



G00002



159159

Techniques for embedding and detecting watermarks in data

Field of the invention

The present invention relates to methods for embedding watermarks in data, and methods for detecting watermarks in data.

5 Background of Invention

In recent decades, many watermarking techniques have been proposed and developed for application in tasks such as copyright protection, data hiding, image captions, and so on. These techniques modify a first set of data
10 (referred to as "host" data) using a second set of data (referred to as "watermark data") to produce a third set of data ("watermarked host data"), such that the fact that the host data has been modified is detectable, e.g. to people having access to the watermark data. In various such techniques the host data is data in the spatial domain, frequency domain, discrete cosine
15 transform (DCT) domain, Mellin-Fourier transform domain, or wavelet domain, etc.. The key techniques that are popularly used for embedding the watermark are referred to as the spread spectrum (SS) technique, and quantization index modulation (QIM). Some techniques for detecting of the watermark in the watermarked data do not employ a knowledge of the host data (these are
20 known as "blind" techniques). Other techniques do employ knowledge of the host data (these are known as "non-blind" techniques).

Some of the most influential works in digital watermarking use the SS technique by embedding the watermark (which may be a modulated
25 pseudorandom sequence) into a "sequence" of the host data (a "host sequence"). The term "sequence" is used here to mean any subset of the elements of the host data. It need not be consecutive elements of the host data; for example, if the host data is an image, with respective elements of the

host data corresponding to the intensities of respective pixels of the image, the term "sequence" is broad enough to cover the intensities of any subset of the pixels, and indeed the order of the elements in a given sequence need not be identical to the order of the corresponding pixels in the image.

5

The embedding may be performed by summing by a linear additive embedding analysis. Specifically, a "0" or a "1" is embedded into a host sequence of length N , denoted by $f(n)$ ($n = 1 \dots N$), by modifying the host sequence using a respective one of two set of N numbers. These sets of numbers are denoted respectively by w_i^0 and w_i^1 ($i = 1 \dots N$) and referred to as "watermark patterns". At the time of detection, it can be determined which of the watermark patterns was used to watermark the host sequence by measuring the correlation of the watermarked host sequence with the watermark patterns [1-3]. Thus, these techniques are analogous to popular CDMA method for encoding data.

15

One known embedding formula is:

$$f'(n) = f(n)(1 + \alpha \cdot w_i^k) \quad (1)$$

where k is the value to be embedded into the host sequence (i.e. "0" or "1"),

α is a strength factor to control the masking threshold and watermark robustness, and $f'(n)$ is the watermarked data.

20

Another embedding formula is:

$$f'(n) = f(n) + \alpha \cdot w_i^k \quad (2)$$

25

In either case (1) or case (2), non-blind watermark detection is conventionally performed by subtracting the host sequence $f(n)$ from the watermarked sequence $f'(n)$ and determining whether the resultant data is correlated with

watermark series w_i^0 or alternatively with watermark series w_i^1 , thus determining whether a 0 or a 1 was embedded in the host sequence. Unfortunately, this detection is accurate only when the subtraction can completely remove the host sequence from the watermarked sequence, which
 5 is hard to be achieve as the amplitudes of the host sequence and the watermarked sequence may be different.

Blind watermark detection is conventionally performed by measuring the correlation of each of the watermark patterns (w_i^0 and w_i^1) directly with the
 10 watermarked sequence f' . For example, in the case of w_i^0 by measuring

$$\delta = \frac{\sum^N f(n)' \cdot w_i^0}{N}. \quad (3)$$

Due to the central limit theorem this mutual correlation is higher in the case of the watermark pattern which was used to produce the data. The presence of a "0" embedded in the data is thus determined by determining whether the
 15 correlation value δ is above or below a detection threshold T which is the average value between the autocorrelation of w_i^0 and the correlation of w_i^0 and w_i^1 (similarly the presence of a "1" in the data can be determined using a δ value which is obtained by modifying (3) by substituting w_i^1 for w_i^0).

20 The robustness of this SS technique is rather dependent on the bias effect from host sequence $f(n)$ and the strength factor α . In blind watermark detection, there always exists the bias effect from the host data, and the effect is hard to predict to properly adjust the detection analysis. A larger watermarking strength would result in a more robust detection result, but it
 25 would also reduce the image quality, so the strength factor α should be kept at a relatively small value. Therefore, how to achieve a good balance point between the good robustness and good quality is always the problem in the

SS scheme.

In QIM watermarking, a set of features extracted from the host signal are quantized by means of a quantizer chosen from a pool of predefined quantizers on the basis of the message to be hidden. In the simplest case, a set of uniform quantizers is used leading to lattice-based QIM watermarking [4].

A simple quantization scheme can be applied to a small sequence of host data. A pair of n -length sequences are chosen as $f_1(1) \dots f_1(n)$ and $f_2(1) \dots f_2(n)$. The absolute difference between two groups is given by

$$\Delta = f_1(\text{total}) - f_2(\text{total}),$$

where $f_1(\text{total}) = |f_1(1)| + |f_1(2)| + \dots + |f_1(n)|$ and

$$f_2(\text{total}) = |f_2(1)| + |f_2(2)| + \dots + |f_2(n)|.$$

15

In order to embed one WM bit, k , which is 0 or 1, the coefficients $f_1(1) \dots f_1(n)$ and $f_2(1) \dots f_2(n)$ are modified such that the absolute distance becomes

$$\Delta = \begin{cases} \leq -Q, & \text{when } k = 0 \\ \geq Q, & \text{when } k = 1 \end{cases}$$

where Q is a parameter controlling the detection robustness or tolerance.

20

In the detection stage, the watermark bit k is simply detected by comparing the coefficient pair or two sequence groups:

$$f_1(\text{total}) \leq f_2(\text{total}) \Rightarrow k = 0$$

$$f_1(\text{total}) > f_2(\text{total}) \Rightarrow k = 1$$

25

The robustness of quantization method depends on the tolerance of the quantization step and the precision of the host sequence and the

watermarked sequence. The robustness decreases quite a lot when the data is compressed and coded.

Several watermarking methods based on ICA (independent component analysis) have also been proposed in the last few years. Noel and Szu [5] were the first to point out that a de-mixing analysis based on an ICA seeking statistically factorized probability density can yield a watermarking technique using unsupervised neural networks. Yu, et. al. [6] and Shen, et. al. [7] employed similar ideas to realize a watermarking system by directly using ICA separation techniques. All these analyses employed a watermarking technique in which the bits of the watermark data are added directly to respective pixels of the host data, which must be an image. An ICA separation technique is used to extract the watermarks directly from the watermarked data.

There are three problems in this scheme: (1) the watermarking scheme is always a non-blind detection because the original host is needed as one of the de-mixing sources to separate the watermarks; (2) the extracted watermark is not exactly same as the original watermark, because the host data and watermark data are not independent statistically, so the two separated signals obtained by ICA may be a mixture of the host data and watermark data; (3) the inherited dilation and permutation in ICA cannot guarantee producing exactly the desired watermarks. Thus, detection error is not avoidable in this scheme, so their robustness is questionable. This watermarking scheme may only work for applications only requiring non-precise watermark detection.

Summary of the Invention

The present invention aims to provide a new and useful watermarking technique using a blind source separation (BSS) method which separates
5 statistically independent components from multiple mixtures.

In general terms the invention proposes that watermark data is modulated onto host data by using bits of the watermark data to select watermark patterns, and then linearly adding the watermark patterns to sequences of the host data. To recognise the watermark patterns within the watermarked host
10 data, an ICA de-mixing process (or more generally another BSS method) is performed taking as inputs (i) the watermarked host data, and (ii) the host data and/or at least one of the watermark patterns. The watermarked patterns can be recognised from the results of this ICA de-mixing process. The watermark data is then regenerated from the recognised watermark patterns.

15 This regenerates the watermark data with a high degree of accuracy.

Note that the term BSS is broader than ICA. BSS is a term for the problem of blindly separating sources from mixture data, and the term "BSS methods" refers to all solutions to achieve BSS. An ICA de-mixing analysis is currently the most suitable and most generic BSS method, but other methods, such as
20 second-order separation (SOS) can also solve the BSS problem under certain circumstances. (note that the term BSS should not be confused with the term "blind detection" as defined here).

Each of the watermark patterns has a plurality of elements, typically at least 1000, but more preferably about 6000. The host data is watermarked by
25 linearly adding the elements of the selected watermark pattern to respective elements of a sequence of the host data.

A first way of selecting the watermark patterns using the watermark data is on a bit-by-bit basis, i.e. for each bit of the watermark data there are two possibilities for the watermark pattern, and one of these is selected according to the value of that bit.

- 5 More generally, the watermark data may be divided into groups of bits (e.g. of two bits each), and each group used to select one of a respective set of watermark patterns. For example, when each group contains two bits of the watermark data (which may take the values 00, 01, 10, or 11), the two bits of the group may be used to make a selection from among 4 watermark
10 patterns. Thus, when one of these watermark patterns is recognised in the watermarked host data, this means that 2 bits of the watermark data are regenerated.

Typically, different respective sequences of the host data are watermarked using the watermark patterns selected using different respective bits (or
15 groups of bits) of the watermark data. Typically, the possible watermark patterns are different for each bit (or groups of bits) of the watermark data.

There are various ways within the scope of the invention in which the presence of the watermark pattern can be recognised from the results of the ICA de-mixing process.

- 20 If the host data is known, a first way of recognising which watermark pattern was used to watermark it is by performing an ICA de-mixing process on the watermarked host data and the host data.

Alternatively (and this can be performed even if the host data is unknown), the ICA de-mixing process can be performed on the watermarked host data and
25 one of more of the watermark patterns. In particular, for each of the possible watermark patterns, a respective ICA de-mixing process can be performed, taking as inputs the watermarked host sequence and that watermark pattern.

When the inputs to the ICA de-mixing process are correlated or dependent in a statistical sense (as they are in the case that the inputs are a host sequence and the watermarked version of the host sequence; or in the case that the inputs are a watermarked host sequence and the watermark pattern which was used to produce that watermarked host sequence), the outputs of the ICA de-mixing process would be respectively the host data and the watermark pattern.

A further possibility, in the case that the host data was unknown (i.e. this is blind detection) is to take the host data extracted by the ICA de-mixing process and to input it together with the watermarked host data into a second ICA de-mixing process. Once again, the outputs should be the watermark pattern and the host data.

In any of these three cases, the recognition of the watermark pattern may be performed by observing the independent components output from the de-mixing process (e.g. whether one of them is correlated with one of the watermark patterns) or by evaluation of the de-mixing matrix which the ICA de-mixing process produces.

Typically, each watermark pattern consists of a set of N values which are each either real numbers or binary numbers (0 or 1). Preferably, the set of watermark patterns have a known correlation. They may for example be uncorrelated. In a particularly preferred case they are statistically independent to every order (they may, for example, be themselves generated by a pre-processing step of ICA de-mixing process).

Alternatively, there may be two possible watermark patterns for each host sequence and they may be opposite to each other. This is equivalent to saying that there is only a single watermark pattern for each host sequence, and that each watermark bit is encoded by selecting a +/- sign for the linear addition of that watermark pattern.

In non-blind watermark detection, the ICA separation can remove the host data from the watermarked data completely so that the detection on separated watermark pattern is much more robust than the conventional methods described above since the bias effect from the host data is decreased dramatically.

In blind watermark detection, the different output components and de-mixing matrices obtained in two ICA analyses using embedded and non-embedded watermark patterns can be used to recognize the embedded watermark bits. One way is by measuring the correlation (or higher order dependence) between the output components and the unused watermark pattern: the smaller correlation value (or smaller higher-order dependence value) indicates the embedded watermark bit. Another way is to use the ratio of the highest singular value of the matrix to the lowest: a larger ratio indicates the embedded watermark bit.

In the case that the two watermark patterns are opposite (i.e. there is only one watermark patterns, which is added to the host sequence with either a positive or negative sign) recognising which sign was used is simple: it can be done by determining whether a certain correlation value is positive or negative, or by examining an element in the separation matrix.

Note that the host data may be any form of data. In particular, it may be data in the spatial domain, or transformed data such as DCT data, wavelet data, ICA domain data, etc. The present technique can be used in any watermarking applications on image, audio and video data, such as copyright protection and so on.

Brief Description of The Figures

Preferred features of the invention will now be described, for the sake of illustration only, with reference to the following figures in which:

Fig. 1 is schematic representation of a mixing process and an ICA de-mixing process, shown as two respective feed-forward neural networks;

Fig. 2, which is composed of Figs. 2(a) to 2(d), is an example of two observed signals (Figs. 2(a)), their statistical dependence, illustrated by their individual and joint probability densities (Fig. 2(b)), their components as obtained by an ICA de-mixing process (Figs 2(c)), and the statistical dependence of these components, illustrated by their individual and joint probability densities (Fig. 2(d));

Fig. 3 is the flowchart of watermark embedding process; and

Fig. 4 is an ICA separation process for watermark detection in an embodiment of the invention by a 2-by-2 linear de-mixing computation with input X , de-mixing matrix W and output Y .

Detailed Description of the embodiments

To begin with, we will provide a mathematical formulation of the problem of blind source separation (BSS) and an explanation of what is meant by an ICA de-mixing process.

Let us assume that there are a set of n original signals denoted by $\mathbf{s}_i$, ($i=1, 2 \dots n$). These signals are written as the respective rows of a matrix $\mathbf{S}$. Let us further suppose that there is a simple linear mixing among these signal components to produce a set of mixture signals,

$$\mathbf{X} = \mathbf{A} \times \mathbf{S} \quad (4)$$

where the matrix $\mathbf{A}$ contains the mixing weights applied to original signals; the sum of weighted original signals generates the new set of signals. The

objective of a BSS problem is to separate the original sources $\mathbf{S}$ from the observed signals $\mathbf{X}$ without having any prior knowledge of $\mathbf{A}$ and $\mathbf{S}$. It is theoretically impossible to achieve this goal as it is an ill-posed problem unless we have some extra assumptions. The assumption we talk about here
 5 is the independence among the original signals. This brings out the technique of independent component analysis (ICA) [8-9] to solve the BSS problem.

The uniqueness of the ICA technique, and its advantage in particular applications, is to transform the observed data into statistically independent
 10 components. As an example, let us treat $\mathbf{X}$ as an observation matrix with each row $\mathbf{x}_i$ as an observed vector. The ICA tries to find a transform matrix $\mathbf{W}$ (by any means, e.g. data-adaptive learning) such that, if

$$\mathbf{Y} = \mathbf{W} \times \mathbf{X} \quad (5)$$

the transformed vectors $\mathbf{y}_i$ on rows of the transformed data $\mathbf{Y}$ are statistically
 15 independent from each other.

Mathematically, this independence means that the correlation of any functions, f , with equal or more than 2nd order on any transformed data $\mathbf{y}_i$ and $\mathbf{y}_j$ in $\mathbf{Y}$ ($i \neq j$) is equal to zero:

$$20 \quad f_1(\mathbf{y}_i) \times f_2(\mathbf{y}_j) = 0 \quad (6)$$

Figure 1 illustrates a feed-forward neural network which performs mixing and de-mixing processes. The input layer contains n neurons labelled by $1i$, where $i=1, \dots, n$. At any time, a given row of $\mathbf{S}$ is considered, and the input neurons $1i$
 25 are set equal to the corresponding values $\mathbf{s}_i$, ($i=1, 2 \dots n$) of that row of $\mathbf{S}$. second layer of n neurons $2i$, where $i=1, \dots, n$, is has values equal to the corresponding row of $\mathbf{X}$. That is, the matrix $\mathbf{A}$ represents the weights between the input layer neurons $1i$ and the second layer of neurons $2i$. The output layer of neurons $3i$ takes the values of the corresponding row of $\mathbf{Y}$. That is, the
 30 weights between the second layer neurons $2i$ and the output layer neurons $3i$

are the matrix $\mathbf{W}$. The task of an ICA de-mixing process is to construct this matrix $\mathbf{W}$ such that the values of the output layer neurons $3i$ are as independent as possible.

- 5 Figure 2 shows an example of observation data $\mathbf{X}$ having two components. Fig. 2(a) shows these two components as respective time series shown in the upper and lower parts of Fig. 2(a) respectively. Fig. 2(b) shows the statistical dependence, illustrated by the individual and joint probability densities, of these two components. Figs. 2(c) and (d) show the results of the ICA. The
- 10 transformed data $\mathbf{Y}$ has two independent components which are time series shown respectively by the upper and lower parts of Fig. 2(c), and the statistical dependence, illustrated by the individual and joint probability densities, of these two components is shown in Fig. 2(d).
- 15 Having defined what is meant by an ICA de-mixing process, we now turn to embodiments of the present invention. Here, we apply the similar ICA technique of mixing and de-mixing in BSS to embed and detect watermarks, respectively.
- 20 Four techniques which are respective embodiments of the invention are illustrated below. The first two embodiments are for use with a type of watermarked data in which each sequence of the host data is watermarked as "0" or "1" by adding to it a respective one of two predetermined watermark patterns. The second two embodiments are for use with a type of
- 25 watermarked data in which each sequence of the host data is watermarked as "0" or "1" by adding to a predetermined watermark pattern with respectively a positive or negative sign.

(a) Generation of the first type of watermark patterns

This is similar to the generation of watermark data described above with reference to Eqn. (2). We denote the sequence of N values of host data into which a single watermark bit is to be embedded by $\mathbf{x} = (x_1, x_2, \dots x_N)$. This is normalized to have zero mean and unit variance. The watermark bits, 0 or 1, are encoded using respective series of "watermark patterns".

In some embodiments of the invention these watermark patterns may simply be N -strings of random numbers: $\omega_0 = (\omega_{01}, \omega_{02}, \dots \omega_{0N})$ and $\omega_1 = (\omega_{11}, \omega_{12}, \dots \omega_{1N})$. Each have zero mean and unit variance.

More preferably, however, the two series of random numbers ω_0 and ω_1 are used as the data matrix $\mathbf{X}$ input to an ICA de-mixing process ("ICA pre-processing"). This generates a pair of completely statistically independent signals, ω_0^I and ω_1^I , which each still have the property of zero mean and unit variance. Any cross-functions between this pair result in zero values as defined in Eqn. (6). These independent signals are more preferably used as the watermark patterns instead of ω_0 and ω_1 . This will be assumed in the following explanation.

(b) Watermarking using the first type of watermark patterns

For the sake of simplicity we will describe in detail a case in which respective bits of the watermark data are encoded by watermarking the host data with a respective one of the two watermark patterns. This produces what is here referred to as a "first type of watermarked data". Note that for each bit of the watermark data the two watermark patterns are different.

Consider bit k of the watermark data. The value k (which may be "0" or "1") is embedded into a sequence $\mathbf{x}$ of the host data, where the term "sequence" is as defined above. In the embodiments described in detail here, the host

sequences do not overlap (i.e. no elements of the host data are present in two sequences), but this possibility is not excluded in the embodiments of present invention.

- 5 The embedding process may be as follows:

$$\mathbf{x}' = \mathbf{x} + \alpha \times \omega_k^l \quad (7)$$

where ω_k^l is the one the watermark patterns ω_0^l or ω_1^l which corresponds to the value k , α denotes the watermarking strength ratio, and $\mathbf{x}'$ is the watermarked data.

10

Figure 3 illustrates this embedding process in a flowchart. Unit 41 generates host data, normally by making a selection from a database of host data. A watermark generation unit 43 generates the watermark data ω_k^l according to which value of k it is desired to embed in each respective sequence of the host data. The watermark data ω_k^l is multiplied by the value α using a multiplier 45, and the result is added to the host sequence $\mathbf{x}$ by an adder 47. The result is watermarked data $\mathbf{x}'$.

15

Normally, a normalisation is performed of the watermarked data $\mathbf{x}'$, to produce normalised data $\mathbf{x}''$. This is represented by $\mathbf{x}'' = \beta \times \mathbf{x}'$.

20

(c) First embodiment: Non-blind detection of the first type of watermarked data

In this first embodiment the host data $\mathbf{x}$ is available to the detector (i.e. the computer system which is arranged to detect the presence of watermarking in the data $\mathbf{x}''$). The detector has a first unit which uses $\mathbf{x}$ and $\mathbf{x}''$ to construct a matrix $\mathbf{X}$ given by:

25

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'' \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \beta(\mathbf{x} + \alpha \omega_k^l) \\ \mathbf{x} \end{bmatrix} \quad (8)$$

The first unit passes the matrix $\mathbf{X}$ to an ICA unit which performs an ICA operation on the matrix $\mathbf{X}$. Ideally, the ICA unit generates two outputs: $y_1 = \mathbf{x}$ and $y_2 = \omega_k^l$ with unit variances. This should be realisable, since $\mathbf{x}$ and ω_k^l are two independent components separable by a linear de-mixing matrix:

$$\mathbf{W} = \begin{bmatrix} 0 & 1 \\ \frac{1}{\alpha} \times \frac{1}{\beta} & -\frac{1}{\alpha} \end{bmatrix} \quad (9)$$

Figure 4 shows such a 2-by-2 linear de-mixing process. The successive values x_1 from the top row of matrix $\mathbf{X}$ are fed to an input neuron 51, while the corresponding values x_2 from the bottom row of matrix $\mathbf{X}$ are fed to input neuron 52. At each instant, the output neurons 53, 54 take the respective top and bottom values of the vector obtained by multiplying a 2-component vector obtained from the values of the input neurons 51, 52 by a 2x2 matrix $\mathbf{W}$. The ICA unit derives the matrix $\mathbf{W}$ (using known ICA methods) such the series of values y_1 successively taken by the output neuron 53 is statistically independent from the series of corresponding values y_2 successively taken by the output neuron 54. The matrix $\mathbf{Y}$ has a top row which is the N values y_1 and a bottom row which is the N values y_2 .

Since the values y_1 are statistically independent from the values y_2 , it is expected from Eqn. (8) that one of the set of values y_1 will take the values $\mathbf{x}$ while the other will take the values ω_k^l . Note that the output neurons 53, 54 have an indeterminacy of permutation, which means that either the output neurons 53, 54 may output ω_k^l while the other outputs $\mathbf{x}$.

To detect the embedded watermark bit, the embodiment uses a two-step process:

- (i) identify which output of the ICA de-mixing process (i.e. y_1 or y_2) corresponds to the watermark pattern. The embodiment does this based on the observation that, as stated in Eqn. (9), one of the two outputs y_1, y_2

corresponds closely to the host data $\mathbf{x}$ and has very little contribution from the watermarked data $\mathbf{x}''$. The β value is close to one since the watermark modification is usually small, so this one of the neurons 53, 54 is characterized in that the contributions from both its inputs are similar, but with different signs. Based on this, we can identify which of the two outputs y_1, y_2 corresponds closely to the host data $\mathbf{x}$.

(ii) watermark detection. The embodiment then obtains the normalized cross-correlation between the output representing watermark data (i.e. the one of the outputs 53, 54 which does not correspond to $\mathbf{x}$) and the watermark data ω_k^l for one of the values of k . Given this value, the embodiment can look up which watermark pattern can be deduced from table 1.

		Watermark Pattern	
		ω_k^l	
		ω_0^l	ω_1^l
Output from whichever of neurons 53, 54 corresponds to watermark data	If $k=0$ so output is ω_0^l	1	0
	If $k=1$ so output is ω_1^l	0	1
Output from the other of the neurons 53, 54	$\mathbf{x}$	ϵ_0	ϵ_1

Table 1: Cross-correlation values between the output 53, 54 and the two independent watermark patterns.

As shown in Table 1, the cross-correlation values between the two independent watermark patterns ω_k^l is zero, and the auto-correlation value of watermark patterns ω_k^l is equal to 1. The cross-correlation values between the host data $\mathbf{x}$ and the watermark patterns ω_k^l are very small, close to zero, and are denoted respectively by ε_0 and ε_1 .

Thus, the embodiment determines whether the output 53, 54 representing watermark data has a high correlation with one of the ω_k . If so, then that sequence of the host data must be watermarked with that value of k . Otherwise, i.e. if the auto-correlation is near zero, it must be watermarked with the other value of k .

Note that the watermark detection can still be performed even if step (i) is omitted (or if step (i) cannot identify the desired output), by calculating the cross-correlation of each of the two outputs 53, 54 with each of the two sets of watermark data ω_0^l . As can be seen from table 1, one of these two sets of watermark data has a low correlation with both of the outputs 53, 54, whereas the other set of watermark data has a high correlation with exactly one of them. This indicates that the host data was watermarked with the latter series of watermark data.

As described above, the conventional subtraction method to remove the host data $\mathbf{x}$ from the watermarked data $\mathbf{x}''$ may be problematic because the amplitudes of two kinds of data may be very different since they may be normalized before embedding and detection. Without knowing their exact amplitudes, there may be a bias effect from the remaining host data and this distorts the detection results. By contrast, in the method according to the first embodiment, the separation done by the ICA de-mixing process can remove the host data $\mathbf{x}$ from the watermarked data $\mathbf{x}''$ completely. The weight value in the de-mixing matrix can be learnt by the ICA de-mixing process performed by

the ICA unit, and this cancels out the amount of host data $\mathbf{x}$ in $\mathbf{x}''$. The independence between the host data $\mathbf{x}$ and watermark patterns ω_0^l and ω_1^l thus ensures the complete separation in non-blind watermark detection.

- 5 Note that although the first embodiment has been explained above on the basis of performing the watermark detection using the cross-correlation of the output representing watermark data and the watermark data itself, the embodiment may be varied to use any higher-order dependence function, thereby obtaining a difference dependence measure. Table 2 shows a
- 10 generalised form of Table 1, in which a different dependence measure is used which is a higher-order dependence than correlation. Such a measure can be used as an alternative to using correlation values (or in principle in combination with the use of correlation values, e.g. to confirm results obtained using correlation values). In this measure the auto-dependence measure of
- 15 ω_0^l is denoted $\Delta(\omega_0^l)$ and the auto-dependence measure of ω_1^l is denoted $\Delta(\omega_1^l)$.

		Watermark Pattern	
		ω_0^l	ω_1^l
Output from whichever of neurons 53, 54 corresponds to watermark data	If $k=0$ so output is ω_0^l	$\Delta(\omega_0^l)$	0
	If $k=1$ so output is ω_1^l	0	$\Delta(\omega_1^l)$

Output from the other of the neurons 53, 54	$\mathbf{x}$	ε_0	ε_1
---	--------------	-----------------	-----------------

Table 2: Generalised form of Table 1 for an alternative dependence measure, which is a higher-order dependence than correlation.

5 (d) *Second embodiment: Blind detection of the first type of watermarked data*

In this embodiment the host sequence $\mathbf{x}$ is not available. The embodiment performs the detection only using the watermark patterns ω_0^I and ω_1^I , representing the watermark bits 0 and 1, and the watermarked host sequence $\mathbf{x}''$.

Without loss of generality, assume that the watermark pattern ω_0^I has been embedded into the host sequence, so the watermarked data to be detected is $\mathbf{x}'' = \beta \times (\mathbf{x} + \alpha \times \omega_0^I)$.

15

In this embodiment the first unit uses ω_0^I and $\mathbf{x}''$ to form a first matrix $\mathbf{X}_0$:

$$\mathbf{X}_0 = \begin{bmatrix} \mathbf{x}'' \\ \omega_0^I \end{bmatrix} = \begin{bmatrix} \beta(\mathbf{x} + \alpha \omega_0^I) \\ \omega_0^I \end{bmatrix} \quad (10)$$

This matrix $\mathbf{X}_0$ is the input to an ICA unit. Furthermore a second matrix $\mathbf{X}_1$ is formed using ω_1^I and $\mathbf{x}''$:

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{x}'' \\ \omega_1^I \end{bmatrix} = \begin{bmatrix} \beta(\mathbf{x} + \alpha \omega_0^I) \\ \omega_1^I \end{bmatrix} \quad (11)$$

20

and this too is input to the ICA unit. The ICA unit processes the matrices $\mathbf{X}_0$ and $\mathbf{X}_1$ separately, to derive respective transform matrices $\mathbf{W}_0$ and $\mathbf{W}_1$.

For the matrix $\mathbf{X}_0$ in Eqn. (10), the watermark pattern ω_0^I is the part of the watermarked data $\mathbf{x}''$, so the ICA de-mixing process is able to separate this

25

pattern ω_0^l from the host data $\mathbf{x}$ in the watermarked data $\mathbf{x}''$, as they are statistically independent. In ideal case, the ICA de-mixing process separates two outputs $\mathbf{y}_1 = \mathbf{x}$ and $\mathbf{y}_2 = \omega_0^l$ with unit variances by a linear de-mixing matrix $\mathbf{W}_0$:

$$5 \quad \mathbf{W}_0 = \begin{bmatrix} 1 & -\alpha \\ \beta & 1 \end{bmatrix} \quad (12)$$

However, for the matrix $\mathbf{X}_1$ given by Eqn. (11), the watermark pattern ω_1^l is completely independent from the other watermark pattern ω_0^l , and thus the top row of $\mathbf{X}_1$ is totally independent from the bottom row. Ideally, since the two
 10 rows in the input matrix are already independent from each other, the output remains the same as the input, $\mathbf{y}_1 = \mathbf{x}''$ and $\mathbf{y}_2 = \omega_1^l$; the linear de-mixing matrix $\mathbf{W}_1$ is an identity matrix, or its permutation with certain dilations:

$$\mathbf{W}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (13)$$

15 In this embodiment, the watermark bit can be detected in either of two ways, based respectively on: (1) the properties of the de-mixing matrix, and (2) the properties of outputs of the ICA unit (i.e. $\mathbf{Y}_0 = \mathbf{W}_0 \cdot \mathbf{X}_0$ and $\mathbf{Y}_1 = \mathbf{W}_1 \cdot \mathbf{X}_1$).

Considering first approach (1), based on the de-mixing matrices given in
 20 Eqns. (12) and (13), we can compare them by obtaining the ratio of the largest singular value of $\mathbf{W}$ to the smallest. The singular values can be computed by a standard singular value decomposition (SVD) method. This ratio is 1 for the identity matrix or its permutation with dilation (which is the lowest value this ratio can take). Any higher ratio indicates that the embedded
 25 watermark bit is the one that the corresponding watermark pattern represents, i.e. bit 0 in this example.

Alternatively, approach (2) can be performed by obtaining the cross-correlations of the output components from the ICA separation with the watermark patterns(i.e. the cross-correlation between ω_1^l and outputs y_{01}, y_{02} from the de-mixing matrix $\mathbf{W}_0$, and the cross-correlation between ω_0^l and outputs y_{11}, y_{12} from the de-mixing matrix $\mathbf{W}_1$), and determining which of the cases shown in Table 3 is present. The small cross-correlation value indicates the embedded watermark bit. In this example, it is bit 0 represented by pattern ω_0^l .

		Watermark Pattern	
De-mixing Matrix	Outputs	ω_0^l	ω_1^l
$\mathbf{W}_0$	$y_{01} = \mathbf{x}$	ε_0	ε_1
	$y_{02} = \omega_0^l$	1	0
$\mathbf{W}_1$	$y_{11} = \mathbf{x}'' = \beta \times (\mathbf{x} + \alpha \times \omega_0^l)$	$\beta(\alpha + \varepsilon_0)$	$\beta\varepsilon_1$
	$y_{12} = \omega_1^l$	0	1

Table 3: Cross-correlation values between the ICA outputs and watermark patterns: the cross-correlation values between two independent watermark patterns is zero, the auto-correlation value of watermark patterns is equal to 1, cross-correlation values between the host data and the watermark patterns are very small, close to zero, denoted by ε .

Because the outputs from the separable de-mixing, $\mathbf{W}_0$ in this example, contain the separated host data and the embedded watermark pattern, ω_0^l , their correlation values with the non-embedded watermark pattern, ω_1^l in this example, are zero or very small, close to zero. But, the other de-mixing matrix, $\mathbf{W}_1$ in this example, does not separate the signals at all; one output contains the embedded watermark pattern ω_0^l , so the cross-correlation value will be much larger. Table 3 shows this detection criterion: that the correlation value $\beta(\alpha + \varepsilon_0)$ is larger than ε_1 . ε_0 and ε_1 are very small correlation between the host data and watermark patterns and β is a value close to 1.

A third approach is to use the separated host data in y_{01} as one input to a further ICA de-mixing process which also receives the watermarked host data. This situation resembles the first embodiment, so that the non-blind detection described above in relation to the first embodiment can be applied. This third approach may be used as an independent way of detecting the watermark patterns, or alternatively it may be used in combination with one of the first and second approaches to verify the watermark pattern detected by that other approach.

As in the case of the first embodiment, the second embodiment can be generalised to the case that an higher-order dependence is used in the watermark detection step than correlation values. In this case, table 3 is generalised to:

		Watermark Pattern ω	
De-mixing Matrix	Outputs	ω_0^1	ω_1^1
W_0	$y_{01} = x$	ϵ_0	ϵ_1
	$y_{02} = \omega_0^1$	$\Delta(\omega_0^1)$	0
W_1	$y_{11} = x'' = \beta \times (x + \alpha \times \omega_0^1)$	$\beta(\alpha\Delta(\omega_0^1) + \epsilon_0)$	$\beta\epsilon_1$
	$y_{12} = \omega_1^1$	0	$\Delta(\omega_1^1)$

Table 4: Generalised form of Table 3 for an alternative dependence measure.

(e) Generation of the second type of watermarked data

In contrast to the two watermark patterns generated in section (a) above which are used in section (b) to modulate one bit of the watermark data, the

second type of watermarked data is produced by encoding each bit of the watermark data (0 or 1) using a single random N-string ω with zero mean and unit variance. (Note that in this case, the ICA pre-processing is not performed.)

5

This N-string ω is linearly added to a respective host sequence with a respective sign (+ or –) according to the value of the bit of the watermark data. (i.e. according to Eqn. (7) but with the + replaced with a – according to the sign of the value of the watermark bit). As in section (b), the resulting

10 watermarked data can be normalised. The normalised watermarked data is represented, as before, by $\mathbf{x}'' = \beta \times \mathbf{x}'$, which is called here the "second type of watermarked data".

(f) *Third embodiment: non-blind detection of the second type of*

15 *watermarked data*

This embodiment is performed exactly in the same way as the first embodiment. Table 5 shows the cross-correlation values between the ICA outputs and the watermark patterns: the auto-correlation value of watermark

20 patterns is equal to 1, cross-correlation values between the host data and the watermark patterns are very small, close to zero, denoted by ϵ .

Output y	Watermark Pattern ω
ω	1
$-\omega$	-1
$\mathbf{x}$	ϵ

Table 5

Thus, detection of the watermark bit can be performed based on the sign of

25 cross-correlation value given in table 5.

As in the case of the first and second embodiments, the third embodiment can be straightforwardly generalised to the use of higher-order dependence values than correlation values.

5

(g) Fourth embodiment: Blind detection of the second type of watermarked data

In this embodiment we form the input matrix $\mathbf{X}$ to an ICA de-mixing process as the watermarked host data together with the watermark pattern, that is

10

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'' \\ \omega \end{bmatrix} = \begin{bmatrix} \beta(\mathbf{x} \pm \alpha\omega) \\ \omega \end{bmatrix}. \quad (14)$$

The ICA de-mixing process should ideally generate two outputs $\mathbf{y}_1 = \mathbf{x}$ and $\mathbf{y}_2 = \omega$ by a linear de-mixing matrix:

$$\mathbf{W} = \begin{bmatrix} \frac{1}{\beta} & \mp \alpha \\ 0 & 1 \end{bmatrix} \quad (15)$$

15 The sign in the de-mixing matrix is the opposite of the embedded watermark bit, since the separation is a kind of signal subtraction. The value of the embedded bit can now be found using analogues of the first and second approaches explained above in relation to the second embodiment.

20

Alternatively, in a way analogous to the third approach in the second embodiment, we can use the host data extracted by this first ICA de-mixing process as one input for a second ICA de-mixing process which also receives the watermarked data. Thus, a non-blind detection corresponding to the third
25 embodiment can be performed. This may be used to verify the detection result.

As in the case of the first, second and third embodiments, the third embodiment can be straightforwardly generalised to the use of higher-order dependence values than correlation values.

- 5 Although the invention has been described above in relation only to four embodiments, many variations are possible within the scope of the invention as will be clear to a skilled reader.

For example, as mentioned above, although in these embodiments the bits of
10 the watermark data are encoded separately, in other embodiments multiple bits can be encoded together (e.g. 2 bits of the watermark data can be encoded by using their value to select one of four watermark patterns, and embedding this selected watermark pattern into a host sequence). In this case, the detection analysis can be varied in straightforward ways. For
15 example, the first embodiment can be varied by testing which of the four watermark patterns is correlated with the output of the ICA de-mixing process which is not the host data. The second embodiment can be varied by performing a respective ICA de-mixing process for each of the watermark patterns together with the watermarked host sequence.

20

Note that in each of the embodiments of the invention discussed above, the length of the watermark patterns has not been specified. Because the independence between signals is a statistical property, the length of the processed data has to be large enough to keep the mutual independence. As
25 noted above, the dependence between two watermark patterns can be completely removed by the ICA pre-processing. The dependence between the host data and the watermark pattern is denoted as ϵ in our previous discussion, and it is usually a very small value compared to the auto-correlation value when the value of N is high enough (typically at least 1000,
30 and more usually about 5000).

References

- [1] I. J. Cox, J. Kilian, T. Leighton and T. Shamon, "Secure Spread Spectrum Watermarking for Multimedia", IEEE Transactions on Image Processing, vol. 6, no. 12, pp.1673-1687, Dec, 1997.
- [2] Geoffrey B. Rhoads, Clay Davidson and Anthony Rodriguez, "Method and Apparatus for Watermarking Video Images", US Patent 6,229,924 B1, May 8, 2001.
- [3] Wenjun Zeng, "Method for Extracting Multiresolution Watermark Images to Determine Rightful Ownership", US Patent 6,373,974 B2, April 16 2002.
- [4] B. Chen and G. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding", IEEE Transactions on Information Theory, vol. 47, no. 4, pp. 1423-1443, May 2001.
- [5] S. Noel and H. Szu, "Multimedia authenticity with independent-component watermarks," in 14th Annual International Symposium on Aerospace/Defense Sensing Simulation, and Controls, Orlando, Florida, April 2000.
- [6] D. Yu, F. Sattar and K.-K. Ma, "Watermark Detection and Extraction using Independent Component Analysis Method," EURASIP Journal on Applied Signal Processing, vol. 2002, no. 1, pp. 92-104, 2002.
- [7] M.-F. Shen, X.-J. Zhang, L.-S. Sun, P. J. Beadle, and F. H. Y. Chan, "A Method for Digital Image Watermarking Using ICA," 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, pp. 209-214, April 2003.
- [8] P. Comon, "Independent component analysis – a new concept?," Signal Processing, vol. 36, no. 3, pp. 287-314, 1994.
- [9] A. Hyvarinen, "Fast and Robust Fixed-point Analysis for Independent Component Analysis", IEEE Transactions on Neural Networks, 10(3):626-634, May 1999.

Claims

1. A method of regenerating watermark data from watermarked host data generated from host data and the watermark data, the watermarked host data having been generated by using portions of the watermark data associated
5 with respective sequences of the host data to select a respective watermark pattern from a set of possible watermark patterns, each of the possible watermark patterns having a plurality of elements, and then adding the elements of the selected watermark pattern to respective elements of the associated sequence of the host data,
10 the method including,

for each sequence of the watermarked host data:

(a) a de-mixing process step comprising at least one de-mixing process for obtaining statistically independent components from a mixture of the components and which takes as inputs (i) the sequence of watermarked
15 host data and (ii) the corresponding sequence of host data and/or at least one of the possible watermark patterns, and

(b) recognising the watermark pattern embedded in the sequence of the watermarked host data from the results of this de-mixing process; and

regenerating the watermark data from the recognised watermark
20 patterns.
2. A method according to claim 1 in which the de-mixing process is an ICA de-mixing process.
3. A method according to claim 1 or claim 2 in which the portions of the watermark data are each respective bits of the watermark data, whereby each

watermark pattern is selected according to the value of a corresponding bit of the watermark data.

4. A method according to claim 1 or claim 2 in which the portions of the watermark data are groups of 2 or more bits of the watermark data, whereby
5 the watermarked host data is selected by selecting each watermark pattern according to the values of the bits of a respective group of bits of the watermark data.

5. A method according to any preceding claim in which, during the generation of the watermark data, for each portion of the watermark data the
10 watermark pattern is selected from a different respective set of possible watermark patterns.

6. A method according to any of preceding claim in which step (a) includes a de-mixing process which takes as inputs (i) the sequence of watermarked host data and (ii) the corresponding sequence of the host data.

15 7. A method according to any of claims 1 to 5 in which step (a) includes a first de-mixing process for at least one the possible watermark patterns, taking as inputs (i) the sequence of watermarked host data and (ii) that possible watermark pattern.

8. A method according to claim 7 in which step (a) includes a respective
20 first de-mixing process for each of the possible watermark patterns, taking as inputs (i) the sequence of watermarked host data and (ii) the corresponding one of the possible watermark patterns.

9. A method according to claim 7 or claim 8 in which a first respective de-mixing process is used to obtain an approximation to the sequence of host
25 data, and step (a) further comprises at least one further de-mixing process taking as inputs (i) the sequence of watermarked host data, and (ii) the approximation to the sequence of host data.

10. A method according to any preceding claim in which step (b) is performed by observing whether one of the independent components output from a said de-mixing process is statistically dependent on one of the possible watermark patterns.
- 5 11. A method according to claim 10 in which step (b) is performed by observing whether one of the independent components output from a said de-mixing process is correlated with one of the possible watermark patterns.
12. A method according to any of claims 1 to 9 in which step (b) is performed by evaluation of the de-mixing matrix which a said de-mixing
10 process produces.
13. A method according to claim 12 in which the de-mixing matrix is evaluated by a process including calculating the ratio of its highest to lowest singular values, and evaluating whether this ratio is above or below a threshold.
- 15 14. A method according to any preceding claim in which for any given host sequence the possible watermark patterns have a known correlation with each other.
15. A method according to any preceding claim in which for any given host sequence the possible watermark patterns are statistically independent.
- 20 16. A method according to claim 15 in which, during the generation of the watermarked host data, the watermark patterns are generated by performing a de-mixing process using two input strings to obtain the statistically independent watermark patterns.
17. A method according to claim 14 in which for each sequence there are
25 two possible watermark patterns which are opposite to each other.

18. A method according to any preceding claim in which the host data is data in the spatial domain.
19. A method according to any of claims 1 to 17 in which the host data is a selected one of DCT data, wavelet data, or ICA domain data.
- 5 20. A method of generating watermarked host data using host data and watermark data, the method comprising:
- (a) associating portions of the watermark data with respective sequences of the host data;
 - (b) for each sequence of the host data generating a set of statistically
10 independent watermark patterns;
 - (c) using each portion of the watermark data to select a respective watermark pattern from the corresponding set of possible watermark patterns, each of the possible watermark patterns having a plurality of elements; and
 - (d) linearly adding the elements of the selected watermark pattern to
15 respective elements of the associated sequence of the host data.
21. A method according to claim 20 in which step (b) includes generating the set of watermark patterns as the statistically independent outputs of a de-mixing process.
22. A watermark regeneration apparatus comprising a computer
20 programmed to perform a method according to any of claims 1 to 19.
23. A watermarking apparatus comprising a computer programmed to perform a method according to claim 20 or claim 21.

24. A computer program product comprising a recording medium storing software readable by a computer to program the computer to perform a method according to any of claims 1 to 19.

5 25. A computer program product comprising a recording medium storing software readable by a computer to program the computer to perform a method according to claim 20 or 21.