Frequency-based Watermarking with Spatial Masking for DRM of MMS Content

A. Santos*, L. Moura e Silva, P. Martins, P. Carvalho
Centre for Informatics and Systems, University of Coimbra
Pólo II, Pinhal de Marrocos, 3030 Coimbra, Portugal
Email: {amancio, luis, pjmm, carvalho}@dei.uc.pt

ABSTRACT
Copyright enforcement is the major constraint to the development of mobile content distribution business. In this paper a low computational complexity watermarking scheme is presented. The algorithm exhibits resilience to the major expected attacks in the MMS context using a canonical insertion procedure in the Fourier domain. Watermark transparency is provided based on spatial texture masking through a constrained frequency perturbation maximization procedure. This watermarking scheme is the core protection mechanism of a Proxy-based DRM solution for MMS, currently under trial with a mobile operator.

KEY WORDS
Watermarking, Digital Rights Management, Telecommunications Applications

1 Introduction

Multimedia Messaging Service (MMS) has opened new business opportunities for content providers to distribute their work in the mobile market. However, high-branded content-providers will not feel motivated to enter and to fully explore this business opportunity if mobile operators lack to deliver a DRM solution to prevent the free distribution of their premium contents among subscribers. One solution might be to restrict access to data using some encryption technique or some private data format. However, these solutions exhibit some drawbacks: (i) once the encrypted data are decrypted, they can be freely distributed and manipulated and therefore encryption does not provide a satisfactory mean of protection; (ii) both encryption and private data format must rely on specific data viewers, leading to major software update constraints. Large scale software updates are very difficult to perform, mainly when they are related to features most users do not explicitly demand or even wish to have installed in their equipments, such as copyright protection agents. It is therefore desirable that copyright information should be hidden in the multimedia data to enable transparent Digital Rights Management (DRM) implementation.

*Acknowledgements: This work was partially financed by AdI (Agência de Inovação) and POSI Program, supported by the Portuguese Government and the European Union.

In this paper we introduce a watermarking scheme designed for DRM in MMS applications. The algorithm is the core protection scheme of a Proxy-based DRM solution, currently being trialed with a mobile operator. This Proxy-based DRM solution has some advantages: (i) its installation does neither require the acceptance of the MMC vendor nor the handset manufacturer; (ii) it does not require any upgrades of the existing MMS Value Added Service (VAS) Applications, since it uses an on-the-fly content protection for all parties involved on the Multimedia Messaging Service Environment (MMSE); (iii) it is mobile phone independent, since it does not require any installation of a DRM agent on the MMS terminal, and, therefore, it guarantees the compatibility with all existing and future MMS terminals. However, this solution implies the installation of a new server in the MMSE, which will have some operational impact, and it does not prevent to illegally copy the purchased contents via e-mail, iRDA or Bluetooth.

Regardless the methodology and application, watermarking algorithms should provide some common desirable properties, being the most important [1]: (i) the perceptual invisibility (often called transparency), i.e., the modifications caused by watermark embedding should not degrade the perceived data quality, and (ii) robustness, that is, watermarks must exhibit robustness to attacks since digital data can undergo a great deal of different modifications that deliberately (piracy attacks) or not (compression, filtering, resizing, color depth reduction) affect the embedded watermark. Ideally, a watermark should be detectable up to the point that the host data quality remains within acceptable limits. Another important aspect of watermarking techniques, that is directly related to their usability in MMS, is blind detection, i.e., detection without access to the original unwatermarked media.

Besides the aforementioned characteristics there are some additional challenges to watermarking solutions for DRM in MMS. Namely, it is desirable to code a reasonably large number of bits into the watermark, to allow computationally affordable provider and content identification. Further, the image size in MMS messages is much smaller than internet content and, therefore, it is much harder to hide transparently a large message into the content itself. On the other hand, the DRM solution has to exhibit low computational complexity during watermarking extraction

to minimize performance impacts in the MMS trajectory.

From the above description it is observed that highly robust watermarking strategies based on statistical measures, such as correlation [2] or the expectation operation [3], are not appropriate for DRM purposes in MMS, since: (i) they often require the unmarked cover media; (ii) they require one statistical test to be performed per possible watermark (zero-bit watermark), i.e., possible copyright owner and content; and finally, (iii) due to the limited size of the data in MMS, statistical testing is highly unreliable.

When thinking about the possible data transformations in the MMS context the most probable attacks are those introduced by the system itself, both due to the limited communication bandwidth and the limited visualization capabilities of terminal equipments. To prevent large MMS messages, MMC servers usually limit the maximum size of each message. Each image, larger than a predefined value, is automatically and proportionally resized by the MMC gateway. Other resize operations may be introduced by terminal equipment, either automatically or on user demand, during storage and message forward operations. Terminal equipments may also introduce other types of data distortion such as compression (usually a small compression is applied) and color depth reduction.

The most damaging attacks under the described scenario are the geometrical resize operations, which may induce considerable smoothing and aliasing, depending on the interpolation method implemented during resampling. Some watermarking methods that are resilient to geometrical attacks were reported in recent papers. These methods can be divided into three categories: (i) template embedding based methods, (ii) invariant transform based methods and (ii) invariant features based methods. In template embedding based methods a known synchronization template is embedded into the image along the watermark [4]. Other template embedding methods rely on autocorrelation peaks detection due to watermark replication [5]. These methods tend to be computationally very expensive while reducing the image fidelity as well as the watermark capacity. Linear geometrical image transformation resilience has also been designed using the Fourier-Mellin transform [6] which, however, suffers from sever implementation difficulties both due its computational complexity and due to the required unstable log-polar mapping. Kim et al. [7] and Lin et al. [8] introduced other invariant methods with respect to linear geometrical distortions based on the Radon transform. This method is for zero-bit watermarking and it is not straightforwardly extensible for multi-bit watermarking. Finally, feature based methods watermarking methods [9][10] rely on the extraction of invariant image features, which tend to be computationally very demanding.

In this paper we propose a low complexity watermarking scheme for MMS content DRM applications, which exhibits robustness to the most probable content attacks. To avoid multiple watermark detection, a message (typically the copyright owner and content identifiers) is transmitted within the multimedia data. A noise visibility function (NVF) for texture masking enables to transparently embed the encoded watermark into the cover data. Due to the scaling property of the Fourier Transform, this task is performed in the frequency domain in perceptual relevant regions and using local features. Resilience to scale changes is obtained by inserting the watermark using a canonical scale. At detection, the image is scaled back to this canonical scale. To avoid cut-off frequencies of the interpolation filters, a set of possible frequencies, depending on the original image size, are applied to embed and to extract the watermark.

The remaining paper is organized as follows: section 2 describes the watermarking strategy devised for MMS content. Finally, in section 3 some major performance results and conclusions are presented.

2 The Watermarking Algorithm

Let $f(k,l)$, $0 \leqslant k \leqslant M$, $0 \leqslant l \leqslant N$, be the image where the watermark is to be inserted into and let $F(u,v)$ be the Fourier transform of $f(k,l)$. Further, let $m = \{m_1, ..., m_L\}$, $m_i \in \{0,1\}$, be the message of L bits to be embedded into $f(k,l)$. In our current system the binary message is composed of 32 bits for provider and content identification and is augmented with an 8 bit CRC. This message is further coded using error correction codes (turbo-codes $\frac{1}{2}$ code rate). Finally, a 12 bit pattern is appended at the beginning of the message for watermark presence identification, leading to a message of a total length of $L = 96$ bits. In our proposed watermarking method the message m is always embedded into $f(k,l)$ in the magnitude of the Fourier frequency domain using a canonically resized version of $f(k,l)$, denoted as $\widetilde{f}(k,l)$, $0 \leqslant k,l \leqslant S_C$. Each bit of m is inserted in two distinct frequencies (u,v) and $(-u,-v)$ to keep the function real valued. Once the DFT watermark has been computed and added to the image, the watermarked image is obtained from the inverse DFT. Finally the watermarked image is resized to its original size.

2.1 Watermark Embedding

Geometrical image transformation implies an interpolation step where the two-dimensional (2-D) continuous signal $f(x,y)$, $f,x,y \in R$, must be reconstructed from its discrete samples $f(k,l)$ using the spatial convolution of a finite approximation of the ideal IIR 2-D $sinc(x,y)$ interpolation filter. Typical interpolators are nearest neighbor, bilinear and bicubic interpolators. These interpolators are far from optimal, introducing blurring due to rapid fall off in their passbands as well as aliasing due to higher cut-off frequencies than the ideal ones [11]. Other frequency smearing effects stem from round-off operations due to the discrete nature of image amplitudes. Most magnitude watermarking schemes rely on high threshold values

to avoid these effects. Unfortunately, in practice, these coding schemes impose sever constraints on image fidelity as well as the watermark capacity. In our coding scheme, local features at each frequency extracted from a local window $W(u,v)$ of size $(2t+1)\times(2t+1)$ are applied to code each bit. The advantage of this insertion strategy is that any signal processing operation tends to induce similar effects on all neighboring frequencies in $W(u,v)$, enhancing, therefore, the robustness of the scheme. Further, smaller magnitude perturbations of $\widetilde{F}(u,v)$, the DFT transform of $\widetilde{f}(k,l)$, can be applied to code each message bit, which leads to better image fidelity and higher capacity. There are several features, such as statistical moments, that can be applied for this purpose. In our present implementation the median magnitude value of $\widetilde{F}(u,v)\in W(u,v)$ is used to code a 1 or a 0. Namely, let $U\equiv\{(u_i,v_i):i=1,..,L\}$ be the set of selected frequencies where message m is to be inserted. For each message bit m_i the median magnitude value $m(u,v)$ of $\widetilde{F}(u,v)$, $(u,v)\in W(u,v)\backslash U$, is computed, i.e.,

$$m(u,v)=\underset{\substack{w_1=u\pm k;w_2=v\pm k\\ k=1,...,t;(w_1,w_2)\notin U}}{median}\left\{\widetilde{F}(w_1,w_2)\right\} \tag{1}$$

To insert message bit m_i the magnitude of $\widetilde{F}(u_i,v_i)$ is changed such that

$$F_w(u_i,v_i)=F_w(-u_i,-v_i)= \tag{2}$$
$$=\begin{cases} m(u_i,v_i)\cdot(1-\delta_i) & \Leftarrow m_i=0\\ m(u_i,v_i)\cdot(1+\delta_i) & \Leftarrow m_i=1\end{cases},\delta_i\in\,]0,1]$$

From equation (2) it is observed that in order to maximize robustness, during embedding, δ_i should be increased as much as possible. However, δ_i has to be limited to avoid watermark visibility. To this end a similar mathematical approach as described in [2] is applied. Namely, the perturbation amplitudes δ_i are computed from the linear constrained optimization problem in (3).

$$\max\sum_{i=1}^{L}\delta_i;G\delta=0;A\delta\leqslant b,\delta\equiv[\delta_i\cdots\delta_L]^T \tag{3}$$

Let IDFT be the inverse discrete Fourier transform, $F_w(u,v)$ be the DFT of the watermarked image, and Δp be the just noticeable pixel perturbation. In (3) matrix A and vector b code the following constraints:

Noise visibility:

$$\left|IDFT\{F_w(u,v)\}-\widetilde{f}(k,l)\right|\leqslant\Delta p(k,l) \tag{4}$$

Image amplitude limits:

$$0\leqslant IDFT\{F_w(u,v)\}\leqslant 255 \tag{5}$$

Variation limits:

$$0\leqslant\delta_i\leqslant 1 \tag{6}$$

From (4) and (5) it is observed that, depending on the canonical image size S_C, the optimization problem formulated in (3) may exhibit a very large number of constraints. To reduce the computational complexity, only a subset of P image points of coordinates$\{(k_i,l_i),i=1,...,P\}$, which exhibit the most restrictive constraints in $\Delta p(k,l)$, are applied in equations (4) and (5).

Using the symmetry property of the Fourier transform of real valued functions and assuming that all frequencies for embedding are selected such that $u_i\neq 0$ and $v_i\neq 0$, it can be shown that matrix A and vector b can be defined as in (7) and (8), respectively, where the intermediate coefficients are defined in (9) through (13).

$$A\equiv\begin{bmatrix} C\times diag(sign(m_i)\cdot m(u_i,v_i))\\ -C\times diag(sign(m_i)\cdot m(u_i,v_i))\\ I\\ -I\end{bmatrix} \tag{7}$$

$$b\equiv\begin{bmatrix} z\\ q\\ 1\\ 0\end{bmatrix},sign(m_i)=\begin{cases}-1 & \Leftarrow m_i=0\\ 1 & \Leftarrow m_i=1\end{cases} \tag{8}$$

$$C\equiv[C_{i,j}]: \tag{9}$$
$$C_{i,j}=\frac{2}{S_C^2}\cos\left(\frac{2\pi}{S_C}u_jk_i+v_jl_i+\alpha_{j,j}\right)$$
$$1\leqslant i\leqslant P;1\leqslant j\leqslant L$$

$$\alpha_{j,j}=\angle\widetilde{F}(u_j,v_j) \tag{10}$$

$$z\equiv[z_i]: \tag{11}$$
$$z_i=\min\{255-g(k_i,l_i),\Delta p(k_i,l_i)+g(k_i,l_i)\}$$
$$1\leqslant i\leqslant P$$

$$q\equiv[q_i]:q_i=-g(k_i,l_i),1\leqslant i\leqslant P \tag{12}$$

$$g(k,l)\equiv\tilde{f}(k,l)+IDFT(m(u,v)) \tag{13}$$

At each pixel position the just noticeable pixel perturbation $\Delta p(k,l)$ is computed from the noise visibility function as introduced in [2] according to equation (14). This function is based on a Stationary Generalized Gaussian image model with an auto-covariance function $R_x=\sigma_xI$.

$$\Delta p(k,l)\equiv(1-NVF(k,l))\cdot S+NVF(k,l)\cdot S$$
$$S,S_1\in\mathbb{R}_0^+ \tag{14}$$

$$NVF(k,l)=\frac{w(k,l)}{w(k,l)+\sigma_x^2} \tag{15}$$

In (15) $w(k,l) = \gamma[\eta(\gamma)]^{\gamma} \frac{1}{\|r(k,l)\|^{2-\gamma}}$, $r(k,l) = \frac{\widetilde{f}(k,l)-\widetilde{f}(k,l)}{\sigma_x}$, $\widetilde{f}(k,l)$ is the non-stationary local mean of $\widetilde{f}(k,l)$, $\eta(\gamma) = \sqrt{\frac{\Gamma(\frac{3}{\gamma})}{\Gamma(\frac{1}{\gamma})}}$ and $\Gamma(x) = \int_0^{\infty} e^{-u}u^{x-1}du$ is the gamma function. The shape parameter γ is computed using the moment matching method described in [12], while S and S_1 control the maximum image perturbation in textured and untextured image regions.

It is well known that the lower the frequencies the higher are the perceivable influences of perturbations in their magnitude. Therefore, if no relation between perturbations δ_i are imposed, the solution of (3) will tend to concentrate most of the magnitude perturbations into higher frequencies. To avoid this effect, we model the magnitude distribution of $\widetilde{F}(u,v)$ using Gaussians and normalize perturbations δ_i according to the variance $\sigma(u,v)$ of $\widetilde{F}(u,v)$ in the direction of the frequency where the perturbation will be inserted, therefore, matrix G of equation (3) is defined according to equation (16).

$$G \equiv \begin{bmatrix} h_1 & -h_2 & 0 & \cdots & 0 & 0 \\ 0 & h_2 & -h_3 & \cdots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & h_{L-1} & -h_L \end{bmatrix}$$
$$h_i = \frac{3\sigma(u_i,v_i)}{\|(u_i,v_i)\|} \quad (16)$$

In our current implementation the set of frequencies U applied to embed the watermark is selected according to the original image size, the expected lossy compression rate and the expected minimum image dimension for the resize operation.

2.2 Watermark Extraction

For watermark extraction, after resizing the image to its canonical size $S_C \times S_C$ and once its DFT has been computed, each of these possible sets of frequencies is inspected to extract the watermark. In particular, for a given set U each bit of the embedded message is extracted according to (17), where $m(u,v)$ is calculated as described in equation (1).

$$m_i = \begin{cases} 0 \Leftarrow \widetilde{F}(u_i,v_i) \leq m(u_i,v_i) \\ 1 \Leftarrow \widetilde{F}(u_i,v_i) > m(u_i,v_i) \end{cases} \quad (17)$$
$$(u_i,v_i) \in U;\ i = 1,...,L$$

To validate the watermark, both the inserted CRC (after decoding using the turbo-codes) and the watermark identification bit pattern are applied.

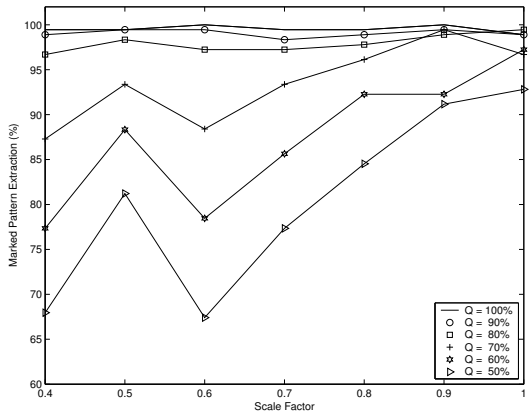

Figure 1. Watermark identification pattern extraction results for combined jpeg-compression and scaling attacks.

3 Results and conclusions

In figures 1 and 2 some statistical results obtained with the described watermarking strategy using 183 outdoor images obtained from the internet are shown. Figure 3 depicts the watermark result of a typical content. The canonical size during the watermark embedding and extracting phases was assumed to be 200×200. To test the robustness of the introduced watermarking scheme a set of 30 combined scale factors (from 1 to 0.4) and jpeg-compression (from a quality factor of 100% to 40%) attacks were performed for each image in the test bed. For each test two types of statistics were performed: correct message extraction and correct watermark identification pattern extraction. As can be concluded from the obtained results, the described scheme is robust to geometrical resize operations even when combined with jpeg-compression. For low compression rates (the most probable case under MMS) the extraction probability of the embedded message is between 80% ($Q = 90\%$) and 99% ($Q = 100\%$) for the worst scaling case.

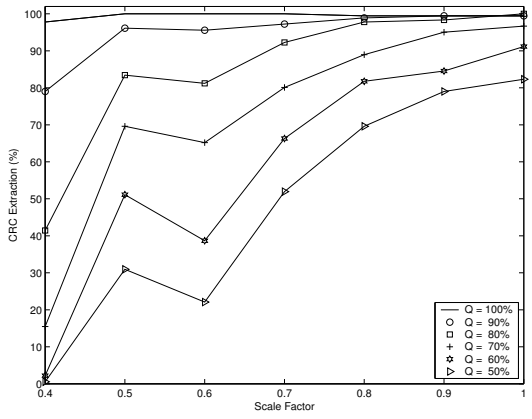

Figure 2. Watermark message extraction results for combined jpeg-compression and scaling attacks.

Figure 3. Watermark result of tipical commercial content. (left) Unwatermarked image. (right) Watermarked image.

Concerning the extraction of the watermark identification pattern, which may serve as a zero-bit watermark, it is observed that the algorithm exhibits above 98% of correct extraction probability for low jpeg-compression. This is a significant result, since, according to [7], the Digimarc watermarking algorithm is able to recover the watermark in about 72% of the situations (although under a different and unknown test). For jpeg-compression attacks ($Q = 50\%$) the algorithm exhibits an extraction probability of around 80% for CRC and 93% for the watermark identification pattern. Furthermore, as can be observed in figure 3, the described scheme is able to preserve the perceivable image quality. These tests were performed with a 2GHz Pentium IV computer with Windows XP running Matlab. Using this system, for the largest images in the test bed each extraction operation took an average of 1.1 seconds.

Major challenges for watermark based DRM in MMS applications are the reduced size of contents, computational efficiency and robustness to geometrical resize operations. From the results presented in figures 1 and 2 it can be concluded that the described watermarking strategy is robust to this type of attacks. Regarding color depth attacks, this watermarking method is almost invariant, since the watermark is embedded into the image's intensity and not into its chromatic channels. In fact some preliminary tests with true color images indicate that the algorithm is not sensible to a 256 indexed color depth reduction.

The results in figures 1 and 2 suggest that the degradation in the extraction performance for small scaling factors combined with large lossy compression rates is mainly induced by the smearing effects introduced by the interpolation filter. Namely, for a scaling factor of 0.5, where interpolation is equivalent to drop half of the image's columns and lines, it is observed that, regardless the compression ratio, extraction results are always better than for neighboring factors, i.e., for scaling factors 0.4 and 0.6. This behavior suggests that the algorithm's performance may be largely improved, if proper frequency magnitude attenuation effects induced by the low-pass interpolation filter are compensated during the watermark extraction phase. Another research direction will be the adaptation of this watermarking strategy to video. A possible strategy is to explore time redundancy in video to further improve the algorithms performance.

References

[1] Voloshynovskiy, F. Deguillaume, and T. Pun, "Content adaptive watermarking based on stochastic multiresolution image modelling," in *10th European Signal Processing Conference*, 2000.

[2] S. Pereira, *Robust Digital Image Watermarking*, Ph.D. thesis, Université de Gèneve, Switzerland, 2000.

[3] W. Zeng and B. Liu, "A statistical watermark detection technique without using original images for resolving rightful ownwerships of digital images," *IEEE Trans. On Image Processing*, 1999.

[4] S. Pereira and T. Pun, "Template matching for affine resistant image watermarks," *IEEE Trans. On Image Processing*, vol. 9(6), pp. 1123–1129, 2000.

[5] M. Kutter, "Watermarking resisting to translation, rotation and scaling," in *Proc. SPIE Multimedia Systems Applications*, 1998, pp. 423–431.

[6] J. O'Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital watermarking," *Signal Processing*, vol. 66, pp. 303–317, 1998.

[7] H.-S. Kim, Y. Baek, and H. K. Lee, "Rotation, scale, and translation invariant image watermark using high order spectra," *SPIE (to be published)*, 2003.

[8] C.-Y. Lin, M. Wu, J. Bloom, I. Cox, M. Miller, and Y. Lui, "Rotation, scale and translation resilient watermarking for images," *IEEE Trans. On Image Processing*, vol. 19(5), pp. 767–782, 2001.

[9] M. Kutter, S. Bhattacharjeea, and T. Ebrahim, "Towards second generation watermarking schemes," in *Proc. IEEE Int. Conf. Image Processing*, 1999, pp. 320–323.

[10] S. Guoxiang and W. Weiwei, "Image feature based second generation watermarking in wavelet domain," *Lecture Notes in Computer Science*, vol. 2251, pp. 16–21, 2001.

[11] T. Lehmann, C. Gönner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Trans. on Medical Imaging*, vol. 18(11), pp. 1049–1075, 1999.

[12] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. PAMI*, vol. 11, pp. 674–69, 1989.