INFORMATICA, 2005, Vol. 16, No. 3, 333–346

Reliable Information Hiding Based on Support Vector Machine

Yong-Gang FU
Department of Computer Science and Engineering, Shanghai Jiaotong University
200030 Shanghai, China
Jimei University
361021 Xiamen, China
e-mail: fyg@mail.sjtu.edu.cn

Rui-Min SHEN, Li-Ping SHEN
Department of Computer Science and Engineering, Shanghai Jiaotong University
200030 Shanghai, China
e-mail: {rmshen,lpshen}@mail.sjtu.edu.cn

Xu-Sheng LEI
Department of Automation, Shanghai Jiaotong University
200030 Shanghai, China
e-mail: xushenglei@mail.sjtu.edu.cn

Received: June 2004

Abstract. In this paper, a reliable information hiding scheme based on support vector machine and error correcting codes is proposed. To extract the hidden information bits from a possibly tampered watermarked image with a lower error probability, information hiding is modeled as a digital communication problem, and both the good generalization ability of support vector machine and the error correction code BCH are applied. Due to the good learning ability of support vector machine, it can learn the relationship between the hidden information and corresponding watermarked image; when the watermarked image is attacked by some intentional or unintentional attacks, the trained support vector machine can recover the right hidden information bits. The reliability of the proposed scheme has been tested under different attacks. The experimental results show that the embedded information bits are perceptually transparent and can successfully resist common image processing, jitter attack, and geometrical distortions. When the host image is heavily distorted, the hidden information can also be extracted recognizably, while most of existing methods are defeated. We expect this approach provide an alternative way for reliable information hiding by applying machine learning technologies.

Key words: information hiding, support vector machine, digital watermarking, BCH coding.

1. Introduction

Digital properties are readily reproduced and redistributed over the Internet and other medias. However these attractive properties lead to problems enforcing copyright protection issues. As a result, the contributor and distributor of the digital properties are hesitant to provide the access to their intellectual properties. It is realized that conventional cryptographic means are not sufficient since the data is without any protection as soon as it is used, e.g., decrypted and displayed in the case of image or video data. A potential approach to solve this problem is information hiding or digital watermarking (Swanson *et al*., 1998). Information hiding is the imperceptible embedding of information bits (signature) into multimedia data, where the information remains detectable as long the quality of the content itself is not rendered useless. As a branch of information hiding, it is commonly assumed that digital watermarking is only one of several measures that have to be combined to build a good copy protection mechanism (Furon and Duhamel, 2000). A significant merit of digital watermarking over traditional protection methods is to provide a seamless interface, so that users are still able to utilize the protected multimedia transparently.

An information hiding scheme should at least meet the following requirements: (1) perceptual invisible (or transparent); (2) difficult to remove without seriously affecting the image quality; (3) robust resistance to image processing, and attacks.

Developing an algorithm capable of producing signature that fulfills all these requirements is not an easy task. On one hand, the information hiding process should not introduce any perceptible artifacts into the host image. On the other hand, for high robustness it is desirable that the mark amplitude is as high as possible. Therefore, the designation of information hiding method always involves a tradeoff between imperceptibility (or transparency) and robustness. A variety of watermarking or information hiding schemes have been reported recently in the literature, and some nice reviews can be found in (Fabien *et al.*, 1999). However, the research on copyright protection of images is still in its early stage and none of the existing methods is totally effective against malicious attacks.

There are a variety of schemes for hiding information into the original image. Typical schemes for the information hiding in images can be broadly classified into two categories: (i) spatial domain methods which embed the data by directly modifying the pixel values of the original image (Nikolaidis and Pitas, 1998); (ii) transform domain methods which embed the data by modulating the coefficients of properly chosen transform domain such as DCT (Cox *et al.*, 1997; Barni *et al.*, 2000), DFT (Barni *et al.*, 2000), and DWT (Xia *et al.*, 1998). Many of the spatial domain techniques provide simple and effective schemes for embedding an invisible watermark into an image but are not robust to common attacks. Information hiding techniques can be alternatively split into two distinct categories depending on whether the original image is necessary for the watermark extraction or not. Although the existence of the original image facilitates watermark extraction (Cox *et al.*, 1997; Swanson *et al.*, 1996; Podilchuk and Zeng, 1998) to great extent, such a requirement raises two problems: (i) owner of the original image is compelled insecurely to share his works with anyone who wants to check the existence of the signature (Barni *et al.*, 1998), (ii) on the other hand, the searching within the database for the original image that corresponds to a given watermarked image would be very time consuming. Thus, methods capable of revealing the information bits presence without comparing the watermarked and original images would be preferable.

In order to design robust information hiding scheme, Cox *et al.* considered watermarking as a problem of communication with side information (Cox *et al.*, 1999). Also, some watermarking algorithm in literature applied error correcting coding(ECC) to improve the bit error rate(BER) performance, such as Bose-Chaudhuri-Hocquenghen (BCH) coding (Huang *et al.*, 1998; Huang and Yun, 2002), Reed-Solomon (R-S) code (Wu and Hsieh, 2000), and Turbo code (Pereira and Pun, 2000). Recently, efforts are made to use artificial intelligence technique for watermark embedding and extraction. Neural networks are introduced into watermarking in (Yu *et al.*, 2001), which makes the watermark extraction more robust against common attacks. Genetic algorithm is proposed for selection of the best embedding positions in block based DCT domain watermarking (Shieh *et al.*, 2004). We have firstly introduced support vector machine for watermark embedding and extraction in (Fu *et al.*, 2004), in which the watermark is embedded into the host by applying the good learning ability of support vector regression machine, and the watermark extraction is finished by the aids of the well trained support vector machine. We can expect that the combination of information hiding and machine learning techniques might be a good solution for reliable information hiding.

From the observations above, in this paper we propose a novel blind reliable information hiding and recovering scheme which makes use of support vector machine and BCH coding. This work can be considered as an extension of some existing research (Kutter *et al.*, 1998; Yu *et al.*, 2001; Fu *et al.*, 2004). In (Kutter *et al.*, 1998), Kutter proposed a spatial domain watermarking scheme for color image. Then Yu (Yu *et al.*, 2001) improved Kutter's method by applying neural networks. Due to the support vector machine's good learning ability in training process, it can memorize the relationship between the embedded information and corresponding watermarked image. Applying SVM's good generalization abilities and error correcting ability of BCH coding, hidden information extraction can be finished well. Experimental results show good robustness of the proposed scheme against common image processing and attacks. This research is much different from our early work. In this research, the support vector machine is only trained and applied in the information extraction procedure, whereas, in (Fu *et al.*, 2004), the support vector regression machine is applied both in the watermark embedding and extraction process.

The paper is organized as follows: in Section 2, basic conceptions for support vector machine are introduced. The embedding and extraction algorithms of our method are described in Section 3. In Section 4, some experimental results are exhibited. The conclusion is stated in Section 5.

2. An Overview of Support Vector Machine

Support Vector Machine (SVM) is a universal classification algorithm developed by Vapnik and his colleagues (Vapnik, 1995; Vapnik, 1998). In recent years, there have been a lot of interests in studying the applications of SVM on function approximation, pattern recognition problems and so on (Campbell, 2002; Christopher, 1998).

Given a training data set of m samples $\{\vec{x}_i, y_i\}$, $i = 1, \ldots, m$, $y_i \in R$, $\vec{x}_i \in R^n$, where $\vec{x}_i$ is the ith input pattern and y_i is the ith output pattern.

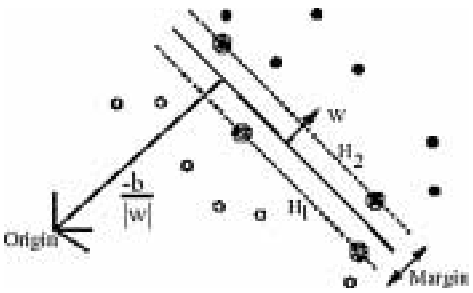

Fig. 1. Linear separating hyperplane.

The support vector machine method supposes we have some hyper-planes that separate the positive samples from negative ones. The point $\vec{x}$ which lies on the hyper-plane satisfies $\vec{w} \cdot \vec{x} + b = 0$, where $\vec{w}$ is normal to the hyper-plane, $|b|/||\vec{w}||$ is the perpendicular distance from the hyper-plane to the origin, and $||\vec{w}||$ is the Euclidean norm of $\vec{w}$. For the linearly separable case, the support vector algorithm simply looks for the separating hyper-plane with the largest margin. This can be formulated as following: suppose that all the training data satisfy the following constraints

$$\begin{aligned} &\vec{x}_i \cdot \vec{w} + b \geqslant 1 \quad \text{for } y_i = 1, \\ &\vec{x}_i \cdot \vec{w} + b \leqslant -1 \quad \text{for } y_i = -1. \end{aligned} \tag{1}$$

This can be combined into one set of inequalities:

$$y_i(\vec{x}_i \cdot \vec{w} + b) \geqslant 1 \quad \forall i. \tag{2}$$

Now consider the points for which the equality in (1) holds. These points lie on the hyper-plane H_1: $\vec{x}_i \cdot \vec{w} + b = 1$ with normal$\vec{w}$ and perpendicular distance from the origin$|1 - b|/||\vec{w}||$. Similarly, the points for which the equality holds in (1) lie on the hyper-plane H_2: $\vec{x}_i \cdot \vec{w} + b = -1$, with normal$\vec{w}$, and perpendicular distance from the origin$| - 1 - b|/||\vec{w}||$. Hence the margin is simply$2/||\vec{w}||$. Thus we can find the pair of hyper-planes that gives the maximum margin by minimizing$||\vec{w}||^2$, subject to constraints (2).

This can be written into a compact form:

$$\text{Minimize } LP = \frac{1}{2}||\vec{w}||^2 \quad \text{s.t. } y_i(\vec{x}_i \cdot \vec{w} + b) \geqslant 1 \quad \forall i = 1, \ldots, m. \tag{3}$$

Since such a problem is difficult to solve, we switch to a Lagrangian formulation of the problem:

$$L = \frac{1}{2}||\vec{w}||^2 - \sum_{i=1}^{m} \alpha_i y_i(\vec{x}_i \cdot \vec{w} + b) + \sum_{i=1}^{m} \alpha_i.$$

Now, we must minimize L with respect to $\vec{w}$, b, and simultaneously require that the derivatives of L respect to all the α_i vanishing, all subject to the constraints $\alpha_i \geqslant 0$. From the dual theory, we can translate the primal problem to the dual form:

$$\text{Maximize } LD = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j \quad \text{s.t. } \alpha_i \geqslant 0, \ \ i = 1, 2, \ldots, m, \quad (4)$$

and the following results are satisfied:

$$\vec{w} = \sum_i \alpha_i y_i \vec{x}_i \quad \text{and} \quad \sum_i \alpha_i y_i = 0. \quad (5)$$

Once we got the solution of (4), they are substituted into (5) and the classifier is finished which can be written as

$$f(x) = \text{sign}\left[\sum_{i=1}^{m} \alpha_i y_i \vec{x} \cdot \vec{x}_i + b\right].$$

For linearly non-separable case, we can find a mapping function φ: $R^n \to R^m$, which map the current space into a higher dimensional one, the data point is separable in the mapped space, and the dot product in the mapped space is avoided by kernel functions $K(\vec{x}, \vec{y})$. For$K(\cdot\,,\cdot\,)$, one typically has the following choices: $K(\vec{x}, \vec{x}_k) = \vec{x}_k \cdot \vec{x}$ (linear SVM); $K(\vec{x}, \vec{x}_k) = (\vec{x}_k \cdot \vec{x} + 1)^d$ (polynomial SVM of degree d); $K(\vec{x}, \vec{x}_k) = \exp(-||\vec{x} - \vec{x}_k||_2^2/\sigma^2)$ (RBF SVM); $K(\vec{x}, \vec{x}_k) = \tanh(k\vec{x}_k \cdot \vec{x} + \theta)$ (two layer neural SVM).

One assumes that the data set in the mapped space is separable, i.e,

$$y_i\big[\vec{w} \cdot \varphi(\vec{x}_i) + b\big] \geqslant 1, \quad i = 1, \ldots, m. \quad (6)$$

Just the same process as linearly separable case, after substituting the kernel function into the formula, one can get the classifier for the linearly non-separable case:

$$f(x) = \text{sign}\left[\sum_{i=1}^{m} \alpha_i y_i \varphi(\vec{x}) \cdot \varphi(\vec{x}_i) + b\right] = \text{sign}\left[\sum_{i=1}^{m} \alpha_i y_i K(\vec{x}, \vec{x}_i) + b\right]. \quad (7)$$

Established on the theory of the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error, SVMs are shown to be very resistant to the over-fitting problem, eventually achieving high generalization performance in solving various classification problems (Vapnik, 1995; Campbell, 2002; Christopher, 1998). Another key advantage of SVM is that the training of SVM is equivalent to solving a linearly constrained quadratic programming problem, so that the solution of SVM is always unique and globally optimal, unlike other networks' training which requires non-linear optimization with the danger of getting stuck into local minima.

3. Information Hiding and Extraction

A watermarking scheme for color image in spatial domain is proposed by Kutter(Kutter *et al.*, 1998), his idea is further developed using neural networks by Yu *et al.* (2001). Here we introduce support vector machine for hidden information bits detection which is based on Kutter's work, the experimental results show convincing reliability. The information hiding procedure can be modeled as a digital communication problem (Huang *et al.*, 2002). As mentioned by Huang, some information bits are embedded into a host image and the host image may go through some manipulations (the signals are transmitted through a noisy channel). Therefore, the detected signals at the receiver end may have some bit errors. To ensure the integrity of the hidden data, we encode the information by using error correcting codes, and applying the good generalization ability of support vector machine. In the experiments of this research, we use the BCH codes because there is an ample selection of block length and code rates. The diagram of the whole system architecture is shown in Fig. 2, where the information hiding procedure is shown in the left part, and the extraction block diagram is depicted in the right.

A. Information hiding scheme

The information bits for our hiding scheme consist of two parts. One is the reference data and the other is the digital signature (logo). They are denoted as $RF = r_1r_2\dots r_K$ and $S = s_1s_2\dots s_L$ respectively, where RF is generated according to a secret key $k1$. We apply the BCH code (n,k) to encode the signature S, where n and k represent the length of BCH codeword and the number of bits in each block, respectively, obtaining a bit stream:

$S' = \{s_i^{'}, i = 1,2,\dots,T\}$, where T is the total number of bits after BCH encoding, i.e., $T = Ln/k$.

Then they are concatenated into a single information sequence $W = RF + S' = w_1w_2\dots w_{K+T}$ for hiding. We further suppose W are binary antipodal sequence, if not, they can be easily mapped into such a sequence by $f(x) = 2x - 1$ from a binary one. The reference data is embedded only for the training of the support vector machine to extract the hidden bits.

Let I be a color image with size $M \times N$, defined by$I = [R, G, B]$, where R, G, B are the three image components corresponding to red, green, and blue channel respectively.

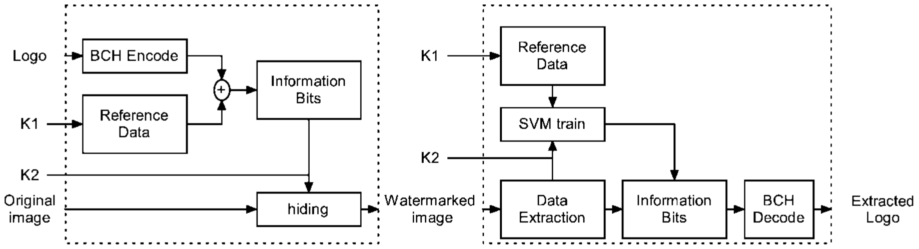

Fig. 2. Information embedding and extracting system architecture.

Suppose $p = (i, j),\ i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$ be the pixel position for further image accessing. In this paper the trained support vector machine is employed to extract the hidden signature S from the tampered image during the extraction stage.

Let $\{p_t = (i_t, j_t)\}|_{t=1,2,\ldots,K+T}$ be the randomly selected position sequence according to another secret key $k2$ provided by the image copyright owner. There are many pseudo random number generators can be applied. Here we adopt the conventional random generator according to uniform distribution. For each selected embedding position, we firstly compute the average pixels intensity on the cross-shaped window with size 5 as following

$$A_{p_t} = \left(\sum_{l=-2}^{2} B(i_t + l, j_t) + \sum_{l=-2}^{2} B(i_t, j_t + l) \right) / 10.$$

To embed the data, the pixel intensities in blue channel are modified as following formula:

$$B_{p_t} \leftarrow A_{p_t} + \alpha_t w_t, \quad t = 1, 2, \ldots, K + T, \tag{8}$$

where α_t is a scaling factor. To embed data bits in the host image as strongly as possible, we vary α according to different characteristics of the host image. This is due to one of the feature of human HVS (Shi and Sun, 1999). Firstly, since the human eyes are not sensitive on the change in the blue channel, the modification of blue channel in RGB image is less detectable by human eyes. Secondly, as to different luminance, human eyes have different sensitivity, i.e., the modification in high luminance is less detectable than that low luminance. Hence the scaling factor α what we selected here is $\alpha_t = cL_{p_t}$, where c is an embedding constant controlling the compromise between robustness and visual quality of the watermarked image, and

$$L_{p_t} = 0.299R_{p_t} + 0.587G_{p_t} + 0.114B_{p_t}$$

is the luminance component of the RGB image at position p_t.

After the hiding procedure, we can obtain the watermarked blue channel image B'. Then, the final watermarked image is reached by integrating the blue channel with the other two color channels, denoted as $I' = [R', G', B']$.

B. Information bits extraction

The information bits extraction diagram is depicted in Fig. 2. Firstly, the reference mark $RF = r_1 r_2 \ldots r_K$ for information bits extraction is generated according to the secret key $k1$. At the same time, the random selected positions for watermark extraction are generated according to copyright owner's key $k2$. These indicate each vector valued pixel hiding the signature message. Then we use extra information extracted from the image and the reference mark to train a support vector machine. Next, the embedded data bits are extracted using the good generalization ability of the trained SVM. Finally, the extracted data is further decoded by BCH decoder (n, k), and the embedded information bits $\widetilde{S}$ are extracted.

The hidden information bits extracting algorithm can be summarized as follows:

Step 1.Generate the reference data $RF = r_1 r_2 \dots r_K$ and the hiding position sequence $\{\rho_t = (i_t, j_t)\}_{t=1,\dots,K+T}$ according to the secret keys $k1$ and $k2$, respectively.

Step 2. For each selected position and a slide cross-shaped window with size 5, the difference between the blue component intensity of central pixel and average intensity of the others within the cross-shaped window is computed $d_{ij} = B'_{ij} - \widetilde{B}'_{ij}$, where $\widetilde{B}'_{p=(i,j)} = (\sum_{r=-2}^{2} B'_{i+r,j} + \sum_{r=-2}^{2} B'_{i,j+r} - 2B'_{ij})/8$. When the watermarked image is undergone some attacks, since the reference mark and signature data is uniformly embedded into the host image, the relationship between the embedded data bits and watermarked image can be memorized by the training process of support vector machine. Using the good generalization ability of the well trained SVM, we can use this information to extract the hidden data bits. Define training dataset:

$$\begin{aligned} Ds &= \big\{d_{i_t-2,j_t}, d_{i_t-1,j_t}, d_{i_t,j_t}, d_{i_t+1,j_t}, d_{i_t+2,j_t}, \\ &\qquad d_{i_t,j_t-2}, d_{i_t,j_t-1}, d_{i_t,j_t+1}, d_{i_t,j_t+2}, r_t\big\}_{t=1,\dots,K} \\ &= \big\{\vec{D}_t, r_r\big\}_{t=1,\dots,K}, \end{aligned} \tag{9}$$

where r_t is the reference label value corresponding to the tth pattern for the training of the SVM and $\vec{D}_t$ is the tth data element in vector form. Applying the dataset and the data labels, we can train the support vector machine, and suppose the trained SVM be

$$f(\vec{x}) = \text{sign}\left[\sum_{t=1}^{K} \lambda_t r_t Ker(\vec{x}, \vec{D}_t) + b\right],$$

where Ker is the selected kernel function, λ_t are the trained coefficients, and b is the bias.

Step 3. The information extraction dataset (excluding the training dataset) can be extracted from the tampered image in a similar way:

$$\begin{aligned} Es &= \big\{d_{i_t-2,j_t}, d_{i_t-1,j_t}, d_{i_t,j_t}, d_{i_t+1,j_t}, d_{i_t+2,j_t}, d_{i_t,j_t-2}, d_{i_t,j_t-1}, \\ &\qquad d_{i_t,j_t+1}, d_{i_t,j_t+2}\big\}_{t=K+1,\dots,K+L} \\ &= \{\vec{D}_t\}_{t=K+1,\dots,K+T}. \end{aligned} \tag{10}$$

Using the well trained support vector machine and the information extraction dataset *Es*, the hidden data bits can be extracted by the output of support vector machine on dataset E, i.e., the extracted data can be denoted as

$$\overline{w}_i = f(\vec{D}_{K+i}), \quad i = 1, 2, \dots, T. \tag{11}$$

Then the final extracted signature is obtained by BCH decoding. To achieve better performance, we adopt the decoding with soft decision, because it is known that about 2-dB improvement in performance can be achieved by using soft decision over hard decision (Huang *et al.*, 2002). Once all the embedded data bits are extracted, the embedded signature $\widetilde{S} = \{\tilde{s}_1\tilde{s}_2 \ldots \tilde{s}_L\}$ is reached after BCH soft decision decoding, and then utilized to identify the copyright of owner's intellectual property by comparing S with $\widetilde{S}$.

4. Experimental Results

Considering the host image capacity, in experiments, we have tried BCH codes (15,5), (15,7), (31,11) and (63,30). As a compromise between the embedded data load and good error correction ability, the BCH code (15, 5) is selected in the final experiments. We have tested the proposed algorithm on RGB images with various content complexities including “Lena”, “pepper”, and

“Baboon”. The experimental results with the “Lena” image of $512 \times 512 \times 3 \times 8$ bits are shown in Fig. 4 and Table 1 respectively. What the signature we adopt is a binary logo image of size 32×25 shown in Fig. 3(b). The logo image is reshaped into line ordered sequence S by row major fashion, then encoded by BCH code, and further modulated into binary antipodal sequence S'.

Some necessary parameters used in our watermarking scheme are determined by experiments. Firstly, there are several common used SVM kernels, including Linear, RBF and Polynomial, and we should select the most suitable one based on many trials on the watermarked image and the attacked ones. From experiments on these three kernels, the Linear and Polynomial can be easily defeated by RBF. As to the kernel parameterσ, all the parameters from 0.1 to 40 with step size 0.2 are tested. We find that when $\sigma \in (5, 10)$, the classified results are all acceptable. Hence we use RBF kernels SVM with width 8 to recover embedded data. The watermark strength is decided by the detection results under attack and the visual quality of watermarked image. Here the embedding strength parameter$c = 0.5$ is used. Finally, the reference watermark length is decided by SVM performance. Since SVM is based on statistical learning, it has good generalization ability. It is not necessary to have a huge dataset to train the SVM. Hence we use 50 randomly generated bits as reference data here.

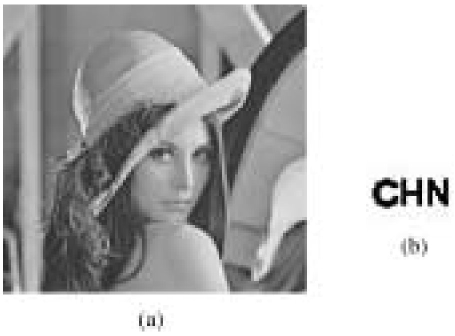

Fig. 3. (a) Original Lena image (512*512); (b) Watermark Logo (32*25 bits).

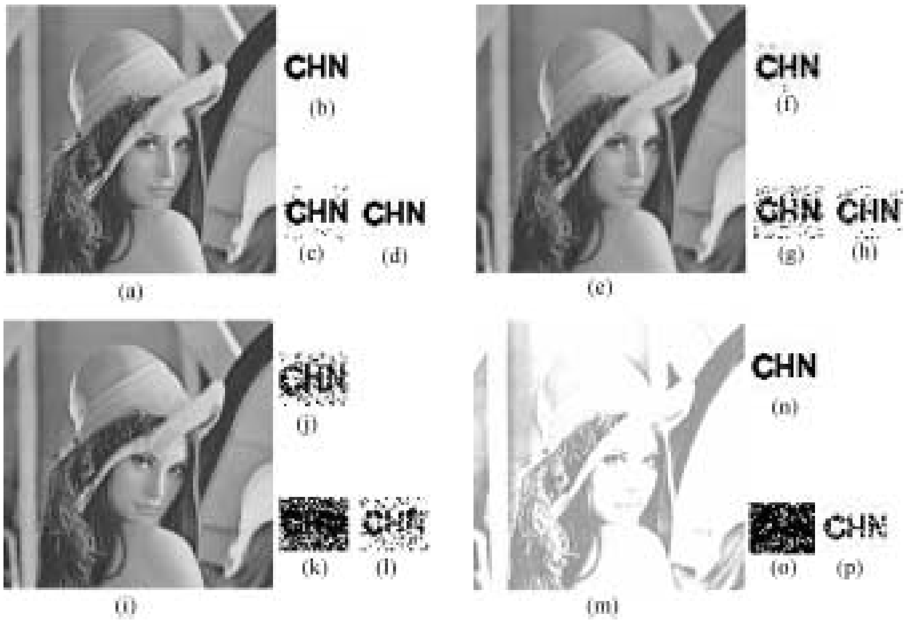

Fig. 4. Experimental results. (a) Watermarked image 512*512; (e) Mosaic; (i) Distorted image; (m) Luminance and contrast enhanced image.(b)(f)(j)(n) Extracted logo using our method from (a)(e)(i)(m), with bit errors 0, 6, 164, 0, respectively.(c)(g)(k)(o) Extracted logo using Kutter's method from (a)(e)(i)(m), with bit errors 22, 77, 387, 598, respectively.(d)(h)(l)(p) Extracted logo using Yu's method from (a)(e)(i)(m), with bit errors 0, 47, 144, 40, respectively.

Table 1

Experimental comparison result on different attacks

Attacks	PSNR(dB)	Bit Error Rate (BER)		
		Our method	Kutter's method	Yu's method
Attack free	38.69	0	0.0275	0
Blurring	37.10	0	0.0275	0
Mosaic(3*3)	26.57	0.008	0.0963	0.0588
Jpeg compression (90)	36.75	0.0036	0.10	0.0462
Jpeg compression (50)	32.55	0.0932	0.2755	0.25
Jpeg compression (20)	31.06	0.3182	0.65	0.42
Filtering (3*3)	25.85	0.0102	0.0775	0.0125
Distortion (15°)	20.57	0.1962	0.4351	0.1625
Distortion (50°)	17.48	0.2925	0.5187	0.2702
Scaling (30%)	27.26	0.0615	0.2375	0.2087
Scaling (300%)	37.34	0	0.0275	0
Jitter (1 row+1 column)	28.35	0.1752	0.4738	0.2211
Rotation (15°)	8.75	0.0675	0.3175	0.1412
75%Luminance+75%Contrast Enhancement	6.92	0	0.6225	0.05
Cropping (25%)	11.19	0.0675	0.2463	0.2212

The original Lena image and watermarked version is shown in Fig. 3(a) and Fig. 4(a) respectively. The peak signal-to-noise-ratio (PSNR) of the marked image by our method and Yu's method with respect to the original image is about 38.69dB and the watermarked image based on Kutter's method is about 2dB higher. Several types of image processing and attacks are simulated to evaluate the performance of our scheme reliability, including blurring, filtering, mosaic, Jpeg compression, scaling, contrast and luminance enhancement, distortion and cropping. Table 1 shows the comparison details of quantity measures for our method, Kutter's method and Yu's one. Due to the limitation of paper space, we exhibit here only three cases of attacks including Mosaic, luminance and contrast enhancement, and distortion attacks for the visual perception.

The watermarked image is shown in Fig. 4(a). When there is no attack, the extracted logo image by our method, Kutter's method and Yu's are shown in Fig. 4(b), (c) and (d), respectively. The embedded information bits can be exactly extracted with no bit errors by our method and Yu's method, but there are 22 bit errors on the extracted information bits using Kutter's method. In the case of Mosaic attack, Fig. 4(e) shows the mosaic version of the watermarked image with parameter3×3, where the PSNR of attacked image is 26.57. Three recovered signature logo from the Mosaic processed image by the three methods are shown in Fig. 4(f)–(h) respectively. The extracted logo images are all distinguishable, and there are only 6 bit errors by our method, whereas there are 77 and 47 bit errors for Kutter's method and Yu's method respectively. In the case of distortion attacks, Fig. 4(i) shows the distorted watermarked image with an angle 30° to the left, the extracted logo image by the three method are depicted in Fig. 4(j), (k) and (l), respectively. From which we can find the logo extracted by our method and Yu's method is still distinguishable, while those extracted from Kutter's is noise like. The processed version after luminance and contrast enhancement is shown in Fig. 4(m). The watermarked image is attacked by a combination of 75% luminance enhancement and 75% contrast enhancement. Three extracted logo image is shown in Fig. 4(n), (o), (p), respectively. The hidden information bits can be successfully extracted from the attacked image without any errors by our method, while there are about 40 bit errors in extracted logo by Yu's method, and there are about 498 bit errors in extracted logo by Kutter's method, i.e., Kutter's method is completely loss.

Some other quantity comparison results are shown in Table 1. Most of the watermarking scheme is defeated by jitter attack because synchronization is the basis of existing watermarking schemes. The bit error rate (BER) of logo extracted by Kutter's method on jitter attacked image with one row and a column shift at position 200 is about 0.4783, which is a noise like logo. Whereas the BER of the logo extracted by our method is only 0.1752, and the logo is still distinguishable. Corresponding to the Jpeg compression attack at different quality factor (QF), Table 1 shows the comparison of the bit error rate (BER), the BER value of our scheme is a little lower than Kutter's method and Yu's method. In the case of rotation, under the condition of rotation left 15°, our method can recover the information bit with BER 0.0675, whereas the value for Kutter's method is 0.3175. When the watermarked image is cropped about 25% at top-left corner, the hidden information can be correctly detected by our method with BER 0.0625, while the extracted information BER recovered by Kutter's method is about 0.24 and the value for Yu's method is 0.2212. From the experimental results in Table 1 and Fig. 4, it shows superior reliability of our proposed scheme against Kutter's method and Yu's method. The main advantages of our proposed scheme come from the good generalization ability of support vector machine. Due to the good learning ability in training process of support vector machine, it can memorize the relationship between the embedded information and corresponding watermarked image. When the watermarked image is tampered, the support vector machine can learn the knowledge and exactly extract the hidden information bits. Furthermore, the BCH coding is helpful for the reliability of the information hiding scheme. While this approach is presented in spatial domain, it can be easily applied to information hiding method in transform domain and video sequences.

5. Conclusions

A novel reliable blind information hiding scheme based on support vector machine and BCH coding is proposed in this paper. The support vector machine can be easily fused with traditional information hiding system to improve the scheme's robustness. By introducing an aiding dataset as reference, the support vector machine can be easily trained and the watermark extraction is finished by the trained support vector machine. Due to the good learning ability of support vector machine, it can learn the relationship between the embedded information and corresponding watermarked image, when the watermarked image is attacked by some intentional or unintentional attacks, the trained support vector machine can successfully recover the right hidden information bits. The experimental result compared with Kutter's and Yu's methods shows outperforming reliability against many different types of attacks. Since the support vector machine training dataset is very small, and it should be trained only once in the whole detection process, this scheme can be easily transplanted to other schemes.

References

Barni, M., F. Bartolini, V. Cappellini, A. Piva (1998). A DCT-domain system for robust image watermarking. *Signal Processing*, **66**(3), 357–372.

Campbell, C. (2002). Kernel methods: a survey of current techniques. *Neurocomputings*, **48**, 63–84.

Christopher, J.C. (1998). A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, Vol. 2. pp. 121–167.

Cox, I.J., J. Kilian, F.T. Leighton, T. Shamoon (1997). Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, **6**(12), 1673–1687.

Cox, I.J., M.L. Miller and A.L. Mckellips (1999). Watermarking as communications with side information. *Proceeding of IEEE*, **87**(7), 1062–1077.

Fabien, A.P., R.J.A. Petitcolas, and M.G. Kuhn (1999). Information hiding – a survey. *Proceeding of IEEE*, **87**(7), 1062–1078.

Fu, Y.G., R.M. Shen, H.T. Lu (2004). Optimal watermark detection based on support vector machine. In *Advances in Neural networks – ISNN'2004, LNCS*, Vol. 3173. pp. 552–557.

Furon, T., and P. Duhamel (2000). Copy protection of distributed contents: an application of watermarking technique. In *Workshop COST 254: Friendly Exchange through the Net*. Bordeaux, France, March 2000.

Huang, J., G.F. Elmasry and Y.Q. Shi (1998). Power constrained multiple signaling in digital image watermarking. In *Proceeding of IEEE Workshop on Multimedia Signal Processing*. pp. 388–393.

Huang, J.W., Q.S. Yun (2002). Reliable information bit hiding. *IEEE Transactions on Circuits and Systems for Video Technology*, **12**(10), 916–920.

Kutter, M., F. Jordan, F. Bossen (1998). Digital signature of color images using amplitude modulation. *J. Electronics Imaging*, **7**(2), 326–332.

Nikolaidis, N., and I. Pitas (1998). Robust image watermarking in the spatial domain. *Signal Processing*, **66**(3), 385–403.

Pereira, S., and T. Pun (2000). Robust template matching for affine resistant image watermarks. *IEEE Transactions on Image Processing*, **9**(6), 1123–1129.

Pereira, S., S. Voloshynovskiy and T. Pun (2000). Effective channel coding for DCT watermarking. In *Proceeding of IEEE International Conference on Image Processing*, Vol. 3. pp. 671–673.

Podilchuk, C.I., and W. Zeng (1998). Image-adaptive watermarking using visual models. *IEEE Journal on Selected Areas in Communication*, **16**(4), 525–539.

Shi, Y.Q., and H. Sun (1999). *Image and Video Compression for Multimedia Engineering*: *Fundamentals, Algorithms, and Standards*. FL: CRC Press, Boca Raton.

Shieh, C.S., H.C. Huang, F.H. Wang, and J.S. Pan (2004). Genetic watermarking based on transform-domain technique. *Pattern Recognition*, **37**, 555–565.

Swanson, M.D., B. Zhu and A.H. Tewfik (1996). Transparent robust image watermarking. In *Proceedings of the International Conference on Image Processing*. pp. 211–214.

Swanson, M.D., M. Kobayashi, and A.H. Tewfik (1998). Multimedia data embedding and watermarking technologies. *Proceedings of IEEE*, **86**(6), 1064–1087.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.

Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley, New York.

Wu, C.F., and W.S. Hsieh (2000). Image refining technique using watermarking. *IEEE Transactions on Consumer Electronics*, **46**, 1–5.

Xia, X.G., C.G. Boncelet, and G.R. Arce (1998). Wavelet transform based watermark for digital images. *Optics Express*, **3**(12), 497–508.

Yu, P.T., H.-H. Tsai, J.-S. Lin (2001). Digital watermarking based on neural networks for color images. *Signal Processing*, **81**, 663–671.

Y.-G. Fu received the BS and MS degree in computational mathematics from Xi'an Jiao Tong University, Xi'an China, in 1995 and 1998 respectively. And now he is a PhD student in computer science from Shanghai Jiao Tong University, Shanghai, China. His main research interests include artificial intelligence, digital watermarking, and multimedia security.

R.-M. Shen is professor and PhD student supervisor in Shanghai Jiao Tong University, Shanghai China. His main research interests includes data mining, E-learning, multimedia retrieval, multimedia security.

L.-P. Shen is a lecturer in Shanghai Jiao Tong University, Shanghai China. Her main research interests include grid computing, and artificial intelligience.

X.-S. Lei is a PhD candidate student in Shanghai Jiao Tong University, Shanghai, China. His main research interests include automatic control, neural networks, and machine learning.

Patikimas informacijos slėpimas taikant atraminių vektorių metodą

Yong-Gang FU, Rui-Min SHEN, Li-Ping SHEN, Xu-Sheng LEI

Darbe siūloma patikimo informacijos slėpimo schema, kurios pagrindą sudaro atraminių vektorių metodas ir klaidas taisantys kodai. Informacijos slėpimas modeliuojamas kaip suspausto vandenženklio vaizdas. Paslėptosios informacijos ištraukimas nagrinėjamas kaip skaitmeninio signalo perdavimo ryšio kanalu uždavinys. Siūlomos schemos patikimumas tikrinamas įvairiomis atakomis. Aprašytų straipsnyje eksperimentų rezultatai rodo gerą pasiūlyto metodo patikimumą.