EURASIP Journal on Applied Signal Processing 2002:1, 92–104

Watermark Detection and Extraction Using Independent Component Analysis Method

Dan Yu
School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798
Email: p141450770@ntu.edu.sg

Farook Sattar
School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798
Email: efsattar@ntu.edu.sg

Kai-Kuang Ma
School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798
Email: ekkma@ntu.edu.sg.

Received 30 July 2001 and in revised form 12 October 2001

This paper proposes a new image watermarking technique, which adopts Independent Component Analysis (ICA) for watermark detection and extraction process (i.e., *dewatermarking*). Watermark embedding is performed in the spatial domain of the original image. Watermark can be successfully detected during the Principle Component Analysis (PCA) whitening stage. A nonlinear robust batch ICA algorithm, which is able to efficiently extract various temporally correlated sources from their observed linear mixtures, is used for blind watermark extraction. The evaluations illustrate the validity and good performance of the proposed watermark detection and extraction scheme based on ICA. The accuracy of watermark extraction depends on the statistical independence between the original, key and watermark images and the temporal correlation of these sources. Experimental results demonstrate that the proposed system is robust to several important image processing attacks, including some geometrical transformations—scaling, cropping and rotation, quantization, additive noise, low pass filtering, multiple marks, and collusion.

Keywords and phrases: watermarking, dewatermarking, independent component analysis (ICA).

1. INTRODUCTION

Digital watermarking technology has evolved very quickly these years. The basic principles of most watermarking methods are applying small, pseudorandom changes to the selected coefficients in the spatial or transform domain. Most of the watermark detection schemes use some kinds of correlating detector to verify the presence of the embedded watermark [1, 2]. The watermark can be extracted with information of the key, and with/without the original (i.e., unwatermarked) image.

Independent Component Analysis (ICA) is probably the most powerful and widely-used method for performing Blind Source Separation (BSS). It is a very general-purpose statistical technique to recover the independent sources given only sensor observations that are linear mixtures of independent source signals [3, 4, 5]. The simplest BSS model assumes the existence of n independent components $s_1, s_2, \ldots, s_n$, and the same number of linear and instantaneous mixtures of these sources, $x_1, x_2, \ldots, x_n$, that is,

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \cdots + a_{jn}s_n; \quad 1 \le j \le n. \tag{1}$$

In vector-matrix notation, the above mixing model can be represented as

$$x = \mathbf{A}s, \tag{2}$$

where $\mathbf{A}$ is the square $n \times n$ mixing matrix. The unmixing process [3, 4, 5] can be formulated as computing the separating matrix $\mathbf{W}$, which is the inverse of the mixing matrix $\mathbf{A}$, and the independent components are obtained by

$$s = \mathbf{W}x. \tag{3}$$

The basic ICA model has been extended in different directions, for instance, more sensors than sources, less sensors than sources, noisy observations, complex signals and mixtures, convolutive mixtures and so on [6].

Applications of ICA can be found in many different areas such as audio processing, biomedical signal processing, image processing and telecommunications [3, 4, 5]. In [7], ICA was firstly applied in digital image watermarking for watermark detection and blind extraction.

The basic idea behind our work is to use some specific images, such as some special patterns or signature images as the key and the watermark for watermark embedding in order to create a watermarked image, which can be considered as an observed mixture image with the original image, the key and the watermark; hence, watermark recovery can be viewed as a blind source separation problem. The motivations and advantages of using ICA technique in watermark detection and extraction could be the following:

• Any of the three source images, that is, the watermark, the key and the original image, can be recovered by the separation process, which is just like a reverse process of watermarking, hence, this process can be defined as "*dewatermarking*."

• The watermark can be embedded either instantaneously or convolutively, that is, the observed watermarked image can be either instantaneous or convolutive mixture.

• The watermark can be extracted with or without the original image. When using both key and original image for extraction, there will be three mixtures that can be generated (see Section 2.2 for detailed explanation), which is the case where the number of observed mixtures are as many as the independent source signals. Considering watermark extraction using only the key or the original image, there will be less observed mixtures than sources which is a special case defined as *overcomplete* ICA [8].

To present the basic principle of this new watermarking technique based on ICA, the paper is restricted to watermarking and dewatermarking with the simplest ICA model: real source signals, linear and instantaneous/convolutive mixtures and as many mixtures as sources (i.e., the system requires both the key and original images for watermark detection and extraction).

A simple subtraction method for watermark detection and extraction has been proposed by Cox et. al. in [9], where the original (unwatermarked) image is to be known and the watermark embedding process has been done by multiplying the watermark samples with coefficients, which can have the same constant value or can be varied with the values of the corresponding original input samples. In [9], several assumptions are made regarding values of the mixing coefficients, distributions of the watermark as well as the mixing process. In our proposed BSS based method, we do not have restrictions on the mixing process as well as the mixing coefficients. The proposed method is more flexible (less restrictions) in the sense that it can work for both the instantaneous and convolutive mixtures. For convolutive mixing case, the coefficients are not only multiplied, but also shifted/delayed. In such cases, it is quite difficult to extract unknown watermark using the existing simple methods, for example using [9].

The objective of this paper is to introduce an efficient ICA based watermark detection and extraction (i.e., *dewatermarking*) scheme for digital image watermarking. A robust, batch ICA algorithm [10] is applied in our efficient dewatermarking processing. The simulation results and performance are shown for various types images which are found quite promising. Experimental results also show the robustness of our method to some attacks like scaling, cropping, rotation, quantization, additive noise, filtering, multiple marks, and collusion.

The paper is organized as follows. Section 2 presents our watermarking system, including both the watermark embedding scheme and the proposed watermark detection and extraction scheme. The robust batch ICA algorithm, which is used for dewatermarking process, is described in Section 3. The simulation results are illustrated in Section 4, and the watermark extraction performance is analyzed in Section 5. Section 6 shows the robustness testing results after image attacking as mentioned above. Finally, conclusions are drawn in Section 7.

2. WATERMARKING SYSTEM

2.1. The watermark embedding scheme

In the generic watermark embedding scheme, the inputs to the system are the original data, the watermark and an optional public or secrete key. The key is used to enforce the security, that is, to prevent unauthorized party from recovering and manipulating the watermark. Our proposed image watermarking system exploits a watermark, **M**, and a secrete key, **K**, for the purpose of conducting two levels of security, by using the special images as the watermark and the key, with the same size as the original image, **I**, to be embedded. Spatial domain is used to perform the hiding operation.

Both the watermark **M** and the key **K** are inserted in the spatial domain of the original image **I**. The watermarked image, **X**, is a linear mixture of the original image **I**, key **K** and watermark **M** with both **K** and **M** having signal energy sufficiently less (at least 10–100 times less) than the energy of the original image **I** in order to make them invisible. That is,

$$\mathbf{X} = \mathbf{I} + a\mathbf{K} + b\mathbf{M}, \tag{4}$$

where a and b are small weighting coefficients, or

$$\mathbf{X} = \mathbf{I} + a\mathbf{K} + \mathbf{b} * \mathbf{M}, \tag{5}$$

where **b** is the small filter coefficients and $*$ denotes *convolution*.

Figure 1 illustrates an example about how a Cameraman image (with size of 64×64) is watermarked (using (4)). The image pattern in Figure 1b is used as the key image. Figure 1c is the watermark image, which is a special sequence called *Gold-like sequence* [11]. The Gold-like sequence is generated by modulo-two addition of a pair of maximal linear sequences, with the output being the same length of the two maximal codes, which gives highly correlated sequence.

Figure 2 shows an example for watermark embedded image using (5). Figure 2a is the frequency response of a 2D filter, **b**, of size $N \times N$ ($N = 4$), which is used to convolve the watermark image shown in Figure 1c. The original and watermarked Cameraman images are displayed in Figures 2b and 2c, respectively.

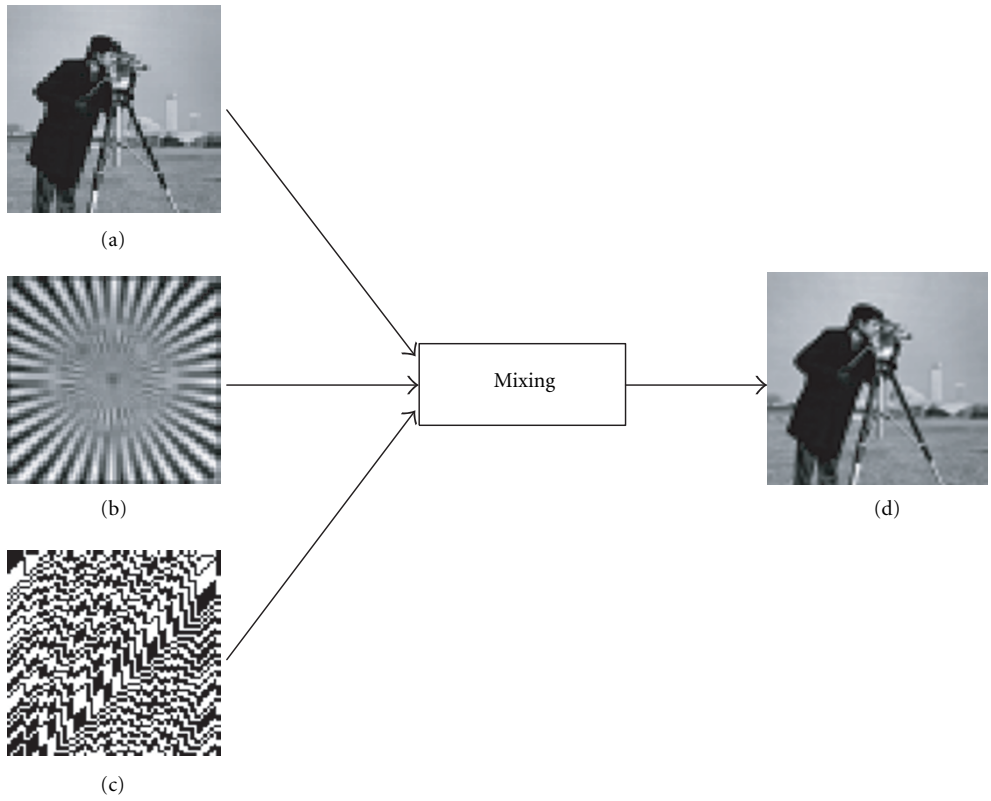


Figure 1: (a) Cameraman (original image), (b) secrete key image, (c) watermark image (Gold-like sequence), (d) watermarked image.

2.2. The proposed watermark detection and extraction scheme

Private watermarking system is adopted in the watermark recovery process, that is, the system requires both the key and the original data for watermark detection and extraction. The authorized party can detect and/or verify the watermark with both the secrete key and the original data.

To assure the identifiability of ICA model, it is required that the number of observed linear mixture inputs is at least equal to or larger than the number of independent sources. For the above-proposed watermark detection and extraction scheme, at least three linear mixtures of the three independent sources are needed. Using the key image **K** and with the help of original image **I**, two more mixed images are generated by adding them into the watermarked image **X**:

$$\mathbf{X}_1 = \mathbf{X}, \qquad \mathbf{X}_2 = \mathbf{X} + c\mathbf{K}, \qquad \mathbf{X}_3 = \mathbf{X} + d\mathbf{I}, \tag{6}$$

where c and d are arbitrary real numbers.

These three images are rearranged into three row vectors $\mathbf{X}_j(k)$, with length of p each, (where $j = 1, 2, 3$ and $k = 1, 2, \ldots, p$), to satisfy the input data requirements of a nonlinear blind extraction algorithm—robust batch ICA algorithm, which is used for dewatermarking process. The proposed watermark detection and extraction scheme is shown in Figure 3, and will be explained in Section 3.

3. ROBUST BATCH ALGORITHM FOR WATERMARK DETECTION AND BLIND EXTRACTION

With the recent increase of interest in ICA, various algorithms have been proposed, based on probabilistic models, information theory, artificial neural networks and so on [3, 4, 5, 12]. In our experiments, a robust batch algorithm which is based on the second-order statistics is used for watermark detection and blind extraction.

The robust batch algorithm is presented in this section, where the source signals are modeled as an autoregressive (AR) process [10]. Hence, the algorithm will be an effective blind separation approach particularly for the temporally correlated sources. Generally speaking, the images are spatially correlated. The pixel in the image is always highly correlated with its neighboring pixels, thus images can be characterized as 1D AR models in time series analysis [13, 14]. The pixel processing will be correlated with its past pixels, in other words, the image sequences will be temporally correlated sequences. That is why in the following batch algorithm, the use of AR model for image source is appropriate.

This method is based on following two stages, that is, PCA whitening process for watermark detection, followed by the robust batch ICA algorithm for watermark extraction.

3.1. PCA whitening—watermark detection

Standard Principle Component Analysis (PCA) is often used for whitening process, since it can compress information op-

(a) (b) (c)

Figure 2: (a) Frequency response of a filter, **b**, in (5), (b) original image, (c) watermarked image using the 2D filter in (a).

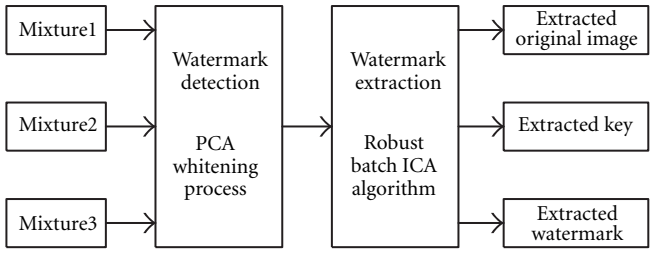


Figure 3: Proposed watermark detection and extraction scheme.

timally in the mean-squared error sense, while filtering possible noise simultaneously [12]. The PCA whitening matrix is given by

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{U}^T, \tag{7}$$

where **D** is the diagonal matrix of data covariance matrix $E[\mathbf{X}_j\mathbf{X}_j{}^T]$, and **U** is its eigenmatrix, and $E[\cdot]$ denotes the expectation operator.

PCA whitening provides a convenient means for estimating the number of sources or independent components, n, from the rank of the diagonal matrix **D** [12]. For example, the rank of **D** is equal to three for watermarked images, meaning there are totally three image sources. On the other hand, if the image is unwatermarked, the three mixtures are actually the combinations of the original and key images only; hence, the rank of **D** will be reduced to two.

3.2. Robust batch ICA algorithm—watermark extraction

After pre-whitening process, the sources are recovered by iteratively estimating the unmixing matrix **W** through a simple batch learning algorithm. The convergence of the learning process is controlled by $\varepsilon_i(k)$ via a FIR filter whose transfer function is $\mathbf{B}_i(z)$ [10]. Figure 4 shows the block diagram of robust batch algorithm.

The output vector $\mathbf{y}(k) = \mathbf{W}\mathbf{x}(k)$. The error from the ith iteration is defined as (sample index k is dropped for simplicity)

$$\varepsilon_i = y_i - \underbrace{\mathbf{b}_i{}^T\tilde{\mathbf{y}}_i}_{\tilde{y}_i}, \tag{8}$$

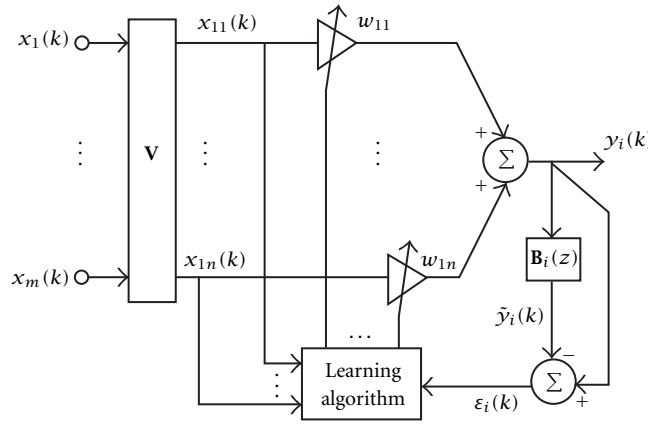


Figure 4: Block diagram of the robust batch ICA algorithm (m: number of mixtures, n: number of sources).

where $\mathbf{b}_i = [b_{i1}b_{i2}\cdots b_{iL}]^T$, $y_i = \mathbf{w}_i^T\mathbf{x}_i$, and $\tilde{\mathbf{y}}_i = [y_{i1}y_{i2}\cdots y_{iL}]^T$. The cross-correlation matrix $\mathbf{R}_{\varepsilon_i\mathbf{x}_i}$ is computed in [10],

$$\mathbf{R}_{\varepsilon_i\mathbf{x}_i} = E[\varepsilon_i\mathbf{x}_i^T] = \mathbf{w}_i^T\mathbf{R}_{\mathbf{x}_i\mathbf{x}_i} - \mathbf{b}_i^T\mathbf{R}_{\tilde{\mathbf{y}}_i\mathbf{x}_i}, \tag{9}$$

where $\mathbf{R}_{\mathbf{x}_i\mathbf{x}_i} = E[\mathbf{x}_i\mathbf{x}_i{}^T]$ and $\mathbf{R}_{\tilde{\mathbf{y}}_i\mathbf{x}_i} = E[\tilde{\mathbf{y}}_i\mathbf{x}_i^T] = \mathbf{R}_{\mathbf{x}_i\tilde{\mathbf{y}}_i}^T$. Based on the principle of decorrelation, the optimum $\mathbf{w}_{i,\text{opt}}$ is obtained when the cross-correlation matrix $\mathbf{R}_{\varepsilon_i\mathbf{x}_i} = 0$,

$$\mathbf{w}_{i,\text{opt}} = \hat{\mathbf{R}}_{\mathbf{x}_i\mathbf{x}_i}^{-1}\hat{\mathbf{R}}_{\mathbf{x}_i\tilde{\mathbf{y}}_i}\mathbf{b}_i, \tag{10}$$

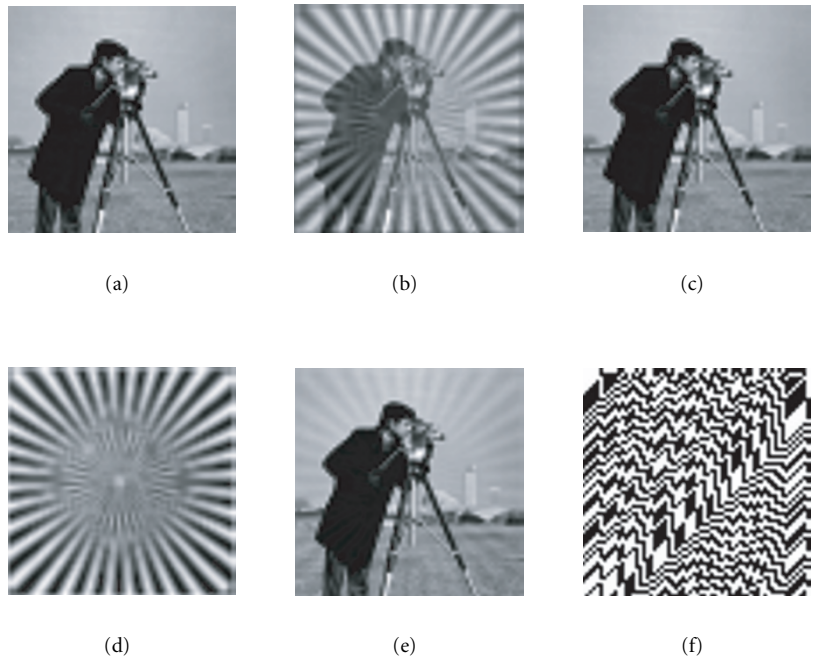
where $\hat{\mathbf{R}}$ denotes the estimated matrix of **R**.

By applying standard PCA and normalization of signals to unit variance after each deflation procedure [10], the autocorrelation matrix $\mathbf{R}_{\mathbf{x}_i\mathbf{x}_i}$ will be the identity matrix. Hence, the optimum $\mathbf{w}_{i,\text{opt}}$ can be further simplified as

$$\mathbf{w}_{i,\text{opt}} = \hat{\mathbf{R}}_{\mathbf{x}_i\tilde{\mathbf{y}}_i}\mathbf{b}_i. \tag{11}$$

For updating the vectors $\mathbf{b}_i$, the optimal minimum mean-squared error $E[\varepsilon_i^2]$ can be written as

$$E[\varepsilon_i^2] = E[y_i^2] + \mathbf{b}_i^T\mathbf{R}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}\mathbf{b}_i - 2\mathbf{R}_{\tilde{\mathbf{y}}_i y_i}\mathbf{b}_i^T. \tag{12}$$



(a) (b) (c)

(d) (e) (f)

Figure 5: Watermarked case: (a) watermarked image, (b) and (c) generated mixture images, (d) extracted key, (e) extracted Cameraman image, (f) extracted watermark.

The gradient of this function related to $\mathbf{b}_i$ will be given by

$$\frac{\partial E[\varepsilon_i^2]}{\partial \mathbf{b}_i} = 2\mathbf{R}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}\mathbf{b}_i - 2\mathbf{R}_{\tilde{\mathbf{y}}_i y_i}. \tag{13}$$

By using the Wiener filtering, we have the optimum vector $\mathbf{b}_{i,\text{opt}}$ [10],

$$\mathbf{b}_{i,\text{opt}} = \hat{\mathbf{R}}^{-1}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i y_i}. \tag{14}$$

Thus, in [10], a heuristical algorithm is obtained that the updating vector $\mathbf{b}_i$ can be further simplified by removing the matrix inverse and obtain

$$\mathbf{b}_{i,\text{opt}} = \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i y_i}. \tag{15}$$

Hence, the optimum updating vectors for $\mathbf{w}_i$ and $\mathbf{b}_i$ adopted in this algorithm are (11) and (15).

4. SIMULATION RESULTS

Simulation experiments are conducted to demonstrate the feasibility and robustness of the proposed ICA method for watermark detection and extraction. Some illustrative results for watermark detection as well as extraction are shown in Figures 5 and 6 for both watermarked and unwatermarked cases, respectively. Figure 5f shows extracted watermark from a watermarked Cameraman image of size 64×64 pixels (Figure 5a). It has been found that the key and the watermark extracted are the reverse of their original images. For unwatermarked case (Figure 6), only two images are extracted after separation—the original image and the key.

Figure 7 shows extraction of watermark for convolutive mixing using the filter $\mathbf{b}$, shown in Figure 2a. Figure 7f is the extracted watermark.

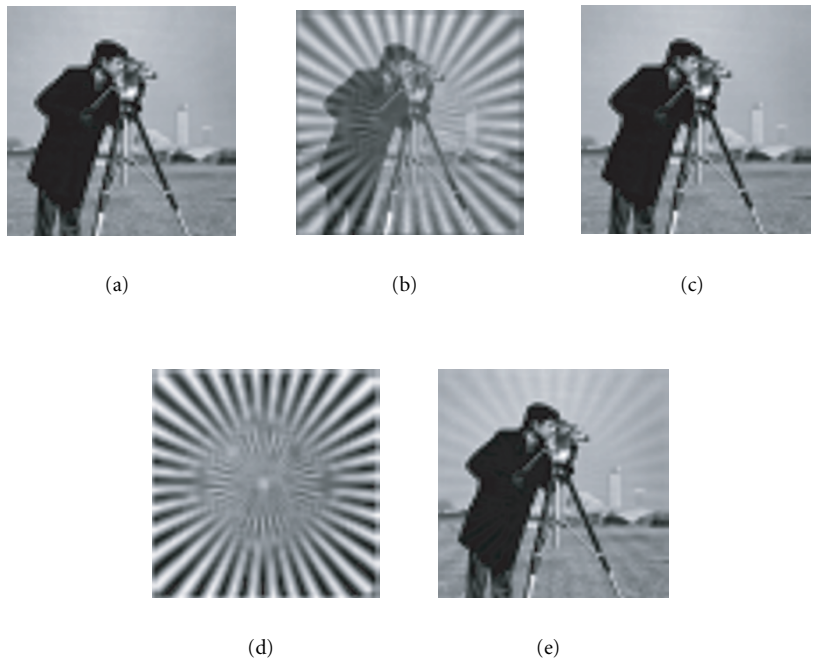
The performance of watermark extraction is evaluated by calculating the *normalized correlation coefficient* r for the extracted watermark and the original embedded watermark as

$$r = \frac{\sum_{k=1}^{p} \mathbf{m}(k) \cdot \hat{\mathbf{m}}(k)}{\sqrt{\sum_{k=1}^{p} \mathbf{m}(k)^2 \cdot \sum_{k=1}^{p} \hat{\mathbf{m}}(k)^2}}, \tag{16}$$

where $\mathbf{m}$ and $\hat{\mathbf{m}}$ are the original and the extracted watermark sequences, respectively, with zero mean each, and p is the total number of pixels of the image [15]. The magnitude range of r is $[-1, 1]$, and the unity holds if the image extracted perfectly matches the original. The minus sign indicates the extracted image is a reverse version of its original image.

Table 1 shows the normalized correlation coefficients between the original and the extracted images for the examples described in Figures 5 and 6. The robust ICA algorithm separates the images from the mixtures successfully. The watermark is perfectly extracted from the watermarked image except that a reversed version of the original watermark image was produced, corresponding to the minus sign to indicate such case. For the watermark extraction result shown in Figure 7f, the correlation coefficient, r, is found as $r = 0.7063$.

The batch algorithm that is based on second-order statistics performs well, when the source signals are temporally correlated and have low kurtosis [10]. This algorithm can be efficiently applied for natural images since they are not purely



(a) (b) (c)

(d) (e)

Figure 6: Unwatermarked case: (a) unwatermarked original image, (b) and (c) generated mixture images, (d) extracted key, (e) extracted Cameraman image.

Table 1: Performance evaluation using Gold-like sequence as the watermark, and measured with respect to the original images.

Test conditions	Correlation coefficient, r		
	Cameraman	Key	Watermark (Gold-like sequence)
Watermarked case	0.9978	−0.9945	−1.0000
Unwatermarked case	0.9978	−0.9946	N/A

random but contain structures. It has been found that the results using Gold-like sequence as watermark gives excellent extraction results, since it gives a highly correlated sequence with very low kurtosis that is nearly to one.

Table 2 shows the performance index, r, for watermark extraction with respect to the size $N \times N$ of **b** filter. As N increases, the filter bandwidth reduces due to sharper cutoff. Therefore, according to Table 2, the values of r become smaller with larger N. It is also noticed that r values are closer as N values increase from 6 to 48.

5. PERFORMANCE ANALYSIS

Extensive experiments have been done for different types of images including the original image, key image, and watermark image. Satisfactory performance (the results are not shown here) indicates the efficiency of the presented method. However, it has been found from the experiments that the watermark extraction performance using the robust batch algorithm is influenced by the following two main factors: the statistical dependence between image sources and the temporal correlation of each source.

5.1. Performance with respect to source dependence

The fundamental assumption for ICA is that the source components are statistically independent. The blind separation performance can thus be determined by the verification of the statistical independence of the recovered sources.

Statistical independence can be defined by the probability densities. The joint probability density function, $f(x, y)$, for statistically independent random variables can always be factored into the product of two marginal density functions as [16]

$$f(x, y) = f_X(x) f_Y(y), \tag{17}$$

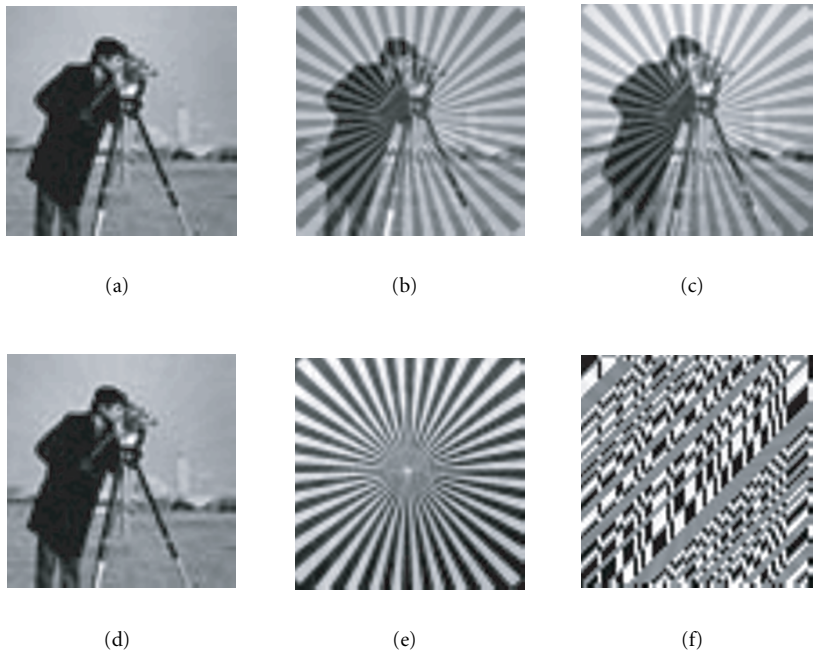
where f_X and f_Y are the two marginal densities for the random variables X and Y. Hence, one of the important consequences to be obtained is

$$E[XY] = E[X]E[Y]. \tag{18}$$

That is, the expected value of the product of two statistically independent random variables is the product of their mean values.

According to (18), if the variables are independent, they are uncorrelated. Variables X and Y are said to be uncorrelated, if their covariance is zero [16]; that is,

$$E[XY] - E[X]E[Y] = 0. \tag{19}$$



(a) (b) (c)

(d) (e) (f)

Figure 7: Results of watermark extraction for convolutive watermark embedding model in (5): (a) watermarked image, (b) and (c) generated mixture images, (d) extracted Cameraman image, (e) extracted key, (f) extracted watermark.

Table 2: Normalized correlation coefficient, r, with respect to the size $N \times N$ of $\mathbf{b}$ filter.

$N \times N$	4×4	6×6	8×8	12×12	16×16	24×24	32×32	48×48
r	0.7063	−0.5775	0.5686	−0.5307	−0.5175	0.5079	−0.5054	−0.4999

A weaker form of independence is uncorrelatedness. On the other hand, uncorrelatedness does not imply independence [16].

Since independence implies uncorrelatedness, many ICA methods constrain the estimation procedure so that it always gives uncorrelated estimates of the independent components [3]. As in robust batch algorithm, the optimum learning rule is derived based on decorrelation principle, where it is assumed that the sources are statistically independent, which implies that the cross-correlation matrix between ε_i and $\mathbf{x}_i$, $\mathbf{R}_{\varepsilon_i \mathbf{x}_i}$, is ideally zero. However, in practice the images are not perfectly independent, hence, it will degrade the extraction performance.

Another extraction example is shown in Figure 8. In this example, a text image instead of Gold-like sequence is used as the watermark. The normalized correlation coefficients are measured against the original images, and shown in Table 3.

Comparing the two examples in Figures 6 and 8, it can be seen that the results using Gold-like sequence as the watermark gives a better extraction results than using text image. This can be explained from the viewpoint of statistical independence of the sources. The degree of independence can be measured in terms of normalized correlation coefficient. If they are uncorrelated, the correlation is zero. Table 4 shows the correlations between sources. The correlation between Cameraman and text image is 0.1505, which is quite high, leading the worse extraction result.

Table 3: Normalized correlation coefficients, r, using text image as the watermark.

Cameraman	Key	Watermark (text)
0.9859	−0.9945	−0.9447

Table 4: Normalized correlation coefficients between sources.

Sources	Watermark	
	Gold-like sequence	Text
Cameraman and Key	−0.0820	−0.0820
Cameraman and Watermark	0.0089	0.1505
Key and Watermark	−0.0041	0.0052

5.2. Performance with respect to source temporal correlation

The robust batch ICA algorithm models the source signals by unknown but stable AR process [10]. In the application of the algorithm on images, an image is characterized by a 1D

Figure 8: Watermark extraction results using text image as the watermark: (a) original Cameraman image, (b) key image, (c) text watermark, (d) watermarked image, (e) and (f) generated mixture images, (g) extracted key, (h) extracted Cameraman image, (i) extracted watermark.

signal that appears at the output of a raster scanner, that is, a sequence of rows or columns. One property of AR model is that the linear prediction of the current sample depends on the previous samples. An estimation of the sources is obtained based on the correlation of this sample with its previous samples. If the sources are more temporally correlated, the adjustment of the updating vectors is easier and faster to converge, hence, to complete the source separation.

To demonstrate the importance of source temporal correlation, another watermark extraction example is illustrated in Figure 9, using the Airplane image as the original image. The performance for using column-wise and row-wise sequencing methods are examined. By using column- or row-wise scanning order, two image sequences with different temporal correlations are obtained. Figures 9d, 9e, and 9f show the extraction results with column-wise scanning order, where the image sources are unable to be well separated. Figures 9g, 9h, and 9i are the extracted images with row-wise scanning order. The performance measured in *normalized correlation coefficient* is shown in Table 5. From this example, it has been observed that the row-wise sequencing makes significant improvements for extracted Airplane and extracted text watermark images comparing with column-wise sequencing. The significant differences in the extraction results for these two sequences with different temporal correlations, show the importance of temporally correlated sources in the robust batch ICA algorithm on blind separation.

6. ROBUSTNESS TESTING

The watermarking system should be robust against data distortions introduced through standard data processing and attacks. It should be virtually impossible for unauthorized users to remove it; and practically the image quality must

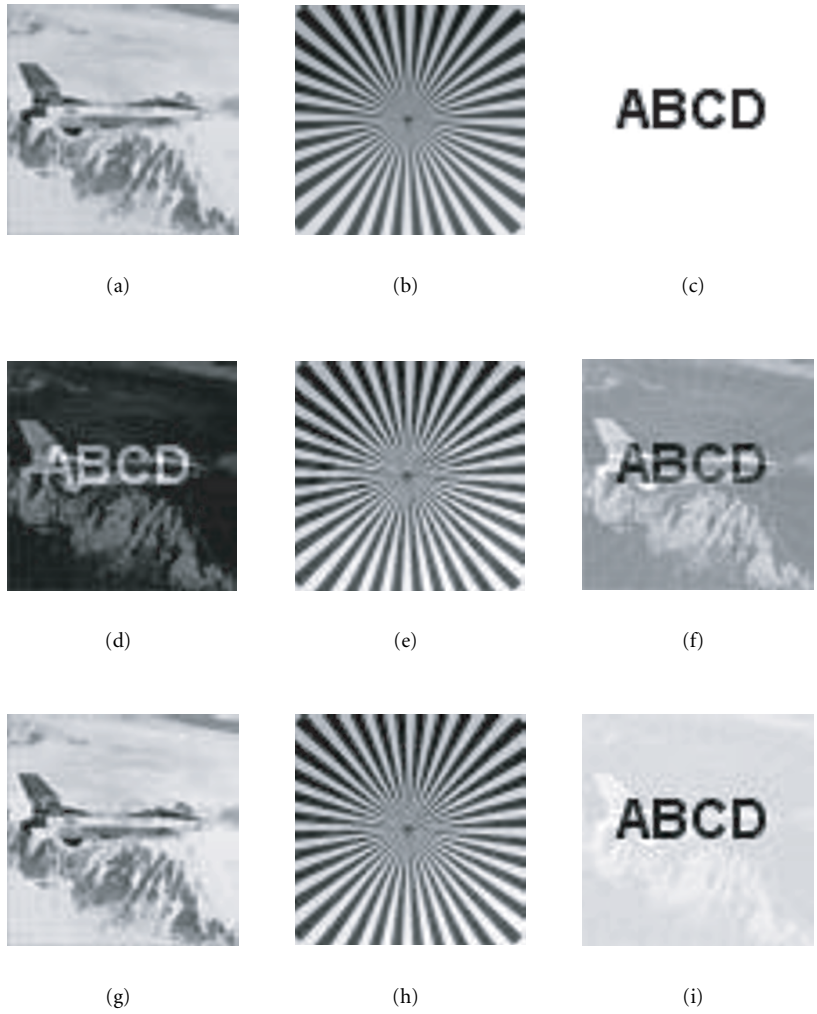


Figure 9: Another watermark extraction example: (a) original image—unwatermarked Airplane image, (b) key image, (c) text watermark, (d)–(f) extracted images using *column*-wise sequencing order, (g)–(i) extracted images using *row*-wise sequencing order.

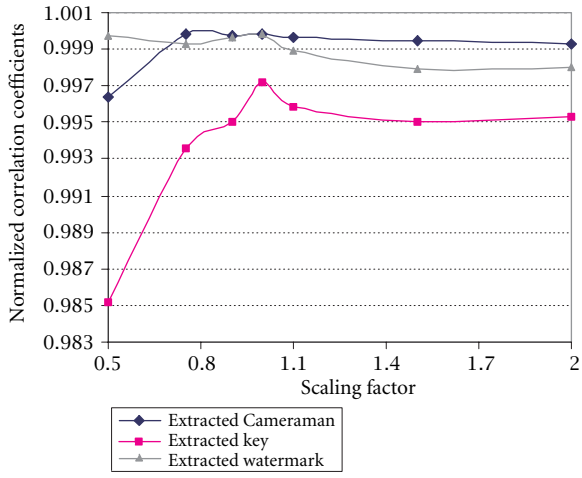
Table 5: Comparison of watermark extraction performance for column-wise or row-wise image scanning using the watermarked Airplane image measured with respect to the corresponding original images.

Test conditions	*Normalized correlation coefficients*, r		
	Airplane	Key	Watermark (text)
Column-wise sequencing	−0.7677	0.9958	0.7052
Row-wise sequencing	0.9978	0.9993	0.9874

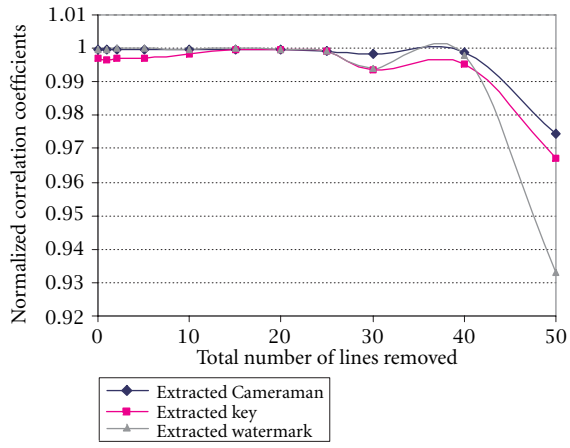
be degraded explicitly before the watermark is lost. Recently, a list of attacks have been proposed, against which image watermarking system could be judged. These attacks include low pass filtering, scaling, cropping, rotation, additive noise, quantization, and so on [17, 18, 19]. These various attacks are applied to the watermarked images to evaluate if the above-proposed dewatermarking system can recover the embedded watermark, thus measuring the robustness of the watermarking system to these types of attacks.

6.1. Geometric transformations

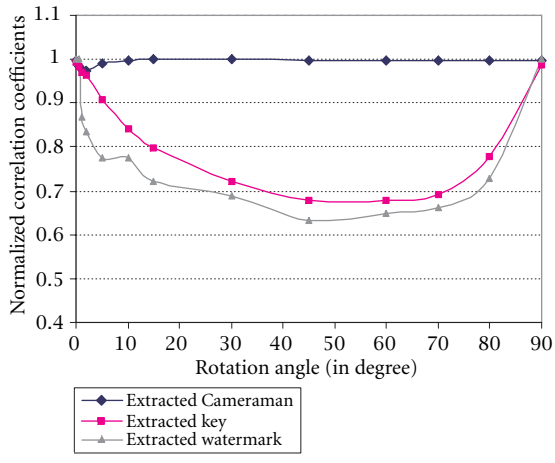
The robustness testing using the watermarking example given in Figure 1 under attacks of some geometric transformations—scaling, cropping, and rotation, are presented in Figures 10a, 10b, and 10c, respectively (using the absolute values of the normalized correlation coefficients). The system requires employing a synchronization template to detect the specific geometrical transformations, because it is needed to apply the same transformations to the key and the original images to generate another two observed mixtures. This is a restriction in handling the geometrical transformations for the proposed watermarking system. Nevertheless, experiments have been carried out with such synchronization template, proving that the information contained in the images after geometric transformations are still sufficient to retrieve the watermark.



(a)



(b)



(c)

FIGURE 10: Robustness tests of the proposed dewatermarking scheme against (a) scaling, (b) cropping, and (c) rotation attacks. (Note that the central portion of the image is kept after cropping, for example, 20 border lines are removed, meaning 10 lines are removed from top, bottom, left, and right sides, respectively.)

6.1.1 Scaling

Figure 10a shows an excellent behavior of the watermark extraction performance against scaling with factors from 0.5 to 2. Only when the scaling factor becomes very small and consequently the total number of pixels are very few, the extraction performance will be degraded. Due to the intrinsic robustness of the proposed watermarking scheme against this particular type of geometric distortion, the watermark extraction turns out to be extremely resistant against all types of practical resizing algorithms.

6.1.2 Cropping

The watermark and the key are permanently embedded into the original image by modifying intensity of all the pixels, therefore, the subparts of the watermark and the key are still present in the cropped watermarked image. Experiments have been carried out by supposing the position of the subimage can be determined properly. The performances are measured by comparing the correlations between the extracted subimages and their corresponding subpart of the original image, respectively. Figure 10b shows that the information contained in a subimage is still sufficient to extract the embedded subparts of the key and the watermark.

6.1.3 Rotation

The watermark extraction performances against rotation angle from 0 to 90 degrees are shown in Figure 10c. The quality of the extracted Cameraman image is almost immune to the rotation angle. The worst case is when the rotation angle is 45 degrees, the normalized correlation coefficients for extracted key and watermark are 0.6794 and 0.6327, respectively, which can still be easily verified visually.

6.2. Color quantization

The operation—color quantization is applied to the watermarked image (using the example in Figure 1), which is accompanied by dithering which diffuses the error of the quantization. Figure 11 illustrates the extraction results of quantization to 256 colors (grayscale) against the mixing coefficients a and b (in (4)). It shows clearly that the performance depends mainly on the weighting coefficient for the watermark, b. The normalized correlation coefficients for the extracted watermark degrades rapidly when b goes as low as 0.003.

6.3. Noise addition

The watermarked Bird image (with size of 256×256) is corrupted by the additive Gaussian noise with zero-mean and variance $\sigma^2 = 1000$, shown in Figure 13a. Although the image appears degraded heavily (comparing to its original image in Figure 12), the watermark is still able to be recovered, which is presented in Figure 13b. The maximum acceptable noise level is limited by the energy strength of the embedded watermark. The tests show that the watermark will become unperceptable, when the additive noise energy level goes up to 40–50 times higher than the energy level of the text watermark.

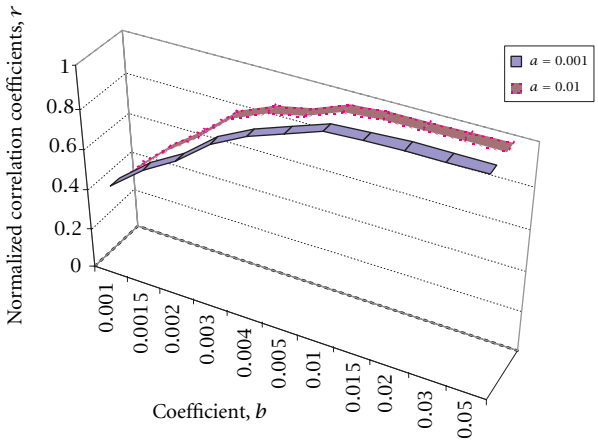
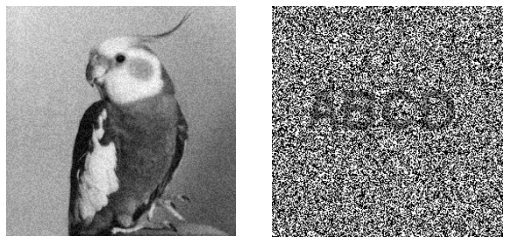


FIGURE 11: Robustness tests of the proposed dewatermarking scheme against quantization to 256 colors (where a, b are weighting coefficients for key **K** and watermark **M**, respectively, as in (4)).

FIGURE 12: The original Bird image.



(a) (b)

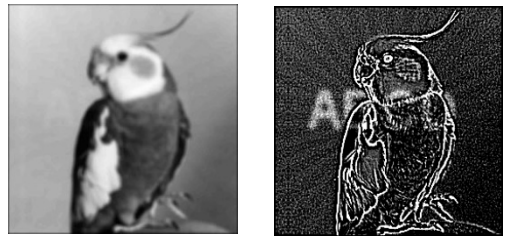
FIGURE 13: (a) Watermarked Bird image with additive Gaussian noise having variance $\sigma^2 = 1000$, (b) extracted watermark "ABCD" from (a).

6.4. Low pass filtering

Figures 14a and 15a are two watermarked Bird images filtered with a 2D low-pass Gaussian and a 2D median filter of size 5×5, respectively. The tests (see Figures 14b and 15b) demonstrate that the watermarking system can survive these types of low pass filtering attacks.

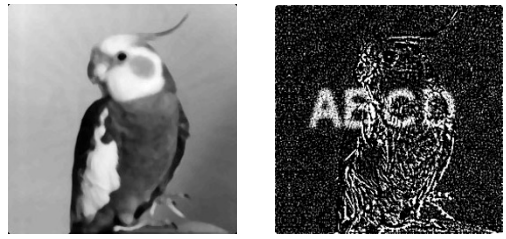
6.5. Multiple marks

Figure 16a shows a Bird image after inserting five different watermarks, which is another form of attack aiming to make original mark unreadable. Figure 16b shows the extracted watermark, from which it is able to clearly recognize the original watermark—the text "ABCD."



(a) (b)

FIGURE 14: (a) Watermarked Bird image filtered with a low pass Gaussian filter (with size of 5×5), (b) extracted watermark "ABCD" from (a).



(a) (b)

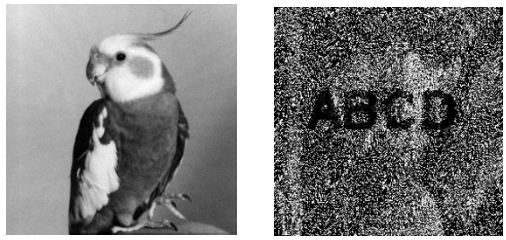
FIGURE 15: (a) Watermarked Bird image filtered with a median filter (with size of 5×5), (b) extracted watermark "ABCD" from (a).

6.6. Collusion

In this experiment in order to simulate collusion attack, five separately watermarked Bird images are generated and averaged them to obtain another Bird image shown in Figure 17a. The original text watermark still exist well in the extracted watermark image shown in Figure 17b.

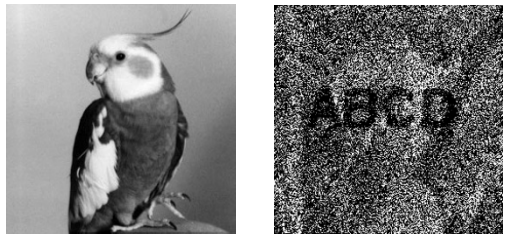
7. CONCLUSIONS

In this paper, a new image watermarking technique based on Independent Component Analysis (ICA) has been proposed. We have shown the efficacy and efficiency in applying ICA method for performing watermark detection and extraction. The watermark is readily detected by Principle Component Analysis (PCA) whitening process. The watermark can be further separated from the mixed source using a robust batch ICA algorithm. The robust batch algorithm is an effective blind source separation approach for temporally correlated sources, hence, the image sources are characterized as 1D temporal correlated sequences. The performance of the proposed method can be evaluated in terms of nor-



(a) (b)

FIGURE 16: (a) The Bird image with five different watermarks, (b) extracted watermark "ABCD" from (a).



(a) (b)

FIGURE 17: (a) The Bird image after averaging together five independently watermarked versions of the Bird image, (b) extracted watermark "ABCD" from (a).

malized correlation coefficient. Simulation results show satisfactory dewatermarking performance. We also have analyzed how the performance of watermark extraction is affected by statistical dependence of the sources and the source temporal correlation. Experimental results demonstrate the proposed watermarking scheme is robust to geometrical transformations (with a proper synchronization template), color quantization, low pass filtering, multiple marks, collusion, and can survive certain high level of additive noise attacks.

Future research work includes reducing the source dependence and increasing the source temporal correlation in order to improve the performance for different types of images as well as to make the proposed watermarking scheme more robust against various possible attacks.

ACKNOWLEDGEMENT

The authors are very thankful to the reviewer for his valuable suggestions, especially for referring the paper in [9] and suggesting us to consider more common and prominent *attacks* for performance analysis.

REFERENCES

[1] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079–1107, 1999.

[2] K. Stefan and F. A. P. Petitcolas, Eds., *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, Boston, 2000.

[3] A. Hyvärinen and E. Oja, "Independent Component Analysis: A Tutorial," http://www.cis.hut.fi/projects/ica/, April 1999.

[4] A. Hyvärinen, "Survey on Independent Component Analysis," *Neural Computing Surveys*, vol. 2, pp. 94–128, 1999.

[5] T.-W. Lee, *Independent Component Analysis—Theory and Applications*, Kluwer Academic Publishers, 1998.

[6] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 9, no. 10, pp. 2009–2026, 1998.

[7] D. Yu, F. Sattar, and K.-K. Ma, "Watermark detection and extraction using a independent component analysis method," in *Proc. of Nonlinear Signal and Image Processing*, Baltimore, Maryland, USA, June 2001.

[8] T.-W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letters*, vol. 6, no. 4, pp. 87–90, 1999.

[9] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[10] A. Cichocki and A. K. Barros, "Robust batch algorithm for sequential blind extraction of noisy biomedical signals," in *5th International Symposium on Signal Processing and its Applications (ISSPA'99)*, vol. 1, pp. 363–366, August 1999.

[11] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, 2nd edition, 1989.

[12] J. Karhunen, E. Oja, L. Wang, R. Vigário, and J. Joutsensalo, "A class of neural networks for independent component analysis," *IEEE Trans. Neural Networks*, vol. 8, no. 3, pp. 486–504, 1997.

[13] R. L. Kashyap and K.-B. Eom, "Robust image models and their applications," in *Advances in Electronics and Electron Physics*, P. W. Hawkes, Ed., vol. 70, pp. 79–157, Academic Press, 1988.

[14] W. A. Fuller, *Introduction to Statistical Time Series*, Wiley, New York, 2nd edition, 1996.

[15] F. Sattar, L. Floreby, G. Salomonsson, and B. Lövström, "Image enhancement based on a nonlinear multiscale method," *IEEE Trans. Image Processing*, vol. 6, no. 6, pp. 888–895, 1997.

[16] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 2nd edition, 1987.

[17] M. Kutter and F. A. P. Petitcolas, "A fair benchmark for image watermarking systems," in *Proc. of Electronic Imaging'99, Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 219–239, San Jose, California, USA, January 1999.

[18] F. A. P. Petitcolas and R. J. Anderson, "Evaluation of copyright marking systems," in *Proc. of IEEE Multimedia Systems'99*, vol. 1, pp. 574–579, Florence, Italy, June 1999.

[19] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems," in *2nd International Workshop on Information Hiding*, vol. 1525 of *Lecture Notes in Computer Science*, pp. 218–238, Portland, Oregon, USA, April 1998.

Dan Yu was born in Jiangxi Province, People's Republic of China, and received her B. Eng. (Electrical and Electronic Engineering) degree in communication with first class honors, from Nanyang Technological University (NTU), Singapore, in June 2000. She is currently pursuing the Ph.D. degree at the School of Electrical and Electronic Engineering, NTU. Her research interests include watermarking, blind source separation, and image processing.

Farook Sattar was born in Dhaka, Bangladesh, and received the B.S. and M.S. degrees in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh. He received his Technical Licentiate and Ph.D. degrees in signal and image processing from Lund University, Lund, Sweden. He is currently an Assistant Professor at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research focuses on signal and image processing with current interests in blind source separation, watermarking, transient signal detection, filter banks, wavelets, time-frequency representation, and image enhancement.

Kai-Kuang Ma received his Ph.D. degree from North Carolina State University, Raleigh, North Carolina, M.S. degree from Duke University, Durham, North Carolina, USA, all in electrical engineering, and B.E. degree (electronic engineering) from Chung Yuan Christian University, Taiwan, Republic of China. He was with IBM Corporation in USA, from 1984 to 1992, before he joined the Institute of Microelectronics, National University of Singapore, and worked on MPEG video research. In early 1995, Dr. Ma joined the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, and he is currently an Associate Professor. His research focuses on digital image/video coding and their standards, content-based image/video indexing and retrieval, joint source and channel coding for wireless image and video, wavelets and filter banks, and other key aspects on multimedia signal processing and communications. Dr. Ma has been serving as the Chairman and Head of Delegation for Singapore in MPEG and JPEG, from 1997 to 2001. In 1999 and 2000, two motion estimation technologies (i.e., Diamond Search and MVFAST) resulted from his research group have been adopted by MPEG-4 standard and integrated into MPEG-4 reference software. Dr. Ma has been serving as an Editor of the IEEE Transactions on Communications and the Chairman of IEEE Signal Processing Chapter, Singapore Section. He has been acting as program committee member and session chair of multiple IEEE international conferences and international standardization working group meetings. Dr. Ma is a Senior Member of the IEEE and a member of Sigma Xi and Eta Kappa Nu.