Multimedia Systems (2003)
Digital Object Identifier (DOI) 10.1007/s00530-003-0083-6

Multimedia Systems
© Springer-Verlag 2003

# Robust image authentication using content based compression

**Ee-Chien Chang**[1], **Mohan S. Kankanhalli**[1], **Xin Guan**[2], **Zhiyong Huang**[1], **Yinghui Wu**[1]

[1] Department of Computer Science, Computational Science, National University of Singapore, ♣
[2] Department of Computational Science, National University of Singapore, ♣

Received: 11 February 2002 / Accepted: 10 August 2002

**Abstract.** Image authentication is becoming very important for certifying data integrity. A key issue in image authentication is the design of a compact signature that contains sufficient information to detect illegal tampering yet is robust under allowable manipulations. In this paper, we recognize that most permissible operations on images are global distortions like low-pass filtering and JPEG compression, whereas illegal data manipulations tend to be localized distortions. To exploit this observation, we propose an image authentication scheme where the signature is the result of an extremely low-bit-rate content-based compression. The content-based compression is guided by a space-variant weighting function whose values are higher in the more important and sensitive region. This spatially dependent weighting function determines a weighted norm that is particularly sensitive to the localized distortions induced by illegal tampering. It also gives a better compactness compared to the usual compression schemes that treat every spatial region as being equally important. In our implementation, the weighting function is a multifovea weighted function that resembles the biological foveated vision system. The foveae are salient points determined in the scale-space representation of the image. The desirable properties of multifovea weighted function in the wavelet domains fit nicely into our scheme. We have implemented our technique and tested its robustness and sensitivity for several manipulations.

**Key words:** Digital signature – Security – Watermarking – Signal processing – Wavelets

## 1. Introduction

With the digitization of media, the question of ownership, copyrights, and integrity has become an important concern. There has been intensive research activity in the area of image watermarking [7,11]. In fact, there is a lot of effort in developing both robust and fragile watermarks. Robust watermarking is geared more toward ownership and copyright concerns, while fragile watermarks are designed for image authentication. Our paper is also concerned with the problem of authenticating image data, i.e., verification of the *genuineness* of the data set. Our goal is to develop a reliable image authentication technique that can be incorporated into all types of digital cameras and scanners in order to make them *trustworthy* [9]. This is already becoming very critical in certain areas like medical imaging. For example, given a medical image that depicts a critical condition like a tumor, we do not want the patient to fraudulently alter the image so that the tumor is removed and thus misrepresent the medical condition to an insurance company. Similarly, we would not like an unscrupulous medical institution to alter the data set in order to introduce artifacts that represent some abnormality and make a patient go through unnecessary expensive medical procedures. In such situations, preserving and checking the veracity of a data set assumes tremendous importance.

We believe that this problem can be addressed by use of a *content-based digital signature* that is robust yet effective. So the idea is that at the time of image creation a content-based digital signature is simultaneously created in the camera/scanner itself. For all further authenticity checks, this image can be verified against its digital signature. If there is a mismatch, then the data are considered unreliable and should not be used. It is clear that traditional message authentication techniques like hashing-based digital signatures or cryptographic authentication [23] cannot be used because of their inherent fragility. We do not propose to use watermarking since in many applications, such as medical imaging, the distortion of pixel values are not allowed due to legal implications. Therefore, a separate digital signature is required for verifying data integrity. In cases where perturbation of pixel values is allowed, our digital signature can be embedded into the image using a robust invisible watermarking scheme. Thus our method would be useful for fragile watermarking techniques as well.

Current image authentication schemes loosely fall into two groups. In the first group, the highly compressed image or the quantized image serves as the signature. Usually the image is divided into equally sized subregions and the signature is a collection of descriptions of all subregions. For example, in [19], the signature is the intensity histogram of each image

*Correspondence to*: M.S. Kankanhalli newline (e-mail: mohan@comp.nus.edu.sg)

block. In [14], the invariant relations between the coefficients of two randomly selected DCT blocks are used. Because we can treat the signature as a collection of individual subsignatures, analysis is usually easier for this approach. However, the drawback is that the signature size is usually large.

In the second group, the content-based signature consists of features detected by a global process. For example, the signature is a set of corners, lines, or other representations obtained from highly nonlinear processes. In [3], the features are the positions of a set of feature points. The signature is very compact, but the analysis is usually difficult. To illustrate, consider a signature that is the set of corners. A small perturbation in the spatial domain might cause significant changes in the configuration of corners.

By comparing the permissible operations and the malefic tamperings on digital images, we make a critical observation that most permissible operations are global operations like resizing, low-pass filtering, and JPEG compression, whereas malefic tamperings are mostly localized operations like cropping of an important region and alteration of local features. Based on this observation, we propose the use of a content-based digital signature that is robust yet effective. The core idea is that a content-based weighting function is obtained using a feature-detection routine. This weighting function is then used to guide a compression process. The highly compressed image, together with the description of the weighting function, forms the signature. This signature is further encrypted for security. For all authenticity checks, the image can be verified against this signature.

In Sect. 2, we first give the desirable properties of an authentication scheme and discuss the setting in which our proposed scheme fits (Sect. 2.1). Then we give the outline of our scheme (Sect. 2.2). From Sect. 3.1 to Sect. 4, we discuss the various components of the method. In Sect. 5, we present the experimental results. Finally, we conclude the paper in Sect. 6.

## 2. Background

### *2.1. Desirable characteristics*

As a design goal, it is important to list the ideal desirable characteristics of a robust content-based digital signature for images. The term *content-based* refers to the fact that important features of the data (whose integrity we are interested in certifying) should be somehow incorporated into the digital signature, the rationale being that if some important content feature is deleted/modified/added, then the digital signature should not match the doctored data set. The term *robust* refers to the fact that any manipulation that does not change the significant features should not affect the veracity of the signature. For such benign operations the digital signature should indeed authenticate the data set. Common types of operations on images are scaling, thresholding, cropping, cut-and-replace a subregion, filtering, addition/removal of noise, and affine transformations. As long as these operations do not change the content features, they are considered benign.

We now list the desirable properties of techniques for robust content-based authentication of images. An authentication technique can be considered effective if it satisfies the following requirements:

1. *Sensitivity*: the authenticator should be sensitive to any malefic operation such as cropping of a significant feature of the image.
2. *Robustness*: the authenticator should be robust against benign operations on the image.
3. *Security*: the technique should not be easy to forge or manipulate. An important property is that the authentication bits should be relatively easy to generate, but inferring the image from the authentication bits should be impossible. This is also known as the one-way function property in cryptography [23].
4. *Identification of manipulated regions*: the authenticator should be able to detect the location of altered regions (if they are localized) and certify other regions as authentic. This is akin to error detection in coding theory.
5. *Recovery capability*: the authenticator should have the ability to recover the lost content (perhaps approximately) in the manipulated regions. This is similar to the error-correction capability of some codes.
6. *Compactness*: the number of authentication bits generated by a technique should be as small as possible while satisfying the other properties.

While the above are the ideal desired characteristics, practical authentication techniques must be designed with a view to minimizing false positives (incorrectly flaging an unaltered image as fake) and true negatives (authenticating a fake image).

### *2.2. Overview of the technique*

We will now provide an overall description of the method for generating the robust content-based digital signature and the method for authenticating an image using this digital signature. In the digital signature creation process, a set of feature points is first extracted. Then a weighting function is constructed from the extracted feature points. The original image data are then lossily compressed under a weighted norm determined by the weighting function. The highly compressed data $S$, together with the description $W$ of the weighting function, forms the signature $(S, W)$. This signature can then be further encrypted. Checking the authenticity involves computing the distortion between the compressed description of the original image with the current image under the weighted norm. We now briefly describe the individual steps.

*Feature points and the weighting function.* Most imaging systems use a norm (usually the Euclidean two-norm) to measure their performance. In real-life data, it is usually possible, either through user interaction or automated detection, to determine regions that are more interesting for the application at hand. For example, through feature detection we can find the significant points in an image. In such cases, a space-variant weighted norm is more appropriate compared to the two-norm, which treats each pixel uniformly. The weighted norm $\|\cdot\|_w$ for the image $I(x, y)$ with a weighting function $w$ is given by:

$$\|I\|_w^2 = \sum_{x,y} w(x,y) I(x,y)^2,$$

where $w(\cdot,\cdot)$ is the weighting function.

The weighting function indicates how the salient information is spatially distributed over the image. In our scheme, the role of the weighting function is to guide the content-based compression and to determine how the distortion (difference) between two images is to be measured.

Because the description of the weighting function is a part of the signature, it must be as concise as possible. Taking this requirement into consideration, our approach is to use a set of feature points that implicitly describe a weighting function. We have found that the multiscale salient points, together with the multifovea weighted function, fit well into our framework. Figure 1b shows the contour plot of a weighting function. This weighting function is determined from the salient points depicted in Fig. 1a. Details of this construction will be discussed in Sect. 3.

*Content-based Compression.* A lossy compression amounts to selectively discarding insignificant information. Normally compression schemes use a two-norm as a guide in the compression process. Unlike the two-norm, which treats each pixel with equal importance, the space-variant weighted norm places differing emphases on different regions. Thus, a content-based compression can be achieved by selecting a weighting function whose weight is higher in the interesting regions. In our scheme, we use the multifovea weighted function to do the compression. The resultant highly compressed image serves as the second part $S$ of the content-based signature.

*Encryption.* For additional security, public-key cryptography [23] is utilized to encrypt the signature derived in the previous step. This also provides the one-way function property. Basically, the secret key of the owner of the image is used to encrypt the signature obtained. For the purpose of authentication, the public key of the owner can be used to decrypt this information and the signature thus recovered. Since this step is well understood, we will not discuss it further in this paper.

*Authentication procedure.* For authenticating a particular image, the following steps need to be performed:

the highly compressed image $\widetilde{I_0}$ is first recovered from the signature. It is then compared with the image in question $I$ (whose integrity we would like to verify). The distortion between $\widetilde{I_0}$ and $I$ is computed. If this value falls below a certain threshold, then the image $I$ is declared to be authentic; otherwise, it is considered to be untrustworthy. The matching process is guided by the same weighting function whose description is available from the signature.

## 3. Creation of the content-based digital signature

In the next few sections, we describe in detail the various steps in the signature creation.

### *3.1. Scale-space salient points*

The salient points of an image $I$ are local maxima in the three-dimensional scale-space representation [15]. Figure 1a gives an example.

The scale-space representation $P(\mathbf{x},\sigma)$ for an image $I$ is determined by a kernel $g : \mathbf{R}^2 \to \mathbf{R}$. The first variable $\mathbf{x}$ has one-to-one correspondence with a spatial location $\mathbf{x}$ in the image. The second variable $\sigma$ is known as the *scale*. The original image $I$ corresponds to the scale at $\sigma = 0$, that is, $P(\cdot, 0)$ is the image $I$. The value of $P(\mathbf{x},\sigma)$ is the inner product of the image with the shifted kernel centered at $\mathbf{x}$ and dilated by $\sigma$. In other words, $P(\cdot,\sigma)$ is the convolution of the image with the dilated kernel $g_\sigma(\cdot) = (1/\sigma^2)g(\cdot/\sigma)$. For example, if $g$ is Gaussian, then

$$\begin{aligned} P(\mathbf{x},\sigma) &= \{p \star g_\sigma\}(\mathbf{x}) \\ &= \frac{1}{\sigma^2 2\pi}\int p(\mathbf{x}-\mathbf{y})\mathbf{e}^{-\|\mathbf{y}\|_2/(2\sigma^2)}\mathbf{dy} \end{aligned}$$

The local maxima (over space and scale) of $P$ are the *salient points*. The value of $P$ at a salient point is the *strength* of the salient point. In general, it is not necessary to take the Gaussian as the kernel. In our implementation, we use $g_2 - g$ as the kernel, where $g$ is Gaussian and $g_2$ is $g$ dilated by a factor of 2, that is, $g_2(\cdot) = (1/4)g(\cdot/2)$. Furthermore, for computational efficiency we use a series of spline functions to approximate $g_\sigma$ for various values of $\sigma$.

Each circle in Fig. 1a indicates a salient point. The radius of the circle corresponds to the scale, and the center is the spatial location of the salient point. Thus, we can view the circle as the region of influence. Not indicated in the figure is the strength of the salient point. Note that a salient point is parameterized by $(\mathbf{x}, s, m)$ where $s$ is its scale, $m$ its strength, and $\mathbf{x}$ its location.

### *3.2. Construction of the weighted norm*

The weighting function for an image is constructed from its salient points. Figure 1b shows the contour plot of a weighting function obtained from the set of salient points in Fig. 1a. In the next two sections, we first give an introduction to wavelet foveation and then describe how to compose a weighting function from a set of salient points.

#### 3.2.1. Wavelet foveation

Our visual system has a space-variant nature whereby the resolution is high at a point (fovea) but falls off as we move toward the periphery [20]. This distribution of resolution provides a fast and simple way of reducing information in the visual field without sacrificing the size of the visual field and the resolution around the fovea. As the biological visual system is highly effective, its space-variant nature has inspired the design of many computer vision systems that resemble the biological foveated vision [1,4,21], video conferencing [2,8], image compression [5], and visualization systems [13].

Figure 2a is a uniform resolution image, whereas Fig. 2b is a *foveated* image (with the center of the right eye as the fovea).

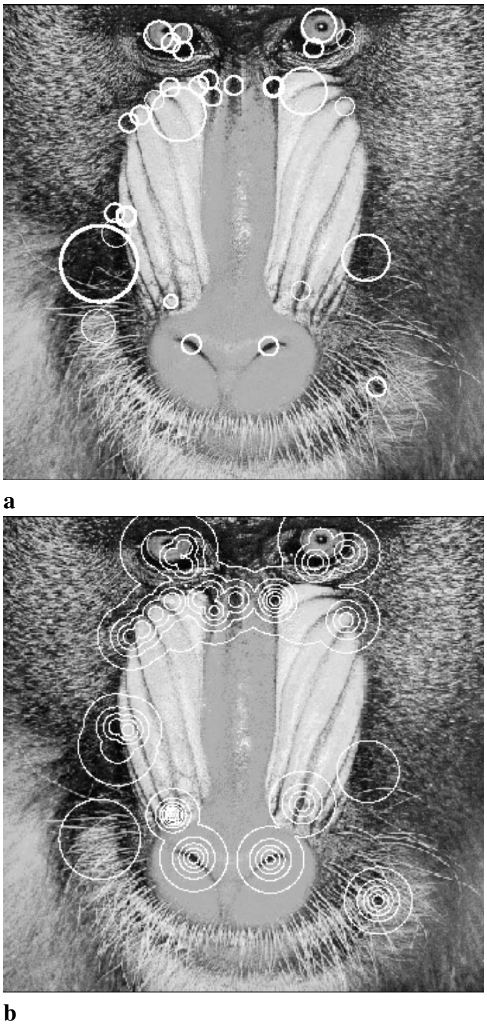


**Fig. 1a,b.** The salient points and weighting function are superimposed onto the image. **a** The top 30 (in strength) salient points, each displayed as a *circle*. **b** The contour plot of the weighting function. This function is a mixture of 30 weighting functions, blended using Eq. 3

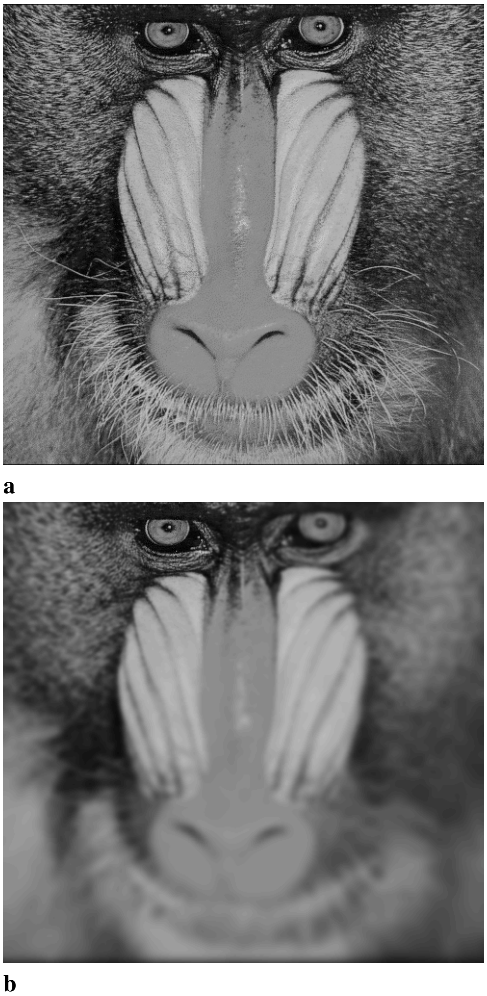


**Fig. 2a,b.** Foveation. **a** The uniform resolution image. **b** The foveated iamge with fovea at the center of the right eye

The foveated image is obtained from the uniform resolution image through a space-variant smoothing process where the width of the smoothing function is small near the fovea but gradually increases toward the periphery. The process of going from a uniform image to a foveated image is known as *foveation*. In this paper, we use the definition and techniques in [5]. The *foveation* of an image $I : \mathbf{R}^2 \to \mathbf{R}$ is determined by a *smoothing function* $g : \mathbf{R}^2 \to \mathbf{R}$ and a *weight function* $w : \mathbf{R}^2 \to \mathbf{R}_{\geq 0}$.

$$\left(T^{\text{fov}}I\right)(\mathbf{x}) := \int_{\mathbf{R}^2} I(\mathbf{t})w(\mathbf{x})g\left(w(\mathbf{x})\|\mathbf{t}-\mathbf{x}\|_2\right)\, d\mathbf{t} \tag{1}$$

The weighting function $w$ depends upon three parameters and takes the form

$$w(\mathbf{x}) = \left(\alpha\|\mathbf{x}-\gamma\|^2+\beta\right)^{-1} \tag{2}$$

We call $\alpha$ the *rate* as it determines how fast the resolution falls off as we go away from the foveal location, $\gamma$ the *foveal location* as it determines the point of highest resolution, and $\beta$ the *foveal resolution* as it determines the resolution at the fovea. Both $\alpha$ and $\beta$ are nonnegative and the smoothing function $g$ is normalized so that $\int_{-\infty}^{\infty} g(\mathbf{x})\, d\mathbf{x} = 1$. In general, we could replace the weighting function by any nonnegative function. This generalization is useful when we are interested in images with multiple foveae. Given two weighting functions $w_1$,$w_2$, the blended weighting function $w_3$ is

$$w_3(\mathbf{x}) = \max\{w_1(\mathbf{x}), w_2(\mathbf{x})\} \tag{3}$$

Foveated images can also be treated as the approximation of an image using a fixed number of bits, using a weighted norm as the underlying measure. This weighted norm can be derived from Eq. 1 and has the form

$$\|I\|_w = \int w(\mathbf{x})I(\mathbf{x})d\mathbf{x}$$

where the weighting function $w$ is the function in Eq. 2.

Wavelet bases have important applications in mathematics and signal processing due to their ability to build sparse representation for large classes of functions and signals [17]. It is a natural choice for the foveated images due to their locality in space and frequency. Note that direct computation of Eq.1

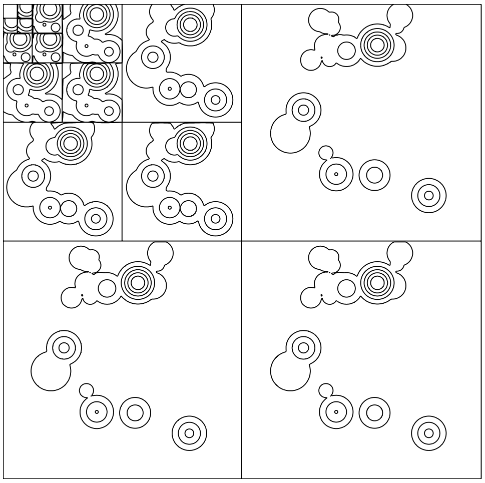


**Fig. 3.** The mask $M$ for the weighting function shown in Fig. 1b

is computationally intensive. Fortunately, there is a fast linear time approximation. This is achieved by multiplying the discrete wavelet transform DWT of the image $I$ by a predetermined mask $M$ followed by the inverse discrete wavelet transform IDWT. That is,

$$(T^{\text{fov}} I) \approx \text{IDWT}(M \ \text{DWT}(I)) \tag{4}$$

Figure 3 shows the mask for a multifovea weighted function. Interestingly, the choice of the weighting function (Eq. 2) gives a self-similarity across scales, which is illustrated in Fig. 3. This fast algorithm plays an important role in both the signature creation and the authentication process. The reader is referred to [5] for the details of the approximation algorithm.

#### 3.2.2. Using a set of foveae for the weighting function

Recall that in our application, we required a weighting function with minimal description. Equation 2 has a constant description and offers a simple yet efficient tradeoff of pixel resolution and coverage of interesting regions. Observe from Eq. 2 that a fovea with lower rate $\alpha$ has a larger region of influence, whereas a fovea with larger foveal resolution $\beta$ is less concentrated around $\gamma$.

We treat each salient point $(\mathbf{x}, s, m)$ as a fovea with foveal location $\gamma = \mathbf{x}$, rate $\alpha = \frac{1}{s}$, and foveal resolution $\beta = \frac{s^2}{m}$. By this choice, a salient point with larger scale has a larger region of influence but is less concentrated around the fovea. Also note that a salient point with larger strength $m$ is more concentrated.

The single fovea weighting function can be generalized to a multifovea weighted function through the blending function Eq. 3. Figure 1b gives an example.

Although we can use any nonzero weighting function, the multifovea weighted function is preferable due to its several desirable properties. It has a short description, it blends well with salient points, there is a fast approximation to compute foveation (Eq. 1), and it provides simplicity in implementation.

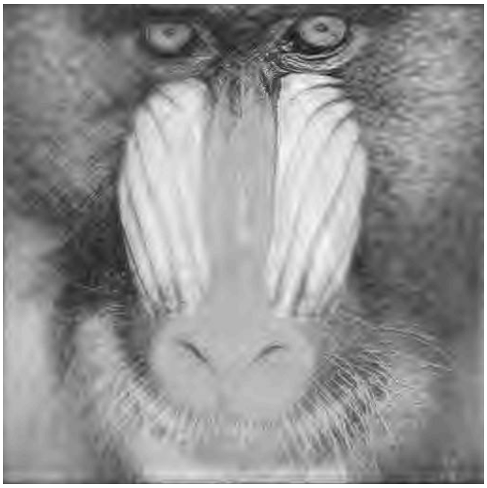


**Fig. 4.** The compressed image using the mask in Fig. 3. This image requires 4 KB, while the original requires 262 KB

### *3.3. Extracting the coefficients*

Recall that the first part of the signature $(S, W)$ is the highly compressed image. To obtain $S$, one could first compute the foveation (Eq. 1) with respect to the multifovea weighted function and then compress the foveated image using a known lossy or lossless compression technique for uniform images. Because computing Eq. 1 directly is computationally intensive, we use the approximation Eq. 4. In our implementation, $S$ is extracted from the image by quantizing the wavelet coefficients $M$ DWT$(I)$ followed by a lossless compression using `gzip` written by J. Gailly and M. Adler [10]. The $(S, W)$ can then be encrypted and stored as the digital signature for that image. Figure 4 shows a lossily compressed image, which can be treated as the information retained in the signature $(S, W)$.

Note that `gzip` uses Lempel-Ziv coding and is a general lossless compression tool. It does not exploit properties of images, especially the coherence of wavelet coefficients across space and scale. Thus it is not the best technique for our application. A possible improvement can be made by incorporating the well-known zero-tree algorithm [22] into our scheme.

## 4. Authentication

Given an image $I$ and its purported digital signature $(S, W)$, we can easily compute and recover the highly compressed image $\widetilde{I}_0$ of the original image (from the signature). It is an interesting question as to which distance function is to be used to measure the distortion of $I$ from $\widetilde{I}_0$. Since most illegal tamperings are localized, the usual two-norm $\|\cdot\|_2$ is not a good choice. To illustrate this point, cropping a $20 \times 20$ pixel subregion from a $512 \times 512$ pixel image is insignificant under the $\|\cdot\|_2$ norm since the cropping energy is averaged over the whole image. On the other hand, the infinity norm $\|\cdot\|_\infty$ is also unsuitable because permissible operations like low-pass filtering might significantly change the value of a single pixel, although the overall distortion is low.

A natural tradeoff is achieved by applying the two-norm locally within a window $u : \mathbf{R}^2 \to \mathbf{R}$ followed by an infinity

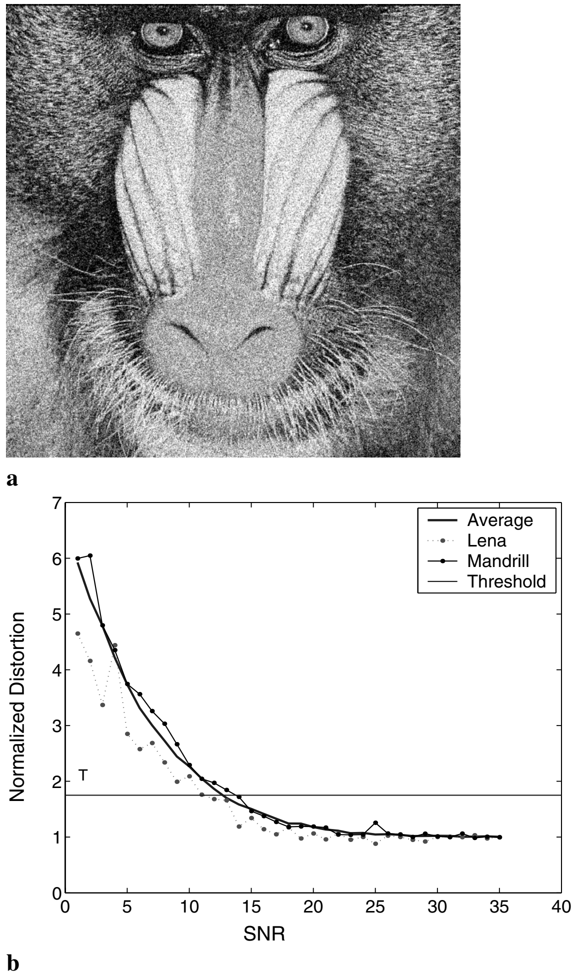


**Fig. 5a,b.** Additive Gaussian white noise. **a** After adding Gaussian noise (SNR = 13 dB). **b** Graph of distortion vs. SNR. The distortion is normalized (divided by the distortion of the original $I_0$ from $\widetilde{I_0}$). The *thin line* is for the image Mandrill, and the *dotted line* is for the image Lenna. The *smooth thick line* is the average of 50 images. The *horizontal line* indicates the threshold $T$ at 1.75

norm. Let

$$U(x,y) = \|I(x,y)\|_{u(\cdot - x, \cdot - y)}$$

where $u(\cdot - x, \cdot - y)$ is the window $u$ shifted by $x$ and $y$. In other words, $U(x,y)$ is the weighted norm with the shifted $g$ as the weighting function. Let us define $\|\cdot\|_{u,\infty}$ to be

$$\|I\|_{u,\infty} = \|U\|_\infty$$

This is equivalent to taking the infinity norm after a convolution of $I$ with the window $u$. Since we assume the image content information is available (through the digital signature), the measurement would be more effective if the windows were wider in the less important regions and narrower in the important regions. This is equivalent to taking the infinity norm after a space-variant smoothing is applied. Recall that the space-variant smoothing process is the foveation operator Eq. 1. Suppose $\widetilde{I_0}$ is the highly compressed foveated image, and $I$ is an image; we then define the distortion to be

$$D(\widetilde{I_0}, I) = \|\widetilde{I_0} - (T^{\text{fov}} I)\|_\infty$$

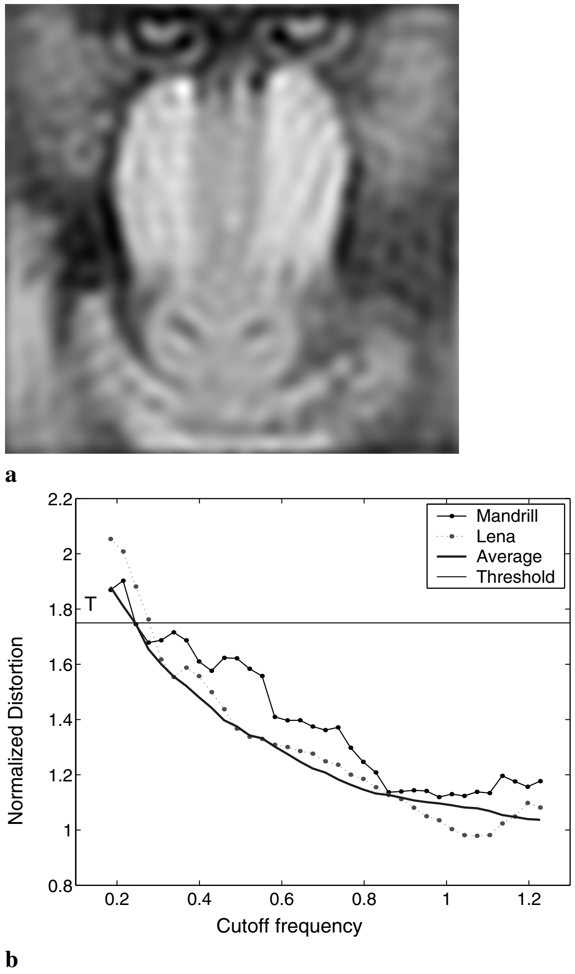


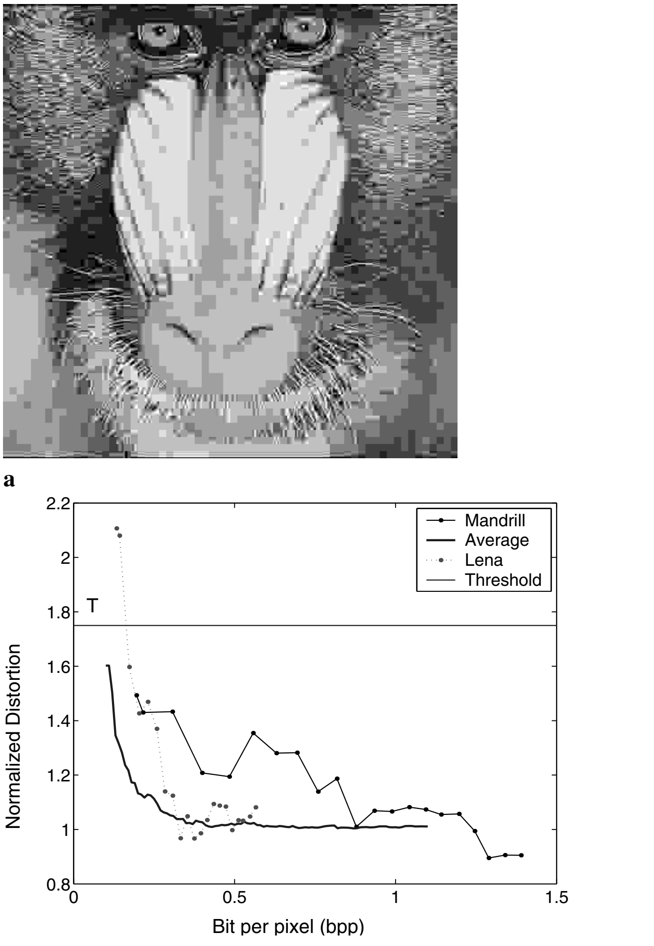
**Fig. 6a,b.** Ideal low-pass filtering. **a** Mandrill after low-pass filtering with cutoff frequency at 0.25 (note that cutoff frequency at $\sqrt{2}\pi$ retains all information). **b** Graph of distortion vs. low-pass filtering with cutoff frequency

During the detection process, the image in question $I$ is declared to be authentic if

$$\frac{D(\widetilde{I_0}, I)}{D(\widetilde{I_0}, I_0)} < T \tag{5}$$

where $T$ is a predetermined threshold. Note that $D(\widetilde{I_0}, I_0)$ is the distortion of the original image $I_0$ from the compressed image $\widetilde{I_0}$. We call the left-hand side of Eq. 5 the *normalized* distortion. In our experiments, we take $T = 1.75$. To compute $D(\cdot,\cdot)$ efficiently, we can use the fast algorithm described in Sect. 3.2.1.

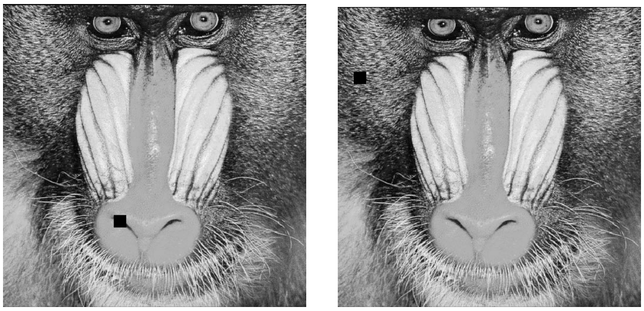
An alternative criterion for Eqr̃efeq:normalizeddistortion takes into consideration the configuration of the feature points. That is, the salient points of the image in question are extracted and compared with the feature points in the signature. Recall that a motivation of our proposed method is to get a tradeoff between the feature-points-based methods and the compression-based methods. Incorporating the feature points configuration into the criterion could be viewed as leaning more toward the feature-points-based methods. A study of how to obtain the right tradeoff is an interesting future work.



a

b

**Fig. 7a,b.** JPEG compression. **a** JPEG compression with rate 0.20 bit per pixel. **b** Graph of distortion $D$ vs. compression rate in bits per pixel (bpp)



a

b

**Fig. 8a,b.** Cropping distortions. **a** Cropping of important region (a square on the nose). Normalized distortion is 1.81. **b** Cropping of unimportant region (a square on the right cheek). Normalized distortion is 1.15

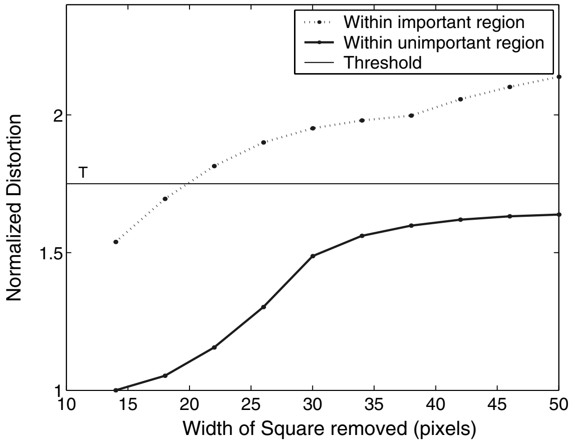
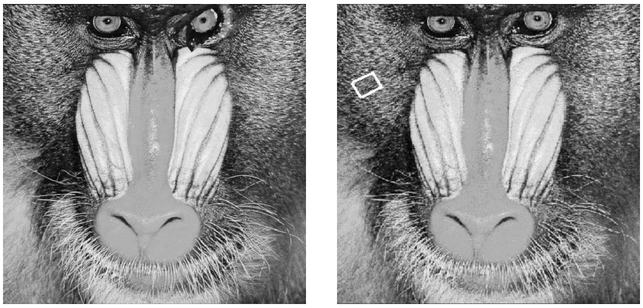


**Fig. 9.** The distortion in Fig. 8 as the size of the square increases in both the important and unimportant region. The location of the square is the same as that in Fig. 8



a

b

**Fig. 10a,b.** Mandrill after rotating different regions. **a** Rotation of important region. Distortion is 112.0 (normalized distortion is 2.37). **b** Rotation of unimportant region. Distortion is 47.3 (normalized distortion is 1.00)

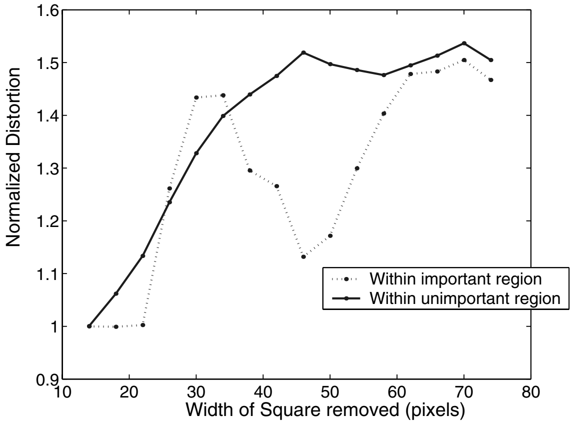


**Fig. 11.** Same as in Fig. 9. Instead of using the proposed weighting function, in this graph, a constant function is used

In our current version, we have not implemented the affine transformation parameter recovery functionality.

## 5. Experimental results

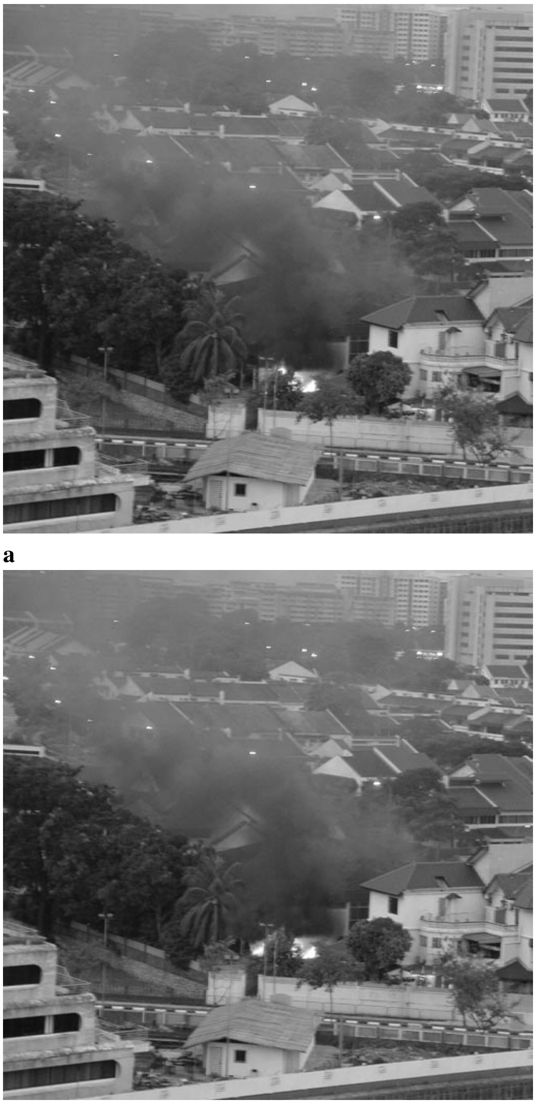
To test the proposed authentication method, we conducted experiments on a database of 50 $512 \times 512$ grayscale images. The value of each pixel ranges from 0 to 255. This database include the images Mandrill (Fig. 2a), Lena, and other natural images, for example, Fig. 12a and c. For each image, its key is generated in the following way. Of all the feature points in the scale space, we chose the top 30 in strength to construct the wavelet mask. This is a tradeoff between the authentication key size and sensitivity to distortions. We used the Villasenor 11/13 biorthogonal wavelet [24] for the wavelet transformation. After applying the wavelet mask (Fig. 3) to the wavelet coefficients, we uniformly quantized the coefficients with a step size of 50. The authentication key was the lossless compression of the quantized coefficients. The average authentication key size is around 2.1 KB. The Mandrill image has the largest key size at about 4.0 KB. Note that fewer/more salient points can be chosen for a smaller/larger signature. Figure 4 shows the Mandrill after content-based compression.

The first part of the experiments deals with the global distortions such as addition of white Gaussian noise, low-pass filtering, and JPEG compression. Figure 5a shows the results after adding Gaussian noise to the Mandrill. For Gaussian noise with a resulting signal-to-noise ratio (SNR) as low as 13 dB, we can still successfully authenticate this image. This experiment is repeated for each of the images in the database. The average distortion is shown in Fig. 5b. Note that for fair comparison among different images, we use the normalized distortion (Eq. 5) rather than the absolute distortion. The low-pass filtering was implemented as ideal low-pass filtering. The blurred Mandrill image after ideal low-pass filtering with cut-off frequency at 0.25 can still be authenticated, as shown in Fig. 6. The experimental results after JPEG compression are shown in Fig. 7. The image in Fig. 7a is declared to be authentic by our method.

The second part of the experiments tests the effect of local tampering of important content (e.g., main features) and unimportant content (e.g., background textures). As seen in Figs. 8 and 10, our method can authenticate the image after tampering in the unimportant regions while raising an alarm in case the tampering is done with the important features. After cropping a $22 \times 22$ region on the nose, the normalized distortion is 1.81, while the same operation on the cheek gives a normalized distortion of 1.15, which is well below the predefined threshold. Figure 9 shows how distortion increases as the size of the cropping region increases. A rotation of the left eye by $72^o$, which makes the Mandrill look angry (Fig. 10a), results in a normalized distortion value of 2.37. On the other hand, rotation of a same-sized area in the cheek region gives a value of only 1.00 (Fig. 10b).

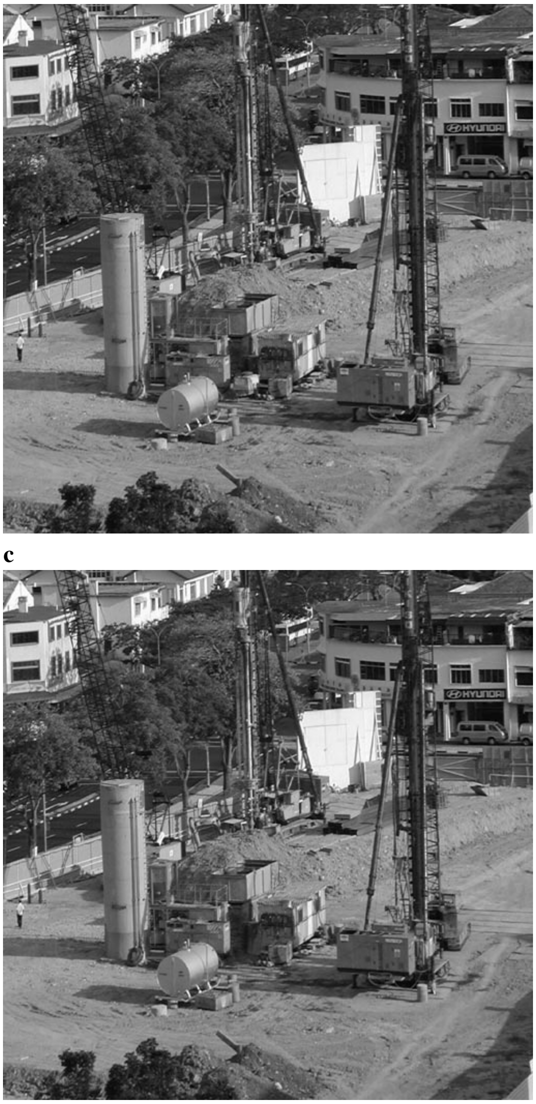
To demonstrate the performance of the proposed content-based weighting function, an experiment similar to the one in Fig. 9 is repeated using a method that uses a *constant* weighting function. More specifically, this method takes the uniformly quantized (step size of 150) wavelet coefficients as the signature. The step size is chosen so that the size of the signature is about 4 KB. Figure 11 shows how distortion increases as the size of cropping region increases. Unlike Fig. 9, the graphs for important region and unimportant region are not distinctly separated. Figure 12 illustrates some more examples of localized tampering.

There is one concern that needs to be discussed. The process of signature extraction is not key-dependent. One can argue that anyone can alter the authenticated image without having access to the signature data by simply tweaking the image data while comparing them with the extracted features using the detection algorithm. As long as the features do not change, the image can be altered. Thus, it may even be possible to modify the image to display a completely different scene, yet it can remain authentic (according to the signature). Actually, any public watermarking scheme is vulnerable to this attack, and this was first pointed out in [18]. However, the analysis and a solution to this was subsequently presented in [16]. The solution basically consists of randomizing the


**a**

**b** $(53.7, 2.10)$.

**Fig. 12.** The image in **a** is the original image. The image in **b** is tampered with by adding new fire. The distortion is 53.7, and the normalized distortion is 2.10. Thus, it is declared to be unauthentic



c

**d** $(35.3, 1.95)$.

**Fig. 13.** The image in **a** is the original image. The image in **b** is tampered with. A generator in the center is removed from the original image. The distortion is $35.3$, and the normalized distortion is $1.95$. Thus, it is declared to be unauthentic

detection process. Thus, instead of using public-key cryptography, we can use a symmetric key cryptographic scheme (like AES) to encrypt the signature. Moreover, we need to randomize the selection of the feature points, which can depend on a key. These additional steps should take care of this security problem. However, the basic ideas of our scheme remain unchanged – only additional protocol steps are now required. These extra steps are probably required of all public watermarking or authentication schemes to be able to resist this attack. Of course, in watermarking, the attacker's goal would be to remove the watermark while making as slight changes as possible. But for an authentication scheme, the attacker's intention would be to make large changes while keeping the authentication check valid.

## 6. Conclusion

We have described a content-based authentication technique for digital images. Many authentication schemes are known. Our work's novelty lies in recognizing that important content information is not uniformly distributed across the image and that illegal operations are usually localized while permissible operations are global in nature. We have developed a scheme based on a weighting function that exploits this observation. Through the use of scale-space salient points and borrowing the idea of foveation from biological vision, we have developed a scheme that extracts a space-variant content-based signature for the image. Foveation also provides desirable properties that allow speeding up the required computation. Taking into consideration the observation that most illegal operations are localized, we give a "space-variant" distortion measurement that is sensitive to localized tampering. Again, this measurement fits nicely with the weighting function used in the signature creation process. The analysis and the experimental results show that this is a promising technique for verification of the genuineness of digital images. Finally, while we have described the technique for grayscale images in this paper, color images can be similarly handled. The space-variant foveation technique needs to be applied independently to each of the channels of the 3D color-space representation, which could be RGB or HSV or any other color space. Thus, the technique should work equally well for color images.

*Acknowledgements.* The authors would like to thank the referees for the very insightful review comments that have substantially improved the content of this paper.

## References

1. Aloimonos J, Weiss I, Bandyopadhyay A (1987) Active vision. First international conference on computer vision, London, pp 35–54
2. Basu A, Wiebe KJ (1998) Video conferencing using spatially varying sensing with multiple and moving fovea. IEEE Transactions on systems, man and cybernetics, 28(2):137–148
3. Bhattacharjee S, Kutter M (1998) Compression tolerant image authentication. IEEE International conference on image processing, vol 1, Chicago, pp 4–7
4. Burt PJ (1988) Smart sensing within a pyramid vision machine. Proceedings of the IEEE, 76(8):1006–1015
5. Chang EC, Mallat S, Yap C (2000) Wavelet foveation. J Appl Comput Harmonic Analysis 9(3):312–335
6. Dittmann J, Steinmetz A, Steinmetz R (1999) Content-based digital signature for motion pictures authentication and content-fragile watermarking. IEEE International conference on multimedia computing and systems, vol II, Italy
7. Fridrich J, Goljan M, Du R (2002) Lossless data embedding for all image formats. Proceedings of SPIE electronic imaging 2002: security and watermarking of multimedia contents, 4675:572–583
8. Eleftheriadis A, Jacquin A (1995) Automatic face location detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates. Signal Processing Image Commun 7(3)231–248
9. Friedman GL (1993) The trustworthy digital camera: restoring credibility to the photographic image. IEEE Transactions on consumer electronics, 39:905–910

10. gzip `http://www.gzip.org/`
11. Hartung F, Kutter M (1999) Multimedia watermarking techniques. Proceedings of the IEEE, 87(7):1079–1107
12. Kundur D, Hatzinakos D (1999) Digital watermarking for telltale tamper proofing and authentication. Proceedings of the IEEE, 87:1167–1180
13. Levoy M, Whitaker R (1990) Gaze-directed volume rendering. Comput Graphics 24(2):217–223
14. Lin CY, Chang SF (1998) A robust image authentication method surviving JPEG lossy compression. SPIE International conference on storage and retrieval of image/video database, vol 3312, San Jose
15. Lindeberg T (1990) Scale-space for discrete signals. IEEE transactions on pattern analysis and machine intelligence, 12:234–254
16. Linnartz JP, van Dijk M (1998) Analysis of the sensitivity attack against electronic watermarks in images. Second international workshop on data hiding, pp 258–272
17. Mallat S (1998) A wavelet tour of signal processing. Academic, New York
18. Perrig A (1997) A copyright protection environment for digital images. Diploma dissertation, EPFL, Lausanne, Switzerland
19. Schneider M, Chang SF (1996) A robust content based digital signature for image authentication. IEEE International conference on image processing, Lausanne, Switzerland
20. Schwartz EL (1994) Topographical mapping in primate visual cortex: history, anatomy, and computation. In: Kelly DH (ed) Visual science and engineering: models and applications, Marcell Dekker, New York, pp 293–360
21. Schwartz EL, Greve DN, Bonmassar G (1995) Space-variant active vision: definition, overview and examples. Neural Netw 8(7–8):1297–1308
22. Shapiro JM (1993) Embedded image coding using zerotrees of wavelet coefficients. IEEE transactions on signal processing, 41(12):3445–3462
23. Stallings W (2002) Cryptography and network security: principles and practice, 3rd ed. Prentice-Hall, New York
24. Villasenor J, Belzer B, Liao J (1995) Wavelet filter evaluation for image compression. IEEE Transactions on image processing, 4:1053–1060
25. Yu GJ, Lu CS, Mark Liao HY, Shen JP (2000) Mean quantization blind watermarking for image authentication. IEEE international conference on image processing