Expert Systems with Applications 40 (2013) 7579–7587

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A low-cost method for reliable ownership identification of medical images using SVM and Lagrange duality

Mir Shahriar Emami *, Khairuddin Omar

Faculty of Information Science and Technology, The National University of Malaysia (UKM), 43600 UKM Bangi, Selangor, Malaysia

ARTICLE INFO

Keywords:
Medical image watermarking
Support vector machine
Ownership identification
False positive error
False negative error
Spatial domain

ABSTRACT

Prevention of false positive and false negative errors is a major challenge for ownership identification and proof of ownership applications using digital image watermarking. Such errors are more critical with sensitive data, such as electronic patient records (EPRs) in medical image watermarking. A false positive error is a watermark detection error, which means that a watermark is detected in a media where there is no watermark. In contrast, a false negative error is an inability of the watermark detector to detect an embedded watermark in a watermarked image. These errors make ownership assessments unreliable, and the incorrect ownership identification of a patient's record could result in failure of the correct diagnostics and treatments. To address this type of problem, a low-cost technique based on a support vector machine (SVM) and Lagrange duality was proposed to achieve reliable approximations for ownership identification in medical image watermarking without requiring the correction of attacked watermarked images. In this technique, the results of the ownership evaluation are categorized into two independent classes, namely watermark-detected and watermark-not-detected, and higher geometric margins between these classes are associated with higher reliability. To address additional situations with false positive and false negative errors, four different situations, including watermarked, unwatermarked, attacked watermarked and attacked unwatermarked images, were investigated. Experiments were conducted on duo-ISB-bit-plane (BiISB) watermarking using the histogram intersection (HI) technique as a testing platform under JPEG2000 and JPEG image compression attacks and using two groups of images: standard image processing images and X-ray medical images. The experimental investigations revealed that the HI technique guarantees that the rightful owner can be reliably identified even after severe attacks and in the face of context similarities between the watermark and the embedding pixels of the host image.

1. Introduction

Digital watermarking as a complementary approach for cryptography has been widely used in recent decades. Ownership identification, proof of ownership, integrity authentication, broadcast monitoring, transaction tracking, and copy control are the most popular applications of digital watermarking. However, one serious problem of many watermarking applications, including ownership identification and proof of ownership, is the false positive probability (Run, Horng, Lai, Kao, & Chen, 2012;Tian, Bloom, & Baum, 2007), i.e., the false positive error. The false positive error or false positive in short is the probability of identifying a watermark in a media that contains no watermark (Cox, Miller, Bloom, Fredrich, & Kalker, 2008). This probability should be as low as possible in a watermarking application. For example, for the proof of ownership application, the false positive probability should be less than 10^{-6} (Cox et al., 2008). Therefore the false positive rate for this application should be approximately zero and that proof of ownership cannot otherwise be reliable. Another problem in ownership identification is the false negative probability (Cox et al., 2008), i.e., the false negative error. The false negative error or false negative in short is the occurrence probability that the watermark detector cannot detect a watermark in the presence of that watermark in a watermarked image. In other words, if the detector fails to identify a rightful owner in a watermarked image, a false negative error occurs. Both false positive and false negative errors bring about unreliable ownership identification where reliability is a necessity, especially for sensitive data such as electronic patient records (EPRs) including X-ray medical images. In such data, an unreliable decision about the ownership regarding an electronic patient record could lead to failures in diagnostics and treatments. False positive and false negative errors can originate from many resources. For example, the context similarity between an unwatermarked host image and a typical watermark can produce such

* Corresponding author. Tel.: +60 127747925.
E-mail addresses: shemami85@yahoo.com (M.S. Emami), ko@ftsm.ukm.my (K. Omar).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.07.062

errors. Watermarking attacks are other resources that can create false positive or false negative errors, and such attacks (including JPEG compression, Gaussian noise, rotation, and skewing) can change or desynchronize the embedded watermark(s) as a result of a change in pixel values or desynchronization of the pixel positions. This problem can be even more severe when the watermark is too small, such as a sub-watermark in the duo-ISB-bit-plane (BiISB) watermarking technique (Emami, Sulong, & Seliman, 2012a, 2012b). A study of the literature shows that a substantial amount of effort has been directed at preventing false positive and false negative errors, regardless of the nature of the watermark. These efforts can be classified as follows:

- Attempts to propose ideal robust watermarking schemes. Visual imperceptibility, capacity limitation, the computational cost, and severe attacks such as non-linear and geometric attacks are the main obstacles to reaching an ideal robust image watermarking scheme that provides high visual quality.
- Watermark embedding/extraction based on learning strategies such as a support vector machine (SVM) (Section 3.1).
- Attempts for the recovery of an attacked watermarked image using learning strategies (Section 3.2).

The remainder of this paper discusses the following components. In Section 2, BiISB image watermarking is revisited; Section 3 investigates the previous applications of SVM in digital image watermarking; the proposed technique is introduced in Section 4; experimental results and discussion are presented in Section 5; and conclusions and future work are presented in Section 6 and Section 7, respectively.

2. BiISB digital image watermarking

The BiISB watermarking technique was proposed by Emami et al. (2012a). This approach introduced an approximation perspective for ownership identification versus a direct correlation view. The BiISB watermarking technique proposed a sub-watermark as an approximation for the main watermark in ownership identification and proof of ownership applications. In fact, the sub-watermark constitutes a bit-pattern histogram that is formed from the main watermark using binary bit patterns (Emami et al., 2012a). Although the sub-watermark is inter-related to the main watermark, it is smaller in size. This provides more visual imperceptibility and more robustness against attacks compared to other spatial domain watermarking techniques, such as LSB (Least Significant Bit) methods (Yang, 2008;Chan & Cheng, 2004) and ISB (Intermediate Significant Bit) techniques (Zeki & Manaf, 2011; Aliwa, El-Tobel, Fahmy, Nasr, & Ei-Aziz, 2010; Zeki & Manaf, 2009). In the BiISB technique, both the main watermark and the sub-watermark are embedded in the host image concurrently to produce a watermarked image (Fig. 1). Later, Emami, Sulong, and Seliman (2012b) proposed a new approach for ownership identification using the histogram intersection (HI) technique (Swain & Ballard, 1991) for the BiISB watermarking approach. This technique uses the statistical information of the watermark for ownership identification exploiting three sub-watermarks including original sub-watermark, extracted sub-watermark, and computed sub-watermark, to be used in the form of the histogram relationship triangle (HRT) (Emami et al., 2012b) to identify the rightful owner. This technique has been improved by Emami and Sulong (2012) using a combination approach including unbiased main watermark and biased sub-watermark strategies. However, further studies should investigate the choice of appropriate thresholds to make ownership identification more reliable.

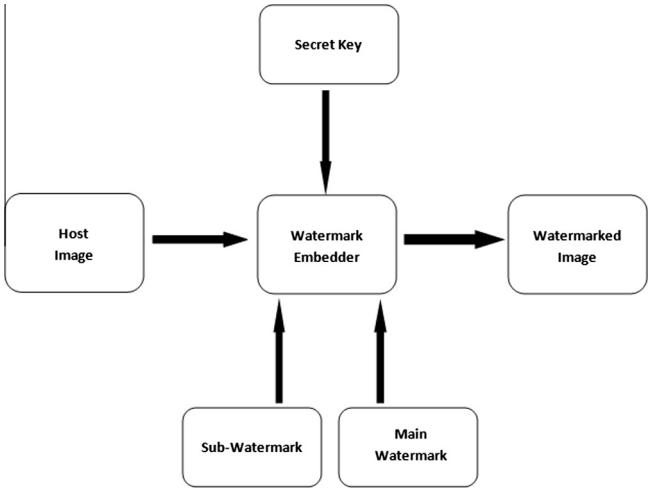

Fig. 1. BiISB watermarking technique. In this technique, both the main watermark and its inter-related sub-watermark are embedded in the host image concurrently within a low bit-plane and a high bit-plane, respectively. The sub-watermark is a bit-pattern histogram. The length of a sub-watermark is determined by a bit pattern length, or β. For example, when $\beta = 4$, there are $2^4 = 16$ bins in the sub-watermark.

3. SVM applications in digital image watermarking

The SVM concept was originally introduced by Vapnik (1995). Among a variety of applications, such as pattern recognition (Les, Kruk, & Osowski, 2013), feature extraction (Sarhan, 2013), medical image analysis (Wang, Zhang, Guo, & Zhang, 2013), concept detection (Lv & Zheng, 2011), function approximation (Almas & Hundewale, 2011), SVM has been employed in the research field of digital image watermarking (Yu, Tsai, & Sun, 2003; Fu, Shen, & Lu, 2004; Fu, 2005; Yen & Wang, 2006; Fuxin, Wei, & Jianjun, 2008; Wang, Xu, & Yang, 2009; Ramly, Aljunid, & Hussain, 2011; Xiang-Yanga, E-Noa, & Hong-Yinga, 2012). A study of the literature shows that there are two main SVM applications in digital image watermarking: SVM for feature extraction in watermark embedding and extraction, and SVM for the recovery of an attacked watermarked image.

3.1. SVM for feature extraction in watermark embedding and extraction

Yu et al. (2003) employed an SVM using 1024 training bits together with the main watermark embedding bits in the embedding stage using the blue channel of color host images. Subsequently, the embedded training data were exploited for SVM training instances, which were used to extract the main watermark. Yen and Wang (2006) exploited the SVM for both the embedding and extraction stages; these authors used 128 extra watermark bits as training bits for training the best hyperplane to classify the bits of the ownership embedding information. Subsequently, the blue channel of the chosen pixel and its neighbor pixels were modified to embed the watermark. Fuxin et al. (2008) proposed a robust image watermarking algorithm based on regression of SVM. This approach splits the host image into two different regions using a blocking approach. The central region, which is the center of each block, is considered for watermark embedding, whereas the second region, which is composed of the neighbors of the central region, is considered to train the SVM based on the relationships among neighboring pixels. Subsequently, the central regions were compared with the predicted pixels using the trained SVM to embed the watermark. Later, Ramly et al. (2011) applied an SVM model for medical image watermarking. These authors applied the SVM

model to a differentiating region of non-interest (RONI) against a region of interest (ROI), regarding the host image prior to the embedding procedure. Subsequently, the watermark was embedded/extracted within/from the host image pixels located in the RONI. Tiwari and Dongre (2012) applied SVM in both watermark embedding and extraction procedures for color images. At first, the host image split into identical size of blocks. Subsequently, a three-level of wavelet transform was performed. Next, for the watermark embedding, a training sequence using the SVM was established to find appropriate quality. Finally, wavelet reverse transforms were performed. Similarly, in the extraction stage, the direct and reverse wavelet transforms and the SVM were used to extract the embedded watermark.

3.2. SVM for recovery of attacked watermarked images

Fu et al. (2004) and Fu (2005) used SVM before the extraction stage to reconstruct a watermarked image from an attacked watermarked image. After the recovery procedure, the extracted watermark was compared to the original watermark using the BER metric for a robustness evaluation. Later, Wang et al. (2009) and Xiang-Yanga et al. (2012) proposed techniques for the classification of a watermarked image from an attacked watermarked image using the SVM, and subsequently, these authors extracted the embedded watermark from the corrected version of the attacked watermarked image. In this method, after the watermark image is extracted, the robustness evaluation is established via a direct correlation approach using the BER metric. Similarly, Tsai, Tseng, and Lai (2010) proposed a robust watermarking scheme using SVM for the recovery of an attacked watermarked image; subsequently, the BCR metric was used for a robustness evaluation via the direct correlation evaluation method.

However, the main drawback of the aforementioned techniques is the need for a time-consuming task, either for the correction of the attacked watermarked image or for feature extraction prior to watermark extraction. In this paper, a low-cost technique based on a support vector machine (SVM) was proposed for watermark detection; this technique requires neither correction nor feature extraction of the attacked watermarked image. Furthermore, this technique investigates the appropriate thresholds to enable a reliable approximation for ownership identification in digital image watermarking. Moreover, the technique introduced considers all possible situations, including watermarked, unwatermarked, attacked watermark, and attacked unwatermarked images, to address additional conditions that could introduce false positive and false negative errors. This technique was experimented on using the HI technique in the BiISB watermarking approach. In the BiISB watermarking (Emami et al., 2012a, 2012b), as mentioned earlier, a sub-watermark that is an approximation of a main watermark is used in both the embedding and extraction stages to provide an effective strategy that results in high robustness and imperceptibility in spatial domain watermarking. However, this approximation should be validated by means of an effective technique. In fact, the decision made for ownership identification with the HI technique (Swain & Ballard, 1991) in BiISB watermarking (Emami et al., 2012b) can bring about false positive and false negative errors due to either the occurrence of watermarking attacks or the presence of similar context features between a host image and its respective sub-watermark (context similarity). Hence, it is important to develop a technique for decision making according to the results provided by this watermarking technique. Thus, the remaining questions involve the best approach to reduce the false positive and false negative probabilities to zero or very close to zero while assessing the ownership probability and which threshold is more reliable for identifying the rightful owner of a digital watermarked image.

4. Proposed approach using SVM and Lagrange duality

The proposed approach was explained in the following sections.

4.1. Description

To address the aforementioned remaining research questions, a supervised learning approach using SVM has been proposed. The novelty of this proposed approach is threefold. First, the proposed approach measures the ownership of the watermarked image instead of evaluating the robustness of the watermarking technique. Second, the proposed approach employs an SVM classifier using the Lagrange dual method for ownership identification of a digital watermarked image to prevent false positive and false negative errors. Third, the approach introduced considers all possible situations, including watermarked, unwatermarked, attacked watermarked and attacked unwatermarked images, during the assessments to increase the reliability of the ownership identification. These advantages make the proposed ownership identification approach more effective, which means that neither the correction nor feature extraction of the attacked watermarked image is needed prior to ownership identification.

4.2. Mathematical analysis

Suppose that n pairs of $\{(X^{(i)},y^{(i)});i=1,2,\ldots,n\}$ represents a training set $T_r = \{(X^{(1)},y^{(1)}),\ldots,(X^{(i)},y^{(i)}),\ldots\ (X^{(n)},y^{(n)})\}$ that comes from four different situations: watermarked, unwatermarked, attacked watermarked, and attacked unwatermarked images. In addition, the function $h(x):X\mapsto y$ dictates that $h(x)$ is a predictor for the corresponding target space $y\in\{-1,1\}$ based on the input space X. Eq. (1) shows the proposed hypothesis. The technique proposed investigates which threshold value would be more reliable for identifying the rightful owner precisely. Specifically, we are looking for a threshold that enables a reliable decision about the ownership probability obtained using the HI technique.

$$f_{w,b,\kappa}(w^TX+b)=\begin{cases}+1; & w^TX+b\geqslant\kappa\\ -1; & Otherwise\end{cases}, \tag{1}$$

where $\sum_{j=1}^{m}w_j=1, w_j\geqslant 0$, and $X=(x_1,x_2)$ is the input feature, which means that $x_1=\beta\in\{1,2,\ldots,m\}$ indicates the bit-pattern length of the sub-watermark, $x_2=H_\beta\in(0,1)$ indicates the value obtained by the HI technique for ownership identification in an m-bit gray-scale host image, and $\kappa\in(0,1)$ is the decision threshold of the hypothesis for the purpose of the watermark detection. In order to represent Eq. (1) in the form of a sign function, we use the following lemma:

Lemma 1. *The term $f_{w,b,\kappa}(w^TX+b)$ can be represented as a sign function,* sgn(z)*, which implies that $z=w^TX+\tau$ where $\tau=b-\kappa$ such that $\forall\tau,b,\kappa\in(0,1)$.*

Hence, the hypothesis mentioned in Eq. (1) can be represented in a canonical form using Lemma 1 as follows:

$$f_{w,\tau}(z)=\begin{cases}+1; & z>0\\ -1; & Otherwise\end{cases}$$
$$Subject\ to: \quad \sum_{j=1}^{m}w_j-1=0, w_j\geqslant 0. \tag{2}$$

Thus, $f_{w,\tau}(z)$ indicates a sign function, where $z=w^TX+\tau$. Here, a positive value for $f_{w,\tau}(z)$ indicates that a member belongs to the class of "watermark-detected", and a negative value for $f_{w,\tau}(z)$ shows that a member belongs to the class of "watermark-not-detected". Fig. 2 demonstrates the proposed approach. In this figure, $\langle w^T,X\rangle+\tau=0$ shows the separating hyperplane, which indicates the decision

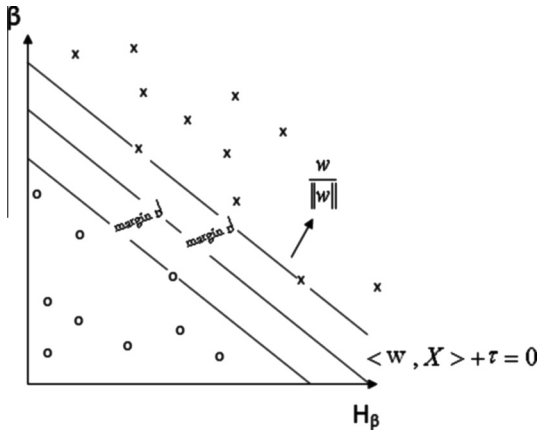

Fig. 2. SVM for the training set regarding ownership identification. The symbol ' × ' indicates a training instance that belongs to the class of watermark-detected, and the symbol 'o' shows a training instance that belongs to the class of watermark-not-detected.

boundary that separates members that belong to the class of watermark-detected from the class of watermark-not-detected. Next, let us describe some definitions:

Definition 1. Let $R \in \Re$ be a set of real numbers. A Supremum for R ("Sup" for simplicity), if it exists, is a least upper bound that can be defined as a number S such that $\{\forall x \in R; x \leqslant S\}$.

Definition 2. Let $R \in \Re$ be a set of real numbers. An Infimum for R ("Inf" for simplicity), if it exists, is a largest lower bound that can be defined as a number I such that $\{\forall x \in R; x \geqslant I\}$.

Next, we can define the geometric margin $d^{(i)}$ as a geometric distance between a training set $(x^{(i)},y^{(i)})$ and the distance given by the separating hyperplane $\langle w^T,X\rangle + \tau = 0$:

$$d^{(i)} = y^{(i)}\left(\left(\frac{w}{\|w\|}\right)^T x^{(i)} + \frac{\tau}{\|w\|}\right), \tag{3}$$

where $w/\|w\|$ is the unit-length vector that is orthogonal to the separating hyperplane.

Thus, for any training set $(x^{(i)},y^{(i)})$, the distance from the separating hyperplane can be computed by Eq. (3) using the parameters w and τ. Thus, the higher the accuracy for the classification is implied, the higher the value of the distance that is required. Here, we can define the geometric margin d with respect to the entire training example $T_r = \{(x^{(1)},y^{(1)}), \ldots, (x^{(i)},y^{(i)}), \ldots (x^{(n)}, y^{(n)})\}$ as the smallest value of the $d^{(i)}$'s:

$$d = \inf_{i=1,\ldots,n}\left\{y^{(i)}\left(\left(\frac{w}{\|w\|}\right)^T x^{(i)} + \frac{\tau}{\|w\|}\right)\right\} = \inf_{i=1,\ldots,n} d^{(i)} \tag{4}$$

The term $\|w\|$ in the above equation nicely satisfies the non-convex constraint mentioned in Eq. (2). Therefore, the term $\|w\|$ no longer appears in this equation as a constraint. Here, to classify the training set $T_r = \{(x^{(1)},y^{(1)}), \ldots, (x^{(i)},y^{(i)}), \ldots (x^{(n)}, y^{(n)})\}$ optimally, an optimization algorithm should be designed in such a way that it maximizes the value of d. Hence, we face an optimization problem as follows:

$$\begin{aligned} &\sup_{d,w,\tau} d = \sup_{d,w,\tau}\left\{\frac{\hat{d}}{\|w\|}\right\} \\ &\text{Subject to}: y^{(i)}(w^T x^{(i)} + \tau) \geqslant \hat{d}; \quad i = 1,\ldots,n, \end{aligned} \tag{5}$$

where $\hat{d}^{(i)} = y^{(i)}(w^T x^{(i)} + \tau)$ is the functional margin of (w,τ) regarding the training set $T_r = \{(x^{(1)},y^{(1)}), \ldots, (x^{(i)},y^{(i)}), \ldots (x^{(n)}, y^{(n)})\}$.

In order to solve Eq. (5) efficiently, let us define the distance between $\langle w^T,X\rangle + \tau = +1$ and $\langle w^T,X\rangle + \tau = -1$ using the following lemma:

Lemma 2. *The Euclidian distance between two hyperplanes $\langle w^T,X\rangle + \tau = +1$ and $\langle w^T,X\rangle + \tau = -1$ is $2/\|w\|$.*

Thus, instead of maximizing $\hat{d}/\|w\|$, the term $(1/2)\|w\|$ can be minimized. Here, to provide possibility of performing a quadratic programming optimization code, the term $(1/2)\|w\|^2$ can be minimized, instead of the term $(1/2)\|w\|$. Hence, we face the following optimization problem, which can be solved more efficiently.

$$\begin{aligned} &\inf_{d,w,\tau}\left\{\frac{1}{2}\|w\|^2\right\} \\ &\text{Subject to}: \quad y^{(i)}(w^T x^{(i)} + \tau) \geqslant 1, \quad i = 1,2,\ldots,n \end{aligned} \tag{6}$$

Next, in an attempt to solve the optimization problem mentioned in Eq. (6), we employ the Lagrange dual method; however, one problem with Eq. (6) is the existing constraint $y^{(i)}(w^T x^{(i)} + \tau) \geqslant 1$, which is different from the condition mentioned in KKT (Karush–Kuhn–Tucker), $g_i(w^*) \leqslant 0$. To address this issue, we can use the following lemma:

Lemma 3. *The constraint of $y^{(i)}(w^T.x^{(i)} + \tau) \geqslant 1$ can be stated in the form of the constraint of $-y^{(i)}(w^T.x^{(i)} + \tau) + 1 \leqslant 0$.*

Using Lemma 3, the Lagrange primal method considering the KKT conditions can serve as a solution for the problem mentioned in Eq. (6), as follows:

$$L_P = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}\lambda_i(y^{(i)}(w^T x^{(i)} + \tau) - 1), \tag{7}$$

where λ is the Lagrange multiplier.

Here, we must minimize L_P by setting the gradient of the Lagrange primal to zero, to find a stationary point for L_P over w and τ. Therefore, we set the Lagrange partial derivatives to zero,

$$\frac{\partial L_P}{\partial w} = 0, \quad \frac{\partial L_P}{\partial \tau} = 0.$$

Subsequently, to solve them for w and τ, we find that:

$$w = \sum_{i=1}^{n}\lambda_i y^{(i)} x^{(i)}, \tag{8}$$

$$\sum_{i=1}^{n}\lambda_i y^{(i)} = 0. \tag{9}$$

Hence, Eq. (7) with Eqs. (8) and (9) can be represented as follows:

$$L_D = \sum_{i=1}^{n}\lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y^{(i)}y^{(j)}\lambda_i\lambda_j(x^{(i)})^T x^{(j)} \tag{10}$$

Eq. (10) gives us the following dual optimization problem:

$$\begin{aligned} &\sup_{\lambda_i} L_D = \sum_{i=1}^{n}\lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y^{(i)}y^{(j)}\lambda_i\lambda_j(x^{(i)})^T x^{(j)} \\ &\text{Subject to} \quad \lambda_i \geqslant 0; \quad i = 1,2,\ldots,n \\ &\sum_{i=1}^{n}\lambda_i y^{(i)} = 0. \end{aligned} \tag{11}$$

Here, we employed a widely used and relatively fast search algorithm, the sequential minimal optimization (SMO) algorithm, to

Fig. 3. Watermark image data.

find the learning parameters that maximize L_D subject to the constraints. Subsequently, we went back to Eq. (8) to find the optimal values for w_i s, w^* s. After determining w^* s, we were able to find the optimal value for τ,τ^*, using the following equation:

$$\tau^* = -\frac{1}{2}\underset{y^{(i)}=+1}{Inf}\{w^T x^{(i)}\} - \frac{1}{2}\underset{y^{(i)}=-1}{Sup}\{w^T x^{(i)}\} \quad (12)$$

5. Experimental results and discussion

Description, analysis, and discussion on the experimental results are presented in the following sections.

5.1. Description

The BiISB watermarking approach mentioned in Section 1 was used as a testing platform using HP = 2 with a bias value (Zeki & Manaf, 2009;Emami et al., 2012b; Emami & Sulong, 2012) of 30 and WP = 5, in which HP indicates the embedding bit-plane for the sub-watermark and WP shows the embedding bit-plane for the main watermark (Emami et al., 2012a, 2012b). Additionally, nine identical copies of an arbitrary gray-scale image logo of 38 × 89 pixels (Fig. 3) were used as the main watermark.

Table 1
Results of an experiment using Fig. 4(d) under a JPEG2000 attack (QF = %90).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	1.00000	1	0.050000	1	No
UW	2	0.89000	−1	−0.056700	−1	No
A	2	0.99000	1	0.040300	1	No
AU	2	0.92000	−1	−0.027600	−1	No
W	3	1.00000	1	0.080000	1	No
UW	3	0.78510	−1	−0.128453	−1	No
A	3	0.99800	1	0.078060	1	No
AU	3	0.78090	−1	−0.132527	−1	No
W	4	1.00000	1	0.110000	1	No
UW	4	0.57260	−1	−0.304578	−1	No
A	4	0.99900	1	0.109030	1	No
AU	4	0.50110	−1	−0.373933	−1	No
W	5	1.00000	1	0.140000	1	No
UW	5	0.50310	−1	−0.341993	−1	No
A	5	0.99500	1	0.135150	1	No
AU	5	0.50270	−1	−0.342381	−1	No

The training examples were separated into four groups based on their corresponding β values: the 1st group with $\beta^{(i)} = 2$, the 2nd group with $\beta^{(i)} = 3$, the 3rd group with $\beta^{(i)} = 4$, and the last group with $\beta^{(i)} = 5$. Moreover, we performed the experiment under JPEG2000 (Quality Factor = %90) and JPEG (Quality Factor = %85) image compression attacks using different 8-bit gray-scale images,

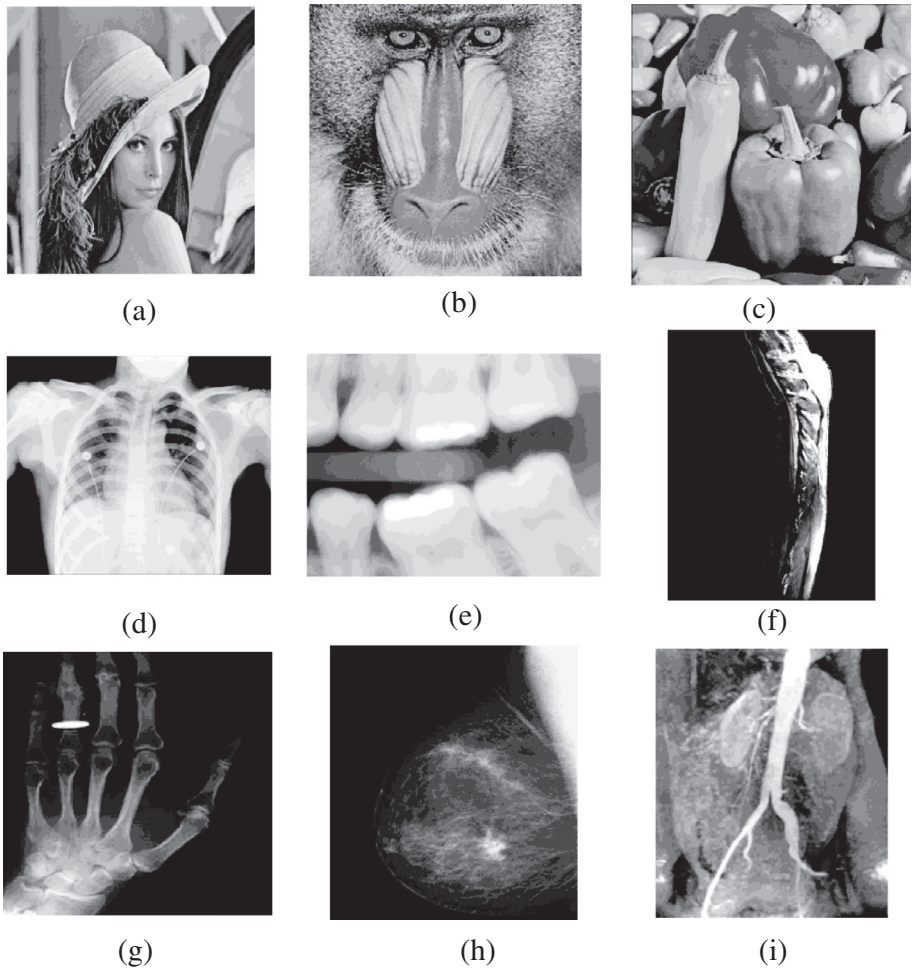

Fig. 4. Images for the training set (4a, 4b, and 4c) and testing set (4d, 4e, 4f, 4g, 4h, and 4i). (a) Lena. (b) Baboon. (c) Peppers. (d) Chest X-ray. (e) Dental X-ray. (f) Spine X-ray. (g) Left hand X-ray. (h) Breast X-ray. (i) Kidney X-ray.

Table 2
Results of an experiment using Fig. 4(e) under a JPEG2000 attack (QF = %90).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	1.00000	1	0.050000	1	No
UW	2	0.94550	−1	−0.002865	−1	No
A	2	0.99100	1	0.041270	1	No
AU	2	0.89080	−1	−0.055924	−1	No
W	3	0.99000	1	0.070300	1	No
UW	3	0.89110	−1	−0.025633	−1	No
A	3	0.98900	1	0.069330	1	No
AU	3	0.80710	−1	−0.107113	−1	No
W	4	0.99900	1	0.109030	1	No
UW	4	0.66340	−1	−0.216502	−1	No
A	4	0.98900	1	0.099330	1	No
AU	4	0.70060	−1	−0.180418	−1	No
W	5	0.99750	1	0.137575	1	No
UW	5	0.64170	−1	−0.207551	−1	No
A	5	0.99100	1	0.131270	1	No
AU	5	0.70990	−1	−0.141397	−1	No

Table 3
Results of an experiment using Fig. 4(f) under a JPEG2000 attack (QF = %90).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	0.99900	1	0.049030	1	No
UW	2	0.77890	−1	−0.164467	−1	No
A	2	0.99010	1	0.040397	1	No
AU	2	0.74220	−1	−0.200066	−1	No
W	3	1.00000	1	0.080000	1	No
UW	3	0.69630	−1	−0.214589	−1	No
A	3	0.99060	1	0.070882	1	No
AU	3	0.58180	−1	−0.325654	−1	No
W	4	0.99750	1	0.107575	1	No
UW	4	0.61400	−1	−0.264420	−1	No
A	4	0.99300	1	0.103210	1	No
AU	4	0.59650	−1	−0.281395	−1	No
W	5	1.00000	1	0.140000	1	No
UW	5	0.58050	−1	−0.266915	−1	No
A	5	1.00000	1	0.140000	1	No
AU	5	0.58880	−1	−0.258864	−1	No

Table 4
Results of an experiment using Fig. 4(g) under a JPEG2000 attack (QF = %90).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	0.99755	1	0.04762	1	No
UW	2	0.88000	−1	−0.06640	−1	No
A	2	0.99900	1	0.04903	1	No
AU	2	0.91100	−1	−0.03633	−1	No
W	3	1.00000	1	0.08000	1	No
UW	3	0.87850	−1	−0.03786	−1	No
A	3	0.99000	1	0.07030	1	No
AU	3	0.80500	−1	−0.10915	−1	No
W	4	0.99100	1	0.10127	1	No
UW	4	0.59440	−1	−0.28343	−1	No
A	4	1.00000	1	0.11000	1	No
AU	4	0.49881	−1	−0.37615	−1	No
W	5	0.99721	1	0.13729	1	No
UW	5	0.51067	−1	−0.33465	−1	No
A	5	0.98990	1	0.13020	1	No
AU	5	0.48955	−1	−0.35514	−1	No

Table 5
Results of an experiment using Fig. 4(h) under a JPEG2000 attack (QF = %90).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	0.99900	1	0.04903	1	No
UW	2	0.89554	−1	−0.05133	−1	No
A	2	0.99900	1	0.04903	1	No
AU	2	0.91100	−1	−0.03633	−1	No
W	3	1.00000	1	0.08000	1	No
UW	3	0.88885	−1	−0.02782	−1	No
A	3	0.99651	1	0.07661	1	No
AU	3	0.88790	−1	−0.02874	−1	No
W	4	1.00000	1	0.11000	1	No
UW	4	0.57799	−1	−0.29935	−1	No
A	4	0.98777	1	0.09814	1	No
AU	4	0.71099	−1	−0.17034	−1	No
W	5	0.99975	1	0.13976	1	No
UW	5	0.48080	−1	−0.36362	−1	No
A	5	0.98655	1	0.12695	1	No
AU	5	0.50000	−1	−0.34500	−1	No

Table 6
Results of an experiment using Fig. 4(i) under a JPEG2000 attack (QF = %90).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	1.00000	1	0.05000	1	No
UW	2	0.84444	−1	−0.10089	−1	No
A	2	0.98486	1	0.03531	1	No
AU	2	0.86050	−1	−0.08532	−1	No
W	3	1.00000	1	0.08000	1	No
UW	3	0.86666	−1	−0.04934	−1	No
A	3	0.98646	1	0.06687	1	No
AU	3	0.82240	−1	−0.09227	−1	No
W	4	1.00000	1	0.11000	1	No
UW	4	0.66231	−1	−0.21756	−1	No
A	4	0.98999	1	0.10029	1	No
AU	4	0.49881	−1	−0.37615	−1	No
W	5	0.99561	1	0.13574	1	No
UW	5	0.49343	−1	−0.35137	−1	No
A	5	0.99005	1	0.13035	1	No
AU	5	0.47433	−1	−0.36990	−1	No

Table 7
Results of an experiment using Fig. 4(d) under a JPEG attack (QF = %85).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	1.00000	1	0.050000	1	No
UW	2	0.94440	−1	−0.003932	−1	No
A	2	0.99860	1	0.048642	1	No
AU	2	0.94680	−1	−0.001604	−1	No
W	3	1.00000	1	0.080000	1	No
UW	3	0.78510	−1	−0.128453	−1	No
A	3	0.99800	1	0.078060	1	No
AU	3	0.80540	−1	−0.108762	−1	No
W	4	1.00000	1	0.110000	1	No
UW	4	0.57260	−1	−0.304578	−1	No
A	4	0.99900	1	0.109030	1	No
AU	4	0.58650	−1	−0.291095	−1	No
W	5	1.00000	1	0.140000	1	No
UW	5	0.50310	−1	−0.341993	−1	No
A	5	0.99500	1	0.135150	1	No
AU	5	0.58450	−1	−0.263035	−1	No

which three of them were standard images in the training set (Fig. 4(a)–(c)), whereas the other were standard X-ray medical images in the testing set (Fig. 4(d)–(i)). These images can be downloaded from www.imageprocessingplace.com. For the standard images, we considered "Lena", "baboon", and "peppers". The reason for selecting these standard images was to cover approximately most of the special features, including the edge and smoothing regions. The baboon image includes several edge areas; the peppers image contains many smooth areas, and the Lena image includes low-frequency, middle-frequency, and high-frequency components. To address many more of the situations that carry false positive and false negative errors, we considered all of the possible situations, including unwatermarked, watermarked, attacked watermarked, and attacked unwatermarked images, in

Table 8
Results of an experiment using Fig. 4(e) under a JPEG attack (QF = %85).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	1.00000	1	0.050000	1	No
UW	2	0.94550	−1	−0.002865	−1	No
A	2	1.00000	1	0.050000	1	No
AU	2	0.94420	−1	−0.004126	−1	No
W	3	0.99000	1	0.070300	1	No
UW	3	0.89110	−1	−0.025633	−1	No
A	3	0.99800	1	0.078060	1	No
AU	3	0.82570	−1	−0.089071	−1	No
W	4	0.99900	1	0.109030	1	No
UW	4	0.66340	−1	−0.216502	−1	No
A	4	0.99000	1	0.100300	1	No
AU	4	0.70110	−1	−0.179933	−1	No
W	5	0.99750	1	0.137575	1	No
UW	5	0.64170	−1	−0.207551	−1	No
A	5	1.00000	1	0.140000	1	No
AU	5	0.70870	−1	−0.142561	−1	No

Table 9
Results of an experiment using Fig. 4(f) under a JPEG attack (QF = %85).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	0.99900	1	0.049030	1	No
UW	2	0.77890	−1	−0.164467	−1	No
A	2	1.00000	1	0.050000	1	No
AU	2	0.78810	−1	−0.155543	−1	No
W	3	1.00000	1	0.080000	1	No
UW	3	0.69630	−1	−0.214589	−1	No
A	3	0.99820	1	0.078254	1	No
AU	3	0.69190	−1	−0.218857	−1	No
W	4	0.99750	1	0.107575	1	No
UW	4	0.61400	−1	−0.264420	−1	No
A	4	1.00000	1	0.110000	1	No
AU	4	0.61550	−1	−0.262965	−1	No
W	5	1.00000	1	0.140000	1	No
UW	5	0.58050	−1	−0.266915	−1	No
A	5	0.99810	1	0.138157	1	No
AU	5	0.61360	−1	−0.234808	−1	No

Table 10
Results of an experiment using Fig. 3(g) under a JPEG attack (QF = %85).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	0.99755	1	0.04762	1	No
UW	2	0.88000	−1	−0.06640	−1	No
A	2	0.99900	1	0.04903	1	No
AU	2	0.91100	−1	−0.03633	−1	No
W	3	1.00000	1	0.08000	1	No
UW	3	0.87850	−1	−0.03786	−1	No
A	3	0.99000	1	0.07030	1	No
AU	3	0.80500	−1	−0.10915	−1	No
W	4	0.99100	1	0.10127	1	No
UW	4	0.59440	−1	−0.28343	−1	No
A	4	1.00000	1	0.11000	1	No
AU	4	0.49881	−1	−0.37615	−1	No
W	5	0.99721	1	0.13729	1	No
UW	5	0.51067	−1	−0.33465	−1	No
A	5	0.98990	1	0.13020	1	No
AU	5	0.48955	−1	−0.35514	−1	No

the experiment. This procedure provided a balanced data set in the experiment and allowed both watermark-detected and watermark-not-detected classes to include an equal number of members.

5.2. Analysis of the results

As mentioned previously, we used the SMO algorithm to find the learning parameters that maximized L_D subject to the constraints mentioned in Eq. (11). Based on the learning data set mentioned in Section 5.1, the findings of this research for the learning parameters were as follows: $w_1^* \cong 0.03, w_2^* \cong 0.97, \tau^* \cong -0.98$ and $d \cong 0.05$. Hence, the proposed classifier mentioned in Eq. (2) can be represented as:

$$f_{w^*,\tau^*}(z) = \mathrm{sgn}(0.03\beta + 0.97H_\beta - 0.98). \quad (13)$$

Tables 1–6 show the results of the experiment using the proposed classifier (Eq. (13)) on the testing data set mentioned in Section 5.1, using the watermarking attack in JPEG2000 format. In these tables, the "W", "UW", "A", and "AU" indicators represent "watermarked image", "unwatermarked Image", "attacked image", and "attacked unwatermarked image", respectively. Additionally, Tables 7–12 show the results of the experiment using the proposed classifier (Eq. (13)) on the testing data set mentioned in Section 5.1, with watermarking attack in JPEG format. The results of the experiment on the training set revealed that the proposed classifier successfully identified the rightful owners even in the presence of context similarities and severe watermarking attacks.

Table 11
Results of an experiment using Fig. 3(h) under a JPEG attack (QF = %85).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	1.00000	1	0.05000	1	No
UW	2	0.90559	−1	−0.04158	−1	No
A	2	0.98999	1	0.04029	1	No
AU	2	0.89000	−1	−0.05670	−1	No
W	3	1.00000	1	0.08000	1	No
UW	3	0.90333	−1	−0.01377	−1	No
A	3	0.99879	1	0.07882	1	No
AU	3	0.89006	−1	−0.02664	−1	No
W	4	0.99850	1	0.10855	1	No
UW	4	0.61111	−1	−0.26722	−1	No
A	4	0.98855	1	0.09889	1	No
AU	4	0.69004	−1	−0.19066	−1	No
W	5	0.98080	1	0.12138	1	No
UW	5	0.56661	−1	−0.28039	−1	No
A	5	0.97900	1	0.11963	1	No
AU	5	0.48733	−1	−0.35729	−1	No

Table 12
Results of an experiment using Fig. 3(i) under a JPEG attack (QF = %85).

Indicator	β	H_β	$y^{(i)}$	Z	Classifier result	Error
W	2	0.99777	1	0.04784	1	No
UW	2	0.87676	−1	−0.06954	−1	No
A	2	0.98776	1	0.03813	1	No
AU	2	0.89003	−1	−0.05667	−1	No
W	3	0.99999	1	0.07999	1	No
UW	3	0.87879	−1	−0.03758	−1	No
A	3	0.99002	1	0.07032	1	No
AU	3	0.80004	−1	−0.11396	−1	No
W	4	1.00000	1	0.11000	1	No
UW	4	0.65005	−1	−0.22945	−1	No
A	4	0.98000	1	0.09060	1	No
AU	4	0.51555	−1	−0.35992	−1	No
W	5	1.00000	1	0.14000	1	No
UW	5	0.48732	−1	−0.35730	−1	No
A	5	0.98977	1	0.13008	1	No
AU	5	0.49033	−1	−0.35438	−1	No

5.3. Discussion

Table 13 shows the features of the data set comparing the technique of Emami et al. (2012b) and the proposed technique. The proposed assessment was deemed more reliable compared to the ownership evaluation method proposed by Emami et al. (2012b) because in the previous method, only the situations of

Table 13
Data set features for different techniques.

Technique	Indicator			
	Watermarked image	Unwatermarked image	Attacked watermarked image	Attacked unwatermarked image
Emami et al. (2012b)	×		×	
Proposed technique	×	×	×	×

Table 14
Ownership evaluation in different approaches after a JPEG2000 attack (QF = %90).

Technique	Measure	Status			
		Correct watermark detection	Correct no-watermark detection	False positive error	False negative error
Emami et al. (2012b)	HI	×	×	×	×
Proposed technique	SVM & HI	×	×		

Table 15
Ownership evaluation using different approaches after a JPEG attack (QF = %85).

Technique	Measure	Status			
		Correct watermark detection	Correct no-watermark detection	False positive error	False negative error
Emami et al. (2012b)	HI	×	×	×	×
Proposed technique	SVM & HI	×	×		

watermarked and attacked watermarked images were considered in the ownership identification. In contrast, in order to guarantee the reliability throughout the ownership identification evaluations, the proposed method utilizes all possible situations, including watermarked, unwatermarked, attacked watermarked and attacked unwatermarked images. Also, in order to cover more varieties of host images, several standard test images were experimented that almost covered most of the special characteristics, including the edge and smoothing regions. Thus, the proposed method provides stability for an ownership identification technique, regardless of the nature of an arbitrary host image and an arbitrary watermark, or in the face of different watermarking attacks.

Tables 14 and 15 show the comparative results of ownership identification between the proposed technique and the technique introduced by Emami et al. (2012b) following the JPEG2000 and JPEG image compression attacks. These results revealed that the previously introduced technique (Emami et al., 2012b) produced false positive and false negative errors, whereas all of the possible situations, including watermarked, unwatermarked, attacked watermarked and attacked unwatermarked images, were used in the ownership identification evaluations. In fact, the inclusion of additional situations, i.e., unwatermarked and attacked unwatermarked, enabled conditions that delivered false positive and false negative errors. Moreover, these results confirmed that the proposed technique did not produce such errors when all of the possible situations were used in the experiments.

6. Conclusions

Reliable ownership identification is a critical need for protecting sensitive data such as electronic patient records (EPRs). To address this need, we proposed an effective method for reliable ownership identification using a BilSB digital image watermarking scheme for medical images; this technique does not require the correction of an attacked watermarked image. The proposed technique employs the SVM and Lagrange dual method to classify the ownership identification into two independent classes, namely watermark-detected and watermark-not-detected, using a balanced data set that encompasses all possible situations, including watermarked, unwatermarked, attacked watermarked, and attacked unwatermarked images. The experimental results for the standard training and testing sets confirmed that the proposed technique was successful at identifying the rightful owner without false positive and false negative errors even in the presence of context similarities between the watermark and the embedding pixels of the host image as well as under severe watermarking attacks.

7. Future research

In future research, additional watermarking attacks should be investigated using the proposed method, and such investigations may result in improvements in the classifier obtained in this work. Moreover, the proposed method should be exploited to investigate appropriate classifiers for other watermarking schemes to provide more reliable assessments on ownership identification and proof of ownership.

Acknowledgments

This work was supported by The National University of Malaysia Research University under research Grant No. AP-2012–019.

References

Aliwa, M. B., El-Tobel, T. E., Fahmy, M. M., Nasr, M. E. S., & Ei-Aziz, M. H. A. (2010). A new novel fidelity digital watermarking based on adaptive pixel-most-significant-bit-6 in spatial domain gray scale images and robust. *American Journal of Applied Sciences, 7*(7), 987–1022.

Almas, A., & Hundewale, N. (2011). *Function approximation using SVM with FCM and slope based partition. Trends in Computer Science, Engineering and Information Technology Communications in Computer and Information Science* (vol. 204). Springer, pp. 704–714.

Chan, C., & Cheng, L. M. (2004). *Hiding data in images by simple LSB substitution. Pattern Recognition* (vol. 37). Elsevier, pp. 469–474.

Cox, I., Miller, M. L., Bloom, J. A., Fredrich, J., & Kalker, T. (2008). *Digital watermarking and steganography* (2nd ed.). Elsevier, pp. 39–40.

Emami, M. S., Sulong, G. B., & Seliman, S. B. (2012a). A novel multiple semi-blind enhanced ISB watermarking algorithm using watermark bit-pattern histogram

for copyright protection. *International Journal of Innovative Computing, Information and Control, 8*(3), 1665–1687.

Emami, M. S., Sulong, G. B., & Seliman, S. B. (2012b). An approximation approach for digital image owner identification using histogram intersection technique. *International Journal of Innovative Computing, Information and Control, 8*(7A), 4605–4620.

Emami, M. S., & Sulong, G. B. (2012). An effective real-time invisible digital asset watermarking technique for copyright protection, © 2012 UTM, Malaysia. All rights reserved.

Fu, Y. (2005). Reliable information hiding based on support vector machine. *Informatica, 16*(3), 333–346.

Fu, Y., Shen, R., & Lu, H. (2004). *Optimal watermark detection based on support vector machine*. *Lecture Notes in Computer Science* (vol. 3173). Springer-Verlog, pp. 552–557.

Fuxin, W., Wei, S., & Jianjun, H. (2008). Regression of SVM based robust watermarking algorithm. In *9th International Conference Signal Processing* (pp. 2197–2200). IEEE Computer Society.

Les, T., Kruk, M., & Osowski, S. (2013). *Automatic recognition of industrial tools using artificial intelligence approach*. *Expert Systems with Applications* (vol. 40). Elsevier, pp. 4777–4784.

Lv, G., & Zheng, C. (2011). *A novel framework for concept detection on large scale video database and feature pool*. *Artificial Intelligence Review*. Springer.

Ramly, S., Aljunid, S. A., & Hussain, H. S. (2011). *SVM-SS watermarking model for medical images*. *Communications in Computer and Information Science* (vol. 194). Springer-Verlag, pp. 372–386.

Run, R., Horng, S., Lai, J., Kao, T., & Chen, R. (2012). An improved SVD-based watermarking technique for copyright protection. *Expert Systems with Applications, 39*, 673–689.

Sarhan, A. M. (2013). Wavelet-based feature extraction for DNA microarray classification. *Artificial Intelligence Review, Springer, 39*(3), 237–249.

Swain, M. J., & Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision, 7*(1), 11–32.

Tian, J., Bloom, J. A., & Baum, P. G. (2007). False positive analysis of correlation ratio watermark detection. In *IEEE International Conference on Multimedia and Expo* (pp. 619–622). IEEE.

Tiwari, S., & Dongre, A. (2012). A superior support vector machine digital watermarking for color images. In *Computer, International Conference on Information and Telecommunication Systems (CITS)* (pp. 1–5). IEEE.

Tsai, H. H., Tseng, H. C., & Lai, Y. S. (2010). Robust lossless image watermarking based on a-trimmed mean algorithm and support vector machine. *The Journal of Systems and Software, 83*, 1015–1028.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York, USA: Springer-Verlag. ISBN:0-387-94559-8.

Wang, X., Xu, Z., & Yang, H. (2009). *A robust image watermarking algorithm using SVR detection*. *Expert Systems with Applications* (vol. 36). Elsevier, pp. 9056–9064.

Wang, X., Zhang, B., Guo, Z., & Zhang, D. (2013). *Facial image medical analysis system using quantitative chromatic feature*. *Expert Systems with Applications* (vol. 40). Elsevier, pp. 3738–3746.

Xiang-Yanga, W., E-Noa, M., & Hong-Yinga, Y. (2012). *A new SVM-based image watermarking using Gaussian–Hermite moments*. *Applied Soft Computing* (vol. 12). Elsevier, pp. 887–903.

Yang, C. (2008). *Inverted pattern approach to improve image quality of information hiding by LSB*. *Pattern Recognition* (vol. 41). Elsevier, pp. 2674–2683.

Yen, S., & Wang, C. (2006). SVM based watermarking technique. *Tamkang Journal of Science and Engineering, 9*(2), 141–150.

Yu, P. T., Tsai, H. H., & Sun, D. (2003). Digital watermarking of color images using support vector machines, 2003 National Computer Symposium (NCS'03).

Zeki, A. M., & Manaf, A. A. (2009). A novel digital watermarking technique based on ISB (intermediate significant bit). *International Journal of Information Technology, 5*(3), 141–148.

Zeki, A. M., & Manaf, A. A. (2011). ISB watermarking embedding: A block based model. *Information Technology Journal, 10*(4), 841–848.