Signal Processing 92 (2012) 1431–1445

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Joint robustness and security enhancement for feature-based image watermarking using invariant feature regions

Jen-Sheng Tsai [a], Win-Bin Huang [a], Yau-Hwang Kuo [a,*], Mong-Fong Horng [b]

[a] Center for Research of E-life Digital Technology, Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan

[b] Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan

ARTICLE INFO

Article history:
Received 16 March 2011
Received in revised form
28 November 2011
Accepted 29 November 2011
Available online 8 December 2011

Keywords:
Digital image watermarking
Feature detector
Knapsack problem
Genetic algorithm
Differential entropy

ABSTRACT

Local image features have been widely applied in feature-based watermarking schemes. The feature invariance is exploited to achieve robustness against attacks, but the leakage of information about hidden watermarks from publicly known locations and sizes of features are often unconsidered in security. This paper, therefore, proposes a novel image watermarking approach, which adopts invariant feature regions to jointly enhance its robustness and security. Initially, circular feature regions are determined by the scale-adapted auto-correlation matrix and the Laplacian-of-Gaussian operation. Leakage of secret information is also controlled carefully during feature detection procedure. An optimal selection process formulated as a multidimensional knapsack problem is then proposed to select robust non-overlapping regions from those circular feature regions to resist various attacks. This process is implemented by a genetic algorithm-based approach, and incorporates randomization to mitigate the security risk. Finally, each selected region is normalized to obtain a geometrically invariant feature region, and embedded with a region-dependent watermark to overcome the weakness of multiple-redundant watermarks. The evaluation results based on the StirMark benchmark present the proposed scheme can tolerate various attacks, including noise-like signal processing and geometric distortions. A security analysis in terms of differential entropy also confirms the security improvement of the proposed method.

1. Introduction

Digital watermarking, which is regarded as a useful approach for copyright protection, content authentication, and transaction tracking, has been widely applied to image, audio, and video. In all of these applications, the effectiveness of a digital watermarking algorithm depends on its ability to resist various attacks. According to different intentions of attacks, there are two criteria, robustness and security, which should be considered in the design of digital watermarking schemes [1–7]. Robustness deals with blind attacks that try to destroy or invalidate hidden watermarks without exploiting knowledge of the watermarking algorithm. The robustness measurement for watermarking schemes is to evaluate their ability to successfully detect the hidden watermark after blind attacks. In general, robust watermarking schemes are developed to resist two types of attacks: noise-like signal processing and geometric distortions. Security, on the other hands, denotes the ability of a watermarking scheme to prevent hidden

* Corresponding author. Tel.: +886 6 275 7575x62522; fax: +886 6 208 8075.
E-mail addresses: asheng@ismp.csie.ncku.edu.tw (J.-S. Tsai), huangwb@ismp.csie.ncku.edu.tw (W.-B. Huang), kuoyh@ismp.csie.ncku.edu.tw (Y.-H. Kuo), mfhorng@ieee.org (M.-F. Horng).

0165-1684/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2011.11.033

watermarks from being accessed by unauthorized users. For the attacks to security, it is usually assumed that the unauthorized users know all knowledge about the watermarking algorithm except the secret key and they try to estimate the hidden watermarks through observing the watermarked images. The security of a watermarking scheme can be measured by analyzing the leakage of information about the hidden watermarks from observations. Until now, most existing watermarking methods have been driven by improvement of robustness, including the spread spectrum [1,2], quantization index modulation [8,9], and resynchronization [10–27] schemes. But there has been little attention to security in watermarking research. However, recent studies [3–6] have shown that security is as important as robustness in developing digital watermarking schemes. Because a successful attack to security can completely break a watermarking system even though it is robust, designers should consider not only the robustness but also the degree of security in their watermarking schemes.

In this paper we use both robustness and security perspectives to investigate current feature-based watermarking methods, which are resynchronization schemes that exploit the invariant features of a medium to resist attacks. Bas et al. used the Harris detector to extract feature points from an image and the Delaunay tessellation to form triangle meshes with theses points for watermarking [16]. Tang and Hang adopted the Mexican Hat wavelet to extract feature regions, and exploited image normalization and FFT to hide watermarks [17]. These two methods exhibit good robustness against most attacks, but features are probably not extracted correctly after suffering from scaling attacks. Therefore, Seo et al. proposed a watermarking method based on the scale-space theory to mitigate this problem [18,21]. Robustness can be also achieved by applying the scale-invariant feature transformation [19], the Harris–Laplacian detector [22,24], or the difference of Gaussian [23]. Recently, Gao et al. [25] used the affine covariant regions to provide good resistance to geometric distortions. However, most of the existing methods usually suffer from inability to resist random or region-of-interest (ROI) cropping attacks and ineffectiveness in security-related applications.

The invalidation for cropping attacks will become more serious due to the phenomenon described in this paragraph. Since the magnitude of pixels in a feature region will be modified when a watermark is inserted into this region, it is preferred to select non-overlapping regions for watermarking to avoid major degradation of image quality. In order to obtain the non-overlapping regions, some reference parameters have been exploited in existing methods. For example, the corner response [18,21] and the number of neighboring feature points within a region [17,22] are used to remove overlapping feature regions. In [24], the minimum spanning tree (MST) clustering algorithm was used to cluster the feature regions into groups according to a distance constraint. The region with the largest corner response in each group is then selected to be watermarked. However, non-overlapping feature regions selected for watermarking by these parameters cannot guarantee that watermark regions are well distributed over an image. Thus, the probability of successful cropping attacks is raised because the selected regions do not always have the maximum cover range.

Next we discuss the issue of ineffectiveness in security-related applications. It is here assumed that the attacker knows the details of watermarking algorithm except the secret key according to the *Kerckhoffs' principle*. The secret key is an input to some mapping functions that outputs secret parameters, such as the watermark sequence [5]. Without knowledge of the key, the secret parameters cannot be forged or estimated. Unfortunately, it is not difficult for attackers to extract the hidden watermark sequence embedded by most of the existing feature-based methods because of information leakage. The information leakage denotes the information about the hidden watermark sequence achieved from the attacker's observation [6]. The leakage is mainly from which the watermarked feature regions' locations and sizes are publicly known and each feature region is embedded with the same watermark. This weakness enables the security attacks to easily break down a watermarking system. For example, the attacker can apply a collusion attack to remove the hidden watermark or a copy attack to fake a watermarked image by collecting and analyzing a set of feature regions with the same watermark from a watermarked image [28,29,39].

This paper proposes a novel feature-based watermarking method that (1) optimizes the cover range of the hidden watermarks for resisting cropping attacks and (2) enhances the security to prevent unauthorized users from accessing the secret parameters. Initially, the Harris–Laplacian detector, which uses the scale-adapted auto-correlation matrix to localize points in the scale space and invokes the Laplacian-of-Gaussian operation to select the points attaining an extremum over scales, is applied to an image to detect its feature points at multiple scale levels [31,32]. Around these feature points, the targeted circular feature regions are determined based on their characteristic scales and a secret key. The high repeatability of these feature regions offers robustness against translation, rotation, scaling, and partial illumination changes, while secrecy of the region size makes it difficult for an attacker to estimate exact range of feature region. Since these feature regions are substantially overlapped and not all stable, we propose a heuristic algorithm to select an optimal non-overlapping region set for watermarking. The region set also features a maximum distribution over the target image to tolerate the cropping attacks. The selection process further incorporates randomization to avoid an attacker correctly identifying the watermarked regions. This work is formulated as a multidimensional knapsack problem and is solved by a genetic algorithm-based procedure. Finally, we perform normalization for each selected region to obtain geometric invariance, and generate a region-dependent watermark sequence, which eliminates the problem of hiding the same signal multiple times, to be additively embedded into the spatial domain of each invariant region.

Comparing the existing feature-based methods [16–27], we investigate the important issues of feature-based

watermarking methods to improve the robustness and security. First, we propose the optimal region selection process to enhance the resistance to cropping attacks. Different from our previous work [26], the method proposed in this paper does not need the time-consuming simulated attacking procedure, and its goal is to make the selected feature regions achieve the greatest distribution over an image to withstand cropping attacks, which is not considered in [26]. Moreover, the detailed implementation of the genetic algorithm-based heuristics is described in this paper. The experimental results have demonstrated that our method has better coverage in the resilience to ROI cropping, random cropping and centered cropping attacks. Also, the evaluation on the StirMark benchmark confirmed the goodness of our method in the resistance to other attacks, including noise-like signal processing and geometric distortions. Second, we propose the region-dependent watermark based on feature descriptor, and incorporate randomization in feature region detection and feature region selection for security enhancement. The region-dependent watermark is derived from the framework of the content-dependent watermark with DCT block-based media hashes [39]. The novelty is that we exploit the feature description as the media hash, which is created by generating the orientation histograms. The feature description has been proved to be distinctive and robust [33], and it is obtained during feature detection without additional DCT operations [39]. Furthermore, the incorporated randomization in determining the feature regions for watermarking makes the secrecy of their locations and sizes to mitigate the security risk. A security analysis in terms of differential entropy for feature region detection and feature region selection is given to demonstrate the effectiveness of this paper.

The rest of this paper is organized as follows. Feature detection with a controlled secret leakage is presented in Section 2 and a novel feature region selection scheme to enhance both robustness and security for watermarking is also proposed. In Section 3, the details of region-dependent watermark embedding and detection schemes based on local features are described. The experimental results, mainly for robustness evaluation and security analysis, are given in Section 4. Concluding remarks are drawn in Section 5.

2. Robust and secure image features for watermarking

Local features representing image structures, ranging from points to regions, have been adopted in many applications, such as object recognition, image retrieval, and camera calibration [30–34]. These features, which are powerful references, have also been applied successfully in feature-based watermarking methods since they can be preserved after suffering distortion such as scaling, rotation, or illumination changes. In general, a feature detector performs a specific transformation on an image to extract local features for watermark embedding and detection. However, a feature region extracted by a detector is not directly applicable to digital watermarking because of the following issues. The locations and sizes of extracted features can be publicly found by the attackers. Embedding watermarks into all regions will also cause heavy image degradation and low robustness since most of features are overlapped. Although many new feature detectors have been proposed to enhance the robustness of feature-based watermarking [16–27], most methods are still vulnerable to security attacks and cropping attacks. Therefore, a qualified feature-based watermarking scheme should examine the robustness of the adopted feature detector, avoid the information leakage of secret parameters, and determine an appropriate non-overlapping feature region set. This section presents two processes, feature region detection and feature region selection, which are important in achieving the desired goal.

2.1. Detection of robust and secure features

In this section, the Harris–Laplacian detector, which consists of scale-adapted auto-correlation matrix and the Laplacian-of-Gaussian operation, is adopted [31,32], while the secret leakage is carefully controlled in order to identify local image features. First, the scale space of an input image I is calculated by the function L at a set of scales to represent different levels of resolutions, which is formulated as

$$L(\mathbf{x},\sigma_D) = G(\mathbf{x},\sigma_D) * I(\mathbf{x}) \tag{1}$$

where $\mathbf{x}=(x,y)$ denotes the image spatial coordinate, σ_D is the differential scale, "$*$" represents the convolution operation, and the uniform Gaussian kernel G is defined by

$$G(\mathbf{x},\sigma_D) = \frac{1}{2\pi\sigma_D^2} e^{-(x^2+y^2)/2\sigma_D}. \tag{2}$$

Then the scale-adapted auto-correlation matrix $\mu(\mathbf{x},\sigma_I,\sigma_D)$ is applied in the scale space to describe the local image structure, and it is formulated by

$$\mu(\mathbf{x},\sigma_I,\sigma_D) = \sigma_D^2 G(\mathbf{x},\sigma_I) * \begin{bmatrix} L_x^2(\mathbf{x},\sigma_D) & L_xL_y(\mathbf{x},\sigma_D) \\ L_xL_y(\mathbf{x},\sigma_D) & L_y^2(\mathbf{x},\sigma_D) \end{bmatrix} \tag{3}$$

where σ_I is the integral scale, and L_i is the first derivative calculated in the i direction that $i\in\{x,y\}$. The corner response estimating principal curvature of the matrix is computed by its trace and determinant as

$$C(\mathbf{x},\sigma_I,\sigma_D) = det(\mu(\mathbf{x},\sigma_I,\sigma_D)) - 0.04 trace(\mu(\mathbf{x},\sigma_I,\sigma_D)). \tag{4}$$

The feature point with large corner response representing significant curvatures has higher repeatability. The candidate points are then determined if their corner response is a local maximum and larger than a threshold T_R used for filtering out unstable feature regions. However, it is difficult to be set as a fixed value for different input images [34]. According to the suggestion in [43], the threshold should be set to 1% of the maximum response value of all extracted feature regions. In order to achieve scaling invariance, the integral scale of all candidate points is compared to the characteristic scale of local image structure. The characteristic scale, which is relatively independent of scale change, is obtained by searching for a local extremum over multiple scale levels of Laplacian-of-Gaussian. Candidate points for a set of scale levels σ_n are identified by setting $\sigma_I=\sigma_n$ and

$\sigma_D=0.7\sigma_I$ where $\sigma_n=\{\delta^i\sigma_0|\sigma_0=1.5,\ \delta=1.1,\ i=1,\ 2,\ \ldots,\ n\}$. The scale step factor δ between two successive levels affect the accuracy of the scale of the candidate point. It should be small to achieve high accuracy and is set to 1.1 in this paper according to the suggestion from [32]. The number of scale levels n depends on the possible scale changes of an image for different applications and is set as 15 in our experiments. The Laplacian-of-Gaussian of the candidate points is calculated as

$$\left|LoG(\mathbf{x},\sigma_n)\right|=\sigma_n^2\left|L_{xx}(\mathbf{x},\sigma_n)+L_{yy}(\mathbf{x},\sigma_n)\right|. \tag{5}$$

A candidate point at the ith scale level is regarded as a feature point with a characteristic scale $\sigma_c(\sigma_c=\delta^i\sigma_0)$ if its Laplacian-of-Gaussian is a local extremum over all scale levels and is higher than a pre-defined threshold as follows:

$$\left|LoG(\mathbf{x},\sigma_i)\right|>\left|LoG(\mathbf{x},\sigma_j)\right|,\quad j\in\{i-1,i+1\} \tag{6}$$

$$\left|LoG(\mathbf{x},\sigma_i)\right|>T_{LoG}. \tag{7}$$

The threshold T_{LoG} is set to 10, which refers to the suggestion from [31]. In order to achieve rotation invariance and incorporate controlled secret leakage in the outputs of the feature detection, each feature point is further used as the center to derive a corresponding circular feature region with a key-dependent radius r determined

$$r=\alpha\cdot\sigma_c \tag{8}$$

which is formulated by a secret key α and the feature point's characteristic scale σ_c. Obviously, the circular feature regions obtained by the scale-adapted auto-correlation matrix and the Laplacian-of-Gaussian operation are highly distinctive and matched with a high repeatability against various image distortions [31,32,34]. The key-dependent radius prevents an attacker from easily accessing a feature region by controlling the uncertainty of its size, and information leakage is also reduced while the watermark is embedded into the region.

2.2. Feature region selection

This work aims to obtain appropriate non-overlapping regions for watermarking since there are serious overlaps and instability in the extracted feature regions. In addition to removing some overlapping regions, the feature regions selected should have maximum distribution over the image to withstand cropping attacks and to incorporate randomization for security. Therefore, this work is formulated as an optimum problem constricted by image quality and regions' overlapping circumstance as follows:

$$maximize\ \sum_{j=1}^{N_R}\beta_j r_j s_j \tag{9}$$

$$subject\ to\ \sum_{j=1}^{N_R} q_j s_j\le T_q \tag{10}$$

and

$$\sum_{j=1}^{N_R} p_{ij}s_i s_j<1,\quad i=1,2,\ldots,N_R \tag{11}$$

where N_R is the number of feature regions extracted, r_j is the radius of region j, $\{\beta_j\}$ are key-dependent pseudo-random numbers with the mean μ and variance σ^2, and s_j is defined as

$$s_j=\begin{cases}1, & if\ the\ region\ j\ is\ selected;\\ 0, & otherwise.\end{cases} \tag{12}$$

The variable q_j denotes the distortion of a watermarked region j compared with its original region, and T_q refers to the limitation of quality degradation of an image after being watermarked. Eq. (11) means that only one region can be selected in each overlapping case. The value of p_{ij} is dependent on the overlapping situation of the two regions of i and j:

$$p_{ij}=\begin{cases}1, & if\ the\ region\ i\ overlaps\ with\ region\ j,\ and\ i\neq j;\\ 0, & otherwise.\end{cases} \tag{13}$$

In order to solve this combinatorial optimization problem, we transform it to a multidimensional knapsack problem (MDKP) by modifying the expression of its constraints as follows:

$$maximize\ \sum_{t=1}^{N_R}\beta_t r_t s_t \tag{14}$$

$$subject\ to\ \sum_{t=1}^{N_R}\omega_{kt}s_t\le T_{\omega_k},\quad k=1,2,\ldots,m. \tag{15}$$

In Eq. (15), the variables ω_{kt} and T_{ω_k} represent the composite weights and constraints of quality distortion and overlapping status specified in Eqs. (10) and (11), respectively. When $k=1$, then $\omega_{1t}=q_t$ and $T_{\omega_1}=T_q$, which is equivalent to the constraint formulated by Eq. (10). Each k greater than 1 correspondingly denotes a specific index pair (i,j), $i\neq j$, in Eq. (11). Then the constraints specified in Eq. (11) can be reformulated into Eq. (15) as follows:

Referring to Eq. (11), it is obvious that

$$p_{ij}s_i s_j<1,\quad \forall r_i,\forall r_j. \tag{16}$$

Since s_i, s_j and $p_{ij}\in\{0,1\}$, Eq. (16) can be rewritten as Eq. (17)

$$p_{ij}s_i+p_{ij}s_j\le 1 \tag{17}$$

that is,

$$0s_1+0s_2+\cdots+p_{ij}s_i+\cdots+p_{ij}s_j+\cdots+0s_{N_R}\le 1. \tag{18}$$

Let $T_{\omega_k}=1$, and $\omega_{kt}=0$ when $t\neq i$ *or* j, otherwise $\omega_{kt}=p_{ij}$. We can get

$$\sum_{t=1}^{N_R}\omega_{kt}s_t\le T_{\omega_k}. \tag{19}$$

Aggregating Eq. (19) for different values of k, we conclude Eq. (15) and $m=(N_R^2-N_R)/2+1$.

Since MDKP is an NP-hard problem [36], the genetic algorithm (GA), a heuristic search approach based on the principles of evolution in nature [36,37], is employed to efficiently obtain a near-optimal solution in this paper. GA exploits the string structures to make an effective search, and works with a population of individuals that represent candidate solutions to a given optimization problem. It is

very likely that the obtained solution is a global solution since there are crossover and mutation operators and diverse individuals in the population being processed. The evaluation and analysis in [38] also demonstrate the solution of MDKP determined by GA is the best approximation to the global optimum among various optimization methods. Firstly, a fixed length and fixed order binary bit string $S \in \{0,1\}^{N_R}$ representing a candidate region set, in which 1 at the jth bit indicates region j is selected, is regarded as a chromosome of an individual in a population for GA operation. The length of S depends on the number of the extracted feature regions. The GA-based search procedure to find a near-optimal solution includes the following steps:

1) *Population initialization*: the initial population is drawn randomly to maintain diversification of chromosomes. Each individual in the population should be a feasible solution without violating the constraints in Eq. (15).
2) *Fitness evaluation*: Fitness is used to evaluate the possibility of an individual to be the best solution. Each individual in current population has its own fitness to represent degree of success as shown by

$$Fitness(S) = \sum_{j=1}^{N_R} \beta_j r_j S[j] \tag{20}$$

where $S[j]$ denotes the jth bit in the chromosome of an individual. The fitness corresponds to the objective function in Eq. (14), and maximization of the fitness leads to the best solution of feature region selection.
3) *Parent selection*: This step is to select individuals from a population for a mating pool to generate new offspring. Based on the natural principle of survival of the fittest, the binary tournament selection is used, which works by forming two tournament pools of individuals, each containing two individuals picked randomly from the population. Two individuals with the highest fitness, each drawn from one of the two tournament pools, are selected to be parents.
4) *Crossover and mutation*: The generation following the selected parent individuals is obtained by two GA operators, crossover, and mutation. First, uniform crossover is adopted for any two parents to generate a single child whose chromosome is determined by copying the corresponding bits in the chromosomes of the two parents. Each copied bit is chosen randomly with equal probability from the two parents using a binary random number generator. If the random number is 1, the bit is copied from the first parent; otherwise it is copied from the second parent. Then the mutation operation is used to flip a small number of bits in the child's chromosome, changing them from 0 to 1 or vice versa. It is noted that the generated child solution by crossover and mutation operators may not be feasible due to the MDKP constraints. A repair operator is used here to overcome this problem [37].
5) *Termination*: An iterative process from step 2 to this step is executed to find the best solution. The termination condition is satisfied when either a user-defined maximum number of iterations is reached or the fittest one is unchanged during a large number of iterations.

The computational cost of the proposed feature region selection method is dominated by the GA-based search procedure for solving MDKP. Furthermore, the computational cost of this procedure mainly depends on the number of feature regions and the constraints (image quality and regions' overlapping circumstances) in MDKP. The number of feature regions is proportional to the length of individual chromosome that affects the execution time in fitness evaluation, crossover, and mutation. As for the constraints, they are related to the execution time for searching feasible solutions in GA. A more detailed complexity analysis of GA for solving MDKP can be found in [37]. Basically, the computational cost of feature region selection in most of the existing feature-based watermarking methods only depends on the regions' overlapping circumstances, for example, the operations for removing region overlapping by comparing the region's corner response [18,21] and the number of neighboring feature points within a region [17,22]. In [24], the operational cost is to cluster the feature regions into groups by the minimum spanning tree clustering algorithm. In order to achieve the desired optimization goal, the execution time of the proposed feature region selection method is longer than the above-mentioned methods. According to our empirical study that the GA-based search procedure was coded in Borland C++ and executed on an Intel Core2Duo 2.4 GHz PC, the execution time spent in searching the near-optimal solution is within 2 min for all test images in our experiments.

3. Proposed watermark embedding and detection schemes

The detailed procedures of watermark embedding and detection are described here, with their block diagrams depicted in Figs. 1 and 2, respectively.

3.1. Watermark embedding scheme

As shown in Fig. 1, the adopted feature regions in a cover image are extracted and selected by the detector and selector described in Section 2. In the step of feature

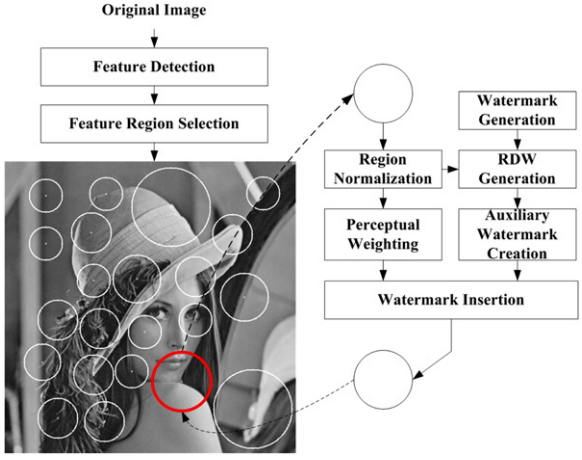

Fig. 1. Block diagram of the watermark embedding scheme.

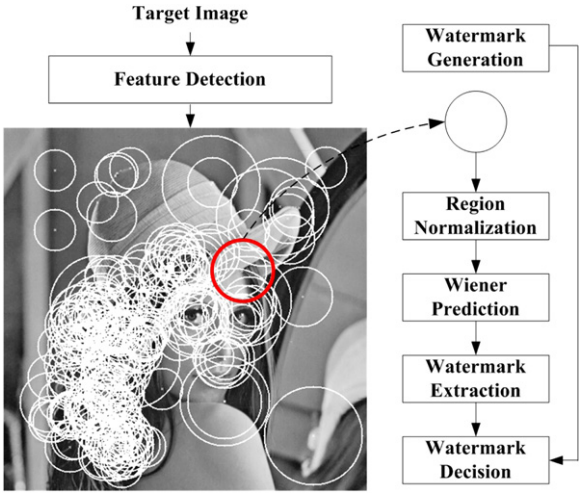

Fig. 2. Block diagram of the watermark detection scheme.

detection, N_R feature points and their characteristic scales are obtained by the scale-adapted auto-correlation matrix and the Laplacian-of-Gaussian operation. Centered on each of the detected feature points, N_R circular feature regions are produced with the radius equal to the product of the center's characteristic scale and a secret key value defined by user. In feature region selection, the proposed selection method is used to determine appropriate circular regions to be watermarked. These regions are intended to cover the maximum range of the target image under the constraints of range overlapping and image distortion.

In Region Normalization, each selected circular feature region is normalized to a canonical form for maintaining rotation invariance and inserting a fixed-length watermark. By this normalization, the circular regions scaled with a user-defined radius are rotated to a consistent orientation based on the gradient histogram within the region [33,34]. Before inserting the watermark into each normalized region, the Perceptual Weighting step is performed to evaluate the watermark embedding strength for avoiding large image degradation according to the noise visibility function (*NVF*) [42]:

$$NVF(\mathbf{x}) = \frac{1}{1+(z/Var_{\max})\cdot Var(\mathbf{x})} \tag{21}$$

where $Var(\cdot)$ denotes the local variance in a window centered on the pixel at coordinate $\mathbf{x}$, $Var_{\max}$ is the maximum local variance in the normalized region, and z is an empirical constant chosen for various cover images in the range from 50 to 100.

In *Watermark Generation*, a watermark sequence, $W=\{w_i \in \{+1,-1\} | i=1,2,\ldots,L_w\}$, is generated for each selected region by a pseudo-random generator with a secret key, and L_w denotes the watermark insertion length. Then in *Region-Dependent Watermark (RDW) Generation*, W is combined into a hash sequence produced according to the descriptor of each normalized region. The descriptor, $D=\{d_i | i=1,2,\ldots,N_p^2 N_o\}$, is created by generating an $N_p \times N_p$ array of orientation histograms with N_o orientations from a normalized region ($N_p^2 N_o$ is set as L_w) [33]. The hash sequence $H=\{h_i | i=1,2,\ldots,L_w\}$ is then obtained by considering the quantized descriptor defined by

$$h_i = \begin{cases} +1, & \text{if } \lfloor d_i/q_{step} \rfloor \text{ is odd;} \\ -1 & \text{otherwise} \end{cases} \tag{22}$$

where q_{step} denotes the quantization step and $\lfloor \cdot \rfloor$ is the floor function that gives the largest integer less than or equal to a real argument. The RDW sequence $W^{RD}=\{w_i^{RD} | i=1,2,\ldots,2L_w\}$ is generated by

$$W^{RD} = \mathcal{S}(W,H) \tag{23}$$

where $\mathcal{S}(\cdot)$ is a key-dependent shuffling function [39].

In order to enhance the robustness of RDW, an *auxiliary watermark sequence* $\check{W}^{RD}=\{\check{w}_j^{RD} | j=1,2,\ldots,2tL_w\}$ is created by repeating each RDW element t times as follows:

$$\check{w}_j^{RD} = w^{RD}_{\lfloor (j-1)/t \rfloor + 1}, \quad j=1,2,\ldots,2tL_w. \tag{24}$$

where $\lfloor \cdot \rfloor$ is the floor function.

In *Watermark Insertion*, the watermark sequence is arranged and embedded into a square region as shown in Fig. 3, where the circular feature region circumscribes the square region. Each element $\check{w}_j^{RD}$ of the RDW sequence is inserted into each pixel of the region by

$$I_w(\mathbf{x}) = I(\mathbf{x}) + ((1-NVF(\mathbf{x}))\cdot b_1 + NVF(\mathbf{x})\cdot b_2)\cdot \check{w}_j^{RD} \tag{25}$$

where $I_w(\mathbf{x})$ and $I(\mathbf{x})$ are the watermarked pixel value and the original pixel value, respectively, in the coordinate $\mathbf{x}$ of the region. The parameters, b_1 and b_2, are empirical values for different cover images [42]. In our method, b_1 is set to 3 according to the suggestion in [42], and b_2 is adjusted to keep the PSNR between the original region and the watermarked one higher than 38 dB. After the insertion procedure, the watermarked image is reconstructed from all normalized regions translated back to original shapes by inverse region normalization.

3.2. Watermark detection scheme

The detection process shown in Fig. 2 extracts the watermark from a target image without an original image, and the detailed operation of each step is explained here.

Similarly, feature regions are detected by the feature detector adopted in the watermark embedding process. Each selected circular feature region is also normalized to a canonical form by Region Normalization. Since the watermark embedded in the spatial domain could be regarded as noise, Wiener filter [16,19,20,21,26,27] is used to blindly extract watermark sequence from the normalized region. Wiener filter is considered as a denoising operation to estimate the watermark from the target image. It is commonly used in watermark detection and regarded as an efficient approach in the existing feature-based watermarking methods [16,19,20,21,26,27]. We first extract hidden information from the normalized region according to

$$\overline{w}_j^{RD'} = Var(\check{w}_j^{RD})\cdot(I'(\mathbf{x}) - \mu(I'(\mathbf{x})))/(Var(\check{w}_j^{RD}) + Var(I'(\mathbf{x}))) \tag{26}$$

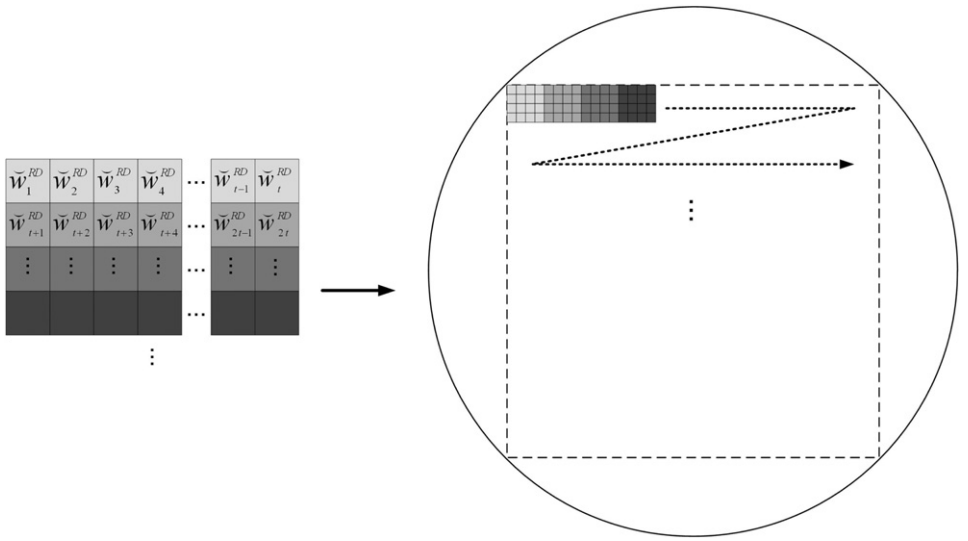

Fig. 3. Watermark sequence is created by repeating each RDW element t times (left), and then it is arranged and embedded into the normalized region (right).

where $I'(\mathbf{x})$ is the pixel value of the normalized region in the target image. $Var(\cdot)$ and $\mu(\cdot)$ denote the local variance and the local mean, respectively. The extracted elements are converted into bipolar ones to obtain the watermark sequence $\breve{W}^{RD'}=\{\breve{w}_j^{RD'}|j=1,2,\ldots,2tL_w\}$ by

$$\breve{w}_j^{RD'}=\begin{cases}+1, & \overline{w}_j^{RD'}>0;\\ -1, & \overline{w}_j^{RD'}\le 0;\end{cases} \tag{27}$$

Then this sequence is translated back to a non-duplicated one according to Eq. (28) in the *Extraction* step:

$$w_i^{RD'}=\begin{cases}+1, & \sum_{j=(i-1)\cdot t+1}^{i\cdot t}\breve{w}_j^{RD'}>0;\\ -1, & \sum_{j=(i-1)\cdot t+1}^{i\cdot t}\breve{w}_j^{RD'}\le 0;\end{cases}\quad i=1,2,\ldots,2L_w \tag{28}$$

A hash sequence is also generated from the normalized region to obtain the final watermark sequence $W'=\{w_i'|i=1,2,\ldots,L_w\}$.

In *Watermark Decision*, the original watermark W is compared with the extracted watermark W'. The presence of a watermark is confirmed when the bit error between W and W' is less than a user-defined threshold T_w, which is determined by the probability of detection error due to false-positive or false-negative detection. The rate of false-positive detection is defined as the probability of successful watermark detection from an un-watermarked image, and the false-negative rate is the probability of failure in detecting watermark from a watermarked image. However, it is difficult to analyze the false-negative rate because a wide variety of attacks could be applied to a watermarked image. The threshold is generally decided by maintaining an extreme low false-positive rate of watermark detection [18,19,22,24,26,27].

We first define $p_{FP\text{-}B}$ as the probability of false-positive detection of a watermark bit from its corresponding repeating bits. Each extracted repeating bit of an un-watermarked region in the detection procedure is treated as an independent random variable with probability 0.5. Based on Bernoulli trials, the probability can be defined by

$$p_{FP\text{-}B}=\sum_{i=\lceil (t+1)/2\rceil}^{t}\binom{t}{i}\cdot(0.5)^i\cdot(0.5)^{t-i} \tag{29}$$

where t is the repeating times and the $\lceil\cdot\rceil$ is a ceiling function that maps a real argument to the smallest following integer. Then the probability of false-positive detection from an un-watermarked region is calculated by

$$p_{FP\text{-}W}=\sum_{i=L_w-T_w}^{L_w}\binom{L_w}{i}\cdot(p_{FP\text{-}B})^i\cdot(1-p_{FP\text{-}B})^{L_w-i}. \tag{30}$$

The detection of watermark for each region is performed by locally searching N_s times to deal with the problem of feature detection errors [18,21,24,26]. If there is at least one successful detection, the region is claimed as watermarked. Therefore, the probability of false-positive detection is

$$p_{FP\text{-}Region}=\sum_{i=1}^{N_s}\binom{N_s}{i}\cdot(p_{FP\text{-}W})^i\cdot(1-p_{FP\text{-}W})^{N_s-i}. \tag{31}$$

The existence of a watermark in an image is determined if at least l regions are successfully detected as watermarked. Finally the probability of false-positive detection of an image can be calculated as follows:

$$p_{FP\text{-}Image}=\sum_{i=l}^{N_R}\binom{N_R}{i}\cdot(p_{FP\text{-}Region})^i\cdot(1-p_{FP\text{-}Region})^{N_R-i} \tag{32}$$

The value of T_w could be decided according to the desired probability of false-positive detection of an image. In Fig. 4, we demonstrate the curves of $p_{FP\text{-}Image}$ (in log scale) versus the watermark detection threshold T_w for $L_w=128$, $N_s=25$, and $l=1$, 2, 3, where the solid line and the dashed line denote $N_R=100$ and $N_R=200$, respectively.

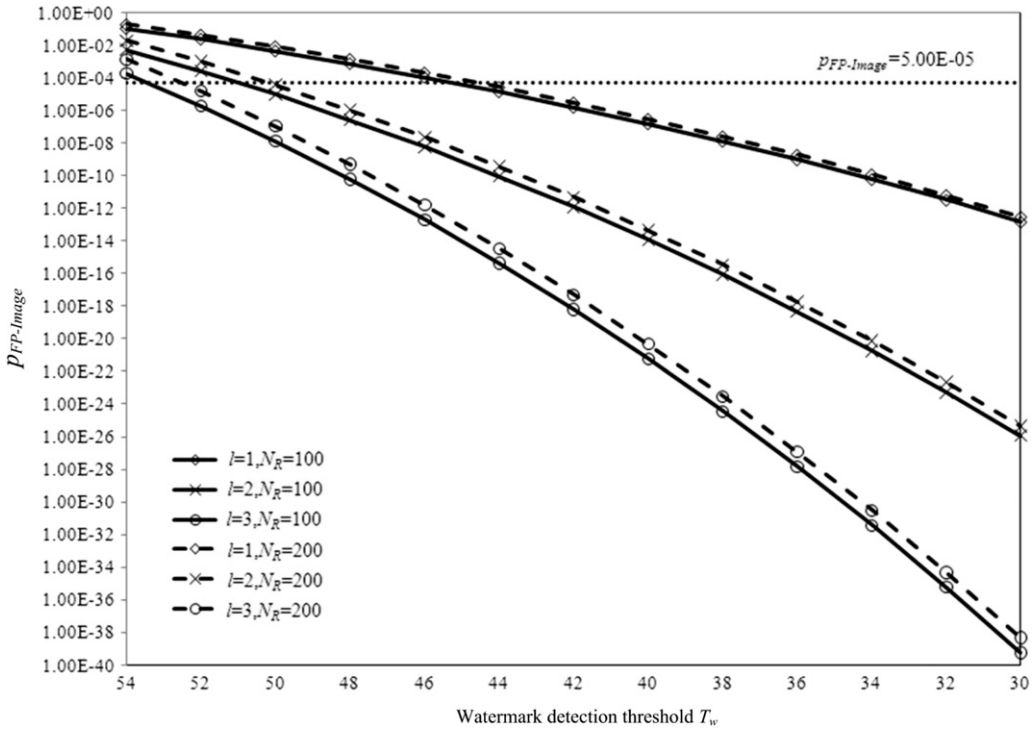

Fig. 4. Curves of $p_{FP\text{-}Image}$ (in log scale) versus the watermark detection threshold T_w for $L_w=128$, $N_s=25$, and $l=1, 2, 3$; the solid line represents the curve for $N_R=100$ and the dashed line represents the curve for $N_R=200$.

4. Experimental results

4.1. Evaluation of robustness

In the experiments a region-dependent watermark sequence is generated with the length 256 ($L_w=128$, $N_p=4$, $N_o=8$) and repeated 16 times. The parameters of initial scale, scale step factor between two successive levels, and the number of scale levels, in feature detection, are set as 1.5, 1.1, and 15, respectively. The local search is performed 25 times (five for orientation and five for location), and the threshold T_w is set at the $p_{FP\text{-}Image}=5.0\times10^{-5}$. In the related feature-based methods [17,18,22,24], $p_{FP\text{-}Image}$ is set to about 5.0×10^{-6}, 10^{-4}, 5.0×10^{-4}, and 3.1×10^{-4}, respectively. Therefore, the probability of false positive detection is low enough to claim that the robustness evaluation is meaningful. The experiments are conducted on three well-known 512×512 images, Lena, Baboon, Peppers, and the other 100 images collected from the Uncompressed Color Image Database (UCID) [45]; these images are converted into gray-level images to test. Fig. 5 shows the three images watermarked by the proposed method. The peak-signal-to-noise-ratio (PSNR) values between the cover image and its watermarked image for Lena, Baboon, and Pepper are 41.51 dB, 39.36 dB, and 42.21 dB, respectively. The PSNR values for those 100 images are between 38 dB and 45 dB. Clearly, it is difficult to visually distinguish the cover image from the watermarked one.

1) *Performance of the feature region selection*: the performance of the feature region selection procedure will affect the watermark robustness. To verify the effectiveness of the proposed selection procedure, we first evaluate the repeatability ratio between the selected feature regions in the original image and the ones in its attacked version for the above-mentioned 100 test images. We conducted the attacks listed in the standard benchmark program, StirMark [35], a print-scan attack, and six cropping attacks that include three centered cropping attacks, two ROI cropping attacks, and one random cropping attack in this evaluation. For the print-scan attack, the images were printed with 300 dpi by HP LaserJet 4350. The printed images were scanned with 300 dpi by the Fuji Xerox DocuPrint C3290FS, and then were cropped and resized using bicubic interpolation to their original size. The centered cropping attacks cut off the surrounding areas and the ROI cropping attacks remove the parts of no interest. The random cropping attack retains a region of a randomly decided size in the target image and removes the rest of that image. In the evaluation, the region-to-region correspondence between a selected region in the original image and a relative one in the attacked image is obtained if their distance in location of center point is less than 1.5 pixels and their surface error in cover area is less than 20% [32]. We calculate the repeatability ratio that denotes the number of region-to-region correspondences to the number of selected feature regions in the original image. The average repeatability ratio of those 100 test images is shown in Table 1. We can observe that the ratio is about 0.55 in average for the noise-like

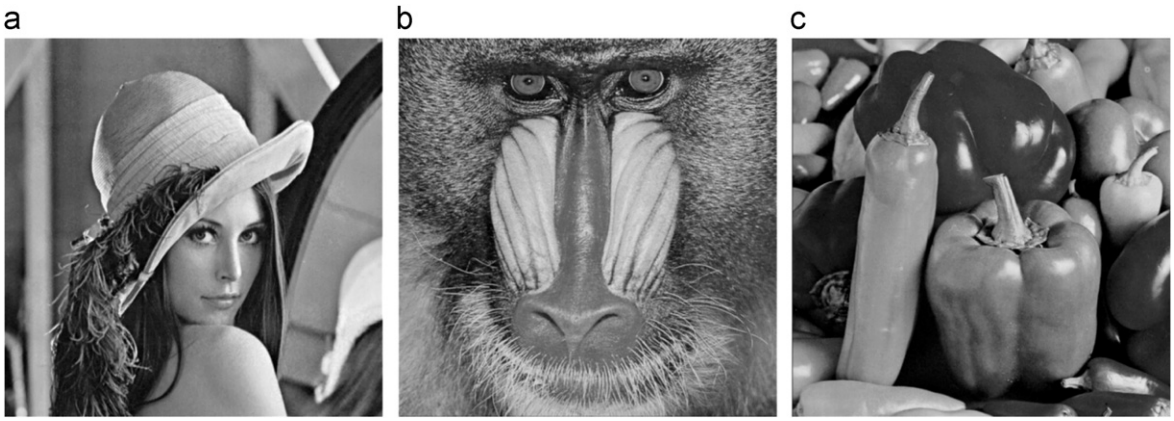

Fig. 5. Watermarked images (a) Lena (41.51 dB), (b) Baboon (39.36 dB), and (c) Pepper (42.21 dB) processed by the proposed watermarking algorithm.

Table 1
Repeatability ratio of the selected regions and watermark detection results for the 100 images collected from the UCID image database [45] against geometric and noise-like signal processing attacks.

Attacks	Repeatability ratio	BER	Correlation coefficient	Detection ratio	Detection failures
Centered cropping 10%	0.50	0.27	0.469	0.37	0
Centered cropping 25%	0.35	0.27	0.467	0.26	0
Centered cropping 50%	0.13	0.28	0.443	0.10	0
Random cropping	0.22	0.26	0.481	0.18	0
ROI cropping-1	0.20	0.26	0.479	0.16	0
ROI cropping-2	0.21	0.26	0.475	0.17	0
Rotation 5+auto-cropping	0.46	0.23	0.550	0.33	0
Rotation 15+auto-cropping	0.33	0.22	0.555	0.23	0
Rotation 30+auto-cropping	0.24	0.23	0.549	0.17	0
Rotation 45+auto-cropping	0.21	0.23	0.538	0.16	0
Scaling 0.75	0.20	0.24	0.515	0.09	0
Scaling 0.9	0.61	0.26	0.473	0.12	0
Scaling 1.1	0.76	0.23	0.549	0.15	0
Scaling 1.2	0.34	0.23	0.543	0.18	0
Aspect ratio change (0.9 1.0)	0.64	0.34	0.321	0.10	5
Aspect ratio change (1.0 1.1)	0.72	0.34	0.319	0.12	2
Line removed-17-row, 5-column	0.74	0.33	0.341	0.24	0
Line removed-5-row, 17-column	0.74	0.33	0.345	0.27	0
Linear (1.013, 0.008, 0.011, 1.008)	0.71	0.27	0.470	0.46	0
Linear (1.010, 0.013, 0.009, 1.011)	0.70	0.27	0.466	0.46	0
Random bending	0.27	0.30	0.403	0.15	0
JPEG 50	0.77	0.31	0.387	0.40	0
JPEG 30	0.71	0.32	0.369	0.26	0
Median 3×3	0.49	0.30	0.403	0.28	0
Median 5×5	0.18	0.31	0.371	0.08	1
Sharpening filter	0.61	0.29	0.423	0.33	0
Gaussian filter	0.52	0.27	0.464	0.34	0
Print–scan	0.59	0.31	0.389	0.24	0

signal processing attacks and 0.44 in average for the geometric attacks. Therefore, the selected regions are stable and can resist most of the attacks.

Secondly, we consider the overall area of the watermarked regions that are selected from all extracted feature regions for the three well-known images and 100 additional images. If the watermarked regions cover most area of an image, it will decrease the risk of failure in detecting a watermark under random and ROI cropping attacks. As shown in Fig. 6, the circular regions are the watermarked regions selected by the proposed method for the three well-known images. Obviously, these regions cover most of the image area without overlapping. We also calculate the ratio of overall area of the watermarked regions over all extracted feature regions as listed in Table 2. The result is compared with those obtained by two conventional feature-based watermarking methods, namely the cornerness-based method and the density-based method. The first method adopts corner responses to select non-overlapping watermarked regions [18]. The second method refers to the number of neighboring feature points insides a region to remove the overlapping regions [17,22]. To make the comparison as fair as possible, the only difference in three methods is the region selection process, while

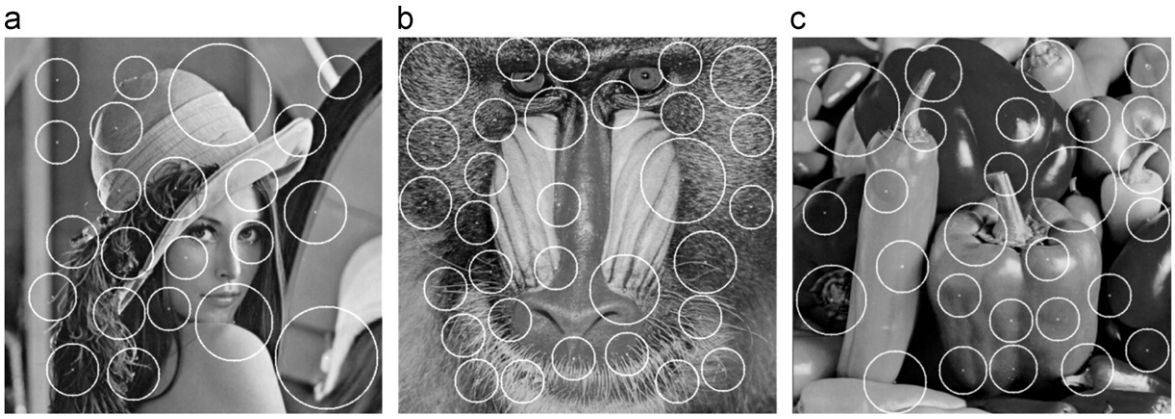

Fig. 6. Watermarked regions selected by the proposed method in (a) Lena, (b) Baboon, and (c) Peppers images.

Table 2
Comparison of the ratio of overall watermarked area over all extracted feature regions on Lena, Baboon, and Peppers images.

Methods	Lena (%)	Baboon (%)	Peppers (%)
Cornerness-based method	44.0	39.6	38.6
Density-based method	46.5	53.7	56.1
Proposed method	63.4	62.4	61.6

the feature detection process is the same. From Table 2, we can observe that the ratio of cover area using our method is about twenty percent higher than one of the cornerness-based method, and about ten percent higher than one of the density-based method. For the additional 100 test images, the average ratios of coverage area of the watermarked regions over all extracted feature regions are 59.9%, 54.6%, and 42.2% using the proposed method, density-based method, and cornerness-based method, respectively. Though it is hard to obtain a region set covering the whole image due to the variety of region sizes and the limitation of non-overlapping between regions, our method achieves the greatest coverage in the comparison with some existing methods [17,18,22]. For the procedure in [24], we think that the MST clustering algorithm could be used to select the regions with the greatest distribution if the distance constraint is set appropriately, and it can remove the region overlaps in each group based on the cover range. However, the issue of feature region selection related to the cropping attacks is not considered in their paper. Comparing with the track-with-pruning selection procedure in [26], the proposed selection procedure is more suitable to achieve the desired goal; because the optimal value is unknown, the track-with-pruning procedure needs to search all possible candidate region sets to determine the best region set (there are $2^{N_R}-1$ possible candidate region sets for N_R feature regions).

However, the coverage of feature regions extracted by the feature detector may affect the distribution of following feature region selection. Basically, this issue could be improved by applying the feature detectors that can extract feature regions with larger coverage over an image, such as the SURF, SFOP, and salient region detectors [44]. Our feature region selector can work in coordination with these feature detectors and determine a non-overlapping feature region set with larger distribution over the whole image. In this paper, the feature region selection work is formulated as an optimum problem, constricted by image quality and regions' overlapping circumstances, and solved by a genetic algorithm-based procedure. According to experimental results, the non-overlapping feature region set selected by the proposed method achieves the greatest coverage in comparison with some existing methods [17,18,22] from the feature regions extracted by the Harris–Laplacian detector. This advantage is also kept when other feature detectors are adopted. Nevertheless, the performance of these detectors is worse than the Harris–Laplacian detector to defend other attacks. This issue deserves further investigation in the future.

2) *Resistance to cropping attacks*: To validate the robustness of our method against cropping attacks, we conducted the six cropping attacks mentioned above on the watermarked images. Fig. 7 shows the cropped versions of the watermarked Lena image after attack. The watermark detection results with respect to the six cropping attacks are illustrated in Tables 1 and 3. The detection ratio refers to the ratio of the number of successfully detected regions with respect to the total number of watermarked regions in an image. Table 1 shows the average detection ratio of the 100 test images, which denotes the mean of detection ratios from those correctly detected images under the conducted attacks. The criterion of detection failures denotes the number of images whose hidden watermarks cannot be detected from any regions among the 100 test images. The coefficient of linear correlation and bit error rate (BER) between the original watermark and the extracted watermark from the correctly detected watermarked regions are also calculated, and their average values are illustrated in Table 1. The experimental results for Lena, Baboon, and Peppers images in Table 3 are also compared with those of several existing feature-based methods, Tang and

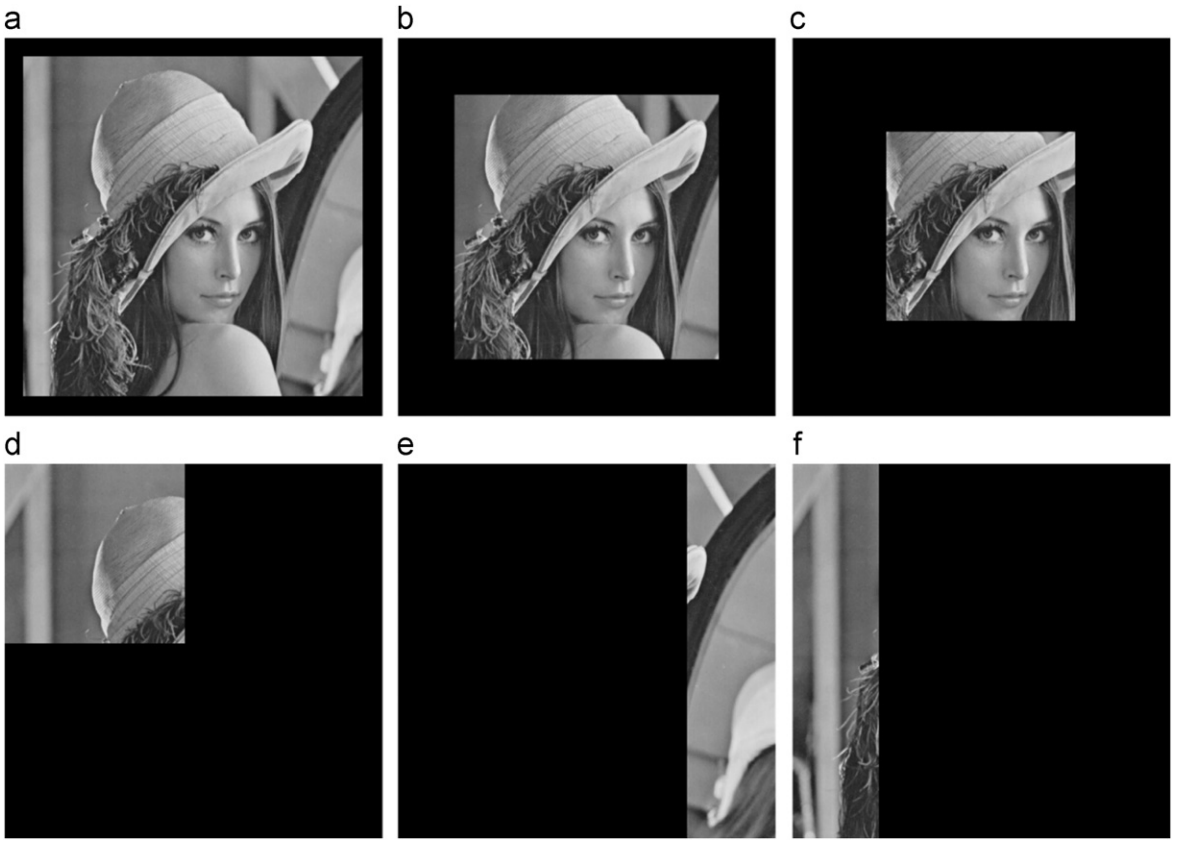

Fig. 7. Watermarked Lena image cropped by (a) centered cropping 10%, (b) centered cropping 25%, (c) centered cropping 50%, (d) random cropping, (e) ROI cropping-1, and (f) ROI cropping-2.

Hang's method [17], Seo and Yoo's method [18], Wang et al. [22], and Deng et al. [24]. The symbol—indicates that the test result was not provided in the reference paper. It can be found that the proposed method is clearly robust against cropping attacks since there is at least one survival region. Referring to Eq. (32), the false-positive rate is low enough to confirm that the robustness of our method is meaningful. Previous methods [17,18,22,24] show robustness against centered cropping attacks, but may fail in ROI or random cropping attacks since their watermarked regions are not sufficiently distributed over an image. For example, Fig. 8 demonstrates the watermarked regions on the Lena image in [17,18,22,24] whose coverage of the regions is less than ours, as shown in Fig. 6(a). The watermarked Lena images in [17,18,22,24] are not able to resist the random cropping or the ROI cropping.

3) *Resistance to noise-like signal processing and other geometric distortions*: the standard benchmark program, StirMark [35], and the print-scan attack mentioned above are adopted to evaluate the robustness of test images against different attacks. Tables 1 and 3 illustrate the detection results against various attacks, including JPEG compression, median filter, Gaussian filter, sharpening filter, print-scan attack, rotation scaling, aspect ratio change, line removed, linear transform, and random bending. We can observe that the images watermarked by the proposed method are robust against most attacks, including noise-like signal processing and geometric distortions. Although the average detection ratio is not exactly high for all attacks, it is enough to prove the existence of the hidden watermarks in images. Referring to the feature-based watermarking methods [16–27], the copyright of an image can be confirmed if its watermark can be detected from at least one feature region under the low probability of the false positive detection. However, the detected feature regions are less resistance to the aspect ratio attacks since the Harris–Laplacian detector use the uniform Gaussian scale-space [31,32]. The repeated watermark embedding is adopted to mitigate this problem, but it still fails on few images since there is a large difference between the embedded region and its corresponding one after attack. Compared with the results in [17,18], our method has better resistance to some attacks. For example, the watermarked Lena and Baboon are undetectable after Rotation 5 degrees in [17], and those in [18] cannot tolerate JPEG 30. The methods presented in [22,24] also demonstrated the robustness on watermarking. But the coverage issue of the selected watermarked regions was not considered in these methods, and it may be fragile against the ROI cropping or random attacks.

Table 3
Comparison of watermark detection results for Lena (L), Baboon (B), and Peppers (P) image against geometric and noise-like signal processing attacks. The detection ratio denotes the ratio of the number of successfully detected regions with respect to the total number of watermarked regions. The symbol – indicates that the test result was not provided in the reference paper.

Attacks	Proposed method			Method in [17]			Method in [18]			Method in [22]			Method in [24]		
	L	B	P	L	B	P	L	B	P	L	B	P	L	B	P
Centered cropping 10%	8/22	6/31	11/27	2/8	2/11	2/4	–	–	–	–	–	–	7/13	10/17	7/18
Centered cropping 25%	5/22	4/31	9/27	–	–	–	4/7	1/7	2/8	–	–	–	–	–	–
Centered cropping 50%	1/22	2/31	2/27	–	–	–	–	–	–	4/6	6/12	5/8	–	–	–
Random cropping	3/22	2/31	5/27	–	–	–	–	–	–	–	–	–	–	–	–
ROI cropping-1	1/22	5/31	4/27	–	–	–	–	–	–	–	–	–	–	–	–
ROI cropping-2	1/22	3/31	9/27	–	–	–	–	–	–	–	–	–	–	–	–
Rotation 5+auto-cropping	3/22	5/31	13/27	0/8	0/11	0/4	–	–	–	4/6	5/12	5/8	8/13	8/17	10/18
Rotation 15+auto-cropping	2/22	1/31	8/27	–	–	–	–	–	–	3/6	4/12	4/8	–	–	–
Rotation 30+auto-cropping	1/22	1/31	3/27	–	–	–	–	–	–	2/6	4/12	2/8	5/13	8/17	7/18
Rotation 45+auto-cropping	1/22	1/31	3/27	–	–	–	2/7	1/7	1/7	–	–	–	–	–	–
Scaling 0.75	1/22	1/31	2/27	–	–	–	3/7	0/7	6/8	–	–	–	–	–	–
Scaling 0.9	1/22	1/31	3/27	–	–	–	4/7	2/7	6/8	3/6	5/12	3/8	–	–	–
Scaling 1.1	2/22	2/31	6/27	–	–	–	–	–	–	–	–	–	–	–	–
Scaling 1.2	1/22	2/31	9/27	–	–	–	–	–	–	–	–	–	–	–	–
Aspect ratio change (0.9 1.0)	1/22	1/31	1/27	–	–	–	–	–	–	–	–	–	–	–	–
Aspect ratio change (1.0 1.1)	1/22	1/31	2/27	–	–	–	–	–	–	–	–	–	–	–	–
Line removed-17-row, 5-column	3/22	3/31	8/27	–	–	–	–	–	–	–	–	–	–	–	–
Line removed-5-row, 17-column	6/22	5/31	8/27	0/8	3/11	1/4	5/7	1/7	5/8	–	–	–	7/13	7/17	8/18
Linear (1.013, 0.008, 0.011, 1.008)	5/22	6/31	13/27	4/8	5/11	0/4	7/7	0/7	5/8	–	–	–	9/13	6/17	7/18
Linear (1.010, 0.013, 0.009, 1.011)	5/22	5/31	12/27	4/8	4/11	1/4	7/7	1/7	7/8	–	–	–	7/13	7/17	10/18
Random bending	4/22	3/31	7/27	–	–	–	4/7	0/7	3/8	3/6	7/12	6/8	7/13	12/17	12/18
JPEG 50	7/22	5/31	9/27	5/8	7/11	3/4	1/7	1/7	4/8	4/6	8/12	6/8	11/13	15/17	15/18
JPEG 30	3/22	4/31	4/27	2/8	4/11	0/4	0/7	0/7	4/8	2/6	8/12	4/8	10/13	14/17	16/18
Median 3×3	7/22	3/31	6/27	1/8	2/11	1/4	–	–	–	3/6	7/12	4/8	7/13	12/17	16/18
Median 5×5	2/22	1/31	3/27	–	–	–	–	–	–	–	–	–	–	–	–
Sharpening filter	5/22	4/31	7/27	4/8	4/11	4/4	1/7	0/7	5/8	3/6	6/12	5/8	–	–	–
Gaussian filter	6/22	5/31	9/27	5/8	8/11	1/4	3/7	1/7	5/8	–	–	–	5/13	8/17	11/18
Print-Scan	2/22	2/31	1/27	–	–	–	–	–	–	–	–	–	–	–	–

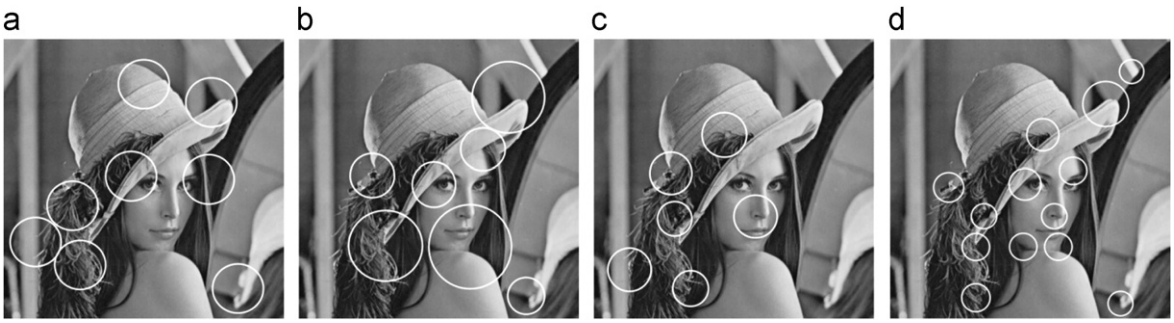

Fig. 8. Watermarked feature regions in Lena image selected by (a) [17], (b) [18], (c) [22], and (d) [24].

4.2. Security analysis

The three tasks of feature detection, feature region selection, and region-dependent watermarking in the proposed method relate to security enhancement. Detailed analysis about region-dependent watermarking has previously been presented [39]. However, there still exist possible security threats to the feature-based watermarking methods while only using the region-dependent watermark, such as the copy attack to similar feature regions. For example, an attacker can estimate the hidden region-dependent watermark from a feature region A in a watermarked image. Then a feature region B, which is similar content to the region A, is searched from the target image. Finally, the estimated watermark from region A is added into region B for forming a counterfeit watermarked image. The copy attack will be successful because the feature regions' locations and sizes are publicly known, and the feature regions with similar contents have the similar media hashes. In order to mitigate the security risk, we incorporate randomization in feature detection and feature region selection. In this paper, we focus on security analysis for feature detection and feature region selection. According to *Kerckhoffs' principle*, it is assumed that malicious attackers know details of the watermarking method except for the secret key. Under this assumption, the probability of a successful attack depends on the randomness of watermarking [40].

Estimating or forging a watermark will be more difficult when the watermarking method has a higher degree of randomness. Therefore, a differential entropy function [41] is employed to evaluate the security capability of the proposed method on feature region detection and selection. The differential entropy, $h(y)$, of a continuous random variable y is defined as

$$h(y) = -\int_{\mathbb{Y}} f(y)\log f(y)dy \tag{33}$$

where $\mathbb{Y}$ is the support set of the random variable and $f(\cdot)$ is probability density function.

The process of feature detection can be separated into deterministic and random parts. The deterministic part is to identify feature points and their characteristic scales, and this part is easily recognized by attackers since its details are all public. The random part is to obtain each circular feature region with a secret radius equal to the product of the characteristic scale of corresponding feature point and a secret key. We assume that a characteristic scale is a constant factor σ_c and a secret key is uniformly distributed over the interval $[\alpha_{\min},\alpha_{\max}]$. Considering various radii for generating feature regions, the probability density function of a secret radius r is given by

$$f(r) = \begin{cases} \frac{1}{\alpha_{\max}\sigma_c - \alpha_{\min}\sigma_c}, & if\ \alpha_{\min} \le \alpha \le \alpha_{\max}; \\ 0, & otherwise. \end{cases} \tag{34}$$

The differential entropy of the feature region detection can be written as

$$\begin{aligned} h(r) &= \int_{\alpha_{\min}\sigma_c}^{\alpha_{\max}\sigma_c} f(r)\log_2\frac{1}{f(r)}dr \\ &= -\int_{\alpha_{\min}\sigma_c}^{\alpha_{\max}\sigma_c} \frac{1}{\alpha_{\max}\sigma_c - \alpha_{\min}\sigma_c}\log_2\left(\frac{1}{\alpha_{\max}\sigma_c - \alpha_{\min}\sigma_c}\right)dr \\ &= \log_2(\alpha_{\max}\sigma_c - \alpha_{\min}\sigma_c). \end{aligned} \tag{35}$$

The randomness in feature region selection is mainly on the key-dependent pseudo-random number, β_j, which are Gaussian distributed with mean μ and variance σ^2. As shown in Eq. (9), those numbers are combined with the radii of extracted feature regions as the fitness of GA-based heuristics. The combined values can be considered as the weighted Gaussian distributed random variables, y, with mean and variance shown by

$$E(y) = \mu \cdot r \tag{36}$$

$$Var(y) = \sigma^2 \cdot r^2. \tag{37}$$

Its probability density function is given by

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2 r^2}} e^{(-(y-\mu r)^2)/2\sigma^2 r^2}. \tag{38}$$

Therefore, the differential entropy of the feature region selection can be written as

$$\begin{aligned} h(y) &= \int f(y)\ln\frac{1}{f(y)}dy = -\int \frac{1}{\sqrt{2\pi\sigma^2 r^2}} e^{-(y-\mu r)^2/2\sigma^2 r^2} \\ &\quad \times \left[\frac{-(y-\mu r)^2}{2\sigma^2 r^2} - \ln\sqrt{2\pi\sigma^2 r^2}\right]dy = \frac{1}{2} + \frac{1}{2}\ln 2\pi\sigma^2 r^2 \\ &= \frac{1}{2}\ln e + \frac{1}{2}\ln 2\pi\sigma^2 r^2 = \frac{1}{2}\ln 2\pi e\sigma^2 r^2. \end{aligned} \tag{39}$$

Changing the base of the logarithm, the differential entropy is

$$h(y) = \frac{1}{2}\log_2 2\pi e\sigma^2 r^2. \tag{40}$$

As a result, the degree of randomness to the feature detection and feature region selection of the proposed scheme can be estimated by Eqs. (35) and (40), respectively. Moreover, as the interval $[\alpha_{\min},\alpha_{\max}]$ or the variance σ^2 is increased, the degree of security will be enhanced. It should be noted that the interval controlling the secret radius is constricted to the characteristic scale, the image size, and the watermark length. The secret radius should not be too small since the capacity of feature region cannot be smaller than the watermark length. On the other hand, it should not be too large since the range of feature region cannot exceed the image size.

Table 4 compares degrees of watermark security of our method and other existing feature-based methods [17,18,22,24]. The symbols ● and ○ denote the method with and without a specific property, respectively. The region-dependent watermarking avoids the problem of hiding the same watermark multiple times, and it was not explicitly considered [17,18,22,24]. The randomness of feature detection and feature region selection prevents an attacker from estimating the exact range and location of the watermarked area. The method in [17] extracts feature regions by Mexican Hat wavelet scale interaction and selects feature regions based on the number of neighboring feature points insides a region; however its two processes are public to adversaries. The watermarking method proposed in [18] also does not include randomness in their feature detection and selection process. In [22], Harris–Laplacian detector is adopted and the local characteristic regions are extracted by a secret integer. Its differential entropy is the same as our analysis here about feature detection, but the limitation in interval values reduces its randomness, as mentioned above. In [24], the distance constraint $\mathcal{D}$ in the feature selection process is regarded as a secret parameter. We assume that $\mathcal{D}$ is uniformly distributed over the interval $[\mathcal{D}_{\min},\mathcal{D}_{\max}]$, so the differential entropy is $\log_2(\mathcal{D}_{\max}-\mathcal{D}_{\min})$. However, the randomness is limited since the interval is constricted by the region size and image size. A too large value of $\mathcal{D}$ may cause that only one feature region is retained.

Table 4
Comparison of watermark security degree between the proposed method and related methods. The symbols ● and ○ denote the method with and without a specific property, respectively.

Property	Proposed method	Method in [17]	Method in [18]	Method in [22]	Method in [24]
Region-dependent watermark	●	○	○	○	○
Randomness of feature detection	●	○	○	●	○
Randomness of feature region selection	●	○	○	○	●

On the other hand, a too small $\mathcal{D}$ makes the clustering ineffective.

4.3. Discussions

In this section, we discuss how to balance the three conflicting factors: robustness, capacity, and imperceptibility in the three main processes of our method: feature detection, feature region selection, and watermark insertion. In the feature detection, there is a trade-off among the three factors while determining the secret key for the key-dependent radius of the feature region. A large value of the secret key would increase the capacity of the region to be watermarked, but the robustness and imperceptibility would be decreased [22,24]. Also, a too large value would cause that the ranges of most feature regions exceed the image size. On the other hand, the value should not be too small since the capacity of feature region cannot be smaller than the watermark length. Therefore, we consider that the secret key should be large enough to make all feature regions' sizes equal or larger than the watermark length, and should be small enough to make the ranges of most feature regions not exceed the cover range of an image in our empirical study.

In the feature region selection process, the capacity is regarded as a constant since the same watermark is embedded into each selected feature region in the feature-based watermarking methods, and its size is determined in the feature detection process. So, there is a trade-off between robustness and imperceptibility in the feature region selection. Selecting more feature regions to watermarking will produce more redundant watermarks for an image. This will increase the robustness but decrease the imperceptibility of the image. In the feature region selection process, we use the threshold T_q to limit the quality degradation of the image to be watermarked and select the regions to achieve the best robustness under the limitation. It decides the level of the imperceptibility in the proposed method, and the robustness of a watermarked image is also determined technically. For example, T_q is set as 40 dB by considering the PSNR between an image and its watermarked image. Then, embedding watermark into the selected regions will not degrade the image quality below 40 dB.

Finally, in the watermark insertion process, there is also a trade-off, related to the watermark embedding strength for each region, between robustness and imperceptibility. The capacity is still considered as a constant since the region's size is determined in the feature detection process. A large embedding strength will increase the robustness but decrease the imperceptibility. Here, we use the noise visibility function (*NVF*) to determine the appropriate embedding strength, which can avoid large degradation of image quality. The *NVF* that characterizes the local image properties can achieve the best balance for the two factors [42].

5. Conclusions

In this paper, we develop a novel method to jointly enhance the robustness and security of feature-based image watermarking schemes. The controlled randomization is incorporated in determining the feature regions of an image for mitigating the leakage of secret information. In addition, an optimal selection process is proposed, formulated as the multidimensional knapsack problem and solved by genetic algorithm-based heuristics. The experimental results of robustness evaluation demonstrate that our method can effectively resist various attacks, including noise-like signal processing and geometric distortions. The security evaluation in terms of differential entropy is also derived, and the performance of the proposed method is confirmed.

Acknowledgments

The authors would like to thank anonymous reviewers for giving constructive and useful comments to improve this paper. This work is supported in part by the National Science Council of Taiwan under Grants 97-2221-E-006-144-MY3 and 98-2221-E-006-222-MY3.

References

[1] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, Secure spread spectrum watermarking for multimedia, IEEE Transactions on Image Processing 6 (12) (1997) 1673–1687.
[2] I.J. Cox, M.L. Miller, J.A. Bloom, Digital Watermarking, Morgan Kaufmann, San Francisco, CA, 2001.
[3] F. Cayre, C. Fontaine, T. Furon, Watermarking security: theory and practice, IEEE Transaction on Signal Processing 53 (10) (2005) 3976–3987.
[4] F. Cayre, P. Bas, Kerckhoffs-based embedding security classes for WOA data hiding, IEEE Transaction on Information Forensics Security 3 (1) (2008) 1–15.
[5] L. Pérez-Freire, P. Comesaña, J.R. Troncoso-Pastoriza, F. Pérez-González, Watermarking security: a survey, Transactions on Data Hiding and Multimedia Security I 4300 (2006) 41–72.
[6] L. Pérez-Freire, F. Pérez-González, Spread-spectrum watermarking security, IEEE Transaction on Information Forensics Security 4 (4) (2009) 2–24.
[7] X. Gao, L. An, Y. Yuan, D. Tao, X. Li, Lossless data embedding using generalized statistical quantity histogram, IEEE Transactions on Circuits and Systems for Video Technology 21 (8) (2011) 1061–1070.
[8] B. Chen, G. Wornell, Quantization index modulation: a class of provably good methods for digital watermarking and information embedding, IEEE Transactions on Information Theory 47 (4) (2001) 1423–1443.
[9] J. Eggers, R. Baüml, R. Tzschoppe, B. Girod, Scalar costa scheme for information embedding, IEEE Transaction on Signal Processing 51 (4) (2003) 1003–1019.
[10] J. ÓRuanaidh, T. Pun, Rotation, scale and translation invariant spread spectrum digital image watermarking, Signal Processing 66 (3) (1998) 303–317.
[11] C.Y. Lin, M. Wu, J.A. Bloom, I.J. Cox, M.L. Miller, Y.M. Lui, Rotation, scale and translation resilient watermarking for image, IEEE Transactions on Image Processing 10 (5) (2001) 767–782.
[12] P. Dong, J.G. Brankov, N.P. Galatsanos, Y. Yang, F. Davoine, Digital watermarking robust to geometric distortions, IEEE Transactions on Image Processing 14 (12) (2005) 2140–2150.
[13] D. Simitopoulos, D.E. Koutsonanos, M.G. Strintzis, Robust image watermarking based on generalized radon transformations, IEEE Transactions on Circuits and Systems for Video Technology 13 (8) (2003) 732–745.
[14] M. Kutter, Watermarking resisting to translation, rotation and scaling, Proceedings of the SPIE Multimedia Systems and Applications 3528 (1998) 423–431.
[15] S. Pereira, T. Pun, Robust template matching for affine resistant image watermarks, IEEE Transactions on Image Processing 9 (6) (2000) 1123–1129.

[16] P. Bas, J.M. Chassery, B. Macq, Geometrically invariant watermarking using feature points, IEEE Transactions on Image Processing 11 (9) (2002) 1014–1028.
[17] C.W. Tang, H.M. Hang, A feature-based robust digital image watermarking scheme, IEEE Transaction on Signal Processing 51 (4) (2003) 950–959.
[18] J.S. Seo, C.D. Yoo, Localized image watermarking based on feature points of scale-space representation, Pattern Recognition 37 (7) (2004) 1365–1375.
[19] H.Y. Lee, H. Kim, H.K. Lee, Robust image watermarking using local invariant features, Journal of SPIE Optical Engineering 45 (3) (2006) 037002-1–037002-11.
[20] C.S. Lu, S.W. Sun, C.Y. Hsu, P.C. Chang, Media hash-dependent image watermarking resilient against both geometric attacks and estimation attacks based on false positive-oriented detection, IEEE Transactions on Multimedia 8 (4) (2006) 668–685.
[21] J.S. Seo, C.D. Yoo, Image watermarking based on invariant regions of scale-space representation, IEEE Transaction on Signal Processing 54 (4) (2006) 1537–1549.
[22] X. Wang, J. Wu, P. Niu, A new digital image watermarking algorithm resilient to desynchronization attacks, IEEE Transaction on Information Forensics Security 2 (4) (2007) 633–655.
[23] D. Zheng, S. Wang, J. Zhao, RST invariant image watermarking algorithm with mathematical modeling and analysis of the watermarking processes, IEEE Transactions on Image Processing 18 (5) (2009) 1055–1068.
[24] C. Deng, X. Gao, X. Li, D. Tao, Local histogram based geometric invariant image watermarking, Signal Processing 90 (12) (2010) 3256–3264.
[25] X. Gao, C. Deng, X. Li, D. Tao, Geometric distortion insensitive image watermarking in affine covariant regions, IEEE Transactions on Systems Man and Cybernetics C Applied Review 40 (3) (2010) 278–286.
[26] J.S. Tsai, W.B. Huang, Y.H. Kuo, On the selection of optimal feature region set for robust digital image watermarking, IEEE Transactions on Image Processing 20 (3) (2011) 735–743.
[27] J.S. Tsai, W.B. Huang, C.L. Chen, Y.H. Kuo, A feature-based digital image watermarking for copyright protection and content authentication, in: Proceedings of the IEEE International Conference on Image Processing, 2007, pp. 469–472.
[28] S. Voloshynovskiy, S. Pereira, V. Iquise, T. Pun, Attack modeling: towards a second generation watermarking benchmark, Signal Processing 81 (6) (2001) 1177–1214.
[29] F. Deguillaume, S. Voloshynovskiy, T. Pun, Secure hybrid robust watermarking resistant against tampering and copy attack, Signal Processing 83 (10) (2003) 2133–2170.
[30] C. Harris, M. Stephen, A combined corner and edge detector, in: Proceedings of the Fourth Alvey Vision Conference, 1988, pp. 147–151.
[31] K. Mikolajczyk, C. Schmid, Indexing based on scale invariant interest points, in: Proceedings of the Eighth International Conference on Computer Vision, 2001, pp. 525–531.
[32] K. Mikolajczyk, C. Schmid, Scale and affine invariant interest point detectors, International Journal of Computer Vision 60 (1) (2004) 63–86.
[33] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.
[34] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L.V. Gool, A comparison of affine region detectors, International Journal of Computer Vision 65 (1–2) (2005) 43–72.
[35] A.P. Fabien, Petitcolas, Watermarking schemes evaluation, IEEE Signal Processing Magazine 17 (5) (2000) 58–64.
[36] H. Kellerer, U. Pferschy, D. Pisinger, Knapsack Problems, Springer, Berlin, 2004.
[37] P.C. Chu, J.E. Beasley, A genetic algorithm for the multidimensional knapsack problem, Journal of Heuristics 4 (1) (1998) 63–86.
[38] R.J. Moraga, G.W. DePuy, G.E. Whitehouse, Meta-RaPS approach for the 0–1 multidimensional knapsack problem, Computers and Industrial Engineering 48 (1) (2005) 83–96.
[39] C.S. Lu, C.Y. Hsu, Near-optimal watermark estimation and its countermeasure: antidisclosure watermark for multiple watermark embedding, IEEE Transactions on Circuits and Systems for Video Technology 17 (4) (2007) 454–467.
[40] A. Swaminathan, Y. Mao, M. Wu, Robust and secure image hashing, IEEE Transaction on Information Forensics Security 1 (2) (2006) 215–230.
[41] T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd edition, John Wiley & Sons, New York, 1991.
[42] S. Voloshynovskiy, A. Harrigel, N. Baumgartner, T. Pun, A stochastic approach to content adaptive digital image watermarking, in: Proceedings of the International Workshop on Information Hiding, LNCS 1768, 1999, pp. 211–236.
[43] C. Schmid, R. Mohr, C. Bauckhage, Evaluation of interest point detectors, International Journal of Computer Vision 37 (2) (2000) 151–172.
[44] S. Ehsan, N. Kanwal, A.F. Clark, K.D. McDonald-Maier, Measuring the coverage of interest point detectors, in: Proceedings of the International Conference on Image Analysis Recognition, LNCS 6753, 2011, pp. 253–261.
[45] G. Schaefer, M. Stich, UCID—an uncompressed colour image database, in: Proceedings of the SPIE Storage and Retrieval Methods and Applications for Multimedia, 2004, pp. 472–480.