2009-01-01

Printed Document Authentication using Texture Coding

Jonathan Blackledge
Technological University Dublin, jonathan.blackledge@tudublin.ie

Khaled Mahmoud
Zarka Private University, Jordan

Printed Document Authentication using Texture Coding

J M Blackledge and K W Mahmoud

Abstract— **The use of image based information exchange has grown rapidly over the years in terms of both e-to-e image storage and transmission and in terms of maintaining paper documents in electronic form. Further, with the dramatic improvements in the quality of COTS (Commercial-Off-The-Shelf) printing and scanning devices, the ability to counterfeit electronic and printed documents has become a widespread problem. Consequently, there has been an increasing demand to develop digital watermarking techniques which can be applied to both electronic and printed images (and documents) that can be authenticated, prevent unauthorized copying of their content and, in the case of printed documents, withstand abuse and degradation before and during scanning. In this paper we consider the background to a novel approach to solving this problem that has been developed into practically realisable system.**

I. Introduction

In this paper, a new approach to digital watermarking is presented and a range of possible applications considered. The process is defined by using analytical techniques and concepts borrowed from Cryptography. It is based on computing a 'scramble image' by diffusing a 'watermark image' with a noise field (a cipher). For e-to-e applications, a cover image (covertext) can be introduced using a simple additive process ('confusion process'). The watermark is subsequently recovered by removing (subtracting) the covertext and then correlating the output with the original (key dependent) cipher. This approach provides the user with a method of hiding image-based information in a host image before transmission of the data. In this sense, the method provides a steganographic approach to transmitting encrypted information that is not apparent during an intercept. Decryption is based on knowledge of the key(s) and access to the host image.

With regard to digital image analysis and e-to-e communications, the method provides a way of embedding information in an image that can be used for authentication from an identifiable source, a method that is relatively insensitive to lossy compression, making it well suited to digital image transmission. However, with regard to document authentication, the use of diffusion and confusion using a covertext is not robust. The reason for this is that the registration of pixels associated with a covertext can not be assured when the composite image is printed and scanned. We therefore consider a diffusion only approach to document authentication which is robust to a wide variety of attacks including print/scan attacks, geometric, soiling and crumpling attacks. This is because the process of diffusion (i.e. the convolution of information) is compatible with the physical principles of an imaging system and the theory of image formation and thus, with image capture devices (digital cameras and scanners, for example) that, by default, conform to the 'physics' of optical image formation.

The diffusion of plaintext (in this case, an image) with a noise field (the cipher) has a synergy with the encryption of plaintext using a cipher and an XOR operation (when both the plaintext and cipher are represented by binary streams). However, decryption of a convolved image (deconvolution) is not as simple as XORing the ciphertext with the appropriate cipher. Here, we consider an approach which is based on preconditioning the original cipher in such a way that decryption (de-diffusion) can be undertaken by correlating the ciphertext with the cipher. The output ciphers generated for printed document authentication are textures of a type that are determined by the spectral characteristics of the plaintext which can be applied using low resolution COTS printers and scanners. In this sense, the approach is based on 'texture coding'. In this paper, we present a method of texture coding that has been developed into a practically vaibale system and present a range of example applications to which it has been applied. Examples of the robustnes of the system to various 'attacks' is provide in an extended Appendix.

II. Transform Domain Watermarking Methods

Like many aspects of digital signal and image processing, watermarking schemes fall into two categories: spatial domain and transform domain techniques [1], [2], [3], [4], [5]. This depends on whether the watermark is encoded by directly modifying pixels (such as simply flipping low-order bits of selected pixels) or by altering some frequency coefficients obtained by transforming the image in the frequency domain. Spatial domain techniques are simple to implement and usually require a lower computational cost. However, such methods tend to be less robust to tampering than methods that place the watermark in the transform domain [6], [14], [15].

Watermarking schemes that operate in a transform space are increasingly common, as they posses a number of desirable features. These include the following: (i) By transforming spatial data into another domain, statistical independence between pixels and high-energy compaction is obtained; (ii) the watermark is irregularly distributed over the entire spatial image upon an inverse transformation, which makes it more

Jonathan Blackledge (jonathan.blackledge@dit.ie) is the Stokes Professor of Digital Signal Processing, School of Electrical Engineering Systems, Faculty of Engineering, Dublin Institute of Technology (http://eleceng.dit.ie/blackledge).

Dr Khaled Mahmoud (k.w.mahmoud@zpu.edu.jo) is Head of the Department of Software Engineering, Zarka Private University, Jordan.

difficult for attackers to extract and/or decode a watermark; (iii) it is possible to provide markers according to the perceptual significance of different transform domain components so that a watermark can be placed adaptively in an image where it is least noticeable, such as within a textured area [16], [17], [18], [19]. In addition, transform domain methods can hide information in significant areas of a covertext which makes them more robust to attacks and distortion while remaining visually imperceptible [20], [38], [22]. Cropping, for example, may seriously distort any spatially based watermark but is less likely to affect a transform-based scheme because watermarks applied in a transform domain are dispersed over the entire spatial domain so that upon inverse transformation, at least part of the watermark may be recovered. Lossy compression is an operation that usually eliminates perceptually unimportant components of an image and most processing of this type takes place in a transform domain. Thus, matching the transform with a compression transform can result in an improved performance (i.e. DCT for JPEG, Wavelet for JPEG-2000). Further, the characteristics of the Human Visual System (HVS) can be fully exploited in a transform domain, e.g. [23], [24].

With transform domain watermarking, the original host data is transformed to produce a matrix of 'coefficients'. These coefficients are then perturbed by a small amount in one of several possible ways in order to represent the watermark. Coefficient selection is based on 'perceptual significance' and/or 'energy significance'. When the watermarked image is compressed or modified by any image processing operation, noise is added to the already perturbed coefficients. Private retrieval operations involve subtracting the coefficients associated with the watermarked image from those of the original image to obtain the noise perturbation. The watermark is then estimated from the noisy data as best as possible. The most difficult problem associated with 'blind-mode' watermark detection (in which the host image is not available) in the frequency domain is to identify the coefficients used for watermarking. Embedding can be undertaken using quantization (thresholding) or image fusion, for example, but in either case, most algorithms consider the HVS to minimize perceptibility. The aim is to place as much information in the watermark as possible such that it is most robust to an attack but least noticeable. Most schemes operate directly on the components of some transform of the 'cover image' such as the Discrete Cosine Transform (DCT), Discrete Wavelet Transforms (DWT) and the Discrete Fourier Transform (DFT) [25], [26] [27].

In general, the HVS is not sensitive to small changes in edges and texture but very sensitive to small changes in the smoothness of an image [28], [29]. In 'flat' featureless portions of the image, information is associated with the lowest frequency components of the image spectrum, while, in a highly textured image, the information is concentrated in the high frequency components. The HVS is more sensitive to lower frequency than high frequency (visual) information. [1], [7], [8], [9]. Taking this into account, the following points are relevant to digital image watermarking in the frequency domain: (i) A watermark should ideally be embedded in the higher frequency range of an image in order to achieve better perceptual invisibility but only on the understanding that high frequency components can be distorted or deleted after attacks such as lossy compression, re-sampling or scanning, for example; (ii) in order to prevent the watermark from being attacked, it is often necessary to embed it into the lower frequency region of the spectrum, which can not be attacked without compromising the image given that the HVS is more sensitive in this region; (iii) given points (i) and (ii), in order to embed a watermark in an image optimally (i.e. so that it can survive most attacks), a reasonable trade-off is to embed a watermark into the intermediate frequency range of the image [10], [11], [12], [13].

III. Diffusion and Confusion based Watermarking

We consider an approch to watermarking plaintext using both diffusion and confusion. The basic method is as follows: Given a *plaintext* image and a *covertext* image, the *stegotext* image is given by

$$stegotext = ciphertext + covertext$$

where

$$ciphertext = cipher \otimes\otimes plaintext$$

and $\otimes\otimes$ denotes the two-dimensional convolution operation. The problem is to find a cipher which provides a ciphertext that, given the equation above, can be well hidden in the covertext.

A. Fresnel Diffusion Watermarking

Consider a watermarking model given by

$$I_3(x,y) = rp(x,y) \otimes\otimes I_1(x,y) + I_2(x,y)$$

with 'Fresnel' Point Spread Function (PSF)

$$p(x,y) = \frac{1}{2}(1 + \cos[\alpha(x^2+y^2)]$$

and where

$$\|p(x,y) \otimes\otimes I_1(x,y)\|_\infty = 1 \quad \text{and} \quad \|I_2(x,y)\|_\infty = 1.$$

Here, r controls the extent to which the host image I_2 dominates the diffused watermark image I_1. In effect, r is like a Signal-to-Noise Ratio, or, in this application a 'Diffusion-to-Confusion' Ratio The output of this process I_3 is the watermarked host image. Recovery of the watermark image is then based on the following process:

$$I_1(x,y) = \frac{1}{r}p(x,y) \odot\odot (I_3(x,y) - I_2(x,y)$$

where $\odot\odot$ denote two-dimensional correlation. The method is implemented numerically using a Fast Fourier Transform and application of the two-dimensional convolution and correlation theorems, i.e.

$$p \otimes\otimes f \Longleftrightarrow PF$$

and

$$p \odot\odot f \Longleftrightarrow P^*F$$

respectively, where $\Longleftrightarrow$ denotes transformation from 'image space' to 'Fourier space'.

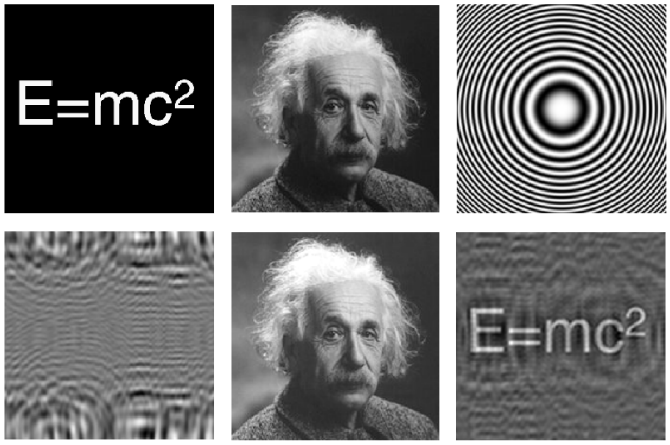


Fig. 1. From top to bottom and from left to right (all images are 512×512): Watermark I_1, host image I_2, PSF p for $\alpha = 0.001$, diffused image $p \otimes\otimes I_1$, host image after watermarking I_3 for $r = 0.1$, recovered watermark.

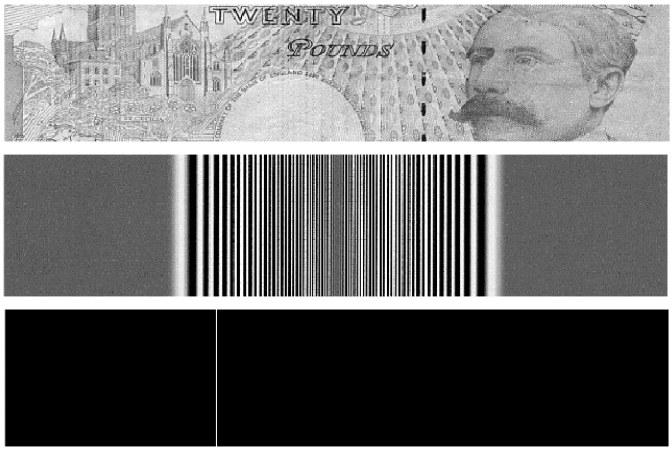


Fig. 2. 600dpi scan of a 20 Pounds Sterling Bank (of England) note (above) whose graphic file includes the addition of symmetric chirp (centre) and recovery of a digital thread.

Figure 1 shows an example result of implementing this watermaking method where a digital image of Albert Einstein (the covertext) is watermarked with a binary image of his most famous of equation (the plaintext). Note that the dynamic range of the diffused field and the reconstruction is relatively low and the images given in Figure 1 are displayed by re-quantisation to an 8-bit grey scale of the data $\min[I(x,y)] \leq I(x,y) \leq \max[I(x,y)]$. On the other hand, the low dynamic range of the diffused field allows the diffused field to be added to the host image without any significant distortions becoming discernable other than increasing the brightness of the image[1].

Fresnel diffusion is only of value when the plaintext is of binary form (i.e. a binary image) and when the covertext is well textured throughout in order to 'hide' the diffused plaintext.

In this application, the host image together with (α, r) form the key where the algorithm is taken to be in the public domain. Given these conditions, the method is useful for application to watermarking digital images provided that the distortion accompanying the restoration of the watermark is acceptable. However, the method is not particularly well suited to document (hard-copy) watermarking accept under special circumstances. One such example is given in Figure 2 which illustrates a method designed whereby, using standard security printing technology, a covert digital thread can be introduced (typically into a print file) that reflects a conventional overt thread. In this example, a one-dimensional Fresnel transform (a symmetric chirp function) is used to encode a single or multiple bar code and the result embedded into an existing print file. Recovery of the 'digital thread' is obtained through correlation of the same symmetric chirp function with a scanned image. This approach is analogous to the application of a matched filter and, in particular, the deconvolution on linear frequency modulated chirps. Applications include currency, bank bonds and other security documents. In this case the 'watermarking model' is based on the following:

$$I_3(x,y) = rp(x,y) \otimes\otimes I_1(x,y) + I_2(x,y)$$

where I_1 is a binary image consisting of a bar code (single or multiple bars),

$$p(x,y) = \frac{1}{2}[1 + \cos(\alpha x^2)]$$

and I_2 is the host image. Recovery of the bar code (i.e. estimation $\hat{I}_1$ of I_1 is then given by

$$\hat{I}_1(x,y) \sim p(x,y) \odot \odot I_3(x,y) + \epsilon$$

where

$$\epsilon = p(x,y) \odot \odot I_2(x,y)$$

such that, provided I_2 does not correlate with p (e.g. I_2 is a textured image), then

$$\|\epsilon(x,y)\| << \|p(x,y) \odot \odot I_3(x,y)\|$$

This condition allows an exact reconstruction of I_1 to be obtained through application of a threshold to binarize $\hat{I}_1$.

B. Noise Diffusion Watermarking

The principal weakness associated with Fresnel diffusion is that the cipher is based on a deterministic function. To overcome this, we consider a noise diffusion approach. Diffusion by noise is compounded in the model

$$I(x,y) = n(x,y) \otimes\otimes I_0(x,y)$$

where n is some stochastic two-dimensional field and I_0 is an input image. There are two approaches to solving the problem: Given I and n, obtain I_0. We can invert or deconvolve by using the convolution theorem giving

$$I_0(x,y) = \mathcal{F}_2^{-1}\left[\frac{\widetilde{I}(k_x,k_y)N^*(k_x,k_y)}{\mid N(k_x,k_y) \mid^2}\right]$$

where $\mathcal{F}_2^{-1}$ is the (two-diemsnional) inverse Fourier transform operator, N is the Fourier transform of n and $\widetilde{I}$ is the Fourier

[1]In each case the data is re-normalised to floating point values between 0 and 1 before application of grey-scale quantisation.

transform of I (with spatial frequencies k_x and k_y). However, this approach requires regularisation in order to eliminate any singularities when $\mid N \mid^2 = 0$ through application of a constrained deconvolution methods. Alternatively, if n is the result of some random number generating algorithm, and since the functional form of n is arbitrary, we can construct the stochastic field

$$m(x,y) = \mathcal{F}_2^{-1}\left[\frac{N(k_x,k_y)}{\mid N(k_x,k_y)\mid^2}\right]$$

where $\mid N(k_x,k_y)\mid^2 > 0$, the diffused field now being given by

$$I(x,y) = m(x,y) \otimes\otimes I_0(x,y).$$

The inverse problem is then solved by correlating I with n, since

$$n(x,y) \odot\odot I(x,y) \Longleftrightarrow N^*(k_x,k_y)\widetilde{I}(k_x,k_y)$$

and

$$N^*(k_x,k_y)\widetilde{I}(k_x,k_y) = N^*(k_x,k_y)M(k_x,k_y)\widetilde{I}_0(k_x,k_y)$$
$$= N^*(k_x,k_y)\frac{N(k_x,k_y)}{\mid N(k_x,k_y)\mid^2}\widetilde{I}_0(k_x,k_y) = \widetilde{I}_0(k_x,k_y)$$

so that

$$I_0(x,y) = n(x,y) \odot\odot I(x,y).$$

The condition that $\mid N(k_x,k_y)\mid^2 > 0$ is simply achieved by implementing the following process:

$\forall k_x, k_y$

if $\quad \mid N(k_x,k_y)\mid^2 = 0$

then $\mid N(k_x,k_y)\mid^2 = 1$

This result can be used to 'embed' one image in another. Consider the case when we have two independent images $I_1(x,y) \geq 0 \forall x,y$ and $I_2(x,y) \geq 0 \forall x,y$ and we consider the case of embedding I_1 with I_2. We construct a stochastic field $m(x,y) \geq 0 \forall x,y$ *a priori* and consider the equation

$$I_3(x,y) = rm(x,y) \otimes\otimes I_1(x,y) + I_2(x,y)$$

where

$$\|m(x,y) \otimes\otimes I_1(x,y)\|_\infty = 1 \quad \text{and} \quad \|I_2(x,y)\|_\infty = 1.$$

By normalising the terms in this way, the coefficient $0 \leq r \leq 1$, can be used to adjust the relative magnitudes of the terms such that the diffused image I_1 is a perturbation of the 'host image' I_2. This provides us with a way of watermarking one image with another, r being Diffusion-to-Confusion of 'Watermarking' Ratio. This approach is of course identical to the Fresnel diffusion method considered earlier but where the Fresnel PSF is replaced with the pre-conditioned stochastic field m. However, for applications in image watermarking, the diffusion of an image with noise provides a superior result because: (i) a stochastic field provides more uniform diffusion; (ii) noise fields can be generated using random number generators that depend on a single initial value or seed (i.e. a private key).

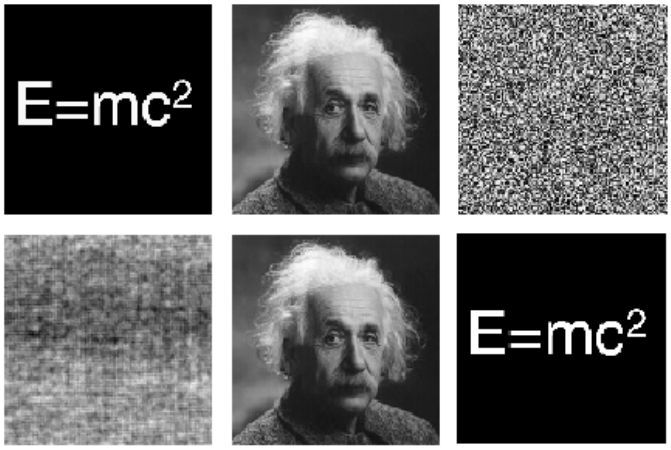


Fig. 3. From top to bottom and from left to right (all images are 512×512): Watermark I_1, host image I_2, PSF m, diffused image $m \otimes\otimes I_1$, host image after watermarking I_3 for $r = 0.01$, recovered watermark.

The noise field can be created using conventional Pseudo Random Number Generators or chaos based iteration schemes [30] for which a PIN (Personal Identity Number) or a password are required. Alternatively, noise fields can be created indirectly through application of various commercial (key dependent) cryptosystems (by encrypting a file consisting of a single plaintext character, for example) or by using genuine random number generating systems such as HotBits [31] to create n (which must then be encrypted).

An example of this approach is shown in Figure 3 which should be compared with Figure 1. Here, the image I_2 (the 'host image' - covertext) is watermarked with another image I_1 (the 'watermark image' - plaintext) to produce output image I_3 with $r = 0.01$. The relatively small perturbation of the term $m \otimes\otimes I_1$ to the host image I_2 for $r = 0.01$ does not affect the output image in any way that is visually significant.

A further advantage of noise diffusion is that it is not limited to watermarking covertexts with binary image plaintexts. However, the effect of adding the diffused greyscale or colour watermark image to the host image yields a different, slightly brighter image because of the perturbation of I_2 by $rm \otimes\otimes I_1$. This effect can be minimized by introducing a smaller Watermarking Ratio such that the perturbation is still recoverable by subtracting the host image from the watermarked image, an example being given in Figure 4.

IV. Hardcopy Steganography

The model

$$stegotext = ciphertext + covertext$$

can be applied for watermarking digital images associated with electronic-to-electronic type communications in which there is no or minimal loss of information. This method can be used to watermark digital images for the purpose of authentication but can also be viewed as a method of covertly transmitting ciphertext when the plaintext is converted to the form of a

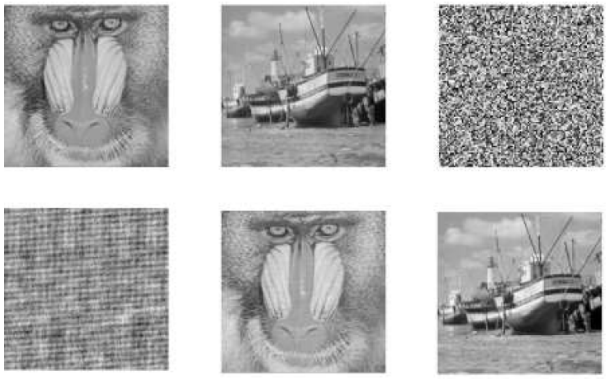


Fig. 4. Example of watermarking an image with another image using noise based diffusion. The 'host image' I_2 (top-left) is watermarked with the 'watermark image' I_1 (top-centre) using the diffuser (top-right) given by a uniform noise field n whose pixel-by-pixel values depend upon the seed used (the private key). The result of computing $m \otimes\otimes I_1$ (bottom-left) is added to the host image for $r = 0.1$ to generate the watermarked image u (bottom-centre). Recovery of the watermark image I_1 (bottom-right) is accomplished by subtracting the host image from the watermarked image and correlating the result with the noise field n.

digital image. Steganography and watermarking techniques are also of value for hardcopy 'data' which has a range of applications for authenticating printed material and copyright validation, for example. However, to be of practical value to the security printing industry the methods must be robust to the significant distortions generated by the printing and/or scanning process. A simple approach is to add information to a printed page that is difficult to see. For example, some modern colour laser printers, including those manufactured by HP and Xerox, print tiny yellow dots which are added to each page. The dots are barely visible and contain encoded printer serial numbers, date and time stamps. This facility provides a useful forensics tool for tracking the origins of a printed document which has only relatively recently been disclosed.

A. Diffusion Only Watermarking

If a stegotext image is printed and scanned back into electronic form, then the print/scan process will yield an array of pixels that will be significantly different from the original electronic image even though it might 'look' the same. These differences can include the size of the image, its orientation, brightness, contrast and so on. Of all the processes involved in the recovery of the watermark, the subtraction of the host image from the watermarked image is critical. If this process is not accurate on a pixel-by-pixel basis and deregistered for any of many reasons, then recovery of the watermark by correlation will not be effective. However, if we make use of the diffusion process alone, then the watermark can be recovered via a print/scan because of the compatibility of the processes involved. However, in this case, the 'watermark' is not covert but overt.

Depending on the printing process applied, a number of distortions will occur which diffuse the information being printed. Thus, in general, we can consider the printing process to introduce an effect that can be represented by the convolution equation

$$I_{\rm print} = p_{\rm print} \otimes\otimes I.$$

where I is the original electronic form of a diffused image (i.e. $I = m \otimes\otimes I_0$) and $p_{\rm print}$ is the point spread function of the printer. An incoherent image of the data, obtained using a flat bed scanner, for example (or any other incoherent optical imaging system), will also have a characteristic point spread function $p_{\rm scan}$ say. Thus, we can consider a scanned image to be given by

$$I_{\rm scan} = p_{\rm scan} \otimes\otimes I_{\rm print}$$

where $I_{\rm scan}$ is taken to be the digital image obtained from the scan. Now, because convolution is commutative, we can write

$$I_{\rm scan} = p_{\rm scan} \otimes\otimes p_{\rm print} \otimes\otimes p \otimes\otimes I_0 = p \otimes\otimes p_{\rm scan/print} \otimes\otimes I_0$$

where

$$p_{\rm scan/print} = p_{\rm scan} \otimes\otimes p_{\rm print}$$

which is the print/scan point spread function associated with the processing cycle of printing the image and then scanning it back into electronic form. By applying the method discussed earlier, we can obtain a reconstruction of the watermark whose fidelity is determined by the scan/print PSF. However, in practice, the scanned image needs to be re-sized to that of the original. This is due to the scaling relationship (for a function f with Fourier transform F)

$$f(\alpha x, \beta y) \Longleftrightarrow \frac{1}{\alpha\beta} F\left(\frac{k_x}{\alpha}, \frac{k_y}{\beta}\right).$$

The size of any image captured by a scanner or other device will depend on the resolution used. The size of the image obtained will inevitably be different from the original because of the resolution and window size used to print the diffused image I and the resolution used to scan the image. Since scaling in the spatial domain causes inverse scaling in the Fourier domain, the scaling effect must be 'inverted' before the watermark can be recovered by correlation since correlation is not a scale invariant process. Re-sizing the image (using an appropriate interpolation scheme such as the bi-cubic method, for example) requires a set of two numbers a and b (i.e. the $a \times b$ array used to generate the noise field and execute the diffusion process) that, along with the seed required to regenerate the noise field n, provides the 'private keys' needed to recover the data from the diffused image. An example of this approach is given in Figure 5 which shows the result of reconstructing four different images (a photograph, fingerprint, signature and text) used in the design of an impersonalized bank card. The use of 'diffusion only' watermarking for print security can be undertaken in colour by applying exactly the same diffusion/reconstruction methods to the red, green and blue components independently. This provides two additional advantages: (i) the effect of using colour tends to yield better quality reconstructions because of the colour combination process; (ii) for each colour component, it is possible to apply a noise field with a different seed. In this case, three keys are required to recover the watermark although it should be noted that, due to the errors associated in the extraction of each colour component from a colour scan, this

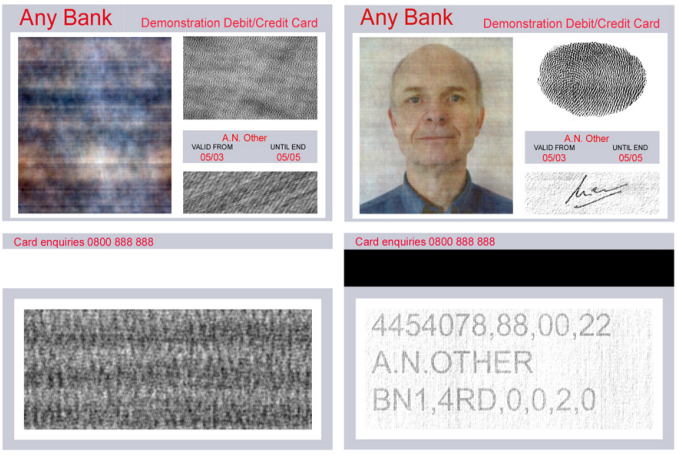


Fig. 5. Example of the application of 'diffusion only' watermarking. In this example, four images of a face, finger-print, signature and text have been diffused using the same cipher and printed on the front (top-left) and back (bottom-left) of an impersonalized identity card using a 600 dpi printer. The reconstructions (top-right and bottom-right, respectively) are obtained using a conventional flat-bed scanner based on a 300 dpi grey-level scan.

Fig. 6. Example of a security card designed to include a texture code of the holders portrate (top). The images (i.e. portrate and texture code) have been printed onto the identity card at 600dpi. An image of this card (bottom-left) has been generated using a mobile phone. After cropping the texture code obtained from this low resolution data, a reconstruction can still be obtained as shown (bottom-right).

approach does not yield reconstructions with the same degree of robustness as in the case when the same key/algorithm is used for each colour component.

Because this method is based on convolution alone and since

$$I_{\text{scan}} = p_{\text{scan/print}} \otimes \otimes I_0$$

as discussed earlier, the recovery of the I_0 will not be negated by the distortion of the PSF associated with the print/scan process, just limited or otherwise by its characteristics. Thus, if an image is obtained of the printed data field $p \otimes \otimes I_0$ which is out of focus due to the characteristics of $p_{\text{scan/print}}$, then the reconstruction of I_0 will be out of focus to the same degree. Decryption of images with this characteristic is only possible using an encryption scheme that is based a diffusion only approach. However, if a covertext image I_{c} is introduced so that

$$I_{\text{scan}} = p_{\text{scan/print}} \otimes \otimes I_0 + p_{\text{scan/print}} \otimes \otimes I_{\text{c}}$$

then because

$$I_{\text{scan}} - I_{\text{c}} \neq p_{\text{scan/print}} \otimes \otimes I_0$$

recovery of the plaintext is not possible which is why we resort to a diffusion only method approach.

Figure 6 illustrates the recovery of a diffused image printed onto a personal identity card obtained captured using a mobile phone camera. In the latter case, the reconstruction is not in focus because of the wide-field nature of the lens used. However, the fact that recovery of the watermark is possible with a mobile phone means that the scrambled data can be transmitted securely and the card holders image (as in this example) recovered remotely and transmitted back to the same phone for authentication. This provides the necessary physical security needed to implement such a scheme in practice and means that specialist image capture devices are not required on site. Applications of this technique to a mobile security spot check environment are clearly possible.

The diffusion process can be carried out using a variety of different noise fields other than the uniform noise field considered here. Changing the noise field can be of value in two respects: first, it allows a system to be designed that, in addition to specific keys, is based on specific algorithms which must be known *a priori*. These algorithms can be based on different pseudo uniform random number generators and/or different pseudo chaotic number generators that are post-processed to provide a uniform distribution of numbers. Second, the diffusion field depends on both the characteristics of the watermark image and the noise field. By utilizing different noise fields (e.g. Gaussian noise, Poisson noise), the texture of the output field can be changed. The use of different noise fields is of value when different textures are required that are aesthetically pleasing and can be used to create a background that is printed over the entire document - texture maps. In this sense, variable noise based diffusion fields can be used to replace complex print security features with the added advantage that, by de-diffusing them, information can be recovered. Further, these fields are very robust to data degradation created by soiling, for example. In the case of binary watermark images, data redundancy allows reconstructions to be generated from a binary output, i.e. after binarizing the diffusion field (with a threshold of 50% for example). This allows the output to be transmitted in a form that can tolerate low resolution and low contrast copying, e.g. a fax.

The tolerance of this method to printing and scanning is excellent provided the output is cropped accurately (to within a few pixels) and oriented correctly. The processes of cropping and orientation can be enhanced and automated by providing a reference frame in which the diffused image is inserted. This is illustrated in Figure 7 which, in addition shows the effect of diffusing a combination of images. This has the

Fig. 7. Example of the diffusion of composite images with the inclusion of a reference frame for enhancing and automating the processes of copping and orientation. In each case the data fields have been printed and scanned at 300 dpi.

effect of producing a diffused field that is very similar but nevertheless conveys entirely different information. Details of the robustness of the method to various 'attacks' are provided in the Appendix.

B. Covertext Addition and Removal

Because diffusion only watermarking is based on convolution/correlation operations it is relatively insensitive to contrast stretching and compression. This provides the opportunity to introduce covertext in the form of the addition of foreground information (e.g. text) to a printed document that has been watermarked *a priori* with a grey scale (or colour) texture map whose brightness and contrast has been adjusted to be unobtrusive with regard to the covertext (i.e. the watermark is made bright compared to black text). Alternative, once the texture field has been designed, it may be introduced into a text editor that provide the inclusion of watermarks. For example, Microsoft Word has the facility to include a printed watermark (Format→Background→Printed Watermark) that provides the option to select a Picture Watermark (Existing Watermark) with options on scale and 'Washout'. In order to extract the watermark, it is then necessary to remove the text after a scan has been undertaken under the assumption that the covertext is not available. This can be accomplished using a median filter which is effective in removing isolated noise spikes, i.e. in this application, foreground text. However, in this case, the median filter is not applied to the image in its entirety. Instead, it is applied only to the neighbourhood of pixels (i.e. a user defined moving window) that exists below a user defined threshold that is specified in order to differentiate between the watermark and those pixels associated with the covertext. After removal of the covertext, the image watermark is reconstructed by correlation with the cipher.

V. Applications of Texture Coding

Some applications of diffusion only coding are already evident from the examples already given to introduce the technique in previous sections. Strictly speaking, the method is not a watermarking technique unless a covertext can be used to hide the ciphertext. However, as discussed previously, application of a covertext is not applicable for the authentication of printed documents due to the degradation of the covertext when printing and scanning a document. Hence, for application to low resolution print security, the method should be referred to as texture coding.

In this section, we consider a range of application to which the method can be applied.

A. Authentication

Authentication of a document image should ensure that the document has not been altered from the time it is created and signed by the author to the time it is received at the destination. Authentication of paper documents is an important concern as the ability of counterfeiters has increased substantially in recent years. This is contributed to by the dramatic improvement in the capability of high resolution scanners and printers. Moreover, digital documents can be accessed and modified by intruders relatively easily. This is especially true in the case of documents that are exchanged over the Internet.

Using the model and methods discussed here, a *selective authentication* approach can be applied in which only significant changes cause authentication to fail. This can be verified by embedding information in a document that can later be verified as to whether it has been tampered with or otherwise.

B. Photo Verification

Figure 5 and Figure 7 show examples of a photo verification application that can be incorporated into an ID card where a photograph of the card holder is texture coded and printed beside the original image. Substantial editing, such as changing the original photo, will be illegitimate because it will completely change the interpretation of the card. Thus, a photo verification system can be designed to do the following:

1) Capture the diffused watermark using any tool (scanner, camera, etc).
2) Read the key. The key might be:
   a) Encoded using a bar code, or
   b) stored in a local Database, or
   c) stored in a database that can be accessed via the Internet.
3) Extract the watermark.
4) Verify the authenticity by comparing the original photo with the extracted photo. This can be done either by:
   a) A subjective test using the judgment of a human (details on the scales that have been suggested for use in evaluation of watermarking quality being given in [32]).
   b) Quality metrics, such as the Mean Square Error or Chi-square test.
   c) Any other matching algorithm including the application of an Artificial Neural Network as required.

Such a system can be modified to include more information in the diffused watermark as required, such as the name of the ID card holder. Moreover, texture coding can be used to generate a de-personalised ID card either on an individual image basis (Figure 5) or in terms of a composite image (Figure 7).

C. Statistical Verification

When a document is prepared using MS Word, for example, or any other major word processing package, statistical information from the document can be gathered; information about the author, date and time, number of characters and spaces and so on. A verification system can use this information to check the authenticity of the document. Any attempt at modification of the file will be reflected in its statistics. The system can either incorporate these data in plaintext or as a diffused code into a patch on the document which is encoded into an indecipherable image. The image needs to be attractively packaged in an appropriate place on the document or can be incorporated as part of a background texture.

It is assumed that the recipient of the document (scanned or electronic) will have the appropriate software available. The encoded image is read into the decoding software and *text-recognition* used to reveal the text which is then compared with the plain-text statistics of the document. The data in the image can alternatively be checked manually against the statistics of the file instead of using text recognition. Each author can have a particular key for encoding the image. Upon receipt, the recipient applies that particular key to decode the image. Alternatively, a separate one-time PIN can be transmitted to the recipient in order to decode the image.

D. Original Copy Verification

When a document is scanned, subject to the scanner type and settings (including the resolution, for example), the output digital image file will have a specific statistical characteristic compounded in the histogram. If the document is copied and scanned again then this characteristic histogram will change because of the copy process. In general, a copied document will tend to have a smoother histogram since it is, in effect, the original document image convolved with a PSF that is characteristic of the copier (a function of the composite scan/print process). By printing a texture code of the histogram of the original document, typically on the back of the document, the document can be scanned and the histogram compared with the watermark, at least within an acceptable tolerance. This application has value in the authentication of high value documents such a Bank Bonds, an example of which is given in Figure 8. Figure 9 shows the texture map and the reconstruction of the plaintext, i.e. a histogram of the luminance of the original colour image together with some basic statistical information. By specifying the type of scanner and operational constraints, statistical information of this type (i.e. the mean, standard deviation and median, for example) can be used to qualify whether or not the Bank Bond or other high value documents has been copied. This statistical verification may include measures relating to the RGB components for the case of high value colour documents. Although each scan (using the same scanner with identical settings) will not output an identical digital image (due to slight differences in the crop, for example, as well as the natural 'jitter' of the scanner) the statistical information should not change significantly unless a copy has been made, acceptable tolerances having been established *a priori*.

Fig. 8. Example of a high value Bank Bond.

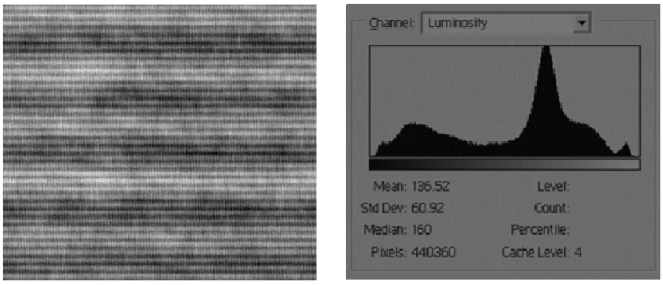


Fig. 9. Texture map (left) and reconstruction of statistical information relating to scan of original document using Adobe Photoshop V5.

E. Component Verification

The method discussed can be extended to include a 'specific parts' from of the text that must be correct, e.g. a sum of money, name of beneficiary etc. An example is shown in figure 10.

After decoding, the results will be as given as shown in Figure 11. Clearly, the diffused code could be placed into the background of each data field (i.e. instead of placing it in the next empty line).

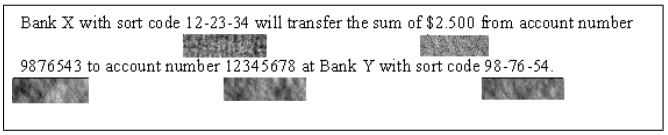


Fig. 10. Coded Specific Part from a Document

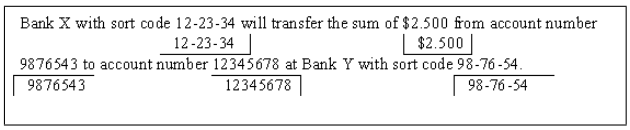


Fig. 11. Revealed Document after Coding Specific Parts.

F. Transaction Tracking

Also called *fingerprinting*, transaction tracking involves the embedding of a different watermark into each distributed copy. This is especially useful for identifying people who obtain a document legally but illegally redistribute.

G. Leaked Document Monitoring

One common method to monitor and discover any 'leak' associated with an important document is to use visible marks. For example, highly sensitive documents are sometimes printed on backgrounds containing large gray digits using a different number for each copy. Records are then kept about who has which copy. Of course, imperceptible watermarks (or at least diffused watermark) are preferable to visible marks. They are easy to remove/replace from a document when it is copied. Using this model for document watermarking, the tracking number is diffused and inserted into the background. The diffused watermark is inseparable from the document. The adversary (a person who attempts to remove, disable, or forge a watermark for the purpose of circumventing its original purpose) does not know the embedded number and can not recognize the difference between copies (it is difficult for human eyes to find a difference between two copies with different watermarks).

H. Owner Identification (Copyright)

Copyright can be undertaken by embedding the identity of a document's copyright holder as a watermark in order to prevent other parties from claiming the copyright of the document. The embedded data can be a biometric characteristic (such as signature). The receiver of the document reconstructs the signature used to watermark the document, which is then used to verify the authors claimed identity.

I. Signature Verification

Handwritten signatures are commonly used to certify the contents of a document or to authenticate legal transactions. A handwritten signature is a well-known biometric attribute. Other biometric attributes, which are commonly used for authentication include iris, hand geometry, face, and fingerprints (e.g. [33] and [34]). While attributes like the iris and fingerprints do not change over time, they require special and relatively expensive hardware to capture the biometric

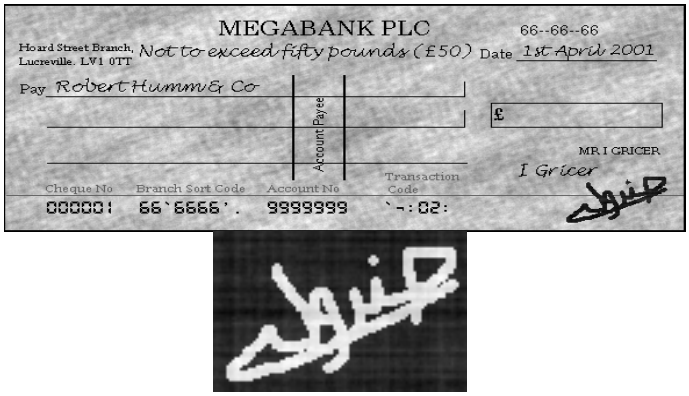


Fig. 12. Watermarked cheque (above) and recovered signature from watermark (below).

data. An important advantage of the signature over other biometric attributes is that it has been traditionally used in authenticating documents and hence is socially accepted. Signature verification is usually done by visual inspection. In automatic signature verification, a computer takes over the task of comparing two signatures to determine if the similarity between the signatures exceeds some pre-specified threshold. There a many similarity measures that can be used for this purpose. Figure 12 shows an example for this approach. The signature of the customer is diffused and inserted into the background of the cheque. Each customer has their own key that is known only to them and their bank. They use the key to generate the background and then print the cheque. The bank then uses the key to extract the customer signature from the cheque. If the extracted and the existing signatures on the cheque are matched (to within a given tolerance), then the cheque is accepted.

J. Binary Data Authentication using Binary Coded Images

The method can be used to encode binary information by applying a threshold to the stochastic field to produce a binary output image. Reconstruction of the information hidden in this binary image is obtained by correlating the scanned image with the original cipher. This provides a facility for printing 'binary texture codes' which have applications a in a range of printing processes that are 'two-tone'. For example, UV sensitive inks can be used to print a binary cipher that encodes the serial number of a bank note. An example of this application is given in Figure 13. The serial number given at the top right-hand corner of a (specimen) 20 Euro note is 'diffused' to produce the binary output shown (centre image). This information is printed onto the bank note with UV sensitive ink making the feature optically covert. A reconstruction of the serial number is then obtained after UV image capture and decryption.

VI. Case Study: Passport Authentication

Like any other security document, ID card and so on, a passport consists of a number of security features depending on the sophistication of the design associated with the authority responsible for an issue. These range from the use of printing complex background, micro-printing, conventional

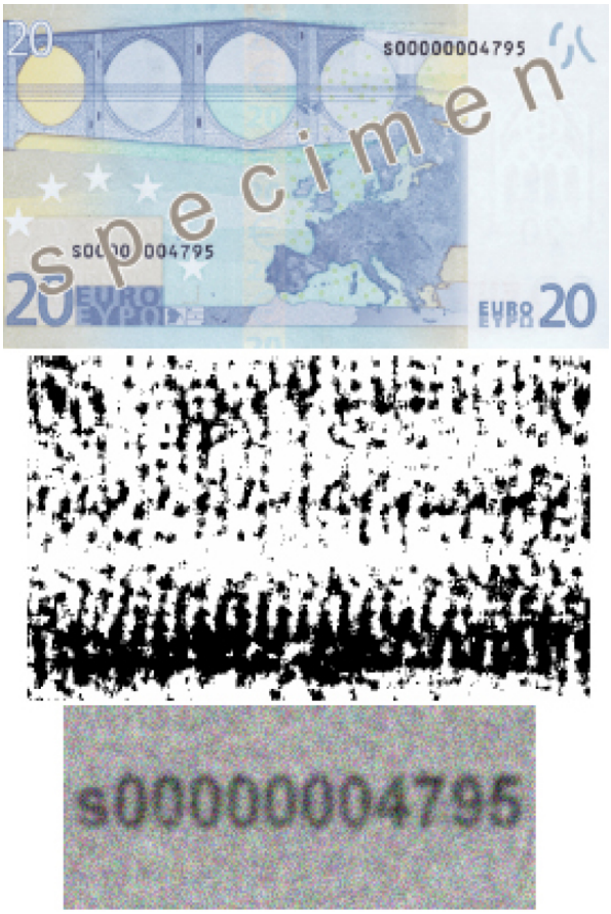


Fig. 13. Example of the application of 'binary texture coding' used to authenticate currency. Specimen Euro bank note (top); binary texture field after diffusing an image of the serial number and printing the result onto the bank note using UV sensitive ink (centre); reconstruction of serial number after capturing the binary texture field under a UV light source (below).

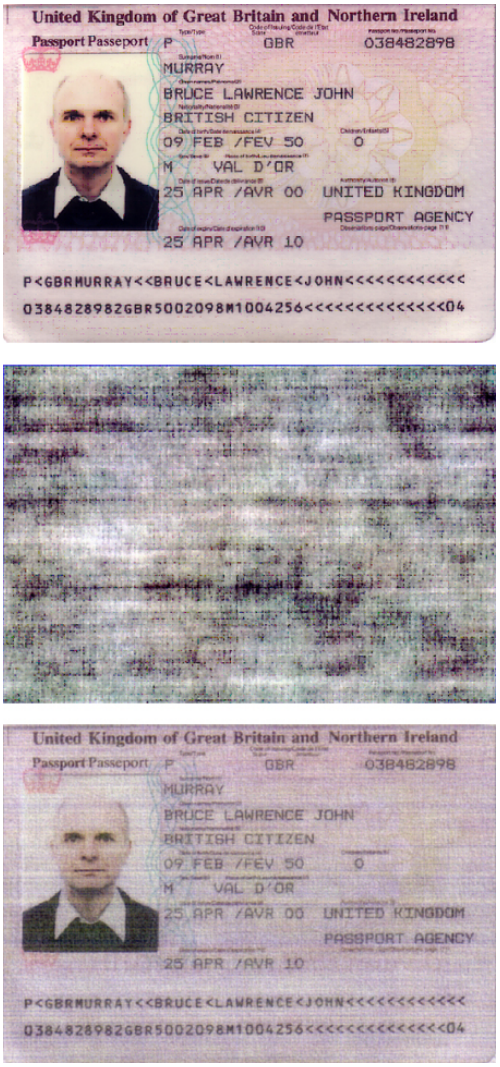


Fig. 14. Example of the stochastic diffusion method applied to passport authentication: Original image scanned from a passport at 400dpi (above), printed image after applying stochastic diffusion (centre) and reconstruction after scanning the printed stochastic field at 300dpi (below).

paper watermarking, UV watermarking, foil holograms, ghost images and so on. Each of these security features may be more or less difficult to counterfeit depending on the sophistication of the feature and the counterfeiter. In this case study, we consider the use of texture coding within the context of authenticating a passport including the protocol associated with a typical 'cycle'. The method is simple and cost effective to implement in terms of the hardware required, i.e. Standard PC, flatbed scanner and printer, all of which are COTS. All that is required is a remote web site hub to which digital scans of the texture code can be emailed and where a decrypt can take place, forwarding the result back to the point of enquiry.

The principal idea is to take a low resolution scan (say 600dpi) of the page (or pages) of a passport that contains the primary information, e.g. Passport number, Name, Date of birth, Signature and Photograph of the passport holder - the plaintext. This plaintext is then forwarded to a designated Hub where it is diffused with a unique noise field that is maintained at the Hub alone to produce the ciphertext. The result is then emailed back to the user, printed and the result (permanently or as required) inserted into the passport, a process that is similar to issuing a Visa, for example. At any point of contact, if the passport requires authentication, the ciphertext is scanned and the digital image emailed to the appropriate Hub where upon it is decrypted and the result (the watermark) sent back to the point of origin. Automation of this cycle would require a new infrastructure to be established which is both time consuming and expensive. Instead, the cycle described above would be best suited for use with regard to spot checks at an airport terminal, for example, especially if the holder of the passport or the passport itself is suspect. The scanning process (using a standard flat bed scanner, for example) can then be undertaken while the holder of the passport is waiting for it to be authenticated (or otherwise) based on a visual comparison between the decrypt and the plaintext.

Figure 14 shows an example of the technique applied to a composite image scanned from a passport, an application which is cheap and simple to implement with regard to authenticating a passport holders personal information. The degradation associated with the reconstruction is due to the low resolution of the printing and scanning rather than the

information hiding method. Unless the correct stochastic field is used (as determined by the keys), it is not possible to reconstruct the image making counterfeiting or forgery improbable. In this case the scan has been emailed as a JPEG attachment where decryption can take place remotely.

VII. Discussion

Valuable paper documents are subject to misuse by criminals. This is largely due to the dramatic improvement in personal computer hardware and peripheral equipment. Embedding watermarks into a printed document is one way to secure them. The ability to extract the watermark from a printed copy is generally useful to help establish ownership, authenticity, and to establish the origin of an unauthorized disclosure. However, finding a robust watermarking technique is a continuing challenge. This is due to extensive amount of noise that is added when a document is printed and scanned. Moreover, printed documents do not maintain their quality over time.

In this paper, we have presented a robust watermarking method for paper security. Unlike traditional watermarking techniques, this approach can extract the hidden watermark after a print/scan attack which is achieved by using the convolution and correlation processes for coding and decoding respectively. This approach is chosen because of its compatibility with the principles of the physical optics involved in scanning a document. The watermark w is diffused (convolved) with a noise field n and placed into the background of a covertext, typically a text document. The watermark can be recovered by removing the covertext f using a modified median filter. We then correlate the diffused watermark with the original noise field. This process (i.e. coding and decoding) is compounded in the following formula:

$$\begin{aligned} w' &= \mathcal{F}\left\{ \overbrace{\left(\underbrace{\left(\overbrace{A_w e^{i\theta_w} A_n e^{i\theta_n}}^{\text{Diffusion}} + f \right)}_{\text{Confusion}} - f \right) A_n e^{-i\theta_n}}^{\text{Covertext removal}} \right\}_{\text{Correlation}} \\ &= \mathcal{F}^{-1}\left\{ A_w e^{i\theta_w} |A_n|^2 \right\} \end{aligned}$$

where w' is the extracted watermark, A_w and A_n are the amplitude spectra of the watermark and cipher respectively, θ_w and θ_n are the respective phase spectra. The extracted watermark is a noisy version of the embedded watermark. This noise is due to the power spectrum $|\ A_n\ |^2$. In order to enhance the extracted watermark, we have to eliminate the power spectrum term or at least minimize its effect. One way to do this is to divide the output over the power spectrum during the diffusion step or correlation step. To avoid singularities, we replace each and all zeros that occur in $|\ A_n\ |^2$ by 1. Alternatively, we can choose n such that it has a homogeneously distributed power spectrum across all frequencies, (such as white noise) or pre-process n by replacing it's amplitude spectrum with a constant value. However, these conditions are restrictive and the regularisation method discussed above to avoid singularities is both simple and effective.

The method is robust to a wide variety of attacks including geometric attacks, drawing, crumpling and print/scan attacks as discussed in the Appendix. Further the method is relatively insensitive to lossy compression, filtering, amplitude adjustments, additive noise and thresholding. The principal weakness of the system is its sensitivity to rotation and cropping. This can be minimized by orienting the document correctly and accurately before scanning and using automatic cropping software which is available with selected scanners (e.g. Cannon scanners). Alternatively, introduction of a frame provides a reference feature from which an accurate crop can be obtained.

The visibility of the diffused watermark and the compatibility of this system with the physical principles of an imaging system, increase the robustness of the system and provides a successful approach to the extraction of the watermark after scanning at low resolution. Moreover, using correlation in the extraction phase increases the robustness of the system to some important attacks such as translation and cropping (most likely to occur during a scan).

The system is secure in that it can not be attacked easily. First, the feature is not 'suspicious' as many documents have a background texture. Second, the attacker does not know the algorithm used to generate the diffused watermark. Even if the attacker does know the algorithm, he/she must still know a significant amount of information before the system can be broken, such as: the correct key, the diffusion operator type, the original image size and so on.

For interested readers, a prototype system is available for trial purposes which can be downloaded from
http://eleceng.dit.ie/arg/downloads/Document_Authentication.zip

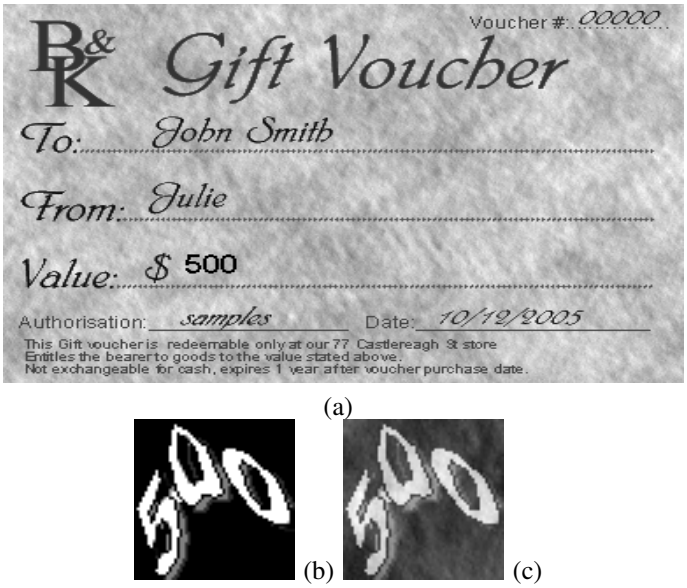
Appendix I
Attack and Robustness Analysis

In practice, a watermarked document may be altered either intentionally or unintentionally but a watermarking system should still be able to detect and extract the watermark. The distortions are limited to those that do not produce excessive degradations otherwise the transformed object is unusable. Examples of processes that a watermark might need to 'survive' are lossy compression, filtering, and geometric distortion. In order to evaluate the watermarking technique, the system is tested against some important attacks.

The results presented in this Appendix[2] are not exhaustive and have been designed to provide the reader with a short overview of an 'attack and robusteness analysis' for the sake of completeness. For this purpose, all tests are applied on the watermarked test document that is shown in Figure 15(a) (277×388). The original and extracted watermarks (60×60) are shown in Figure 15 (b) and (c) respectively. The diffusion is undertaken using a Gaussian noise field [39]

To measure the effect of the attacks on the extracted watermark, pixel-based visual distortion metrics are used. The quantitative distortion metrics allow for fair comparison

[2]Based on an edited version of material given in [39]



(a)

(b) (c)

Fig. 15. Attack experiment parameters (a) Watermarked document used to test attacks; (b) Original watermark; (c) Extracted Watermark.

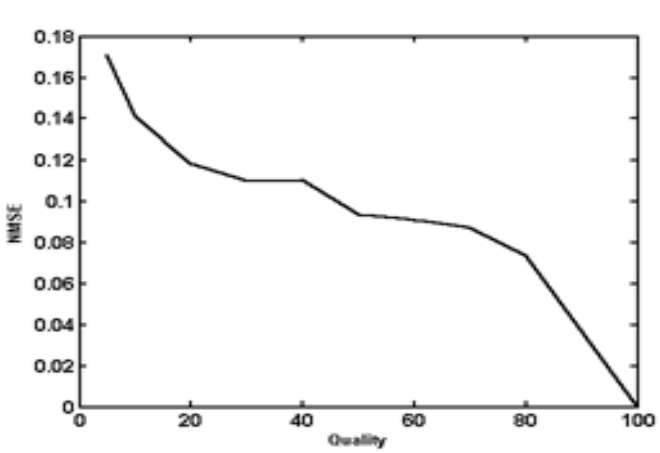
between different images. One way to test the similarity between two images is the Normalized Mean Square Error (NMSE) given by

$$\mathrm{NMSE} = \frac{\sum_{\mathrm{ij}} \left(\mathrm{P_{ij}} - \mathrm{Q_{ij}}\right)^2}{\sum_{\mathrm{ij}} \left(\mathrm{P_{ij}} + \mathrm{Q_{ij}}\right)^2}$$

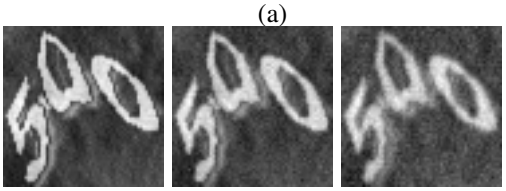
where P_{ij} represents a pixel, whose coordinates are (i, j), in the original undistorted image, Q_{ij} represents a pixel, whose coordinates are (i, j) in the distorted image. If the two images are identical, NMSE returns zero. If the difference between the two images are significant, then NMSE returns a number close to one.

A. Lossy Compression

Lossy compression techniques try to reduce the amount of information by removing imperceptible signal components. The loss of information can be acceptable because the computer representation of the signal contains redundancy with respect to what is needed for human perception and interpretation. One of the most popular lossy compression is the JPEG method. The compression itself is performed in three sequential steps: block-DCT computation, quantization, and variable-length code assignment [10]. A 'Quality' measure, a number between 0 and 100, is usually bounded with this kind of compression. Higher numbers imply better quality (i.e. less image degradation due to compression), but the resulting file size is larger. Due to the fact that JPEG is designed to disregard redundant perceptually insignificant information, and that the watermark is a visibly diffused watermark inserted into background, it is expected that this coding system is robust to JPEG. In this experiment, the watermarked document is compressed using different quality values. The extracted watermarks are compared to the original watermark, and the results are plotted in Figure 16. The extracted watermarks for



(a)



(b)

Fig. 16. JPEG test results. (a) Results of JPEG test. (b) Extracted watermark for JPEG quality = 80, 50 and 20 (from left to right).

quality = 80, 50 and 20 are shown in Figure 16 (a). The results show that the method is robust for JPEG even with a very small quality - Figure 16 (b). It is very important that the watermarking technique is robust to JPEG, since most images are vulnerable to compression, especially if they are to be uploaded to the Internet, when the scanner (or any other image capture device) and the extractor program are in two different locations and the Internet is used to transfer the images.

B. Linear Filtering

Another common type of process that changes images in a deterministic fashion is linear filtering, i.e.

$$I' = I \otimes\otimes f$$

where I is an image, f is a filter, and $\otimes\otimes$ denotes 2D convolution. Many normal operations on images are explicitly implemented with linear filters. The blurring and sharpening effects in image editing programs apply simple filtering operations [35]. In addition, a scanning process can be modeled by convolution. To understand the effect of linear filtering on the system, the following analysis is considered. Convolution in the spatial domain corresponds to a multiplication in the Fourier domain. Thus, we can think of each frequency component as being either attenuated or amplified by a real number. A diffused watermark can be represented in the frequency domain using polar coordinates by

$$\mathrm{PW} = A_p e^{i\theta_p} A_w e^{i\theta_w}$$

where A_p is the amplitude spectrum of the diffused operator, A_w, is the amplitude spectrum of the watermark, θ_p and θ_w are the phase spectra of the diffused operator and the

watermark respectively. The filtered diffused watermark can be represented by

$$\mathrm{FPW} = A_f e^{i\theta_f} A_p e^{i\theta_p} A_w e^{i\theta_w}$$

where A_f and θ_f are the amplitude spectrum and the phase spectrum of the filter respectively. The watermark extraction (i.e. correlation) process, can then be modelled in terms of

$$\begin{aligned} W' &= A_f e^{i\theta_f} A_p e^{i\theta_p} A_w e^{i\theta_w} A_p e^{-i\theta_p} \\ &= |A_p|^2 A_f e^{i\theta_f} A_w e^{i\theta_w}. \end{aligned}$$

Hence, the extracted watermark is a filtered version of the original watermark. In other words, in the frequency domain, the watermark is spread across the whole range of frequencies (*Spread Spectrum Coding*). Thus, if the watermarked document is distorted by the some process that affects only a fraction of the frequency spectrum, as with linear filtering, then the watermark is still identifiable.

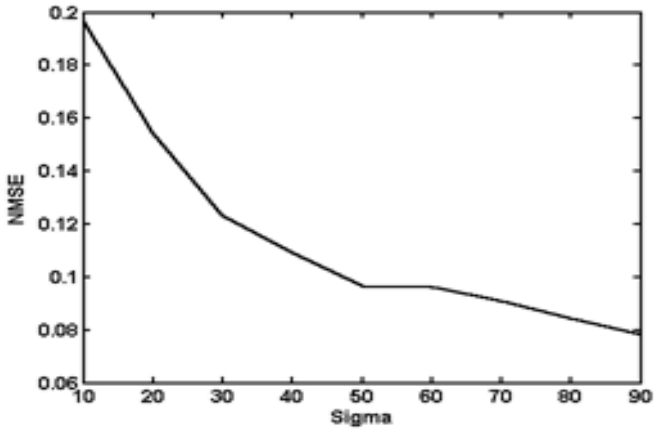


Fig. 17. GLPF test results. a) Results of GLPF test. b) Extracted watermark for GLPF with sigma = 80, 50 and 20 (from left to right).

C. Low-Pass Filtering

To test the effect of low-pass filtering, the watermarked document is distorted with a Gaussian Low-Pass Filter (GLPF) of varying degrees of the standard deviation σ (set to radii values). The extracted watermarks are compared to the original and the results are plotted in Figure 17. The extracted watermarks for σ = 80, 50 and 20 are shown in Figure 17 (a). The results show that the method is robust for a GLPF, even for very small values of σ - Figure 17 (b).

D. High-Pass Filtering

In this test, the watermarked document is distorted with an Ideal High-Pass Filter (IHPF) of varying Half-Band Width

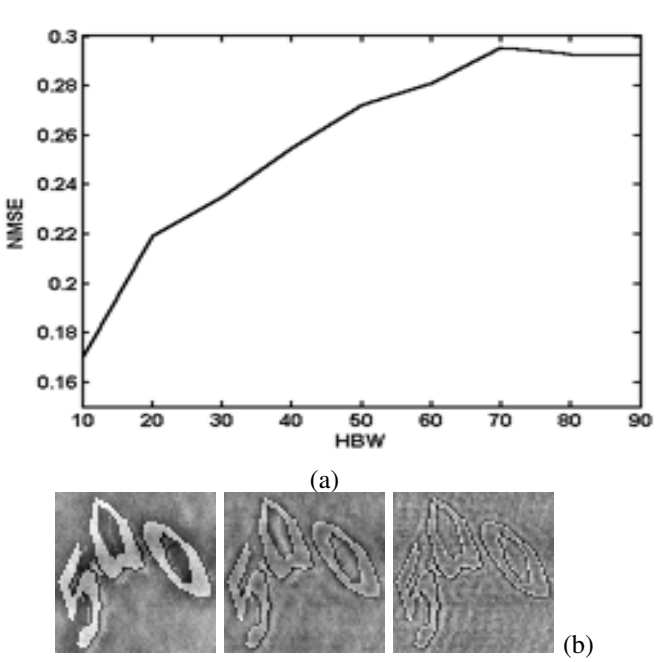


Fig. 18. IHPF Test Results. a) Results of IHPF test. b) Extracted watermark for IHPF HBW = 20, 50 and 80 (from left to right).

(HBW). The difference between the extracted watermark and the original are plotted in Figure 18 (a). The extracted watermarks for HBW = 20, 50 and 80 are shown in Figure 18 (b). The results show that the method is robust for a IHPF even with a large HBW (the watermark can be recognized from its edges).

E. Additive Noise

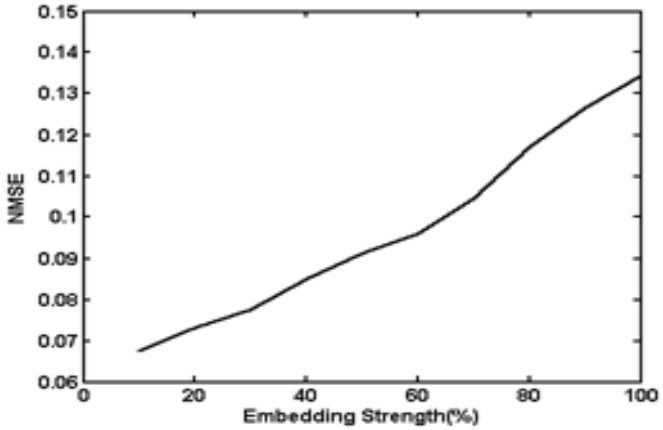
Many processes that may be applied to an image have the effect of introducing additive noise [35], i.e.

$$c_n = c + \alpha n$$

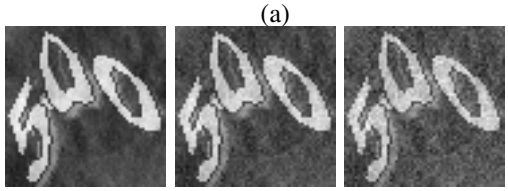
where c is the original image, n is the noise, and α is the embedding strength. For example, Gaussian noise arises in an image due to factors such as electronic circuit noise and sensor noise due to poor illumination and/or high temperature, for example [10]. Such noise processes are cases of *additive noise*. Most watermarking techniques are specifically designed to survive this type of distortion. In this test, we investigate the effect of additive noise on the system. The watermarked document is distorted with additive Gaussian noise with different embedding strengths α. The extracted watermarks are compared to the original watermark and the results are plotted in Figure 19 (a). The extracted watermarks for α = 0.20, 0.50 and 0.80 are shown in Figure 19 (b). The results show that the method is robust for additive noise. Note that if the same noise field as that used for diffusion with $\alpha = 1$ is applied, then the extracted watermark will be a spike.

F. Amplitude Effects

Many processes applied to a watermarked document are *deterministic* functions. A simple, but important example is the



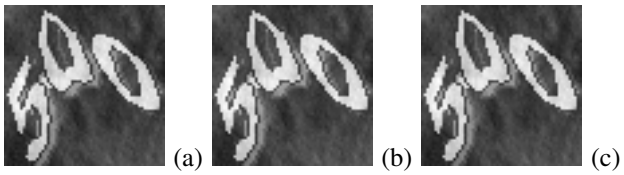
(a)



(b)

Fig. 19. Additive noise test results. a) Results of additive noise test. b) Extracted watermark for embedding strength = 0.2, 0.5 and 0.8 (from left to right).

change in amplitude, i.e. $c_n = vc$ where c is the original image and v is a scaling factor. For music, this simply represents a change of volume. In an image, it represents a change in brightness and contrast. In this test, the contrast of the watermarked document is modified by scaling the amplitude of the image by several scaling factors between $v = 1$ and $v = 0$. Extracted watermarks for $v = 20,\ 50$ and 80 are shown in Figure 20 (a), (b) and (c) respectively. This test shows that *'amplitude effects'* have no influence on the model. This is because any change in the contrast can be reversed by removing the foreground with a suitable threshold.



(a) (b) (c)

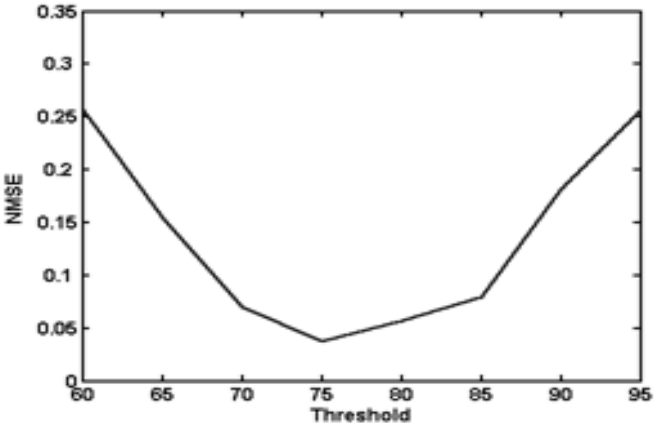
Fig. 20. Amplitude change test results. a) Extracted watermark for scale = 0.2. b) Extracted watermark for scale = 0.5. c) Extracted watermark for scale = 0.8.

G. Thresholding

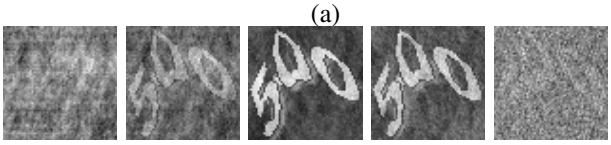
A thresholded image $g(x, y)$ can be defined as

$$g\ (x,y) = \begin{cases} 1 & \text{if } f(x,y) > T \\ 0 & \text{if } f(x,y) \leq T \end{cases}$$

where $f(x, y)$ is the original image and T is the threshold. This is the simplest of all thresholding techniques where we partition the image histogram by using a single global threshold T. Image thresholding plays a central role in the applications of *image segmentation*. Segmentation subdivides an image into its constituent regions or objects. In this system, thresholding can be applied on the scanned watermarked document (i.e. instead of scanning the document using a grey level scale). The test shows that the watermark can be extracted



(a)



(b)

Fig. 21. Thresholding test results. (a) Results of thresholding test. (b) Extracted watermark for threshold = 0.60, 0.65, 0.75, 0.85 and 0.95 (from left to right).

after thresholding in a specific range $(65 \leq T < 95)$. Most of the data outside this range are 'cleared' (i.e set to one or zero). The results are shown in Figure 21.

H. Geometric Distortion

Geometric distortions that affect image data include rotation, spatial scaling, translation, and changes to the aspect ratio. Although many different approaches have been investigated, robustness to geometric distortion remains one of the most difficult outstanding areas of watermarking research [36]. The geometric distortions that are applied to the watermarked document must be identified first. Then the distortion is inverted before the extraction proceeds. Most suggested approaches fall into one of the following categories:

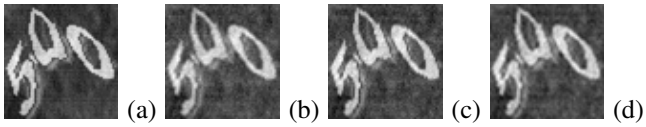
1) Exhaustive search. After defining a range of likely values for each distortion parameter, every combination of parameters are examined.
2) Registration. When the original non-watermarked document is available at the extractor side, regions suspected of containing a watermark are aligned (using techniques from the pattern recognition literature) prior to a single application of the extractor. A common approach for public extraction is the embedding of a dedicated synchronization pattern in addition to the embedded watermark [37]
3) Autocorrelation. Autocorrelation (i.e. peaks in the autocorrelation surface) can be used to identify and invert geometric distortion.
4) Invariant watermarks. Rather than detect and invert the geometric distortions, an alternative approach is to

design watermarks that are invariant to such distortions [38].

In the following subsections, the system is tested against these kind of distortions.

I. Scaling

The size of any image captured by a scanner depends on the resolution used. The size of the image is usually different from the original because the resolution of the scanner and the printer may be different. Since scaling in the spatial domain causes inverse scaling in the Fourier domain (i.e. as the spatial scale expands, the frequency scale contracts and the amplitude increases linearly), the scaling must be reversed before extracting the watermark. The original document size is part of the key and hence, it can be used to rescale the document to its original size before correlating since correlation is not a scale invariant process. Figure 22, shows the extracted watermarks

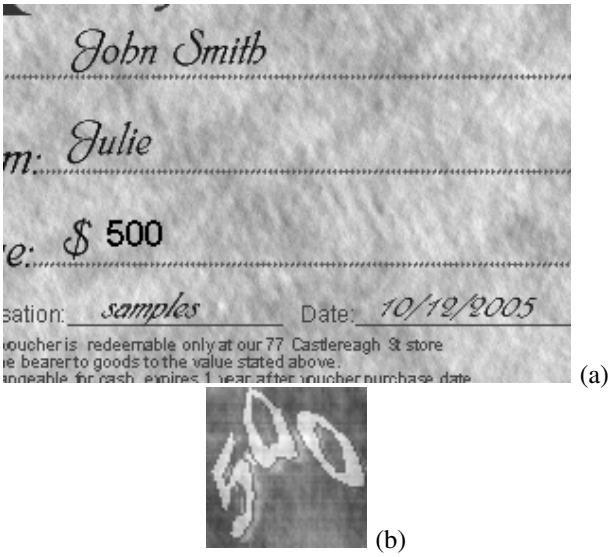


(a) (b) (c) (d)

Fig. 22. Scaling test results. (a), (b) Extracted watermark after scaling up/down while reserving the aspect ratio. (c), (d) Extracted watermark after scaling without reserving the aspect ratio.

after: (a) scaling up to 348×488 while preserving the aspect ratio; (b) scaling down to 206×288 while preserving the aspect ratio. (c) and (d) scale to 200×400 and 200×300 respectively without retaining the aspect ratio. In all cases, the watermark can be extracted.

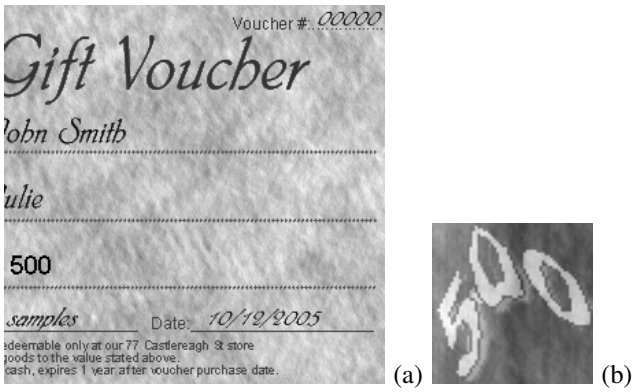
J. Cropping



(a) (b)

Fig. 23. Cropping test results. a) Cropped document (top=70, left=50, height=199, width=299). b) Extracted watermark after cropping.

Cropping is a typical effect of scanning. The scanned image may be cropped (i.e. to include only a part of the original image). Figure 23 and Figure 24 show two examples of cropped documents and the corresponding extracted watermarks. The test shows how robust the system is to cropping.



(a) (b)

Fig. 24. Cropping test results. a) Cropped document (top=0, left=100, height=288, width=277). b) Extracted watermark after cropping.

This robustness is justified by the following analysis.

1) Redundant Effects of Convolution: Physically, convolution can be thought of as a *'blurring'* or *'smearing'* of one function (image) by another, an effect that is clear from the definition of convolution. The convolution of two functions is the same as the product of their Fourier transforms in Fourier space *(convolution theorem)* where the convolution between two images of size $M \times N$ can be defined by the following expression [10]:

$$f(x,y) \otimes\otimes h(x,y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) h(x-m, y-n) \qquad (1)$$

Equation 1 is just an implementation of (i) flipping one image (mask) about the origin; (ii) sliding that function (image) past the other by changing the values of (x, y); (iii) at each displacement (x, y), the entire summation in equation 1 is carried out.

Convolution can be thought of in terms of a *Redundant Embedding* of one image with another. To understand this, we introduce the definition of the *impulse function*. An impulse function of strength A, located at coordinates (x_0, y_0), is denoted by $A\delta(x-x_0, y-y_0)$ and is defined by the expression (for any function f)

$$\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) A\delta(x-x_0, y-y_0) = Af(x_0, y_0) \qquad (2)$$

This equation represents one of the most important properties of the impulse function, namely, the *shifting property*. Using the definition of convolution (Eq. 1), equation 2 states that the convolution of a function with an impulse *copies* the value of the function (multiplied by the strength of the impulse) at the location of the impulse [10]. Since any image $(M \times N)$ used for diffusion is a set of $(M \times N)$ impulses with different strength, then the convolution between this image and the watermark can be thought of as an addition of $(M \times N)$ images. Each image is a shifted version of the watermark with a strength equal to the corresponding impulse. The addition of images in the spatial domain is equivalent to addition

in Fourier space (i.e. The DFT is a linear operator). From this analysis, the texture produced from diffusion is nothing more than a redundant embedding of the watermark. This redundancy provides robustness to cropping.

K. Spatial Translation

A translation or shift implies zero padding of the image, such as would occur if an image were placed on a scanner and scanned (i.e. scanning the whole image with an additional background). Zero-padding in the spatial domain is equivalent to increasing the sampling resolution in the frequency domain. Figure 25, shows the extracted watermark after applying different translation shifts on the watermarked document. Watermarks (a), (b), (c) and (d) are extracted after zero-padding the watermarked document. The remaining images (i.e. e, f, g, and h) are extracted from a watermarked document padded by 1s. The shifting is undertaken as follows:

- (a) and (e): $top = 10$, $left = 10$, $bottom = 10$ and $right = 10$ pixels.
- (b) and (f): $top = 1$, $left = 1$, $bottom = 1$ and $right = 1$ pixels.
- (c) and (g): $top = 10$, $left = 10$, $bottom = 0$ and $right = 0$ pixels.
- (d) and (h): $top = 0$, $left = 0$, $bottom = 10$ and $right = 10$ pixels.

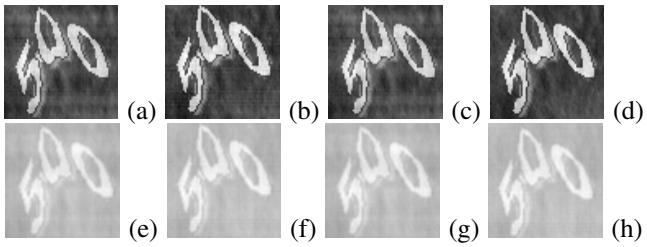


Fig. 25. Extracted watermark after translation. (a,b,c,d) from the zero-padded watermarked document. (e,f,g,h) from one-padded watermarked document. Note: refer to the text for more details.

Cropping the translated document to its original size is important before applying correlation.

L. Rotation

Almost all applications involving printing and scanning results in some degree of rotation. On the other hand, robustness to rotation is still one of the most difficult forms of attack. Most watermarking algorithms are robust to a small degree of rotation while they fail to extract the watermark for large degrees of rotation. If document verification is based on a flat-bed scanner, it is possible to place the image in the corner of the scanner in order to minimize the effect of rotation.

Rotating an image by an angle θ in the spatial domain, rotates the frequency space data by the same angle. Hence, in the watermarking system considered here, the rotation must be reversed before applying the correlation. Figure 26 shows the extracted watermarks for documents that have been:

- Rotated by one degree (a).
- Rotated by one degree and then cropped to its original size (b).
- Rotated by one degree, rotation inverted and then cropped to its original size (c).
- Rotated by two degrees (d).
- Rotated by two degrees and then cropped to its original size (e).
- Rotated by two degrees, rotation inverted and then cropped to its original size (f).

This test shows that the system can not extract the watermark after rotation, unless, the tested document is rotated by one degree or less ($\theta \leq 1$) and then cropped to its original size. For larger angles, the rotation must be reversed first and then cropped to its original size.

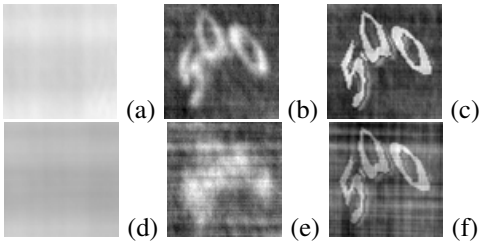


Fig. 26. Rotation test results. (a, b, c) Watermarked document rotated by one degree. (d, e, f) Watermarked document rotated by two degrees.

M. Document Damage

Printed documents do not maintain their quality over time. They are subject to aging, soiling, crumpling, tearing, and deterioration. Figure 27, shows a document that has been crumpled, teared and 'scribbled' on.

Figure 28, shows an acceptable watermark recovered from a damaged document. Thus, even though the damage to a document may be extensive, this test shows that a watermark is still recoverable.

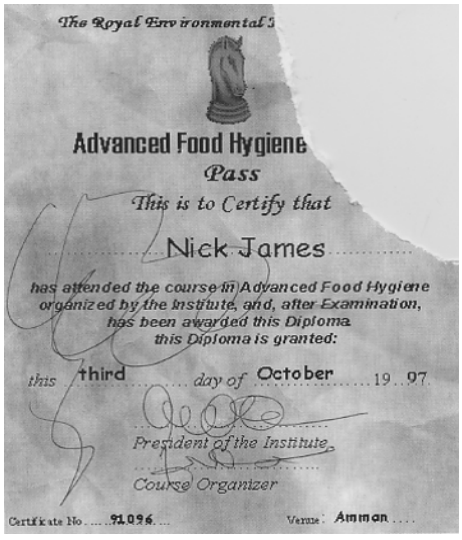


Fig. 27. Damaged document.

Fig. 28. Extracted watermark from a damaged document.

ACKNOWLEDGMENTS

J M Blackledge is supported by the Science Foundation Ireland (Stokes Professorship Programme).

REFERENCES

[1] C. T. Hsu and J. Ling, *Hidden digital watermarks in images*, IEEE Transactions on Image Processing, Vol. 8, 58-68, 1999.

[2] J. Zaho and E. Koch, *Embedding robust labels into images for copyright protection*, Proceedings of the International Conference on Intellectual Property Rights for Information, Knowledge and New Techniques, 242-251, (München, Wien: Oldenbourg Verlag), 1995.

[3] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, *A secure, robust watermark for multimedia*, First International Workshop on Information Hiding, Ed. R. Anderson, 1174 of Lecture Notes in Computer Science, 183-206, Springer-Verlag, 1996.

[4] M. Barni, F. Bartolini, V. Cappellini, and A.Piva, *A DCT-domain system for robust image watermarking*, Signal Processing (EURASIP), Vol. 66, 357-372, 1998.

[5] J. J. K. O. Ruanaidh, W. J. Dowling, and F. M. Boland, *Watermarking digital images for copyright protection*, IEE Proceedings on Vision, Signal and Image Processing, Vol. 143, 250-256, 1996.

[6] M. D. Swanson, B.Zhu, and A. H. Tewfik, *Transparent robust image watermarking*, International Conference on Image Processing, IEEE,**3**, 211–214, 1996.

[7] J. J. Chae and B. Manjunath, *A technique for image data hiding and reconstruction without host image*, Proc. Of SPIE Electronic Imaging'99, Security and Watermarking of Multimedia Contents, Wong and Delp, eds., **3657**, (San Jose, California), January 1999.

[8] A. Bors and I. Pitas, *Image watermarking using block site selection and d.c.t. domain constraints*, Optics Express **3**, 512-523, Dec 1998.

[9] B. Tao and B. Dickenson, *Adaptive watermarking in the DCT domain*, International Conference on Acoustics and Signal Processing, 1997.

[10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, New Jersey, 2nd ed., 2002.

[11] X. Xia, C. Boncelet, and G. Arce, *A multi-resolution watermark for digital Images*, in Proc. IEEE Int. Conf. On Image Processing, **1**, 548-551, Oct 1997.

[12] A. Lumini and D. Maio, *A wavelet-based image watermarking scheme*, Proceedings of the IEEE International Conference on Information Technology: Coding and Computing, 122-127, March 2000.

[13] D. Kundur and D. Hatzinakos, *Digital watermarking using multiresolution wavelet decomposition*, Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, **6**, 2969-2972, 1998.

[14] D. Kundur and D. Hatzinakos, *A robust digital image watermarking method using wavelet-based fusion*, International Conference on Image Processing, IEEE, 544-547, (Santa Barbara, California, U.S.A.), Oct 1997.

[15] M. D. Levine, *Vision in Man and Machine*, McGraw-Hill, Toronto, 1985.

[16] J. Ohnishi and K. Matsui, *Embedding a seal into a picture under orthogonal wavelet transform*, Proc. Int. Conference on Multimedia Computing and Systems, 514-521, June 1996.

[17] M. Barni, F.Bartolini, V.Cappellini, A. Lippi, and A. Piva, *A DWT-based technique for spatio-frequency masking of digital signatures*, Proc. Of SPIE Electronic Imaging'99, Security and Watermarking of Multimedia Contents, **3657**, 31-39, (San Jose, California), January 1999.

[18] A. S. Lewis and G. Knowles, *Image compression using the 2-d wavelet transform*, IEEE trans. Image Processing, **1**, 240-250, April 1992.

[19] H. Inoue, T. Katsura, A. Miyazaki, and A. Yamamoto, *A digital watermark technique based on the wavelet transform and its robustness on image compression and transformation*, IEICE Transactions on Fundamentals of Electronics, **E82-A**, 2-10, Jan 1999.

[20] J. Shapiro, *Embedded image coding using zerotrees of wavelet coefficients*, IEEE Trans. Signal Processing, **41**(12), 3445-3462, 1993.

[21] J. J. K. O. Ruanaidh and T. Pun, *Rotation, scale and translation invariant spread spectrum digital image watermarking*, Signal Processing(EURASIP), **66**, 303-317, May 1998.

[22] J. J. K. O. Ruanaidh, W. J. Dowling, and F. M. Boland, *Phase watermarking of digital images*, Proceedings of the IEEE International Conference on Image Processing, **3**, 239-242, sep 1996.

[23] V. Solachidis and I. Pitas, *Circularly symmetric watermark embedding in 2-d DFT domain*, International Conference on Acoustics, Speech and Signal Processing, IEEE Signal Processing Society, 1563-1565, (Phoenix, Arizona, USA), March 1999.

[24] W. Kim, J. Lee, and W. Lee, *An image watermarking scheme with hidden signature*, IEEE Proceeding of the International Conference on Image Processing, 206-210, (Japan), Oct 1999.

[25] H. Raymound, F. Chan, and K. Yeung, *A frequency domain watermarking scheme*, Jan 2001.

[26] A. Herrigel, J. O'Ruanaidh, H. Petersen, and S. Pererira, *Secure copyright protection techniques for digital images*, Information Hiding of Lecture Notes in Computer Science, D. Aucsmith, ed., **1525**, 169-190, Springer-Verlag, 1998.

[27] C.-Y. Lin, M. Wu, J. Bloom, I. Cox, M. Miller, and Y. Lui, *Rotation, scale, and translation resilient public watermarking for images*, Security and Watermarking of Multimedia Contents II, Proceedings of SPIE, P. W. Wong and J. E. Delp, Eds., **3971**, 90-98, 2000.

[28] M. Kankanhalli and R. Ramakrishnan, *Content based watermarking of images*, 6th ACM International Multimedia Conference, 61-70, (Bristol, England), Sep 1998.

[29] R. J. Anderson and F. Petitcolas, *On the limits of steganography*, IEEE: Selected Areas in Communications, **16**, 474-481, May 1998.

[30] J. M. Blackledge, *Multi-algorithmic Cryptography using Deterministic Chaos with Applications to Mobile Communications*, ISAST Transactions on Electronics and Signal Processing (ISSN 1797-2329), **1**(2), 23 - 64, 2008

[31] http://www.fourmilab.ch/hotbits/

[32] M. Kutter and F. Hartung, *Introduction to watermarking techniques*, Information Hiding: Techniques for Stegangraphy and Difgital Watermarking, S. Katezenbeisser and F. A. P. Petitcolas, Eds., Ch. 5, 97-120, Artech House, Boston, 2000.

[33] A. K. Jain, S. Pankanti, and R. Bolle, eds., *BIOMETRICS*: Personal Identification in Networked Society, Kluwer, 1999.

[34] A. K. Jain, F. D. Griess, and S. D. Connell, *On-line signature verification*, Pattern Recognition **35**, 2963-2972, December 2002.

[35] I. J. Cox, M. Miller, and J. Bloom, *Digital Watermarking*, Morgan Kaufmann, 2002.

[36] J.-L. Dugelay and F. A. P. Petitcolas, *Possible counter-attack against random geometric distortions*, Security and Watermarking of Multimedia Content, **3971**, 338-345, SPIE, 2000.

[37] S. Pereira, J. K. O. Ruanaidh, and T. Pun, *Secure robust digital watermarking using the lapped orthogonal transform*, Security and Watermarking of Multimedia Content, **3657**, 21-30, SPIE, 1999.

[38] J. J. K. O. Ruanaidh and T. Pun, *Rotation, scale and translation invariant spread spectrum digital image watermarking*, Signal Processing(EURASIP), **66**, 303-317.

[39] K. Mahmoud, *Novel Methods for Print Security and Anti-counterfeiting Technology* PhD Thesis, Loughborough University, 2006.

Jonathan Blackledge graduated in physics from Imperial College in 1980. He gained a PhD in theoretical physics from London University in 1984 and was then appointed a Research Fellow of Physics at Kings College, London, from 1984 to 1988, specializing in inverse problems in electromagnetism and acoustics. During this period, he worked on a number of industrial research contracts undertaking theoretical and computational research into the applications of inverse scattering theory for the analysis of signals and images. In 1988, he joined the Applied Mathematics and Computing Group at Cranfield University as Lecturer and later, as Senior Lecturer and Head of Group where he promoted postgraduate teaching and research in applied and engineering mathematics in areas which included computer aided engineering, digital signal processing and computer graphics. While at Cranfield, he co-founded Management and Personnel Services Limited through the Cranfield Business School which was originally established for the promotion of management consultancy working in partnership with the Chamber of Commerce. He managed the growth of the company from 1993 to 2007 to include the delivery of a range of National Vocational Qualifications, primarily through the City and Guilds London Institute, including engineering, ICT, business administration and management. In 1994, Jonathan Blackledge was appointed Professor of Applied Mathematics and Head of the Department of Mathematical Sciences at De Montfort University where he expanded the post-graduate and research portfolio of the Department and established the Institute of Simulation Sciences. From 2002-2008 he was appointed Visiting Professor of Information and Communications Technology in the Advanced Signal Processing Research Group, Department of Electronics and Electrical Engineering at Loughborough University, England (a group which he co-founded in 2003 as part of his appointment). In 2004 he was appointed Professor Extraordinaire of Computer Science in the Department of Computer Science at the University of the Western Cape, South Africa. His principal roles at these institutes include the supervision of MSc and MPhil/PhD students and the delivery of specialist short courses for their Continuous Professional Development programmes. He currently holds the prestigious Stokes Professorship in Digital Signal Processing for ICT under the Science Foundation Ireland Programme based in the School of Electrical Engineering Systems, Faculty of Engineering, Dublin Institute of Technology.

Khaled Walid Mahmoud received a BSc in Computer Science from Jordan University in 1992, an MSc in Computer Science (Artificial Intelligence) from Jordan University in 1998 and a PhD in Print Security and Digital Watermarking from Loughborough University in 2004 under the supervision of Professor J M Blackledge. This was followed by academic appointments at Zarka Private University, Jordan, including Assistant Professor of Computer Science. He is currently Head of the Department of Software Engineering at Zarka University and has research interests that include Information Security, Digital Watermarking, Image Processing, Artificial Intelligence and Arabic Language processing.