A PIXEL-BASED ROBUST IMAGE WATERMARKING SYSTEM

Wenming Lu, Wanqing Li, Rei Safavi-Naini, Philip Ogunbona

University of Wollongong, Australia

ABSTRACT

Robust image watermarking systems are required to be resistant to geometric attacks in addition to common image processing tasks, such as JPEG compression. However, robustness against geometric attacks, such as rotation, scaling and translation, still remains one of the most challenging research topics in image watermarking. We propose a new pixel-based watermarking system in which a binary logo is embedded, a bit per pixel, in the pixel domain of an image. The encoder of the proposed system is based on a sliding window embedding scheme that applies the local average quantization index modulation (QIM), to achieve geometric attack robustness. The decoder employs a maximum *a posteriori* (MAP) estimation supported by Markov Random Field (MRF) model to achieve robust decoding. Additionally, we demonstrate that the proposed scheme is also robust against possible watermark removal due to JPEG compression.

1. INTRODUCTION

For digital image watermarking systems, geometric attacks, such as rotation, scaling and translation, do not distort or remove the embedded watermark, but instead geometrically and globally modify the watermarked image to make the watermark decoder (or detector) unable to re-synchronize the received image. Most existing robust watermarking systems are block based and/or rely on the correct synchronization of the image to extract the embedded watermark. Geometric attacks destroy the synchronization and render the extracted embedded watermark incorrect or the extraction process impossible, thus making the watermark undetectable.

Robust image watermarking systems, which are used to address security concerns, such as copyright protection or copy control, should guarantee resistance to geometric attacks. Several systems have been proposed to address the problem of robustness against geometric attacks[1]. Exhaustive search techniques try all possible combinations of the geometric distortion and can be computationally costly or infeasible. Methods that embed a reference pattern into an image in addition to the robust watermark for aligning the received image at the decoder can impair either the fidelity or general robustness of the system. Invariant watermarking systems are designed to be robust only against certain geometric attacks and the approaches of autocorrelation and implicit synchronization also suffer from variant problems. Therefore, robustness against geometric attacks still remains one of the difficult challenges in image watermarking research [1, 2].

We suggest that pixel-based watermarking is ideal for addressing the problem of geometric attacks. The watermark embedding takes place on each individual pixel - in a binary system, a single bit is embedded in each pixel. Since the embedding unit is a pixel, the smallest block possible in images, the watermark extraction (decoding) is not affected by geometric attacks because re-synchronization problem is obviated. However, we now face another problem, namely, pixel-based watermarking is usually very fragile even to common image processing. For example, an attacker can use high quality JPEG compression to compress the image and remove the embedded watermark. To make pixel-based watermarking systems practically applicable, robustness against common image processing tasks must be improved.

In this paper, we propose a novel pixel-based watermarking system that is robust against both JPEG image compression and geometric attacks. For a given image, a binary logo, with the same size as the host image, is embedded. At each pixel we embed the corresponding bit of the logo and this embedding scheme makes the embedded logo resilient to geometric attacks. A sliding window with the predefined shape visits each pixel and the embedding is achieved by utilizing the quantization index modulation (QIM) [3, 4] and exploiting the idea from the local average QIM [5]. The applied embedding technique makes the logo robust against the JPEG compression. The possibly distorted watermark (the binary logo) is extracted and recovered by a MAP decoder (The original image is not known to the decoder since QIM as a host signal interference rejecting system does not need the original signal in the decoding process). Compared with the default Minimum Distance decoder of QIM, MAP decoder greatly reduces the error-decoding rate and improves quality of the decoded watermark. The proposed system can be used to address copyright protection of images where the distinctive logo can be used to identify the owner or establish claim of ownership.

2. THE PROPOSED SYSTEM

The proposed system consists of a sliding window based local average QIM (SW-LAQIM) embedding component and a maximum *a posterior* probability (MAP) decoder.

2.1. SW-LAQIM embedding

For a given image I with width w and height h, we select a registered binary logo Z of the same size as I; Z may be constructed through tiling, scaling or cropping from a logo that is smaller or larger than I.

Assume that $s \equiv (i, j)$, $0 \leq i < w$ and $0 \leq j < h$, represents a site in the image I; I_s and Z_s then represent respectively the corresponding pixel in I and the bit in the logo Z. The process of embedding Z into the host image I is as follows.

For pixel I_s, we choose a local window centered at s, as shown in Fig.1. Let $\mathbf{G}_s$ be the set of pixels in the local window, $\mathbf{G}_s \in I$. Let μ_s represent the average value of the pixels in $\mathbf{G}_s$. For the given

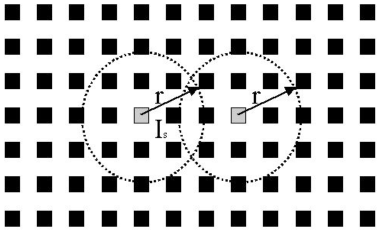

Fig. 1. Sliding window embedding (the gap between pixels is exaggerated for display purpose)

embedding bit Z_s, QIM embedding function modifies μ_s as,

$$\mu_s' = \mathbf{q}(\mu_s + d(Z_s)) - d(Z_s) \tag{1}$$

where $\mathbf{q}(.)$ is the scalar quantization function and $d(.)$ is the corresponding dither vector (Readers are referred to [3, 4] for details on QIM). The corresponding watermarked pixel I_s', is given by,

$$I_s' = I_s + (\mu_s' - \mu_s) \tag{2}$$

The embedding is repeatedly applied at each pixel site of the image I to produce the watermarked image I'.

Notice that windows $\mathbf{G}_s$ and $\mathbf{G}_{s+1}$ centered at two neighboring sites, s and $s+1$, are overlapped. This is particularly different from the Local Average QIM (LAQIM) in [5] where none-overlapping windows are selected. Therefore, we name the proposed embedding method as the sliding window based LAQIM (SW-LAQIM). Also, the modification on μ_s only passes to I_s in SW-LAQIM, not to all pixels in the selected window as in LAQIM. However, we expect that there is a strong spatial coherency of the pixels in both I and Z and the bit embedded at I_s can be decoded from the local average at I_s' of the embedded image I'. In addition, We choose $\mathbf{G}_s$ as a disk-shape window centered at s with radius r, as shown in Fig.1, in order to make the embedding resistant to rotation and geometric scaling.

2.2. MAP decoding

2.2.1. *Models*

Let I' be the image arrived at decoder. For pixel at site s, $I_s' \in I'$, the average intensity, μ_s', of a local window $\mathbf{G}_s$ centered at I_s' is used to calculate the distances between μ_s' and the nearest bit 0 and bit 1 quantizers, denoted as d_s^0 and d_s^1 respectively, where $d_s^0 + d_s^1 = \frac{\Delta}{2}$, Δ is the quantization step for $\mathbf{q}(.)$ in QIM. We refer either $d_s^0\,\forall s$ or $d_s^1\,\forall s$ as a $d-map$.

In conventional QIM decoder, the bit embedded at site s is decoded by comparing d_s^0 with d_s^1. The embedded bit is 0 if $d_s^0 \leq d_s^1$ and 1 otherwise. Since every bit is decoded from the corresponding pixel independently, high error decoding rate is inevitable. While the error rate can be reduced significantly by embedding bits into local averages instead of original pixels [5], further improvement is possible by exploiting the contextual information when the bits in the embedded message (logo here) have some spatial coherency, e.g. a bit is more likely to be 1 (0) if its neighboring bits are 1(0). In this paper, we propose a maximum *a posteriori* probability (MAP) decoding method to utilize such contextual information.

Let $D = \{D_t : 1 \leq t \leq M\}$ be a random field defined on lattice L, where $M = w*h$ is the number of pixels and $t = j + w*i$ is the index of the pixel at (i,j). We consider a $d-map$ that is calculated from the received image as a realization of the $D = \{D_t : t \in L\}$, $d^* = \{d_t : t \in L\}$. We also consider the embedded bitmap (logo Z) as a true but unknown labelling of the $d-map$ and assume that the labelling is a realization of a random field $X = \{X_t : t \in L\}$, $x^* = \{x_t : t \in L\}$, where $x_t \in \{0,1\}$. Then the problem of decoding the embedded logo can be formulated as an estimation of the labels, $\hat{x}$, that maximizes the *a posteriori* probability [6, 7] given the observed $d-map$. Using Bayes rule

$$p(X = x|D = d) \quad \propto \quad p(D = d|X = x)p(X = x) \tag{3}$$

where $p(\cdot)$ is a probability density function (pdf). Assuming X is a Markov Random Field (MRF) [6] defined in a neighboring system η and $\{D_t, t = 1, 2, \cdots, M\}$ is conditionally independent and each D_t has the same conditional pdf, $f(d_t|x_t)$, dependent only on x , then Eq.(3) becomes

$$p(X = x|D = d) \propto \prod_{t \in L} f(d_t|x_t)p(x_t|x_{\partial t}) \tag{4}$$

where $p(d_t|x_t)$ is known as data model and

$$p(x_t|x_{\partial t}) = \frac{e^{-u(x_t|x_{\partial t})}}{Z_t} \tag{5}$$

is the prior pdf , known as prior model, of x_t given its neighbors, $x_{\partial t}$, defined in a neighborhood system η_t [6, 7]. Z_t is a partition function and $u(\cdot)$ is usually referred as an energy function.

We empirically define the prior model, or specially the energy function $u(\cdot)$, over cliques in a second-order neighborhood system [8] as

$$u(x_t|x_{\partial t}) = -\beta \sum_{s \in \partial t} \delta(x_t - x_s) \tag{6}$$

where β is a parameter to encourage (big value) or discourage (small value) the spatial coherency of the decoded bits and $\delta(\cdot)$ is a delta function.

The conditional pdf $f(d_t|x_t)$ is modelled as a Gaussian process with variance σ in this paper. Assume the $d-map$ is composed of $d_t^0, \forall t$, then

$$f(d_t|x_t = 0) \quad = \quad \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{d^2}{2\sigma^2}} \tag{7}$$

$$f(d_t|x_t = 1) \quad = \quad \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\frac{\Delta}{2} - d)^2}{2\sigma^2}} \tag{8}$$

2.2.2. *MAP estimation*

With the models defined in the previous section, the MAP estimation of the embedded logo is

$$\hat{x} = \arg\max_{x \in \Omega} \prod_{t \in L} f(d_t|x_t)p(x_t|x_{\partial t}) \tag{9}$$

It is not computationally feasible to find an optimum solution of Eq.(9). However, there are three well-know algorithms for local optimum solutions: simulated annealing (SA) [8], maximizer of posterior marginal (MPM) [7] and Iterated Conditional Modes (ICM) [9].

Fig. 2. Original(left) and the watermarked $lena$, PSNR=38.8

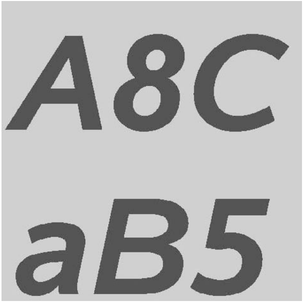

Fig. 3. The original logo

We adopted ICM due to its simplicity and reasonable performance. ICM is a deterministic algorithm and iteratively updates the current decoding $\hat{x}_t$ at pixel t on the basis of the observation $d - map$ and the decoding of its neighbors. The locality property of MRF has led to the updating rule for the ICM algorithm

$$\hat{x}_t = \arg\max_{x_t} f(d_t|x_t)p(x_t|x_{\partial t}) \tag{10}$$

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Experimental setup and results

In our experiments, we set the radius for the sliding window r=2, quantization step Δ=10, $\beta = 1$ and σ is set to the half of the quantization step Δ. ICM was initialized by minimum distance (MD) decoding. In most cases, ICM converged within 6 iterations. Notice that MD decoding is equivalent to maximum likelihood (ML) decoding (i.e. without exploiting any prior information in the MAP decoding).

Experiments were repeatedly carried out on 25 512×512 images with 10 different logos. Due to the limited space we only show the results of embedding a logo (as shown in Fig.3) that is composed of characters into the well-known image *lena* (as shown in Fig.2). It has to be pointed out that the PSNR of the watermarked image mainly depends on the quantization step Δ and is about $38.8db$ at $\Delta = 10$.

To demonstrate the robustness of the proposed SW-LAQIM against JPEG compression, we embedded the logo into the image using both conventional QIM (one bit per pixel as well) and SW-LAQIM. Both watermarked images were gone through JPEG compression at 50% quality level. Then we employed MD and MAP decoding to extract the logo from the compressed watermarked images. Fig.4 shows the decoded logo from the watermarked image using conventional QIM and Fig.5 shows the decoded logo from watermarked image using SW-LAQIM. It is obvious that SW-LAQIM improves the robustness against JPEG compression compared to conventional QIM and MAP decoding brings substantial further improvement. The scheme of SW-LAQIM plus MAP decoding produces clearly identifiable logo.

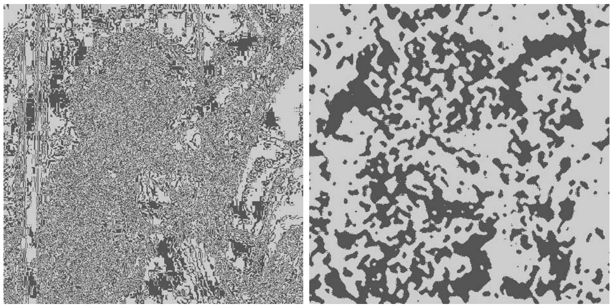

Fig. 4. MD(left) and MAP(right) decoded logo on JPEG 50% compressed QIM watermarked $lena$

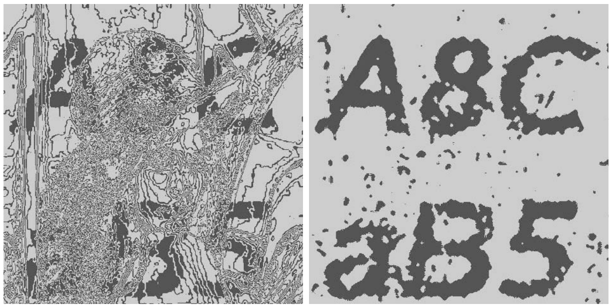

Fig. 5. MD(left) and MAP(right) decoded logos on JPEG 50% compressed SW-LAQIM watermarked $lena$

Fig.6 plots the error-decoding rates of MAP and MD decoding against the JPEG compression from quality level 90% to 20% on a SW-LAQIM watermarked $lena$. The rates were averaged over 10 logos. At all compression quality levels, MAP decoding consistently outperforms the MD decoding. The minimum error rate of MD decoding is above 30% at quality level of 80% and the error rate of MAP decoding is only around 5% at the same quality level. Experiments on other images presented similar results.

Fig.7(left) shows the MAP decoded logos on half-sized SW-LAQIM watermarked $lena$. We also further scale the JPEG 50% compressed SW-LAQIM watermarked $lena$ to 90% of its original size, rotate (10 degree counter-clockwise), and cropped to an image of size 384×384. The decoded logo is shown in Fig.7(right). Although the logo is more distorted and incomplete, it is still clearly identifiable.

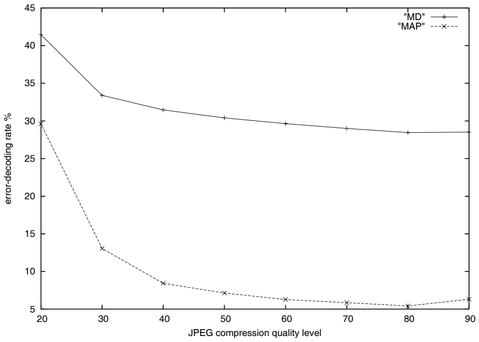

Fig. 6. MAP decoding is more robust against JPEG than the Minimum Distance decoding

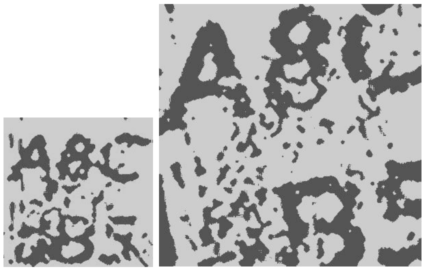

Fig. 7. MAP decoded logos: on watermarked image scaled-down to half of its original size(left); and on JPEG 50% compressed, cropped, rotated and scaled watermarked $lena$ (right)

3.2. Discussion

Watermarked images in practice often go through some noise channels before arriving at the decoder. The noise can usually be modelled as either Gaussian or uniform noise. The proposed system is robust against such noise channels. [5] proved that applying QIM on the local averages (LAQIM) results in very high robustness against white Gaussian and uniform noise, and showed that the embedding was also robust against the JPEG compression. This property still holds for SW-LAQIM in most cases except in the busy areas of the host image and at the sites where the bit change occurs on the logo. However, most embedded bits will be recovered by the MAP decoding as shown in the experimental results.

The proposed system is robust against rotation due to the pixel based embedding and adoption of disk-shape windows. When cropping happens to the watermarked image, at the sites where the cropping occurs, error decoding is likely to happen because the integrity of the sliding windows for those pixels has been damaged. However, the integrity of the embedded logo is kept if the main integrity of the host image is kept. The error will be corrected by the MAP decoding and thus will not affect the justification of the copyright claims.

The watermark is also robust against the scaling attack. When the host image is scaled by factor α, the decoder could set up the disk-shape sliding window **G** with radius $r'=\alpha\, r$. Consequently, μ'_s is same or very close to the original one. The decoding is not compromised due to the precise extraction of μ'_s.

4. CONCLUSION

A pixel-based watermarking system has a great advantage of surviving geometric attacks since the synchronization of the watermark embedding and decoding is always held. However, a pixel-based system is usually too fragile for practical application: even the high quality JPEG compression can remove or disrupt the watermark. Our proposed system overcomes this problem by utilizing a sliding window based QIM embedding together with MAP decoding. Experimental results have demonstrated that MAP decoding performs substantially better than the Minimum Distance decoding against JPEG compression and the proposed system presents strong robustness against JPEG compression and geometric attacks as well.

5. REFERENCES

[1] I.J. Cox, M.L. Miller, and J.A. Bloom, “Digital watermarking,” *Morgan Kaufmann Publishers*, 2002.

[2] J.L. Dugelay and F.A.P. Peticolas, “Possible counter-attacks against random geometric distortions,” in *Proceedings of the SPIE Conference on Security and Watermarking of Multimedia Content II*, 2000, vol. 3971, pp. 338–345.

[3] B. Chen, *Design and analysis of digital watermarking, information embedding, and data hiding systems*, Phd. dissertation, MIT, Cambridge, MA, June 2000.

[4] B. Chen and G.W. Wornell, “Quantization index modulation: a class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. Information Theory*, vol. 47, pp. 1423–1433, May 2001.

[5] W. Lu, W. Li, R. Safavi-Naini, and P. Ogunbona, “A new qim-based image watermarking method and system,” *2005 Asia-Pacific Workshop on Visual Information Processing, Hong Kong*, pp. 160–164, 2005.

[6] C.S. Won and R.M. Gray, “Stochastic image processing,” *Kluwer Academic/Plenum Publishers*, 2004.

[7] R.C. Dubes and A.K. Jain, “Random field models in image analysis,” *Journal of Applied Statistics*, vol. 16, no. 2, pp. 131–163, 1989.

[8] S. Geman and D. Geman, “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, 1984.

[9] J. Besag, “On the statistical analysis of dirty pictures,” *J. R. Statist. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.