euclidean metric on ψ .

Let us define \mathbf{f} as a simplification of the Naka–Rushton function: the transduction of the light by the photoreceptor [3]:

$$\begin{aligned} \mathbf{f}(x) &= \alpha \frac{x}{x + 1}, \\ y = \mathbf{f}(x/x_0) &= \alpha \frac{x}{x + x_0} \end{aligned} \tag{5}$$

with y being the transduction level, i.e., represents the electrical responses of a cone in function of x the excitation level of the cone produced by the light, x_0 is the adaptation state, and α is the gain of the cone channel. x_0 is modified in function of the average excitation level of the photoreceptor. Equation (5) represents a photoreceptor behavior in function of two constants α and x_0 fixed by environmental conditions and the HVS state.





In fact, the adaptation process of a photoreceptor is only partially understood and this model is probably a simplification.

The function saturates at α when $x \rightarrow \infty$ and x_0 gives the half maximum of the curve. Depending on x_0 , the graph of the function changes. Low x_0 gives high steepness and high x_0 gives low steepness of the graph. This nonlinear function is illustrated in Fig. 4.

For the three-dimensional space of physical stimuli φ , we can write the correspondence between the physical space and the perceptual space as follows:

$$\begin{aligned}
 \mathbf{f} : \varphi &\rightarrow \psi \\
 x \mapsto y = \mathbf{f}(x) &= \left(\frac{\alpha_L L}{L + L_0}, \frac{\alpha_M M}{M + M_0}, \frac{\alpha_S S}{S + S_0} \right), \tag{6}
 \end{aligned}$$

where $x = (L, M, S)$ is the coordinate of the physical stimulus in φ space, $x_0 = (L_0, M_0, S_0)$ and $\alpha = (\alpha_L, \alpha_M, \alpha_S)$ are respectively the adaptation of the observer and the gain of color channels. We implicitly assume here without loss of generality that the coordinates in φ correspond to the responses of the L, M, and S cones. Then $y = (\ell, m, s)$ is the encoding of the physical stimulus x by the visual system adapted to x_0 with gain α .

According to this mapping, the metrics in the two spaces are related to each other. g^ψ in the space of physical stimuli is given by the pull-back of the metric in the perceptual space g^ψ through the mapping \mathbf{f} and writes:

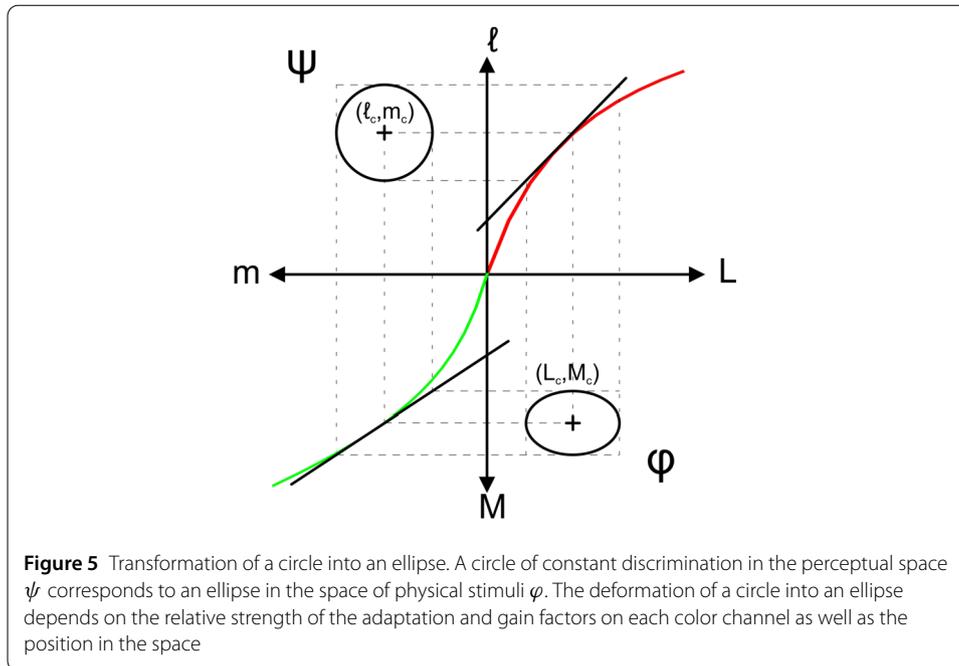
$$g^\varphi(u, v) = g^\psi(D\mathbf{f}(u), D\mathbf{f}(v)), \tag{7}$$

where u and v are two vectors of the tangent space (Fig. 5) of ψ and $D\mathbf{f}$ is the Jacobian matrix of \mathbf{f} which writes:

$$D\mathbf{f}(x) = \begin{bmatrix} \frac{\alpha_L L_0}{(L+L_0)^2} & 0 & 0 \\ 0 & \frac{\alpha_M M_0}{(M+M_0)^2} & 0 \\ 0 & 0 & \frac{\alpha_S S_0}{(S+S_0)^2} \end{bmatrix}. \tag{8}$$

Because we consider the perceptual space to be Euclidean, the metric g^ψ is the identity matrix, thus the metric g^φ writes:

$$g^\varphi = \begin{bmatrix} \frac{\alpha_L^2 L_0^2}{(L+L_0)^4} & 0 & 0 \\ 0 & \frac{\alpha_M^2 M_0^2}{(M+M_0)^4} & 0 \\ 0 & 0 & \frac{\alpha_S^2 S_0^2}{(S+S_0)^4} \end{bmatrix}. \tag{9}$$



We now propose to use this model in order to determine for each color of the image to be watermarked the direction optimizing the robustness/invisibility compromise.

4 Application for color watermarking

In the previous section, we have introduced a color vision model to develop a more invisible watermarking method. In the LMS colorspace, we built ellipsoids that concretely illustrate how to control psychovisual distortion and then enable robustness improvements for a given psychovisual distortion level. Indeed, the notion of psychovisual distortions represents the color difference perception at embedding as opposed to numerical distortions. Then, since psychovisual distortions describe the watermark invisibility for the human better than numerical distortion measures, the hypothesis is that it becomes possible to obtain a better tradeoff between invisibility and robustness: for the same psychovisual distortions level, there are different levels of numerical distortions, and the choice of the more important numerical distortion is interesting for robustness improvements (maximizing numerical distortions increases robustness).

In other words, we propose in this section to extract from every color pixel P a direction vector u_P which has the largest norm in its ellipsoid. u_P is a vector in the RGB space where we have the maximum numerical distortions level allowed.

4.1 Conversions

We give color space conversion matrices to work from psychovisual spaces XYZ and LMS to the RGB space. Transformations are chosen following the 1931 CIE standard.³ Conversion between color spaces is represented by the following matrices:

$$M_1 = M_{\text{RGB} \rightarrow \text{XYZ}} = \begin{pmatrix} 0.488718 & 0.310680 & 0.200602 \\ 0.176204 & 0.812985 & 0.010811 \\ 0.000000 & 0.010205 & 0.989795 \end{pmatrix},$$

$$M_2 = M_{XYZ \rightarrow LMS} = \begin{pmatrix} 0.38971 & 0.68898 & -0.07868 \\ -0.22981 & 1.18340 & 0.04641 \\ 0.0 & 0.0 & 1.0 \end{pmatrix}.$$

Let P_{RGB} , P_{XYZ} , and P_{LMS} be color pixels in the respective spaces RGB, XYZ, and LMS. We have

$$\begin{aligned} P_{LMS} &= M_2 P_{XYZ}, \\ P_{XYZ} &= M_1 P_{RGB}, \\ P_{LMS} &= M_2 M_1 P_{RGB}. \end{aligned} \tag{10}$$

Distortion properties discussed earlier can now be transferred through the space thanks to the conversion formulas. In the next subsection, we detail the direction vector extraction process.

4.2 Direction vector extraction

In Sect. 2, we saw that the choice of a direction vector has a huge impact on watermark invisibility. When we choose a fixed direction for every color at fixed numerical distortions, we see strong psychovisual distortions in some areas of the image. However, the invisibility is significantly improved when the best direction vector is chosen for every color pixel. Since we suppose that psychovisual distortions are the same for all elements in ψ , we can choose the furthest point from P denoted by P_f to benefit from the maximum numerical distortions allowed at embedding to improve the robustness. Direction vector u_p is defined so that $u_p = \overrightarrow{PP_f}$.

We denote by \mathcal{E} the set of points belonging to the perception volume associated with P . More precisely, \mathcal{E} is the set of points in the RGB space associated with a constant discrimination. Then we have the furthest point P_f from P :

$$P_f = \max_{P' \in \mathcal{E}} \|P - P'\|_2. \tag{11}$$

A direct method of extracting an optimal direction with the based Jacobian estimation is to use the fact that the principal axes of the ellipses can be calculated from eigenvectors of the metric, and in this case the eigenvector associated with the lowest eigenvalue of the metric is u_p . To use this model, constants L_0 , M_0 , and S_0 and gains α_L , α_M , and α_S should be given. We use the parameter calculated in [2] for the MacAdam ellipses [14]: ($\alpha_L^2 = 1665, \alpha_M^2 = 1665, \alpha_S^2 = 226$), ($L_0 = 66, M_0 = 33, S_0 = 0.16$).

Finally, we deduce from equation (9) that the model is parametric, and to select the lowest eigenvalue (the value $\min[\frac{\alpha_L^2 L_0^2}{(L+L_0)^4}, \frac{\alpha_M^2 M_0^2}{(M+M_0)^4}, \frac{\alpha_S^2 S_0^2}{(S+S_0)^4}]$) and to give the expression of the direction u_p for each color, the calculation shall be based on the parameters $\{x_0, \alpha, M_{RGB \rightarrow LMS}\}$ and the studied pixel $P = (R, G, B)$.

In the last step, it is necessary to ensure that the detection step is possible after a modification of the marked image (we consider the case with acceptable damages). In other words, direction vectors computed at embedding and detection must be close enough to ensure an error free detection. The stability of the retrieval direction is linked with the color of the pixel and it also depends on the spatial environment of the pixel for some

attacks (JPEG for example). To be robust, it is necessary that the direction of the watermarking between insertion and detection changes little or not at all. Given the model, this can happen when the color of the pixel is at the border of one of the three direction cases. In this case the color would have to be changed before watermarking to move away from this border. However, we see in the results of Sect. 5 that the robustness is not impacted and that it is not judicious to make the algorithm more complex.

5 Psychovisual algorithm for color images

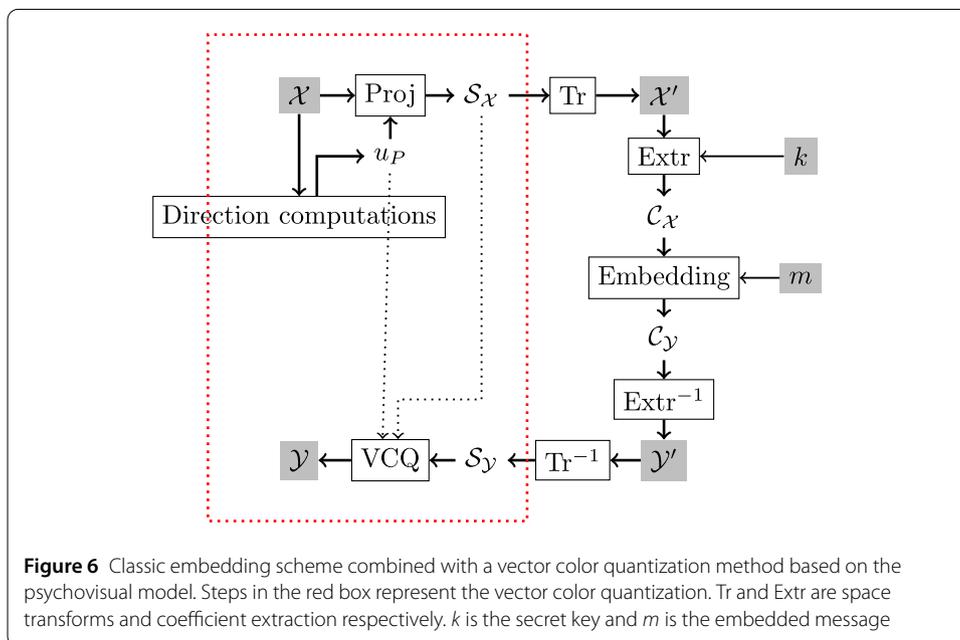
In this section, we introduce an embedding algorithm which improves the psychovisual quality of a marked image. Compared to the classic embedding scheme for grayscale images, we propose to add a step to adapt the embedding process for color images (Fig. 6).

Before choosing the quantization space, a direction vector u_p is computed using the psychovisual model for every color pixel P of host image \mathcal{X} . A grayscale image $\mathcal{S}_\mathcal{X}$ is obtained by computing a scalar product for every P (Proj step in Fig. 6):

$$s = \langle P, u_p \rangle. \tag{12}$$

By considering the knowledge of computed direction vectors, we obtain a bijective relation between \mathcal{X} and $\mathcal{S}_\mathcal{X}$. In Fig. 7, we can see examples of grayscale images $\mathcal{S}_\mathcal{X}$ obtained from \mathcal{X} .

The next step consists in choosing the quantization space (transform function denoted by Tr) of $\mathcal{S}_\mathcal{X}$ for coefficient extraction and embedding. For example, it is possible to apply directly the embedding process on the pixel value (the spatial domain) or to compute a discrete wavelet transform (DWT) of the image and apply the embedding process on the wavelet coefficient. In general, an image representation is chosen according to its properties and the robustness against some attacks (for example, it is well known that the wavelet domain is adapted to increase the robustness against compression).



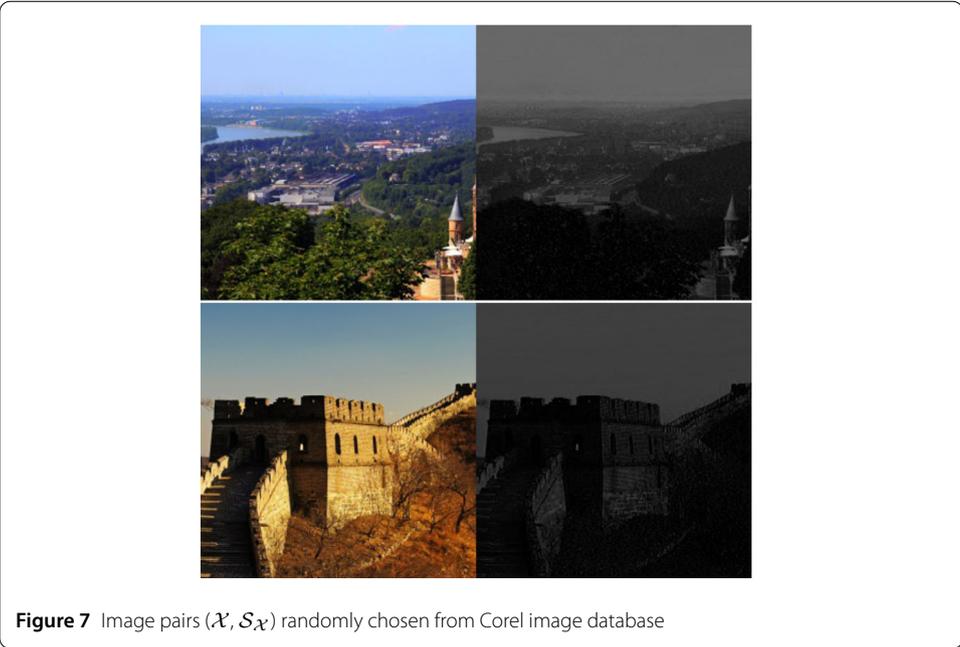


Figure 7 Image pairs $(\mathcal{X}, \mathcal{S}_{\mathcal{X}})$ randomly chosen from Corel image database

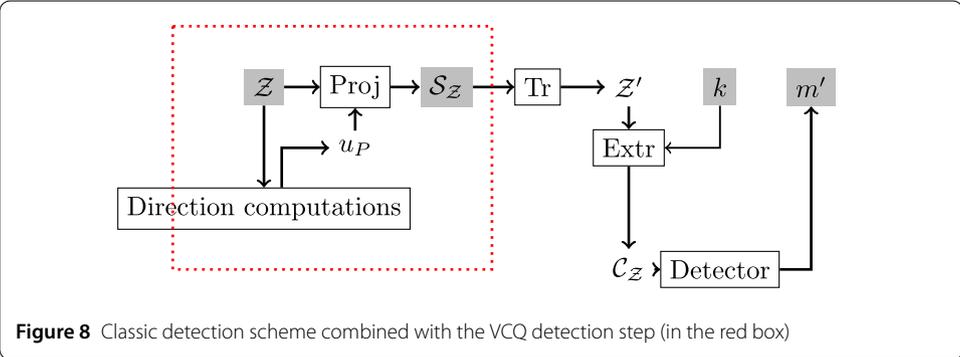


Figure 8 Classic detection scheme combined with the VCQ detection step (in the red box)

Then, coefficients are extracted using the function *Extr*. In our work, we used a random selection of coefficients in the scalar image $\mathcal{S}_{\mathcal{X}}$. The extraction step is done in the watermarking space only accessible with the secret key k (only known to the embedder and the receiver).

Once coefficients $\mathcal{C}_{\mathcal{X}}$ are selected, they are modified by an embedding method, which is the lattice QIM in our case, according to the message m . The inverse transformation is performed on the modified scalar image $\mathcal{S}_{\mathcal{Y}}$. At last, using the following equation (VCQ), modified colors are computed, and we obtain the watermarked image \mathcal{Y}' composed of modified pixels P'_w :

$$P'_w = P + (s'_w - s)u_P \tag{13}$$

with s'_w being the result of LQIM quantization when s is extracted.

Detection step is illustrated in Fig. 8. Using key k , coefficients $\mathcal{C}_{\mathcal{Z}}$ are extracted from the scalar image $\mathcal{S}_{\mathcal{Z}}$ obtained from the received image \mathcal{Z} thanks to equation (12). Then, a message estimation m' is obtained with LQIM detector. If the attack is weak enough, then we have $m = m'$.

As we said before, the computation of direction vectors is very important for the detection process. Supposing colors are reasonably modified, the ability to successfully detect a message is guaranteed by the slow variation of direction vectors from one color to another. In our experiments, we show that this hypothesis is verified, and we obtain robustness improvements.

We have proposed a color quantization algorithm using a psychovisual model of the HVS to watermark color images. The lattice QIM method was chosen for our experiments but it is also possible to adapt other methods. The classic grayscale embedding and detection schemes can be adapted for color images by adding an extra layer containing the vector color quantization method proposed in Sect. 2. In the next section, we propose to evaluate invisibility and robustness performances.

6 Experiments

For our tests, we used images from the Corel database where 1,000 images are randomly chosen among the 10,000 available. Image coefficients are randomly chosen. To evaluate the watermark invisibility, parameters such as the message length or the lattice dimension L are computed so that we obtain an adequate image quality and embedding rate.

6.1 Psychovisual invisibility

In this subsection, we first propose to evaluate the psychovisual invisibility of several marked images with two methods denoted by GA and AA defined as follows:

- GA: color quantization with a constant direction vector $u = (1, 1, 1)$. u is the best choice in terms of invisibility for a constant direction vector approach according to our experiments.
- AA: color quantization using an adaptive axis u_p in function of the color.

Both watermarking methods are modified versions of the LQIM method previously introduced in this article. For every method, direction vector norms are fixed to 0.5. The goal of this experiment is to show that the AA method has better psychovisual invisibility than the GA method.

In those experiments, we evaluate the numerical distortion with the help of a signal to noise ratio between the host image \mathcal{X} and \mathcal{Y} also known as DWR (document to watermark ratio).

In Fig. 9, we show color image examples marked with the GA method. For each image, we easily see that colors are saturating towards gray. Then it is easier to perceive color differences with a naked eye. This can be explained by the random selection of quantized coefficients, which adds a salt and pepper noise texture to the image. Then the HVS will focus its attention on the noise by visual saliency.

With the AA method (Fig. 10), the quantization noise makes colors saturate toward blue and green, i.e., it adapts itself in function of the color so it is harder to perceive it. Compared to pixels in homogeneous image regions, detecting a color difference is more difficult.

At equal numerical distortion level, marked images with both methods are not affected by the same psychovisual quantization noise. We observe in Figs. 9 and 10 that it is more difficult to perceive the psychovisual quantization noise with the AA method compared to the GA method.

For a better visual observation of the embedding noise, we have chosen a small image size of 60×60 pixels with a strong embedding rate of 0.5. In practice, images are larger: for

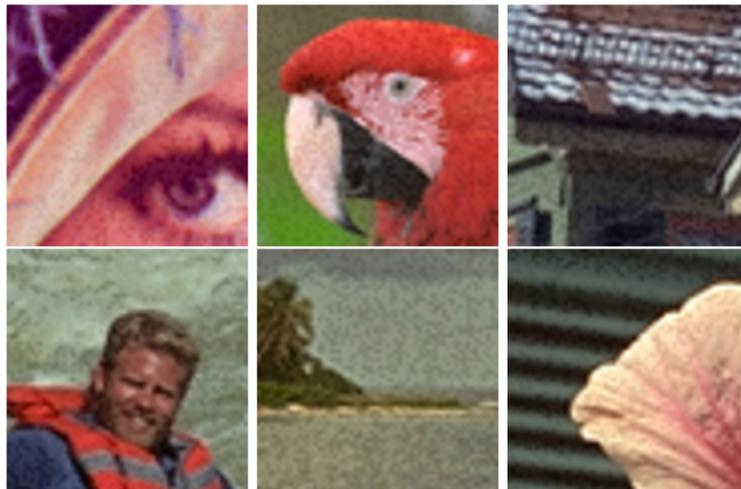


Figure 9 Cropped color images (Lenna and Kodak database images with size 60×60) with the GA method. Average DWR is $\simeq -5.5$ dB and ER = 0.5



Figure 10 Cropped color images (Lenna and Kodak database images with size 60×60) with the AA method. Average DWR is $\simeq -5.5$ dB and ER = 0.5

example, HD image size is 1080×1920 and the embedding rate is smaller giving a more imperceptible quantization noise.

Secondly, to validate these visual observations, we have repeated this comparison experiment with 15 human test subjects. 24 images from the Kodak database with the size equal to 768×512 have been selected. And, for each image, a pair of watermarked images with both methods are proposed to every test subject. Each observer has to decide which image (marked with GA and AA methods) is less noisy. Parameters are set to DWR = 20 dB, ER = 1/2 for each image.

Obtained results are given in Table 1. We can see that only 4% of images were described as less noisy in average with the GA method. Based on those results, we confirm the pre-

Table 1 Psychovisual experiments with marked image comparisons. This table shows the percentage of images described as less damaged for each method

Methods	GA method	AA method
Average votes	4% \pm 3%	96% \pm 3%

Table 2 Table of maximum quantization steps in average before a mark becomes visible

Δ /DWR	GA	AA
SP	16/29.4	24/38.8

vious observation: watermarked images with the AA method provide better psychovisual invisibility than those with the GA method at the same numerical distortion level.

Now, we propose to analyze the robustness performances of both approaches at equal psychovisual distortion level.

6.2 Robustness

6.2.1 Protocol

For our measures, we propose two embedding spaces for a message length of $n = 128$ bits, a lattice dimension $L = 2$, and the spatial domain is used (we modify directly the pixel value).

Before presenting the robustness results, we explain how to fairly compare GA and AA methods. First, we maintain satisfying psychovisual image quality/watermark invisibility for both methods. More precisely, we have experimentally determined the maximum average quantization step Δ before a mark becomes visible for each method and embedding domain (see Table 2). Notice that actually the threshold associated with the fact “the mark becomes visible” is set by a very simple method: the same image with different values of quantization step is proposed to some human observers, and each observer has to decide which image (marked with GA or AA methods) is degraded and which is not.

In Table 2, we can see that we can use a greater quantization step with the AA method (more robustness allowed) compared to the GA method. We have an improvement of 5 dB from GA to AA. The AA method allows one to be more robust against several attacks.

We have measured the robustness performance of both methods using the quantization steps defined in Table 2: two average binary error rates (based on 100 embedding repetitions with random payloads and extractions) are computed in function of different attack strength parameters (± 0.01). We propose now to analyze each result.

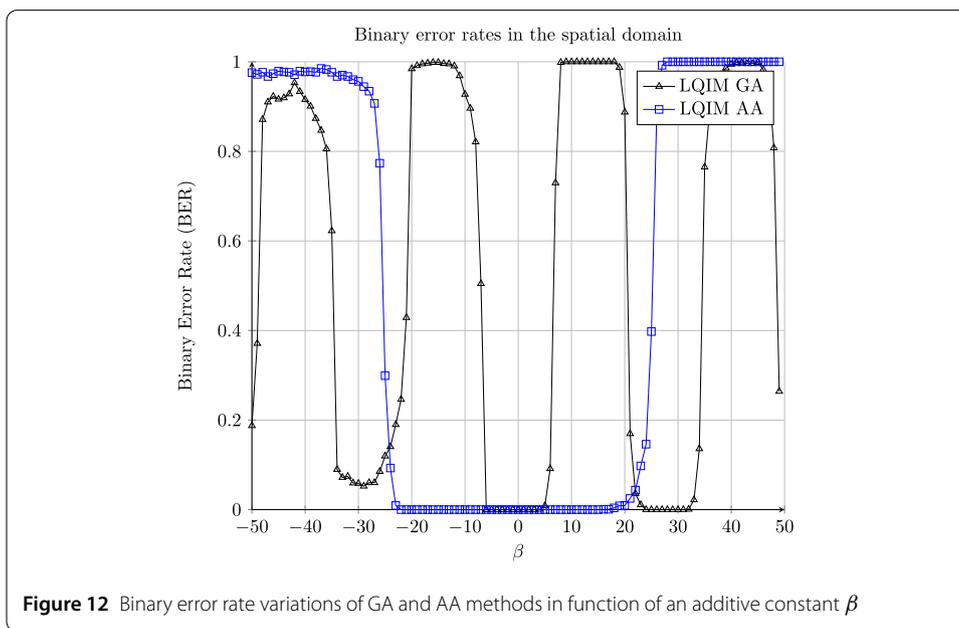
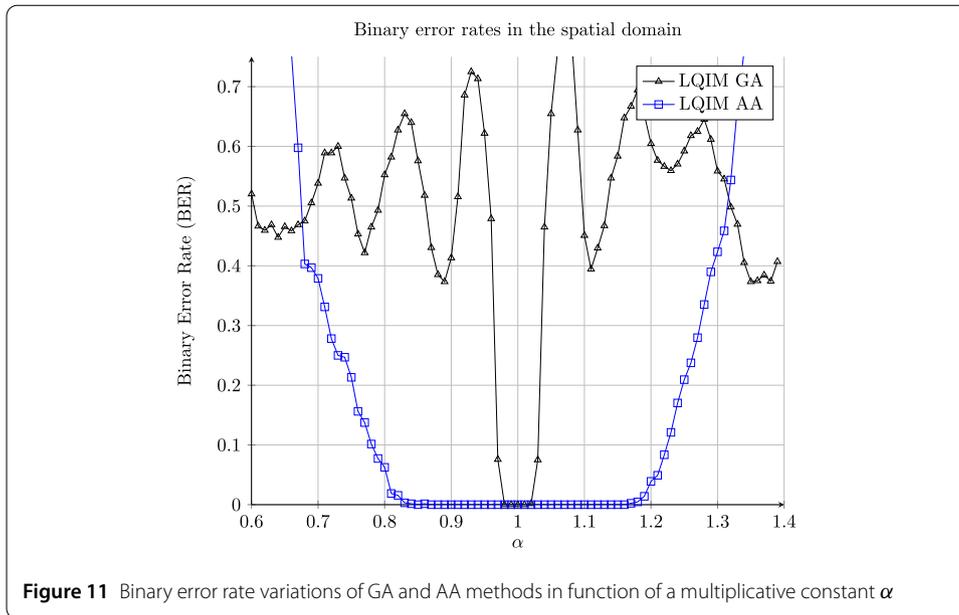
6.2.2 Contrast modification

We define the contrast modification by a multiplicative constant change of parameter α . Let x be a pixel value. We have

$$y = \alpha x \quad (14)$$

with y being the modified version of pixel x . Errors are stronger when the difference between α and 1 is larger. This image processing also depends on the image content.

Robustness results against this attack are illustrated in Fig. 11. There are more errors when α is not close to 1 and tends towards 0.5. We can see that LQIM AA curve is at 0 for a longer interval around value $\alpha = 1$ ($\alpha \in \{0.81, \dots, 1.19\}$) than LQIM GA curve ($\alpha \in$



{0.98, ..., 1.02}). In fact, LQIM AA is computed with a larger quantization step than LQIM GA.

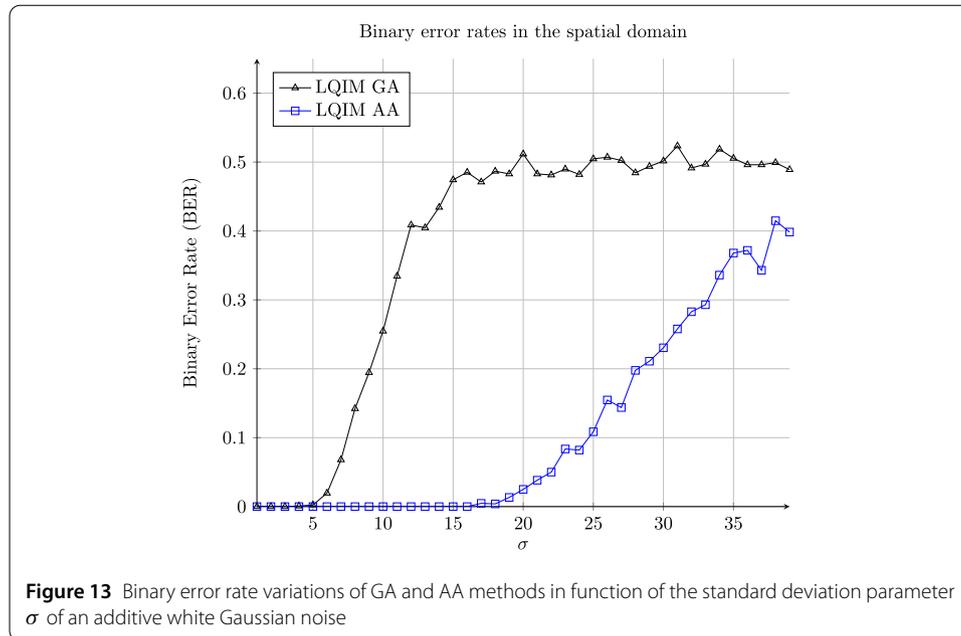
6.2.3 Luminance modification

We model this attack by the following equation:

$$y = x + \beta \times (1, \dots, 1) \tag{15}$$

with y being the result of the luminance modification of a pixel value x .

Results are shown in Fig. 12. In the spatial domain, error curves oscillate between 0 and 1 like a step function. This image processing is less dependent on the image content than



the contrast modification. Again, LQIM AA is obtained with a larger quantization step than LQIM GA which explains why LQIM AA is at 0 for a larger interval.

6.2.4 Additive white Gaussian noise

We also made robustness experiments against additive white Gaussian noise (Fig. 13). This image processing modifies every pixel of an image with a random value following a Gaussian distribution and is parametrized by parameters $\mu = 0$ and σ . We can see that error rates are closer to BER = 0 for LQIM AA than for LQIM GA.

6.2.5 Modifications of hue, saturation, and value

Finally, we propose to show results about HVS color space component modifications which produce color distortions and how this affects our watermarking methods. Watermarked images are converted in HVS space and color components h , s , and v are modified using the image library OpenCV (component variations between 0 and 255). Obtained results are illustrated in Fig. 14.

In this figure, we can see that LQIM AA SP curves are closer to BER = 0 than LQIM GA SP curves.

Since the results are better with the adaptive strategy, we can deduce that the system fails to decode the message correctly when the pixel modification due to the attack becomes too large, although with the adaptive approach we were able to define cells for each larger word (for constant visual degradation). In this case the decoding of the bit becomes random.

To conclude, we propose to discuss a generalization that can be made for your process by using a wavelet decomposition. This first step changes the representation of an image in terms of variations, i.e., in frequency bands, and explains why both methods are robust in this embedding domain. Moreover, we can have better robustness results using this image decomposition compared to the spatial domain. To illustrate this fact, we propose an illustration with the JPEG attack. We choose to embed a watermark at level 2 in the diagonal details plane to obtain better detection performances. Compared to spatial do-

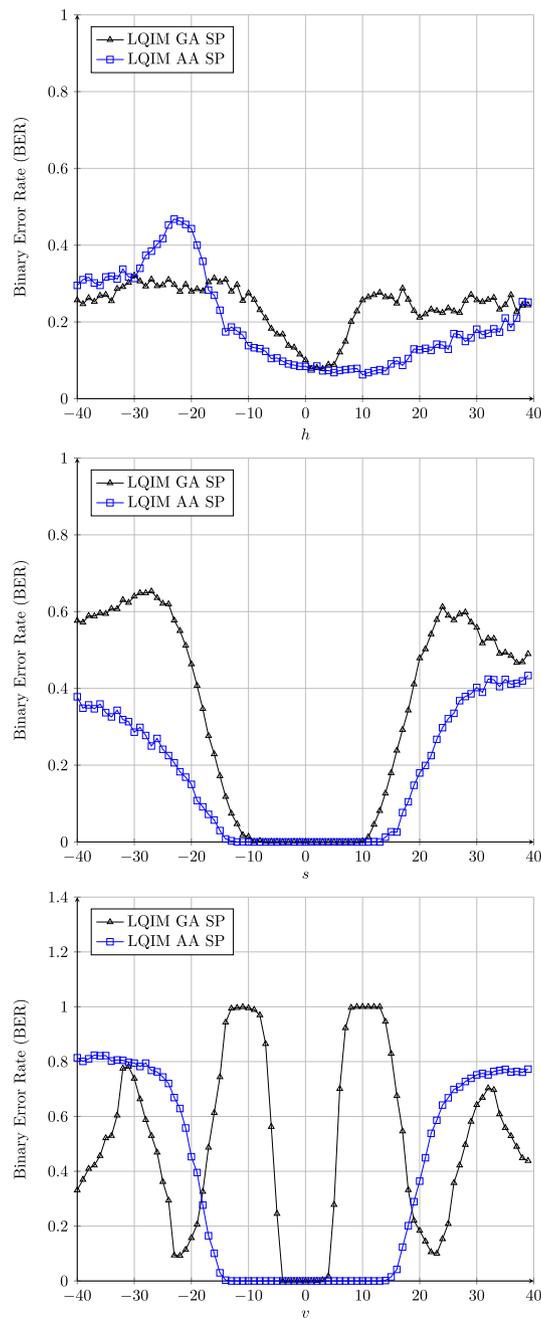
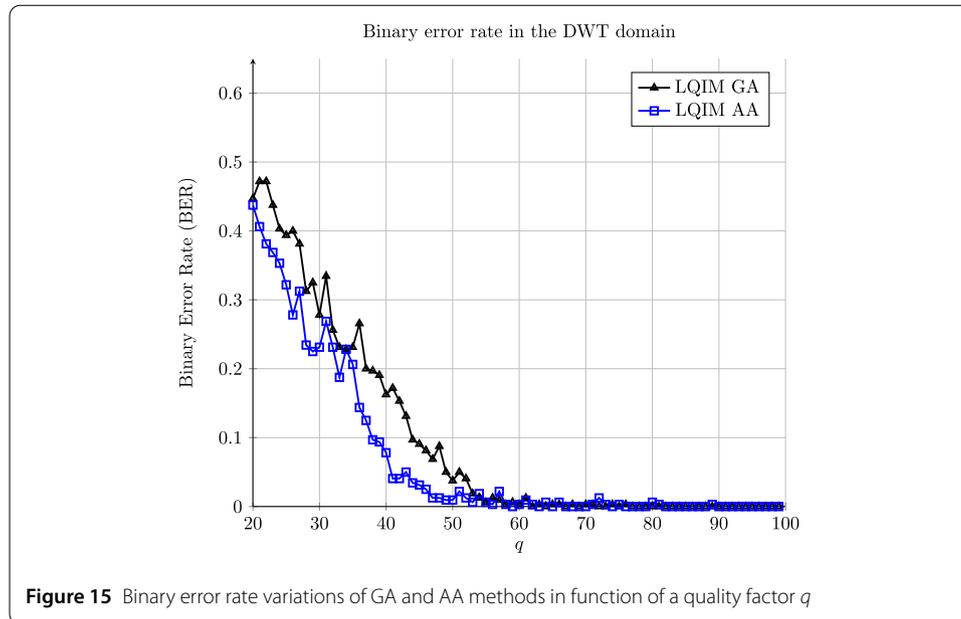


Figure 14 Error rate variations of GA and AA methods in spatial domains for the modifications of hue, saturation, and value

main, DWT coefficients are floating numbers, and it is important to note that this does not change the embedding algorithm.

6.2.6 JPEG compression

JPEG compression is a lossy compression algorithm for color images. It allows one to reduce the amount of data to encode an image in function of a quality factor. A high quality factor means that the image quality is close to its original version and a low factor means



that the compressed image is strongly damaged but requires much less data. For color images, JPEG compression deletes in majority the color information.

In the spatial domain, error curve LQIM AA is close to 0.5: the spatial domain is not adapted.

However, we observe a different result in Fig. 15 in the DWT domain. Error rates are rapidly reaching 0 with $q = 95$ for both methods. Modifying wavelet coefficients have mathematical properties that allow our methods to survive JPEG compression. Moreover, LQIM AA curve is reaching 0 faster than LQIM GA: we have a robustness improvement by working with wavelet coefficients.

7 Conclusion

In this article, we proposed a color vector quantization algorithm for color image watermarking. It allows embedding watermark information along a direction vector in the RGB color space. Compared to an approach where the vector direction is fixed, our approach uses an adaptive direction calculated for every pixel in the image and provides better invisibility of the mark.

The adaptive method uses a Riemannian model of the representation of light by photoreceptors including their dynamical responses. Since the model is parametric, we use real experiment (MacAdam ellipses) to compute parameters. The model is then able to guess what is the direction of less discrimination around a color.

For a watermarking application, we used this model to extract direction vectors in order to minimize the perception of color differences at embedding and proposed to adapt the well-known lattice QIM method for color images.

Our experiments showed good improvements for the adaptive method in terms of psychovisual invisibility, compared to the method which uses a constant approach, at equal numerical distortions. After adjusting the invisibility/robustness tradeoff (comparing at equal psychovisual distortions), we compared the robustness of both methods against several attacks and concluded that the adaptive method is more robust than the constant

method. Moreover, results were improved on both sides by using the wavelet transform as an embedding domain for particular attacks (like compression).

In this first study of watermarking invisibility using HVS model, we use a simple model taking into consideration only the encoding by photoreceptors and their dynamics. A natural extension of the model would be the use of a three-layer model accounting for the other cells (bipolar, ganglion) in the retina that also regulates light information. Indeed, it has been shown that the three-layer model allows a better account for the color discrimination data [3]. Another extension would be to use a pseudo-Riemannian model of color vision that preferentially describes color attribute such as hue, saturation, and brightness.

Acknowledgements

PC and DA were funded by the project INFINITI of the Mission pour L'Interdisciplinarité of CNRS and the project 80primes of the CNRS. DA would like to thank Jeanny Hérault for his proposal of the representation Fig. 5.

Funding

This work was supported by the CNRS.

Abbreviations

HVS, Human visual system; LQIM, Lattice quantization index modulation.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Authors' contributions

PL and PC did the watermarking process and the experiment, DA did the discrimination model, PL, PC, and DA wrote the paper. All authors read and approved the final manuscript.

Author details

¹Laboratoire XLIM CNRS UMR 7252, Université de Poitiers, Bât. H1 - SP2MI, 11 Bd Marie et Pierre Curie, TSA 41123, 86073 Poitiers Cedex 9, France. ²Laboratoire LPNC CNRS UMR5105, Université Grenoble Alpes, Bâtiment Michel Dubois, 1251 Avenue Centrale, 38400 Saint-Martin-d'Hères, France.

Endnote

^a White point of equal energy E .

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 24 September 2019 Accepted: 13 October 2020 Published online: 23 October 2020

References

1. Abadpour A, Kasaei S. Color PCA eigenimages and their application to compression and watermarking. *Image Vis Comput.* 2008;26(7):878–90.
2. Alleysson D. Le traitement du signal chromatique dans la rétine: un modèle de base pour la perception humaine des couleurs. PhD thesis. Université Joseph Fourier-Grenoble 1 Sciences et Géographie; 1999.
3. Alleysson D, Hérault J. Variability in color discrimination data explained by a generic model with nonlinear and adaptive processing. *Color Res Appl.* 2001;26(S1):S225–9.
4. Alleysson D, Méary D. Neurogeometry of color vision. *J Physiol (Paris).* 2012;106(5–6):284–96.
5. Chareyron G, Macq B, Tremeau A. Watermarking of color images based on segmentation of the XYZ color space. In: *Conference on colour in graphics, imaging, and vision.* vol. 2004. 2004. p. 178–82.
6. Chareyron G, Trémeau A. Color images watermarking based on minimization of color differences. In: *Int. workshop on multimedia content, representation, classification and security, MRCS'2006.* 2006.
7. Chen B, Wornell GW. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. In: *IEEE trans. On information theory.* vol. 47. 1999. p. 1423–43.
8. Coltuc D, Bolon P. Robust watermarking by histogram specification. In: *Image processing. ICIP 99. Proceedings. 1999 international conference on.* vol. 2. New York: IEEE; 1999. p. 236–9.

9. Dadkhah S, Abd Manaf A, Hori Y, Hassanien AE, Sadeghi S. An effective SVD-based image tampering detection and self-recovery using active watermarking. *Signal Process Image Commun.* 2014;29(10):1197–210.
10. Kim H-S, Lee H-K, Lee H-Y, Ha Y-H. Digital watermarking based on color differences. vol. 4314. 2001. p. 10–17.
11. Koenderink JJ, Van de Grind WA, Bouman MA. Models of retinal signal processing at high luminances. *Kybernetik.* 1970;6(6):227–37.
12. Kutter M, Jordan FD, Bossen F. Digital signature of color images using amplitude modulation. vol. 3022. 1997. p. 518–526.
13. Li J, Lin Q, Yu C, Ren X, Li P. A QDCT- and SVD-based color image watermarking scheme using an optimized encrypted binary computer-generated hologram. *Soft Comput.* 2016. 1–19.
14. MacAdam DL. Visual sensitivities to color differences in daylight*. *J Opt Soc Am.* 1942;32(5):247–74.
15. Moulin P, Koetter R. Data-hiding codes. *Proc IEEE.* 2005;93(12):2083–126.
16. Parisis A, Carre P, Fernandez-Maloigne C, Laurent N. Color image watermarking with adaptive strength of insertion. In: *Acoustics, speech, and signal processing, 2004. Proceedings. (ICASSP'04). IEEE international conference on.* vol. 3. New York: IEEE; 2004. p. iii–85.
17. Petitot J. The neurogeometry of pinwheels as a sub-Riemannian contact structure. *J Physiol (Paris).* 2003;97(2):265–309.
18. Petitot J. *Neurogéométrie de la vision: modeles mathematiques et physiques des architectures fonctionnelles.* Editions Ecole Polytechnique; 2008.
19. Voyatzis G, Pitas I. Digital image watermarking using mixing systems. *Comput Graph.* 1998;22:405–16.
20. Yu P-T, Tsai H-H, Lin J-S. Digital watermarking based on neural networks for color images. *Signal Process.* 2001;81(3):663–71.
21. Zrenner E, Stett A, Weiss S et al. Can microphotodiode arrays replace degenerated retinal photoreceptors to restore vision. *Vis Res.* 1999;39(15):2555–67.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
