DWT-based high-capacity blind video watermarking, invariant to geometrical attacks

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

Abstract: The 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 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°, 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. The system is also robust to MPEG2 compression, even when combined with shifting and cropping.

1 Introduction

One of the most difficult problems in digital video watermarking is watermark recovery in the presence of geometric attack. Typical attacks are frame shift (translation), cropping, scaling, rotation, and change of aspect ratio, and recovery is particularly difficult when these are combined together [1]. The work presented in this paper was carried out in the context of uncompressed video where 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, because it must be carried out blind, due to the difficulty of storing the original. In this case, for the typical spread spectrum 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.

A partial solution is to employ fast correlation. A symmetrical-phase-only matched filter (SPOMF), as proposed in [2–4], is used to solve the frame shift problem, leading to a shift-invariant watermarking system. Unfortunately, this technique can be applied only to frame shifts.

It is well known in image processing that transformation of Cartesian co-ordinates into log-polar co-ordinates prior to the FFT gives scale and rotation invariance [5, 6]. This is the Fourier–Mellin transform (FMT), and it was first used by O'Ruanaidh [7] to achieve invariance for image watermarking. Unfortunately, marking in the FMT domain has two major drawbacks: the need to compute the inverse log-polar transform (a lossy operation that drastically reduces system performance) and the need to maintain the FFT symmetry, which halves the watermark capacity. These drawbacks make the system impractical, although an improved technique was proposed by Lin [8].

This paper combines the advantages of an algorithm based on the FMT image registration techniques, with watermarking in the discrete-wavelet-transform (DWT) domain. The idea is to first undo geometric attack 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. The main watermark, which carries multi-bit data, is inserted in the DWT domain according to a human visual system (HVS) model. The system can be regarded as a noisy communications channel, and so it 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 bit error rate (BER) of 10^{-8}. It offers higher capacity and robustness compared to other watermarking systems described in the literature [1].

2 Combating geometric attack using log–polar and log–log transformation

Ideas developed by Casasent [5, 6] were later adopted for image processing in the context of image registration

IEE Proceedings online no. 20030159

DOI: 10.1049/ip-vis:20030159

Paper first received 12th December 2001 and in revised form 12th September 2002

C.V. Serdean was with the University of Plymouth and is now with the Electronics Department, University of Kent at Medway, Horsted Centre, Maidstone Road, Chatham, Kent ME5 9UQ, UK

M.A. Ambroze and M. Tomlinson are with the Department of Electronic and Communication Engineering, University of Plymouth, Plymouth, PL4 8AA, UK

J.G. Wade is with the Department of Electrical & Computer Engineering, University of Newcastle, Callaghan, NSW 2308, Australia

[2, 9, 10]. This involves two images: the original and an attacked copy, and the objective of image registration is to determine the parameters of the geometric distortion. The attack can then be inverted to give geometric alignment of two images. When registering two images, the noise is relatively small, and so the correlator usually performs very well.

However, this approach cannot be used for blind watermarking because the original video frame is not available. To overcome this problem, we suggest a 'blind registration' technique. The unavailability of the original is circumvented by using a spatial spread-spectrum reference watermark. In this case, the PN sequence used to embed the reference watermark plays the role of the 'original image', and the attacked watermarked video frame (watermark plus significant noise, arising from the video itself) represents the 'attacked image'. Therefore, for 'blind registration' the signal-to-noise ratio (SNR) is very low relative to that for image registration. In the proposed system, we embed two different watermarks. The first is a 1-bit watermark used exclusively for geometric reference, and for simplicity 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, which are then easily recovered by a SPOMF. Performing a log–log transform (LLT) of the input permits recovery from arbitrary scale changes. Similarly, a log–polar transform (LPT) converts rotation and scaling to spatial shifts, and permits recovery from rotation and scaling. The theory behind the LPT/LLT and its application to image registration is well represented in the literature [2, 5–7, 9, 10].

As 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 LPT/LLT module. The Fourier magnitude is shift-invariant, and so the attack parameters can be found even in the presence of shift. After undoing the attack, the shift is then recovered by performing a simple SPOMF correlation. This technique works well in the particular case of image–image registration [9], because the correlation peaks are relatively large and the phase loss can be tolerated. Unfortunately, this technique fails for 'blind registration' and, therefore, this approach cannot be used for retrieval under combined attack.

An LPT permits recovery over a wide range of scale changes, rotation, or even combined scale–rotation attack. If an LLT 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/rotation, or shift/aspect-ratio change) requires a comprehensive search, for all of the possible shifts [10], and is computationally intensive.

3 Robust system

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. One advantage of using two watermarks is now apparent: if the reference cannot be found, it is assumed that either the video is not marked, or that the mark is destroyed, and recovery of the main watermark payload is abandoned (saving computation time). Also, in the proposed scheme, the two watermarks are embedded in different domains, and each watermark is embedded at the full strength dictated by its own HVS model.

As stated, the reference watermark is used exclusively as a reference and can be regarded as just one-bit. This is embedded in the spatial domain using spread spectrum, together with a simple visual model that inserts a stronger watermark in those regions where it is less easily observed. The same reference watermark is embedded in all the frames in order to increase the robustness (the SNR at the correlator input is increased via frame averaging) and the detection speed of the algorithm (the registration process takes place only once, not for each separate frame). This is possible because attacks must be identical for each frame in order to avoid temporal artefacts. Moreover, the three registration modules can work in parallel to increase the speed of the algorithm.

Fig. 2 shows implementation detail of the LPT/LLT registration module. The role of the Laplacian high-pass filter (HPF) is to remove low- and medium-frequency video components (which represent noise) and pass only the high-frequency components, which contain the spread-spectrum, noise-like watermark. This significantly improves the correlator performance.

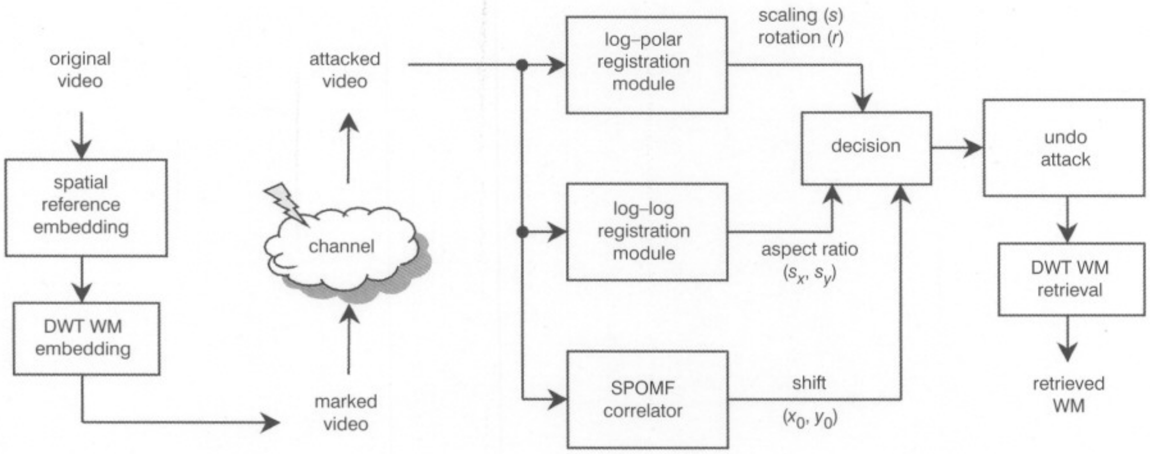

Fig. 1 *Block schematic of the geometric invariant video watermarking system*

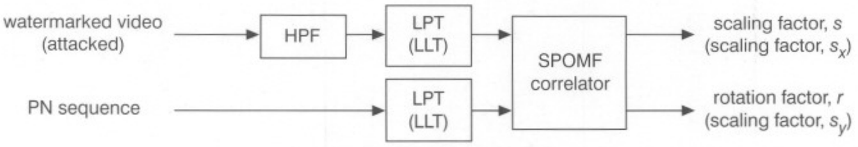

Fig. 2 *The log–polar/log–log module*

4 DWT video watermarking scheme

The wavelet watermarking offers many advantages compared with FFT or DCT watermarking, as shown in [11]. The wavelet transform itself has been widely described in the literature [12–15]. For watermarking, we selected the Antonini 7.9 wavelet, as being one of the best wavelets available [12–14]. The reasons for choosing this basis and its important properties are discussed in [11].

The proposed DWT video watermarking scheme is shown in Fig. 3, where the noisy channel corresponds to the video sequence. The watermark energy is maximised by embedding the main payload according to an HVS model, and the BER is minimised through the use of turbo coding. This significantly increases the operational capacity of the system [11]. In the recovery process, because the cross-correlator performs a sequence of correlation sums, it follows from the central limit theorem that the cross-correlation peaks have a normal distribution [16]. This is convenient for turbo decoding as the latter generally assumes a Gaussian input. Thus, for any particular system, the mean μ and variance σ^2 of the correlation peaks defines the SNR of the channel: $SNR = (\mu/\sigma)^2$. The corresponding BER for an uncoded system is simply $BER_u = Q[\mu/\sigma] = Q[\sqrt{(SNR_u)}]$. For a coded system, the decoded BER is $BER_c = f(SNR_c)$, where f is a known function for a particular iterative decoder. For this work, we used a rate 1/4 multiple parallel-concatenated convolutional code (3PCCC) rather than the basic turbo code (2PCCC) to improve performance [16].

Watermark embedding is shown in Fig. 4, where we use 3 levels of decomposition. Embedding uses the spread-spectrum approach and retrieval is via cross-correlation (matched filtering). The security of such a system relies in the secret watermarking keys, K_1 and K_2. The interleaver (key K_2) provides a random distribution of the data bits within each sub-band. Watermark retrieval is shown in Fig. 5. The video sequence is filtered using a Laplacian 3 × 3 filter prior to cross-correlation, in order to improve the performance of the correlator. It is advantageous to have a self-contained watermark (all data bits) in each sub-band, because an 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, orientation and level, obtaining each time a set of cross-correlation peaks (one peak for each embedded data bit). An SNR is then computed for each set of cross-correlation peaks, and retrieval is carried out for the sub-band with the highest SNR.

4.1 HVS-based embedding

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 as follows:

$$C_i^M = \begin{cases} C_i + \underbrace{\alpha \dfrac{\boldsymbol{Q}(\lambda,\theta)}{\boldsymbol{Q}_{\min}} \cdot \dfrac{|C_i|}{\mathrm{mean}(|C_i|)}}_{S} \cdot W_i & \text{(details)} \\ \text{if} \quad S > 24, \quad \text{then} \quad S = 24 & \\ C_i + \alpha \dfrac{\boldsymbol{Q}(\lambda,\theta)}{2} \cdot \dfrac{|C_i|}{\mathrm{mean}(|C_i|)} \cdot W_i & \text{(approximation)} \end{cases} \tag{1}$$

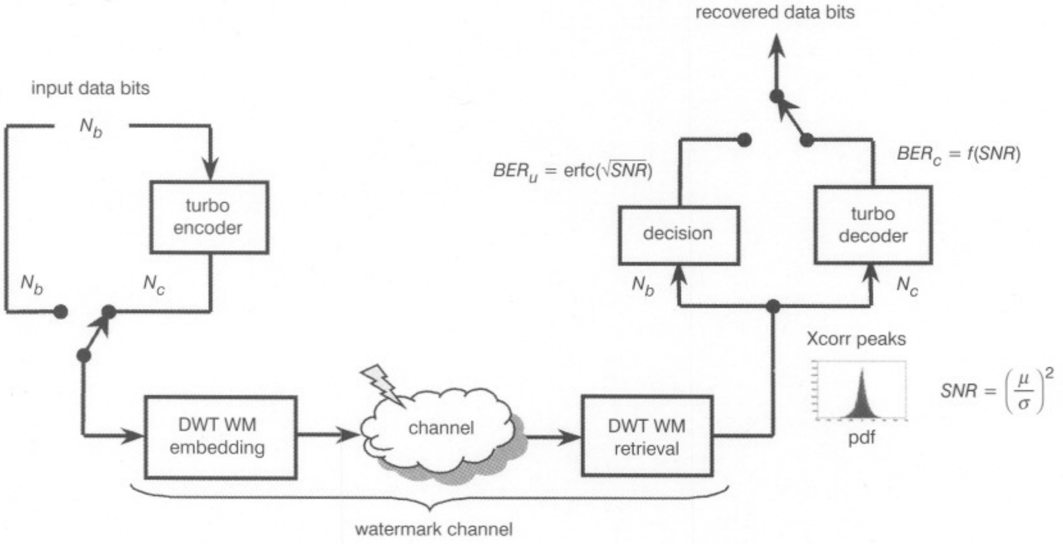

Fig. 3 *The DWT video watermarking scheme*

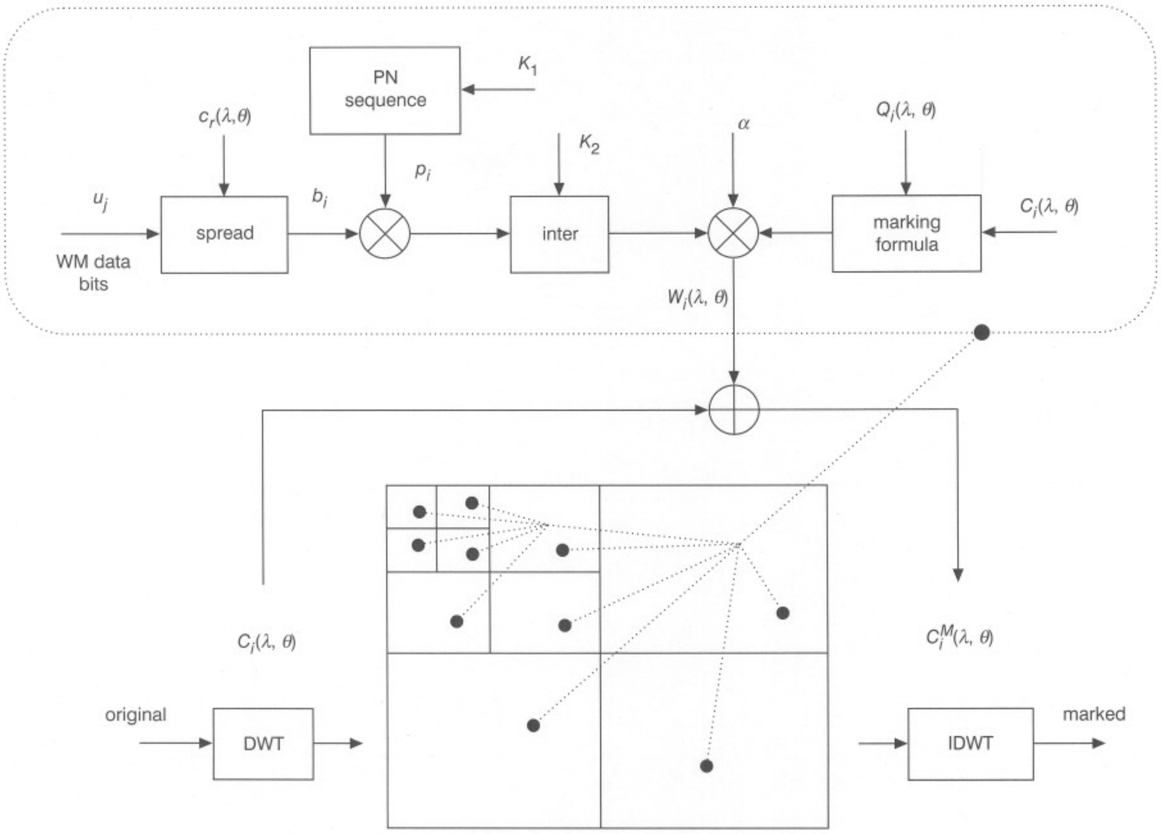

Fig. 4 *Spread spectrum watermark embedding in the DWT domain*

where $\boldsymbol{Q}_{\min}$ is the minimum value from the quantisation matrix $\boldsymbol{Q}(\lambda,\theta)$, W_i is the watermark, C_i is the original wavelet coefficient, C_i^M is the marked coefficient, λ is the level and θ denotes orientation. Note that (1) incorporates media dependence ($|C_i|$), essential for robust watermarking. The high-frequency sub-bands and the largest coefficients are marked more heavily, because modification of these coefficients is less likely to incur visible artefacts. The HVS is incorporated in $\boldsymbol{Q}(\lambda,\theta)$ which is computed according to [15].

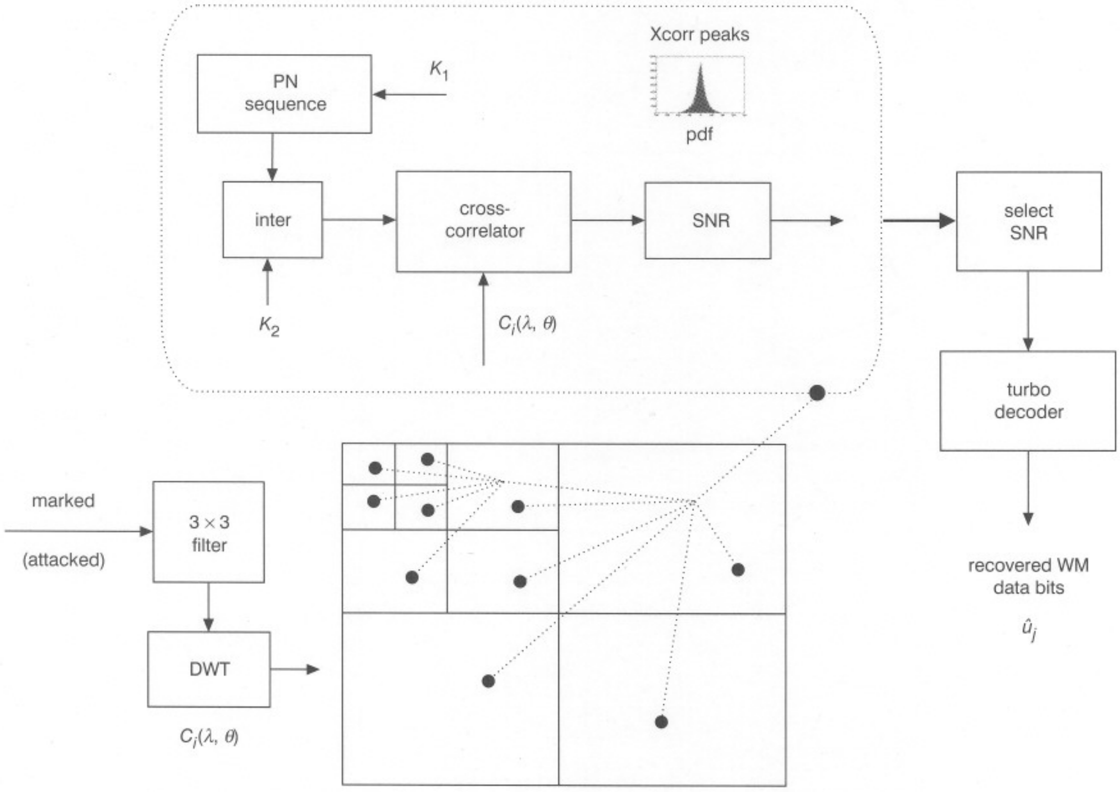

Fig. 5 *Spread spectrum DWT watermark retrieval*

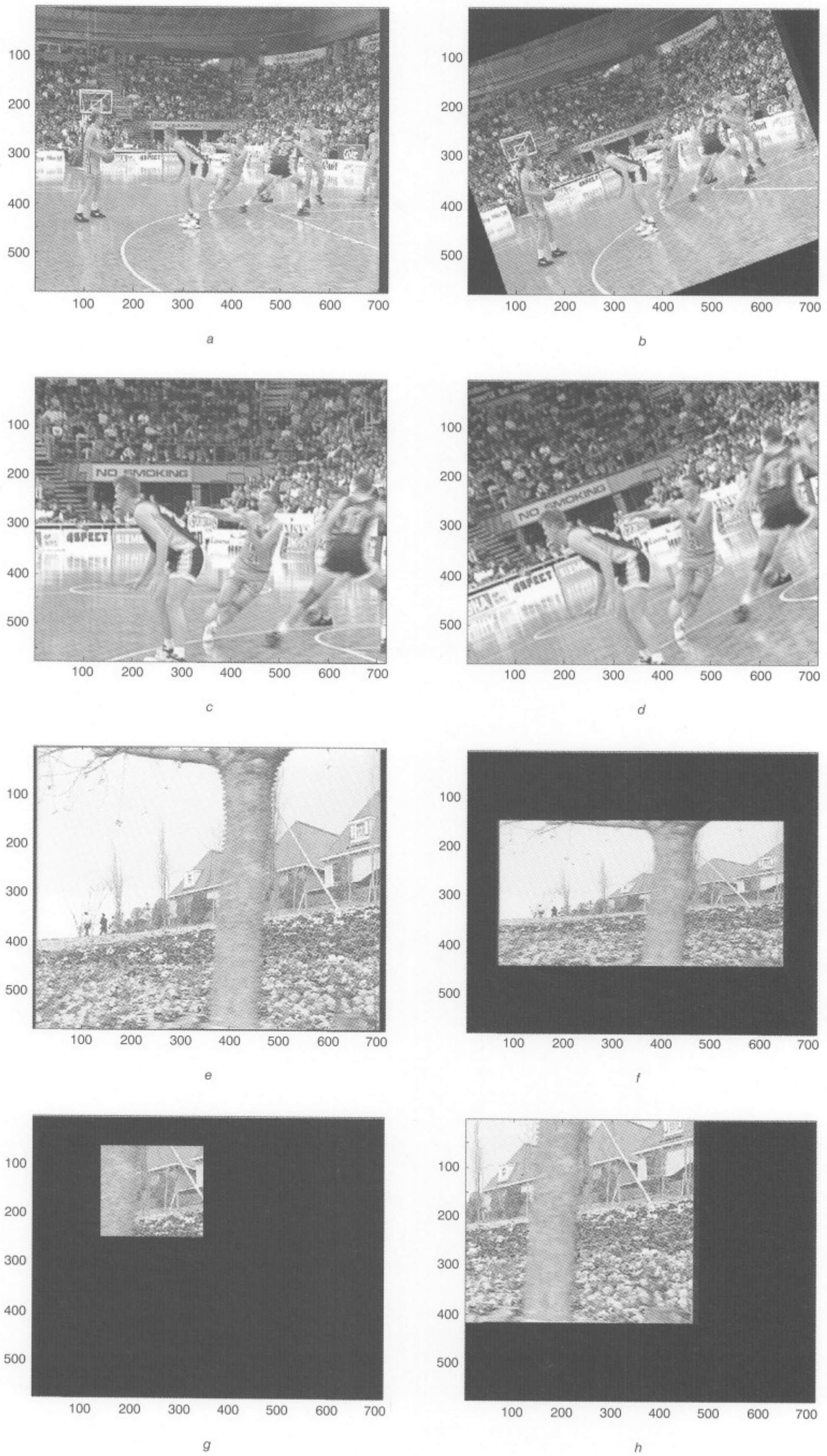

Fig. 6 *Effects of different attacks*
a Original 'Basketball'
b 20° rotation
c 100% scaling
d 20° rotation combined with 100% scaling
e Original 'Flower'
f Arbitrary scaling, image is rescaled from [576 × 720] to [300 × 600]
g Cropping [400, 200, 208, 196] combined with shift [140, 240]
h MPEG2 (2 Mbits/s) combined with shift [160, 240]

5 System performance and false detection probability

The registration module provides invariance to frame shift, rotation, scaling, rotation combined with scaling, and aspect ratio change. The system can also handle a range of other attacks, such as cropping, shift + cropping, MPEG compression, compression + shift + cropping. These attacks are illustrated in Fig. 6, for test sequences 'Basketball' and 'Flower'. For watermarking, 'Basketball' represents a typical average sequence, while 'Flower' is known to be a very difficult sequence. The invisibility of the mark was subjectively assessed using specialised hardware.

Fig. 7 shows the performance of the system for different degrees of rotation, when n frames are averaged in order to improve the robustness of the system. As the minimum watermarking segment is 25 frames, then $n \leq 25$. Compared with the $n = 1$ case, the cross-correlation peak for $n = 25$ is about four times larger. Similar results are presented in Fig. 8 for scaling. Figs. 9 and 10 show the system performance for $n = 25$ and different degrees of rotation and scaling. Finally, Fig. 11 presents the case of rotation combined with scaling for the 'Basketball' sequence ($n = 25$).

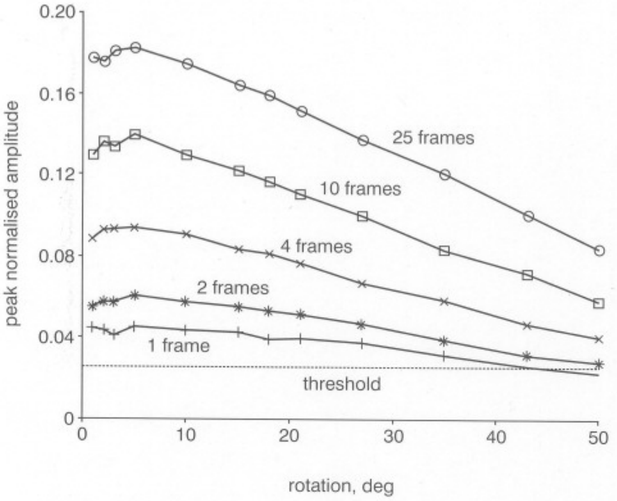

Fig. 7 *Performance of the system for rotation when averaging frames*

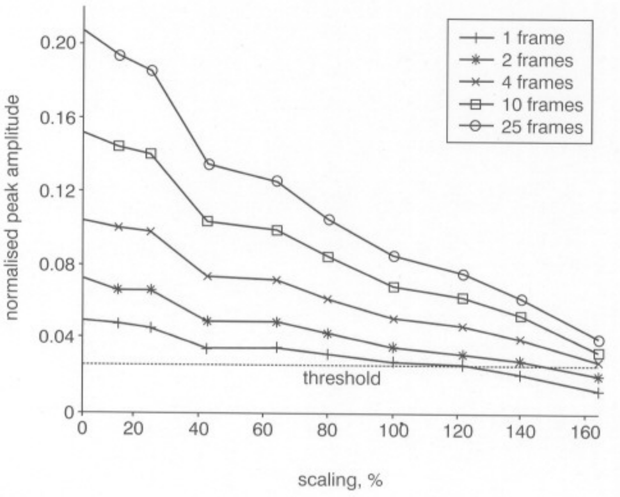

Fig. 8 *Performance of the system for scaling when averaging frames*

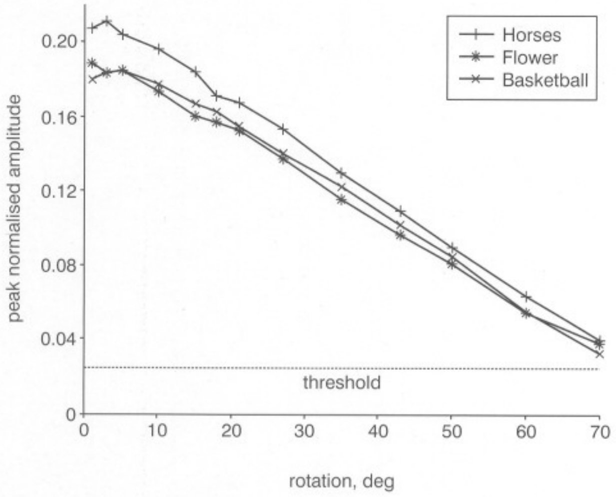

Fig. 9 *Peak normalised amplitude for different video sequences under rotation attack*

A threshold value of 0.025 can be observed in each Figure (Figs. 7–11 and 13). This guarantees a false positive detection, probability better than 10^{-8} when the correlation peak exceeds the threshold. The value was

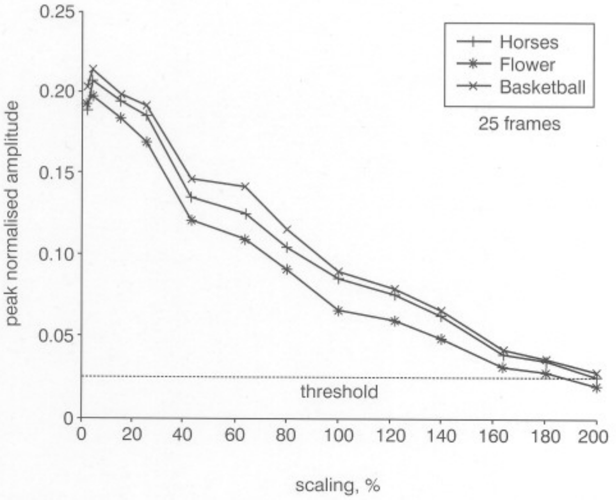

Fig. 10 *Peak normalised amplitude for different video sequences under scaling attack*

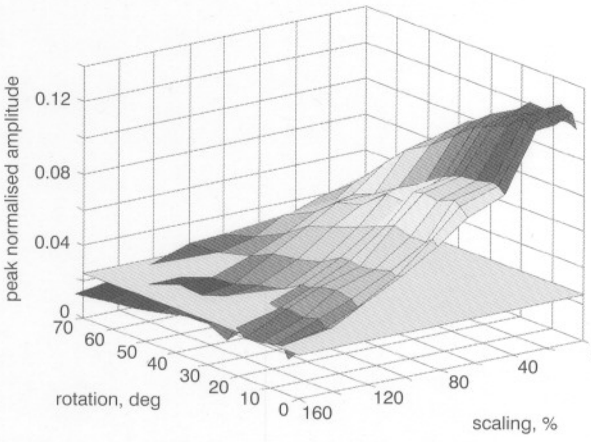

Fig. 11 *Performance of the system for rotation combined with scaling (25 frames)*

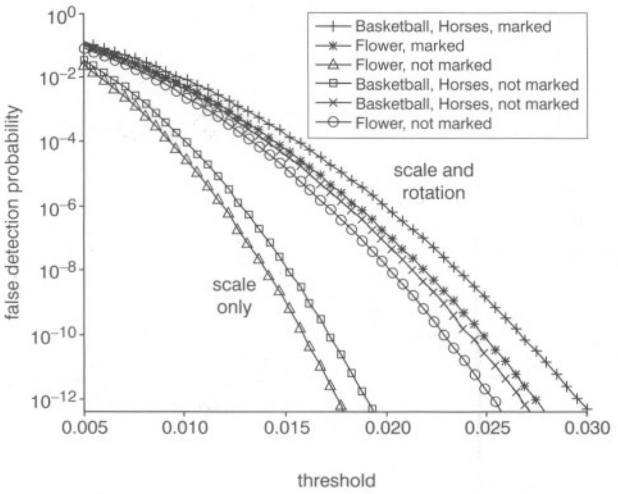

Fig. 12 *Threshold selection for a desired probability of false positive detection*

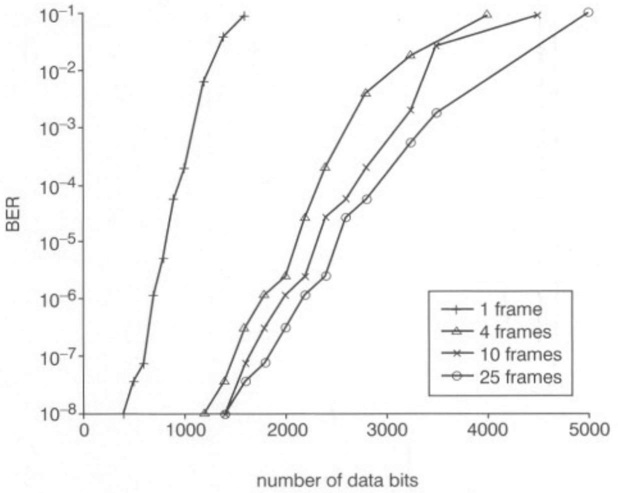

Fig. 14 *Capacity of the DWT system under 2 Mbit/s MPEG2 compression attack when using frame averaging, for the 'Flower garden' video sequence*

experimentally derived for a set of 3 test sequences and a wide range of scaling and rotation attacks: the pdf (probability distribution function) of the peaks was computed for each case and the worst-case scenario determined. The resulting pdf is not Gaussian due to the large number of very small peak values, but, by fitting a zero-mean Gaussian distribution with the same standard deviation as the experimentally determined pdf, the resulting Gaussian distribution can be used to determine the optimum threshold for a given false error probability. The Gaussian distribution fits very well the worst-case scenario pdf in the zone of interest (at the extremities), and is actually chosen to be quite pessimistic. We have investigated several hypotheses: when the sequence was marked with the correct watermark, when the sequence was not marked and when the sequence was marked with a wrong mark, for different attacks and different strength of the attacks, and finally for 3 different video sequences. The results (Fig. 12) suggest that the worst-case scenario is when the sequence is marked with the correct mark, and show that the 0.025 threshold is appropriate for a false detection probability of 10^{-8}.

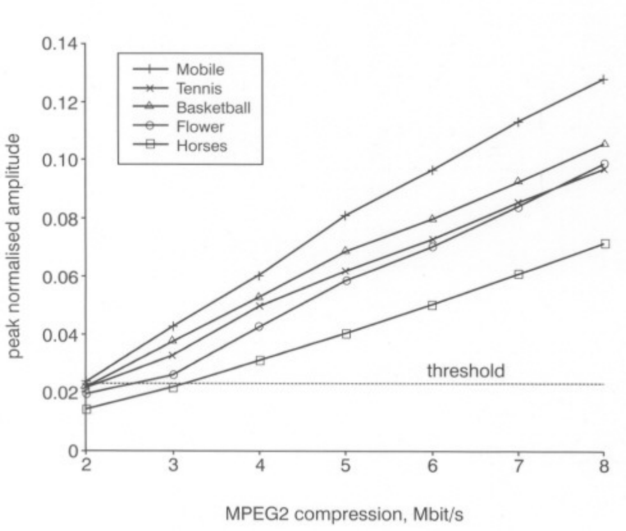

Fig. 13 *Performance under combined attack: MPEG2 compression combined with frame shift, for different video sequences (25 frames averaging)*

5.1 Scaling, rotation and cropping attack

As can be seen in Fig. 9, the system is invariant to any amount of rotation smaller than 70°. Fig. 10 shows that the system can handle any degree of scaling up to 180% (it can also handle scaling up to −50%, i.e. smaller frames). The system therefore exceeds the EBU recommendation [1] for both rotation and scaling.

When rotation is combined with scaling, up to 120% scaling and up to 20° rotation can be tolerated, even for the 'Flower' sequence. All simulations assume a bilinear interpolation in the log-polar module. Our tests show that bilinear interpolation leads to a substantial performance increase (almost double) compared with a simple nearest neighbour interpolation. A combined attack of 20° rotation plus 100% scaling is shown in Fig. 6*d*.

The system copes very well with cropping attack. Even under severe cropping (as in Fig. 6*g*, where the useful frame area is only 208 × 196) the capacity is approximately 1500 bits/frame with turbo coding, reducing to 850 bits/frame without coding [15].

5.2 Compression attack

We have investigated system performance for MPEG2 and JPEG [15] compression attacks. The registration module can cope with MPEG2 compression as low as 2–3 Mbit/s. The system can handle even combined attacks like MPEG2 compression (as low as 3–4 Mbit/s) combined with frame shifts, as illustrated in Fig. 13. In terms of capacity, the DWT watermark survives MPEG2 compression at 2 Mbit/s, for a capacity higher than 1200 bit/s (Fig. 14).

6 Conclusions

Robustness to geometric attack is one of the most important requirements for a watermarking system. To achieve this, an approach based on the image registration techniques and LPT/LLT of the video frames has been developed.

An additional spatial reference watermark compensates for the unavailable original video sequence and makes possible the geometrical 'blind registration'. 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.

Table 1: Performance of proposed system compared with EBU recommendations

Parameter	EBU recommendations	Proposed system
General parameters of the system:		
Watermarking minimum segment (WMS)	1 s, 5 s	min 1 s
Data capacity	64 bits/WMS	$\geq$ 1200 bits/WMS @ 2 Mbit/s MPEG2
Probability for error-free payload per WMS	$> 10^{-8}$	10^{-8}
False positive probability per WMS	$< 10^{-8}$	$< 10^{-8}$
Format of original and watermarked signals	ITU-R 601 (ITU-T BT.656)	ITU-R 601 (ITU-T BT.656)
Watermark recovery	blind	blind
Robustness to attacks:		
MPEG2 compression	2–6 Mbit/s MPEG2	2–6 Mbit/s MPEG2
Colour-space conversion	yes	invariant
Shift	up to 320 × 288	higher than 320 × 288
Scaling	desired: 200%, −50% best achieved: 140%, −70%	180%, −50%
Aspect-ratio conversion	16:9 ↔ 4:3	16:9 ↔ 4:3 (easy), 200%, −100%
Small rotation	up to 2°	up to 2°
Noticeable rotation	up to 10°	up to 70°
Small bend/shear	up to 2° (10°)	no
Cropping	minimum size: 320 × 288	even smaller than 200 × 200
Combined attacks	not specified	yes (wide range)

In this way, the advantages of both techniques are preserved: the speed and efficiency of image registration techniques and the robustness and high capacity of the wavelet system. The performance of the proposed system compared with the EBU recommendations [1] is summarised in Table 1.

7 References

1 CHEVEAU, L., GORAY, E., and SALMON, R.: 'Watermarking—summary results of EBU tests', *EBU Tech. Rev.*, March 2001, (286)

2 CHEN, Q.-S., DEFRISE, M., and DECONINCK, F.: 'Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1994, **16**, (12), pp. 1156–1168

3 KUGLIN, C.D., and HINES, D.C.: 'The phase correlation image alignment method'. Proceedings of IEEE Intl. Conference on Cybernetics Society, September 1975, pp. 163–165

4 HORNER, J.L., and GIANINO, P.D.: 'Phase only matched filtering', *Appl. Opt.*, 1984, **23**, (6), pp. 812–816

5 CASASENT, D., and PSALTIS, D.: 'Scale invariant optical correlation using Mellin transforms', *Opt. Commun.*, 1976, **17**, (1), pp. 59–63

6 CASASENT, D., and PSALTIS, D.: 'Position, rotation and scale invariant optical correlation', *Appl. Opt.*, 1976, **15**, (7), pp. 1795–1799

7 O'RUANAIDH, J.J.K., and PUN, T.: 'Rotation, scale and translation invariant spread spectrum digital image watermarking', *Signal Process.*, 1998, **66**, (3), pp. 303–317

8 LIN, C.-Y., WU, M., BLOOM, J.A., COX, I.J., MILLER, M.L., and LUI, I.M.: 'Rotation, scale and translation resilient public watermarking for images', *Proc. SPIE-Int. Soc. Opt. Eng.*, 2000, **3971**, pp. 90–98

9 REDDY, B.S., and CHATTERJI, B.N.: 'An FFT-based technique for translation, rotation, and scale-invariant image registration', *IEEE Trans. Image Process.*, 1996, **5**, (8), pp. 1266–1271

10 WOLBERG, G., and ZOKAI, S.: 'Robust image registration using log-polar transform'. Proceedings of Intl. Conf. on Image processing, Vancouver, Canada, September 2000

11 SERDEAN, C.V., TOMLINSON, M., WADE, J.G., and AMBROZE, M.A.: 'Protecting intellectual rights: digital WM in the wavelet domain'. Proccedings of IEEE Intl. Workshop on Trends and recent achievements in information technology, Cluj-Napoca, Romania, 16–18 May 2002, pp. 70–77

12 KINGSBURY, N.G., and MAGAREY, J.F.A.: 'Wavelet transforms in image processing'. Proceedings of First European Conf. on Signal analysis and prediction, Prague, 24–27 June 1997, pp. 23–34

13 ANTONINI, M., BARLAUD, M., MATHIEU, P., and DAUBECHIES, I.: 'Image coding using wavelet transform', *IEEE Trans. Image Process.*, 1992, **1**, (2), pp. 205–220

14 VILLASENOR, J.D., BELZER, B., and LIAO, J.: 'Wavelet filter evaluation for image compression', *IEEE Trans. Image Process.*, 1995, **4**, (8), pp. 1053–1060

15 WATSON, A.B., YANG, G.Y., SOLOMON, J.A., and VILLASENOR, J.: 'Visibility of wavelet quantization noise", *IEEE Trans. Image Process.*, 1997, **6**, pp. 1164–1175

16 AMBROZE, A., WADE, G., SERDEAN, C., TOMLINSON, M., STANDER, J., and BORDA, M.: 'Turbo code protection of video watermark channel', *IEE Proc., Vis. Image Signal Process.*, 2001, **148**, (1), pp. 54–58