ADDING ROBUSTNESS TO GEOMETRICAL ATTACKS TO A WAVELET BASED, BLIND VIDEO WATERMARKING SYSTEM

C.V. Serdean, M.A. Ambroze*, M. Tomlinson* and J.G. Wade***

*Department of Communication & Electronic Engineering, University of Plymouth,
Plymouth, PL4 8AA, UK
**Department of Electrical & Computer Engineering, University of Newcastle,
Callaghan, NSW 2308, Australia

ABSTRACT

This paper describes a high capacity blind video watermarking system invariant to geometrical attacks such as shift, rotation, scaling and cropping. A spatial domain reference watermark is used to obtain invariance to geometric attacks by employing image registration techniques to determine and invert the attacks. A second, high capacity watermark, which carries the data payload, is embedded in the wavelet domain according to a human visual system (HVS) model. This is protected by a state-of-the-art error correction code (Turbo code). For a false detection probability of 10^{-8}, the proposed system is invariant to scaling up to 180%, rotation up to 70^{0}, and arbitrary aspect ratio changes up to 200% on both axes. Furthermore, the system is virtually invariant to any shifting, cropping, or combined shifting and cropping attack, and it is robust to MPEG2 compression as low as 2-3Mbps.

1. INTRODUCTION

One of the most difficult problems in digital video watermarking is watermark recovery in the presence of geometric attacks like frame shift, cropping, scaling, rotation, and change of aspect ratio, especially when some of these are combined together. The work presented in this paper was carried out in the context of uncompressed video (ITU-R 601 as found in TV studios), and so geometric attacks tend to be less severe compared to those for image watermarking [1]. On the other hand, the recovery problem is compounded for video since it must be carried out blind due to the difficulty of storing the original. In this case, for the typical spread spectrum (SS) watermarking system, blind retrieval is performed via cross-correlation between the marked video and the secret pseudo-noise (PN) sequence used to spread the watermark at the embedding stage. Recovery is straightforward given perfect synchronisation between the attacked video and the PN sequence, but is difficult when geometric attacks destroy the synchronisation. In this case it is possible to perform some form of sliding correlation in order to re-establish synchronisation i.e. multiple cross-correlations over a specified search space. Unfortunately the search space grows very quickly, making it difficult to recover the watermark in a reasonable time. Clearly, given that retrieval in a video context must be done in near real time, the computational problem is very significant in the presence of attacks. One way to overcome this is to use image registration techniques for resynchronisation.

This paper combines the advantages of an algorithm based on the Fourier-Mellin transform (FMT) image registration techniques [2], with watermarking in the Discrete Wavelet Transform (DWT) domain. The idea is to first undo geometric attacks using the FMT approach and an additional spatial reference watermark used only for registration purposes. Once the attack parameters are determined, the geometric attacks are undone and the resulting frame is passed to the main watermark decoder where the embedded data bits are recovered.

The main watermark, which carries multi-bit data, is inserted in the DWT domain and capacity is maximised by embedding based on a HVS model. The complete system can be regarded as a noisy communications channel and so is protected by turbo coding. The net result is a system that can withstand severe geometric attack, the limiting attack being defined by a threshold yielding a false detection probability of 10^{-8}, and capacity being defined by a BER of 10^{-8}.

2. COMBATING GEOMETRIC ATTACK USING LOG-POLAR AND LOG-LOG TRANSFORMATIONS

When registering two images, the noise is relatively small, and so the correlator usually performs very well. The problem is more difficult for video watermarking since the original video frame is not available. To overcome this, we use a SS reference watermark. The PN sequence used to embed this watermark corresponds to the "original image" and the watermark embedded in the unsynchronised marked video corresponds to a noisy "attacked image" (the video itself represents the noise).

In the proposed system we embed two different watermarks. The first is a 1-bit watermark used only for geometric reference, and is embedded in the spatial domain. The second, multi-bit watermark is used for the data payload, and is embedded in the DWT domain.

The desired geometric invariance can be achieved by using the FMT to convert rotation and scale to spatial shifts [2], which are then easily recovered by a Symmetrical Phase Only Matched Filter (SPOMF).

If frame f_2 is the scaled and rotated replica of frame f_1 with a scaling factor a and angle θ_0 , then:

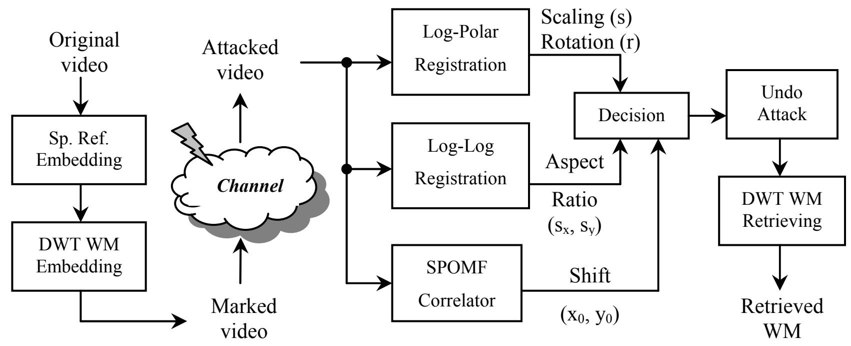


Figure 1. Block schematic of the geometric invariant video watermarking system

$$\begin{aligned}
&f_2(x,y) = \\
&\quad f_1\left(a\left(x\cos\theta_0 + y\sin\theta_0\right), a\left(-x\sin\theta_0 + y\cos\theta_0\right)\right) \\
&F_2(u,v) = \\
&\quad \frac{1}{a^2}F_1\left(\left(u\cos\theta_0 + v\sin\theta_0\right)/a, \left(-u\sin\theta_0 + v\cos\theta_0\right)/a\right)
\end{aligned} \tag{1}$$

In order to convert both scaling and rotation to shifts, it is necessary to convert the Cartesian coordinates into log-polar coordinates:

$$\begin{aligned} x &= e^{\log\rho}\cos\theta \\ y &= e^{\log\rho}\sin\theta \end{aligned} \tag{2}$$

The result is:

$$F_2\left(\log\rho,\theta\right) = F_1\left(\log\rho - \log a, \theta - \theta_0\right) \tag{3}$$

where the scale and rotation factors can be retrieved by SPOMF correlation.

Since the FMT is not shift invariant, it is necessary to apply the Fourier magnitude of the frame (rather than the frame itself) to the input of the log-polar conversion module. The Fourier magnitude is shift invariant and so the rotation and scaling parameters can be found even in the presence of shift. After undoing rotation and scaling, the shift is then recovered by performing a simple SPOMF correlation. This technique works well for image-image registration, since the correlation peaks are relatively large and the phase loss can be tolerated. Unfortunately, for video watermarking, the loss can make cross-correlation unreliable, and this approach cannot be used for retrieval under combined attack.

A log-polar map permits recovery over a wide range of scale changes, rotation, or even combined scale-rotation attack. If a log-log map is used, then it is possible to recover arbitrary aspect ratio changes (different scale factors for x and y axes). The shifts alone are easily recovered using a SPOMF module. However, shift recovery from a combined attack (e.g. shift + scaling) requires a comprehensive search for all of the possible shifts, and is computationally intensive.

Fig. 1 shows a schematic of the proposed system. The decision block determines if the reference watermark is present (to within a desired false detection probability), and if present it automatically determines the attack parameters. In the proposed scheme, the two watermarks are embedded in different domains in order to minimise crosstalk, and each watermark is embedded at the full strength dictated by its own HVS model. The reference watermark is embedded in the spatial domain using SS, together with a simple visual model that inserts a stronger watermark in those regions where it is less easily observed (at edges and in high texture regions). The same reference watermark is embedded in all the frames in order to increase the SNR at the correlator input via frame averaging. As a result, the registration takes place only once, and not for each separate frame. This is possible because attacks must be identical for each frame in order to avoid temporal artefacts.

3. THE DWT VIDEO WATERMARKING SCHEME

The hierarchical property of the DWT offers the possibility of analysing a signal at different resolutions (levels) and orientations. This multiresolution analysis gives both space and frequency localisation, and different orientations extract different features of the frame, such as vertical, horizontal, and diagonal information. Generally speaking, edges and textures will be represented by large coefficients in the high frequency sub-bands, and they are well localised within the sub-band. In practice, wavelet analysis is performed using multilevel filter banks. Essentially this comprises a succession of filtering and sub-sampling operations and has been widely described in the literature. For watermarking, we selected the Antonini 7.9 wavelet, as being one of the best wavelets available for image compression [3, 4]. Watermarking in the DWT domain has many advantages compared with FFT or DCT marking [5]. In particular, the multiresolution property provides both local and global spatial support, it is compatible with the HVS, there are no blocking artefacts, and it has lower computational cost and better energy compaction properties than the FFT and DCT.

The information capacity of the channel is maximised by embedding the main payload according to an HVS model, and through the use of turbo coding. The embedding and retrieving of the watermark are shown in Fig. 2, where we use 3 levels of decomposition. The security of such a system relies in the secret watermarking key, K_1, and in order to improve the system's overall security we use an interleaver to provide a random distribution of the data bits within each sub-band. The interleaver uses a separate key, K_2.

The hierarchical nature of the DWT is exploited by inserting a self-contained watermark in each sub-band, i.e. all payload bits are inserted into each sub-band. The watermark is embedded using amplitude modulation:

$$C_i^M = \begin{cases} C_i + \alpha\underbrace{\frac{Q(\lambda,\theta)}{Q_{\min}}\cdot\frac{\left|C_i\right|}{mean\left(C_i\right)}}_{S}\cdot W_i, & \text{(details)} \\ if \quad S > 24, \quad then \quad S = 24. & \\ C_i + \alpha\frac{Q(\lambda,\theta)}{2}\cdot\frac{\left|C_i\right|}{mean\left(C_i\right)}\cdot W_i, & \text{(approximation)} \end{cases} \tag{4}$$

where Q_{min} is the minimum value from matrix Q, W_i is the watermark, C_i is the original wavelet coefficient and C_i^M is the

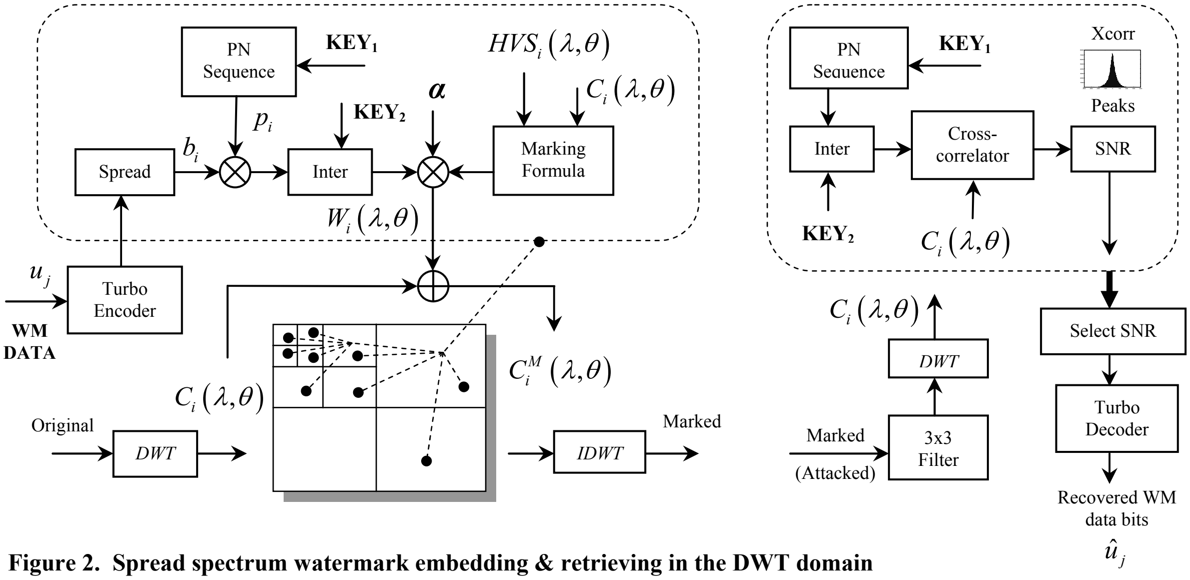


Figure 2. Spread spectrum watermark embedding & retrieving in the DWT domain

marked coefficient. Note that (4) incorporates media dependence, which is essential for robust watermarking. This marks more heavily the high frequency sub-bands and the largest coefficients, since modification of these coefficients is less likely to incur visible artefacts. The HVS is incorporated in the quantisation matrix, $Q(\lambda,\theta)$ where λ is the level and θ is the orientation. For computing $Q(\lambda,\theta)$ we use a visual model developed by Watson for the Antonini 7.9 DWT [3]. Although this is a much simpler HVS model than those usually used in DCT schemes and incorporates only limited information about the HVS (only the modulation transfer function of the eye - MTF), the overall performance of the system is actually better.

For retrieval, the video sequence is filtered using a Laplacian 3x3 filter prior to cross-correlation in order to improve the performance of the correlator.

It is advantageous to have a self-contained watermark in each sub-band, since a SNR can be determined for each sub-band as an indicator of sub-channel quality. Different types of attack affect different levels and orientations in different ways, and so it is always possible to select an optimal sub-band via SNR. Correlation is therefore performed separately for each sub-band, obtaining a set of cross-correlation peaks (one peak for each embedded data bit) for each sub-band. A SNR is then computed for each set of cross-correlation peaks, and retrieval is carried out for the sub-band with the highest SNR.

4. RESULTS

Fig. 3a,b,c shows how the system performs for different degrees of rotation and scaling, when *n* frames ($n \le 25$) are averaged in order to improve the robustness of the system. A threshold value of 0.025 can be observed in each figure. This guarantees a false detection probability better than 10^{-8} when the correlation peak exceeds the threshold (Fig. 3d).

For a false detection probability of 10^{-8}, the proposed system is invariant to scaling in the range –50% to 180%, invariant to rotation up to 70^0, and invariant to any arbitrary aspect ratio changes in the range –100% to 200% on both axes. When rotation is combined with scaling, up to 120% scaling and up to 20^0 rotation can be tolerated.

Furthermore, the system is virtually invariant to any shifting, cropping, or combined shifting and cropping. Even under severe cropping (when the useful image is only 200x200) the capacity is approximately 1500 bits/frame with turbo coding, reducing to 850 bits/frame without coding.

The system can cope very well with compression like JPEG and MPEG2. The results under MPEG2 compression are presented in Fig. 4. Combined attacks like MPEG2 compression plus arbitrary frame shifts can be handled as long as the MPEG2 compression is at least 3-4Mbps.

5. CONCLUSIONS

Robustness to geometric attack is one of the most important requirements for a watermarking system. To combat this type of attack, an approach based on the FMT and log-polar/log-log representations of the video frames has been developed. This is combined with the advantages of the DWT, HVS-based marking, and turbo coding to produce a very robust, high capacity video watermarking system. It outperforms many current schemes in terms of geometric invariance and channel capacity. For a wide range of attacks, the system presented in this paper meets and even exceeds the EBU watermarking recommendations [1].

6. REFERENCES

[1] L. Cheveau, E. Goray and R. Salmon, "EBU Technical Review – March 2001"

[2] B.S. Reddy and B.N. Chatterji, "An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration", *IEEE Trans. Image Processing,* Vol.5, No.8, August 1996.

[3] A.B. Watson, G.Y. Yang, J.A. Solomon, and J.D. Villasenor, "Visibility of Wavelet Quantization Noise", *IEEE Trans. Image Proc.*, Vol.6, 1997.

[4] J.D. Villasenor, B. Belzer and J. Liao, "Wavelet Filter Evaluation for Image Compression", *IEEE Trans. Image Proc.*, Vol.4, No.8, August 1995.

[5] C.V. Serdean, J.G. Wade, M.A. Ambroze, and M. Tomlinson, "Video Watermarking in the DWT domain", accepted to *Watermarking 2002*, Paris, France, 5-8 March 2002.

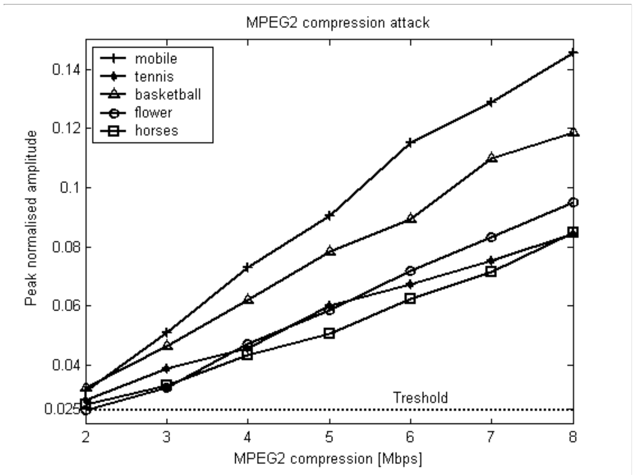
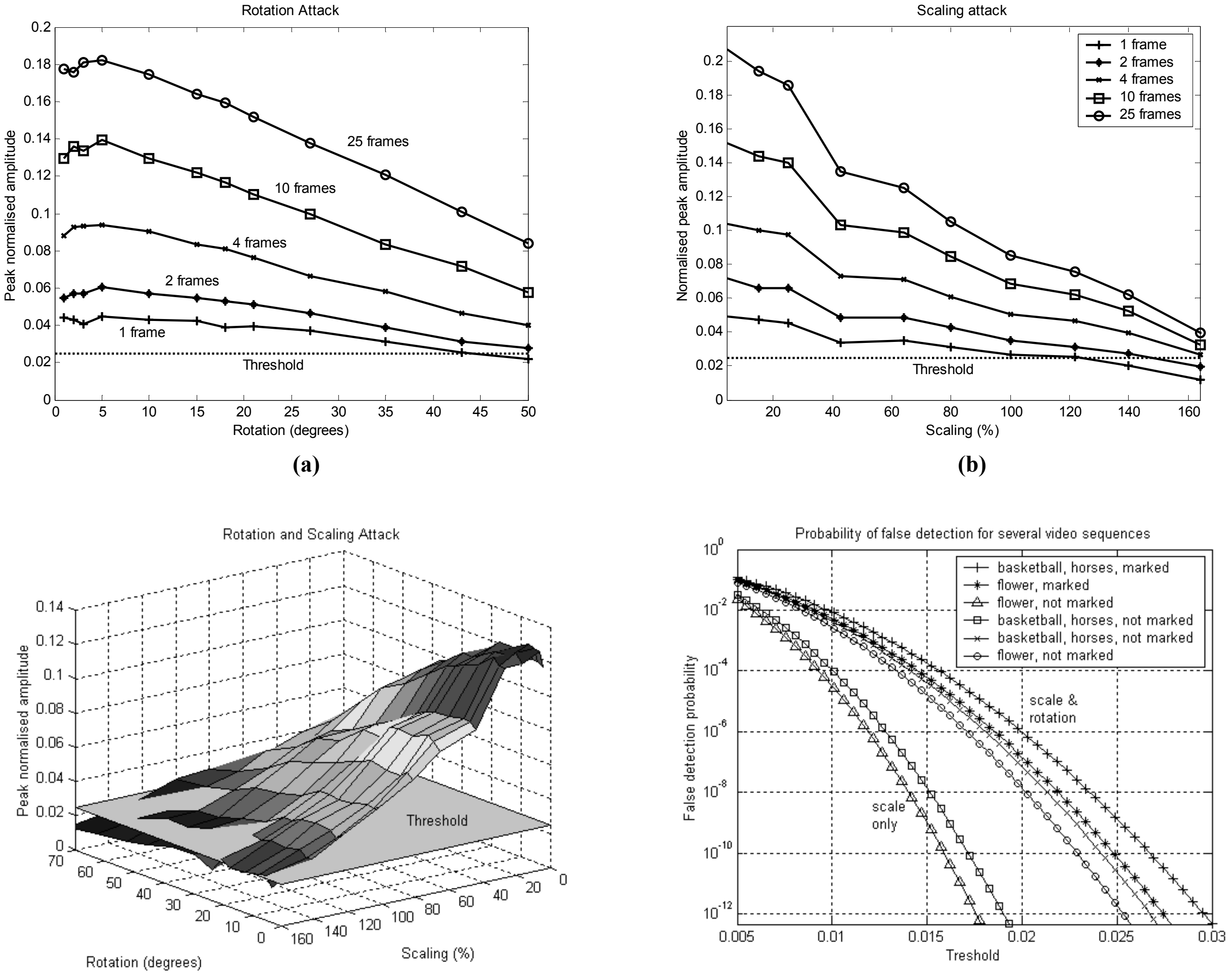


Figure 4. Performance of the system for MPEG2 attack



(a)

(b)

(c)

(d)

Figure 3. Performance of the system for (a) rotation and (b) scaling for basketball sequence when averaging frames; (c) Rotation combined with scaling attack (25 frames avg.); (d) Threshold selection for a desired probability of false detection.