Turbo code protection of video watermark channel

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

Abstract: The operational capacity of a watermark data channel is deduced by examining the correlation distribution at the retrieval end of a spread-spectrum-based watermarking system. This enables capacity to be determined given MPEG-2 compression, geometric attack, visual thresholds and channel coding. Watermarking itself is carried out in the DCT domain using video-dependent and visual perception concepts. An objective of the paper is to determine the capacity improvement provided by advanced FEC. It is found that FEC based on multiple parallel concatenated convolutional codes (3PCCCs) can give an order improvement in capacity for compressed video, and typically gives 0.5 kbit/s capacity under a combined compression-geometric attack.

1 Introduction

Hidden data or a watermark is inserted into a video sequence for the purposes of copyright protection and video 'fingerprinting'. It can be performed either on uncompressed video (ITU-R 601) or MPEG compressed video [1], although only the former is discussed here. For studio working, e.g. editing, it is desirable to embed the watermark in as short a video segment as possible, typically only 1 s [2]. A (rather large) watermark payload of, say, 100 bytes would then require the hidden data channel to have a capacity around 1 kbit/s in the presence of both unintentional and intentional attack. Unintentional attack arises from normal signal processing, such as MPEG compression or frame shift due to video mixing, whilst a serious intentional attack could be an 'ambiguity' attack [3] or a 'detection-disabling' attack. The latter form of attack is of interest here and it could take the form of line or frame cuts, or more general geometric distortion [4].

The system under consideration is shown in Fig. 1. This is a blind watermarking system in that the unmarked host video is not available for retrieval. It uses the well known spread-spectrum approach, but also protects marking using channel coding. N_b coded or uncoded bits are embedded over a sequence of frames, giving N_b normally distributed crosscorrelation peaks. The paper first overviews video watermarking based on spread-spectrum techniques, with the aim of defining a signal-to-noise ratio (SNR) for the watermark channel. We then consider the information capacity of the channel, with emphasis on practical, measurable parameters. An objective of the paper is to determine actual capacities given typical compression and geometric attacks. In particular, it is of interest to determine the improvements in channel capacity that can be achieved through the application of turbo-like codes.

IEE Proceedings online no. 20010109

DOI: 10.1049/ip-vis: 20010109

Paper first received 17th March and in revised form 19th October 2000

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

J. Stander is with the School of Mathematics and Statistics, University of Plymouth, Plymouth, Devon PL4 8AA, UK

M. Borda is with the Communications Department, Technical University of Clui-Napoca, Romania

2 SNR of a spread-spectrum watermark channel

We concentrate on transform domain marking since it is easy to avoid marking high video frequencies (which tend to be attenuated by compression), and because it is naturally suited for perceptual marking based on the human visual system (HVS). Also, from an information theoretic argument, transform domain marking can give increased channel capacity compared to spatial domain marking [5].

Fig. 2 shows a spread-spectrum watermarking and retrieval system based on the discrete cosine transform (DCT). For chip rate c_r, each data bit u_j is spread as $b_i = u_j$, $jc_r \leq i < (j+1)c_r$, and the product $b_i p_i$ is formed, where $\{p_i\}$ is a binary $\{\pm 1\}$ PN sequence. This sequence spreads u_j over many 8×8 pixel blocks distributed over a number of video frames, the random locations of the blocks being determined by the key used to generate the sequence.

Video-dependent marking is an essential component of a successful marking scheme [6] if only that it ensures that marking energy is low in low detail areas of a video frame. A simple video-dependent watermark is

$$w_i = \alpha b_i p_i |C_i| \tag{1}$$

Here, C_i is a DCT coefficient and α is selected to give a mark below the threshold of visual perception ($\alpha \ll 1$). If there is no attack or filtering, and assuming $u_j = 1$, the normalized crosscorrelation is

$$d_j = \frac{1}{c_r} \sum_{i=jc_r}^{(j+1)c_r-1} p_i C_i + \frac{\alpha}{c_r} \sum_{i=jc_r}^{(j+1)c_r-1} |C_i| p_i^2 \tag{2}$$

We are interested in the distribution of these crosscorrelation peaks since it determines the detected bit error rate

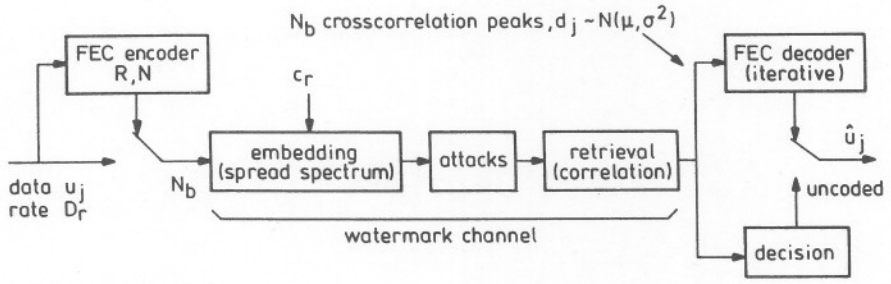

Fig. 1 *Channel coding in a watermark channel*

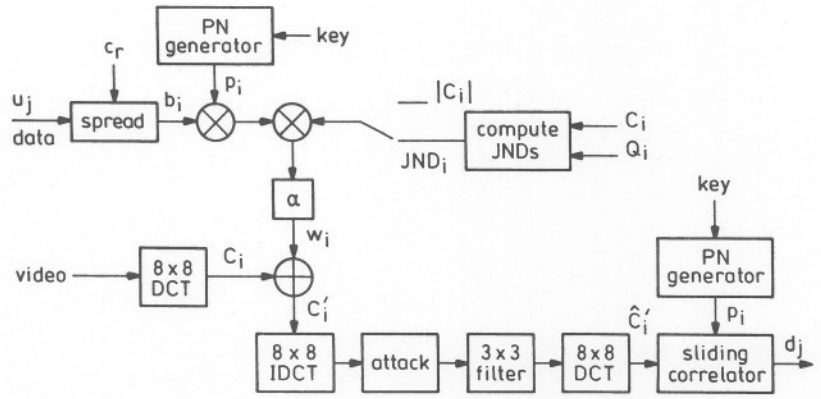

Fig. 2 *Transform domain watermarking and retrieval*

(BER). If, for simplicity, we assume C_i is i.i.d. with $C_i \sim \mathrm{N}(0, \sigma_c^2)$, then d_j has the approximate distribution

$$d_j \sim \mathrm{N}\left(\alpha\mu_{|C|}, \frac{\sigma_c^2}{c_r}\right) \tag{3}$$

where $\mu_{|C|}$ is the mean of $|C_i|$, $\alpha \ll 1$ and $p_i \in \{\pm 1\}$. The normal distribution follows from the central limit theorem since the crosscorrelator performs a sequence of correlation sums. Clearly, the BER will decrease as c_r increases (as expected) due to reduced distribution variance. The variance of d_j arises mainly from the first term in eqn. 2 and so we conclude that the nonzero crosscorrelation of the PN sequence with the DCT coefficients is a source of noise in the channel.

In practice, the distribution will depend on other factors (including attacks). For example, it is widely recognised that crosscorrelation can be significantly improved by inserting a 3×3 spatial filter in the video path (Fig. 2). This removes low frequency video components prior to crosscorrelation and gives a distribution with larger mean and smaller variance. In practice, compared to filtering, balancing the PN sequence to ensure that it has zero mean gives only a relatively small improvement. Since marking is video-dependent, the distribution will also depend on the choice of sequence. Fig. 3 illustrates the point for two standard MPEG video test sequences (ITU-R 601 format); it is apparent that sequence 'flower garden' will have a larger BER than sequence 'mobile'. The underlying normal distribution having the same mean and variance is shown dotted. Generalising, for any particular system the distribution mean μ and variance σ^2 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, μ and σ define a signal-to-noise ratio SNR_c at the decoder input, and the decoded bit error rate is $BER_c = f(SNR_c)$, where f is a known function for a particular iterative decoder (Fig. 4).

Rather than using eqn. 1, a better approach is to base marking on both the video sequence and the HVS. In this paper we compute a perceptual threshold or just noticeable difference, JND_i, for each DCT coefficient [7, 8], and watermarking is applied as

$$C_i' = C_i + \alpha b_i p_i JND_i \tag{4}$$

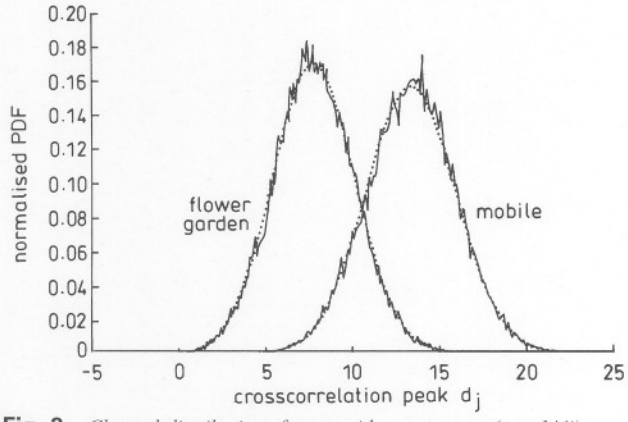

Fig. 3 *Channel distributions for two video sequences ($u_j = 1\ \forall j$)*

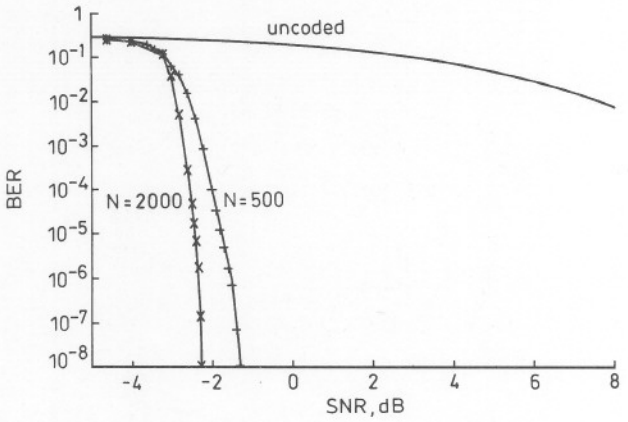

Fig. 4 *Simulated performance of an iterative decoder for interleaver sizes of 500 and 2000 for rate 1/4 3PCCC FEC*

(see Fig. 2). In practice, the theoretical *JND* values have been found to be within a factor 2 or 3 of the actual perceptual threshold values, and this is accounted for by the factor α in eqn. 4. Marking is HVS based since JND_i is computed from the frequency sensitivity, and luminance/contrast masking of the eye, and it is video-dependent since JND_i is also a function of C_i. Human perception is also incorporated by making JND_i a function of the MPEG-2 default quantisation matrix elements Q_i [9].

3 Channel capacity for uncoded video

If we regard the watermark channel as a communications system with input X (the watermark data) and output Y, the channel capacity is defined as the maximum mutual information:

$$C_{chan} = \max_{p(x)} I(X; Y) = \max_{p(x)}[h(X) - h(X|Y)] = \max_{p(x)}[h(Y) - h(Y|X)] \quad (5)$$

where the maximum is taken over all possible distributions $p(x)$. Term $h(X|Y)$ represents information loss due to channel 'noise', which will be a combination of the host video and signal processing (compression/attack). If the loss is modelled as the addition of an independent Gaussian noise source, $Z \sim \mathrm{N}(0, \sigma_z^2)$, i.e. $Y_i = X_i + Z_i$, where Z is a continuous random variable, then eqn. 5 reduces to [10]

$$C_{chan} = \frac{1}{2}\log_2\left(1 + \frac{\sigma_x^2}{\sigma_z^2}\right) \quad \text{bits/symbol} \quad (6)$$

providing $X \sim \mathrm{N}(0, \sigma_x^2)$. In [5], σ_x^2 was estimated from acceptable JPEG performance, and an equivalent Gaussian noise variance was computed for the host and JPEG compression. In [11] the noise was restricted to an AWGN attack, the host noise being virtually eliminated by using it as side information during embedding.

In this paper we invoke *JND*s to maximise the signal power. From the discussion in Section 2, the channel in Fig. 1 can be modelled as a gain factor cascaded with a Gaussian noise source $Z \sim \mathrm{N}(0, \sigma_z^2)$ (the gain and variance depending on the host video, MPEG compression and geometric attack). For example, we could estimate a basic operational capacity as follows. Suppose that all N_p pixels in the frame are transformed via DCT blocks, and that the channel noise is simply that of the host video (in the following tests, marking is restricted to the luminance channel, and $N_p = 720 \times 576$ pixels). Assuming C_i is i.i.d. for simplicity, the noise power per video frame is $N_p\overline{C^2}$, where $\overline{C^2}$ is the mean coefficient power for a particular video sequence. If only one data bit is embedded per video frame (corresponding to $c_r = N_p$), and there is no FEC, the SNR is

$$SNR = N_p \frac{(\alpha\overline{JND})^2}{\overline{C^2}} = N_p\overline{SNR} \quad (7)$$

where $\overline{SNR}$ is a measured mean SNR for the video sequence and $\overline{JND}$ is the mean JND for the sequence. If N_b data bits are embedded into a frame the signal to noise ratio per uncoded data bit reduces to $SNR_u = SNR/N_b$, and the data rate or capacity for an uncoded system of frame rate F_r is

$$D_r = N_b F_r = \frac{kN_p\overline{SNR}F_r}{SNR_u} \quad \text{bits/s} \quad (8)$$

In eqn. 8 we account for the fact that, in general, only a fraction k of the coefficients in each DCT block are marked. Also, in the following work we will define the operational capacity as the maximum value of D_r for which the *BER* does not exceed a tolerable level; typically this is 10^{-8} for video watermarking [2].

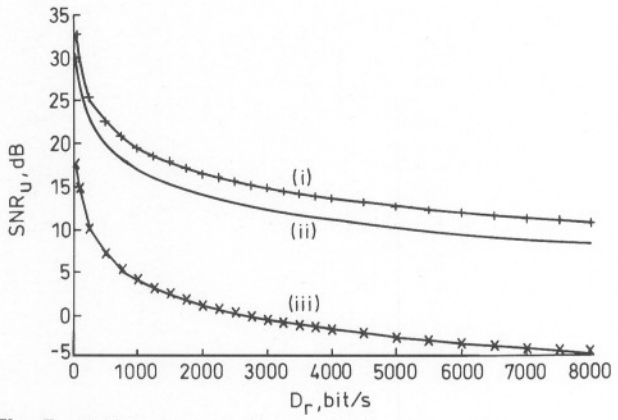

Fig. 5 *SNR for sequence 'flower garden'*

(i) Uncompressed filtered; (ii) uncompressed, unfiltered, eqn. 8; (iii) MPEG-2 compressed to 6 Mbit/s, filtered

Eqn. 8 has been used to estimate the capacity for the uncoded, uncompressed video test sequence 'flower garden' (graph (ii) in Fig. 5), assuming the *JND*-based marking in eqn. 4. Coefficient α in eqn. 4 was set to give marking safely below the threshold of visibility (visible *JND* marking appearing as fine grain noise). Graph (i) in Fig. 5 shows the corresponding simulated capacity, obtained by embedding N_b bits over N_f frames ($D_r = N_b F_r/N_f$) using eqn. 4. This gives N_b correlation peaks and the resulting distribution gives $SNR_u = (\mu/\sigma)^2$. The discrepancy between graphs (i) and (ii) is attributed to highpass filtering prior to crosscorrelation. Graph (iii) shows the simulated result for MPEG-2 compression to 6 Mbit/s. The reduced capacity due to compression is more clearly shown in Fig. 6. Using $BER_u = Q(\sqrt{SNR_u})$, graph (i) shows that the uncompressed capacity is ~ 3 kbit/s, and graph (ii) gives experimental confirmation of (i) by directly counting data errors. Graph (iii) shows that MPEG-2 compression reduces the capacity to the order of 100 bit/s. A capacity increase is to be expected if the DCT is replaced by the discrete wavelet transform (DWT), since the hierarchical decomposition permits the construction of a spatially local and spatially global watermark. For images, DWT marking has shown more robustness, especially in the presence of combined errors, such as compression plus geometric attack [12].

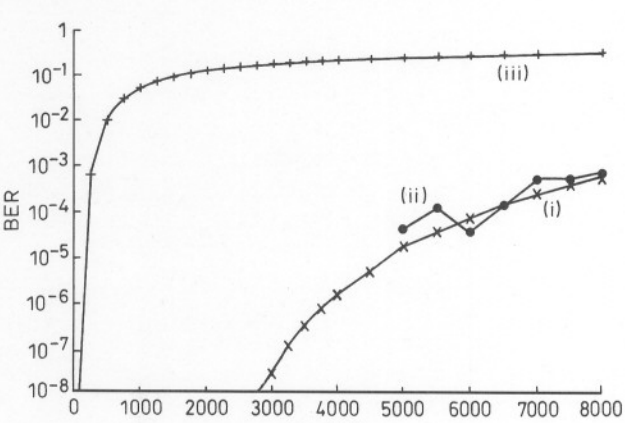

Fig. 6 *BER for sequence 'flower garden'*

(i) Uncompressed; (ii) uncompressed simulation; (iii) MPEG-2 compressed to 6 Mbit/s, filtered

4 Capacity for coded video

Consider a coded system of rate R (Fig. 1) and assume that N_b coded bits are embedded over a sequence of N_f video frames. The watermark data rate is now

$$D_r = \left(\frac{N_b R}{N_f}\right) F_r \quad \text{bit/s} \tag{9}$$

Assuming as before that all N_p pixels in a frame are transformed, the spread-spectrum chip rate is

$$c_r = \frac{kN_f N_p}{N_b} = \frac{kN_p R}{D_r} F_r \tag{10}$$

According to eqn. 10, for a fixed D_r, the use of FEC reduces the chip rate by a factor R. As indicated in eqn. 3, this increases the variance of the channel distribution, resulting in increased *BER*, and the FEC decoder must more than compensate for this increase to provide coding gain.

For a channel with a potentially large BER (due to attack) it is essential to use soft decision decoding, and in practice this restricts the choice of FEC to convolutional codes. Viterbi decoding is the usual ML decoder for an AWGN channel, and so has been used to protect watermarked video [13]. Here, we are interested in the recovery of channel capacity that can be achieved by using turbo-like codes in an attacked watermarking channel. These codes offer better coding gain and could be used at low rate in this application (with consequent improvement in performance). We consider a code of rate R and interleaver size N, and treat the FEC as block coding (block length N). Note from Fig. 1 that coding precedes spread-spectrum so that the input to the iterative decoder will be the output of the spread-spectrum correlator. This scheme ensures that the watermark channel up to the decoder input approximates to a Gaussian channel, and the latter is the usual assumption for turbo code systems. An alternative scheme is to place the FEC encoder after the spread-spectrum process. This has the potential to provide large block lengths for the encoder and so improve its performance, but has the disadvantage that the channel at the decoder input is poorly defined and will have a very low SNR.

In this paper a multiple parallel concatenated convolutional code (3PCCC) has been used to protect the watermark channel, and the encoder is shown in Fig. 7. The use of two interleavers (I_1 and I_2) rather than one as in the basic turbo code reduces the error rate floor and so gives improved performance [14]. Each recursive systematic code (RSC) is an optimum (5, 7) code [15], giving an unpunctured code rate $R = 1/4$. Fig. 4 shows the simulated performance of the overall code for several interleavers.

Fig. 5 shows that an uncoded watermark channel at 1 kbit/s corresponds to $SNR_u \approx 4$ dB when 6 Mbit/s MPEG-2 compression is used. Whilst this low SNR is unusable without FEC (see Fig. 6), Fig. 4 shows that that iterative decoding should be effective against this sort of attack. Fig. 8 illustrates the effectiveness of iterative decoding against a 6 Mbit/s MPEG-2 attack, and also compares the performance of the heuristic marking scheme in eqn. 1 (Fig. 8*a*) with that of the *JND* scheme in eqn. 4 (Fig. 8*b*). The BER of the uncoded system is computed as before, whilst that of the coded system is computed as $BER_c = f(SNR_c)$ using Fig. 4. Without FEC the attack reduces the 10^{-8} capacity to ~ 300 bits/s, but with FEC the capacity can be > 3 kbit/s. In each case, α was selected to give marking just below the threshold of visibility. Note that *JND* marking can give a capacity > 8 kbit/s for uncompressed video (not shown).

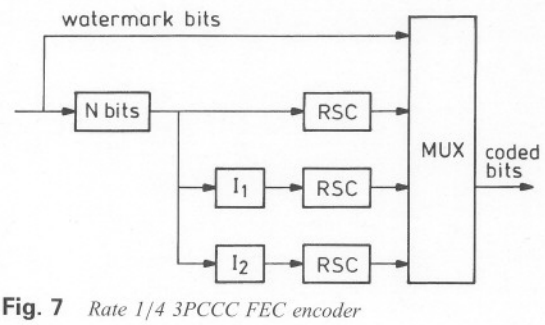

Fig. 7 *Rate 1/4 3PCCC FEC encoder*

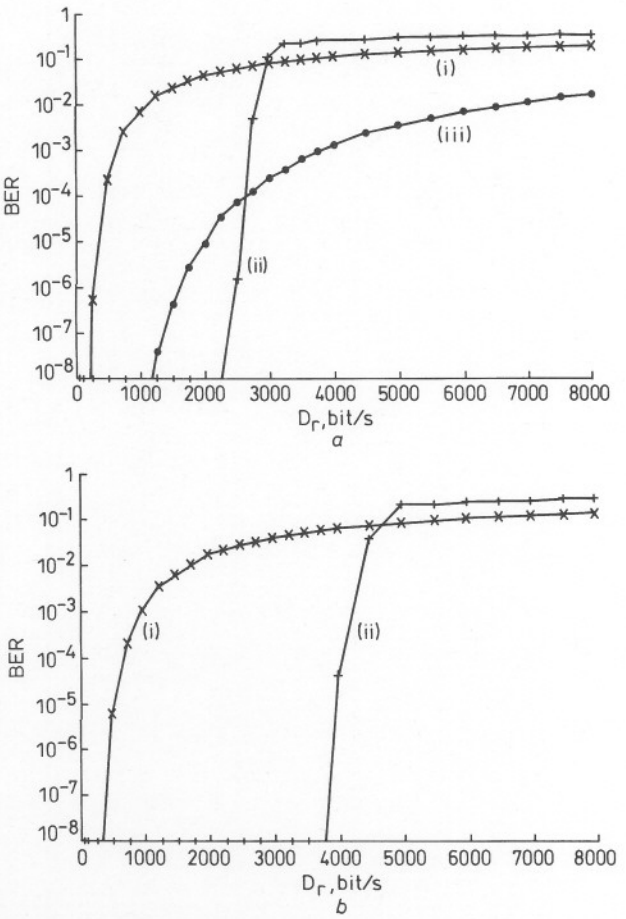

Fig. 8 *Combating a 6 Mbit/s MPEG-2 attack with iterative decoding ($N = 2000$) for sequence 'basketball'*

(i) Uncoded, compressed; (ii) coded, compressed; (iii) uncoded, uncompressed
a Heuristic marking, $\alpha = 0.004$
b JND marking, $\alpha = 3$

4.1 Synchronisation and combined attacks

Spread-spectrum systems are vulnerable to synchronisation error, as can occur in a geometric attack. A 2-D attack, such as the deletion of a scan line on all frames (and the use of line repeat to maintain frame size) can be combated by using a 2-D sliding correlator to search for the correlation peak. A temporal attack (such as a frame cut) can be combated by using a 3-D sliding correlator, which essentially searches for the peak in a temporal window over several frames. The use of a sliding correlator has the disadvantage of decreasing the effective chip rate since a local crosscorrelation peak is computed for small blocks (the overall crosscorrelation being the sum of the local correlations). This amounts to 'correlator loss' and is illustrated in Fig. 9 for a frame cut. The loss is ~ 3 dB

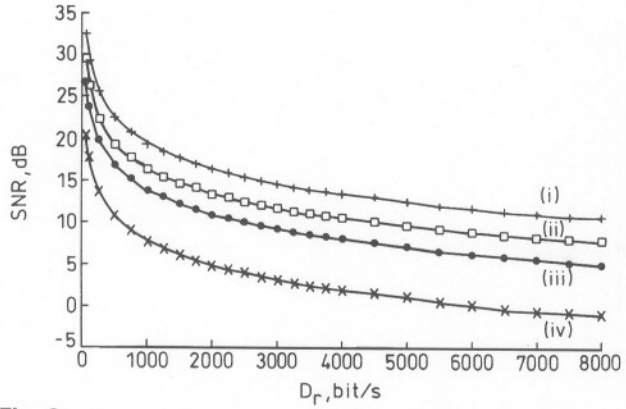

Fig. 9 *Correlator loss for sequence 'flower garden' (120 frames)*

(i) No attack, no sliding; (ii) 3-D correlator with/without frame cut attack; (iii) 2-D correlator, frame cut attack after 60 frames; (iv) 2-D correlator, frame cut attack after 30 frames

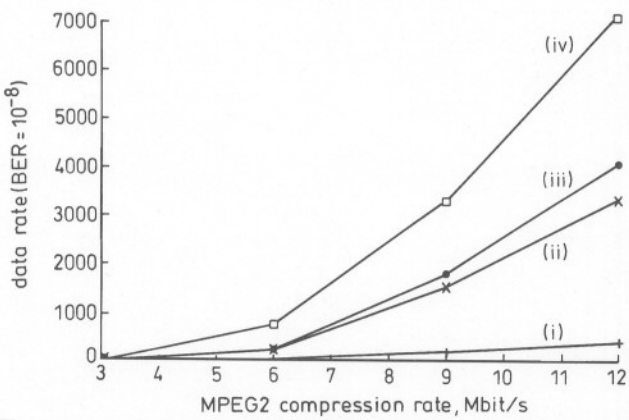

Fig. 10 *Combating a combined compression and line cut attack on sequence 'basketball' with iterative decoding*

(i) Uncoded; (ii) coded, $N = 500$; (iii) coded, $N = 2000$; (iv) Shannon limit

for a 3-D correlator (graph (ii)), irrespective of the location of the cut within the sequence. The loss for a 2-D correlator can vary significantly, e.g. 12 dB if the frame cut occurs at frame 30 within a 120 frame sequence (graph (iv)).

In practice watermarked video is likely to suffer from a combination of attacks, such as MPEG-2 compression and geometric distortion, and an attack of this nature can defeat many watermarking schemes [16, 4]. Fig. 10 shows the effect of a combined compression and line cut attack on the *JND*-based marking scheme in eqn. 4. The 10^{-8} capacity is relatively low for an uncoded system (graph (i)), and a possible explanation is that compression reduces the watermark amplitude to below the performance threshold of the sliding correlator. On the other hand, graphs (ii) and (iii) show that FEC can still give a significant improvement in capacity.

5 Conclusions

A spread-spectrum-based video watermark data channel is conveniently characterised by the Gaussian distribution at the output of the sliding correlator. This distribution defines an SNR for the channel, from which can be deduced an operational channel capacity for a practical video sequence subjected to HVS-based marking, combined attacks and FEC. The Gaussian input to the FEC decoder, and the fact that low code rates can be tolerated, makes iterative decoding particularly appropriate for the protection of a watermarked channel. The computational complexity of such decoding is still relatively small compared to that of the sliding correlator.

As expected, channel capacity increases through the use of perceptual marking (*JND*s) and FEC, and decreases when watermarked video is subjected to attacks. MPEG-2 compression to 6 Mbit/s reduces the 10^{-8} capacity from 8 kbit/s (uncompressed video, *JND* marking) to ~300 bits/s, although an order improvement is achieved through FEC. A combination of MPEG-2 compression and simple geometric attack can severely reduce the ability of FEC to recover channel capacity, although useful improvements can still be made. Under such an attack, the use of FEC enables the current watermarking scheme to achieve a typical capacity of 0.5 kbit/s, depending on the video sequence. Higher capacity requires more robust watermarking, e.g. by replacing the DCT with the DWT, and corresponding improvements to the sliding correlator. For example, a 3-D sliding correlator has proved effective against a frame cut attack.

6 References

1 HARTUNG, F., and GIROD, B.: 'Digital watermarking of raw and compressed video', *Proc. SPIE-Int. Soc. Opt. Eng.*, October 1996, **2952**, pp. 205–213

2 Union Europeenne de Radio-Television (EBU-UER) Watermarking Working Group, February 2000

3 CRAVER, S., MEMON, N., YEO, B., and YEUNG, M.: 'Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications', *IEEE J. Sel. Areas Commun.*, 1998, **16**, (4), pp. 573–586

4 PETITCOLAS, F., ANDERSON, R., and KUHN, M.: 'Attacks on copyright marking schemes', *Lect. Notes Comput. Sci.*, vol. 1525, Portland, OR, USA, 14–17 April 1998, pp. 218–238

5 RAMKUMAR, M., and AKANSU, A.: 'Information theoretic bounds for data hiding in compressed images'. Proceedings of IEEE 2nd Workshop on *Multimedia signal processing*, Redondo Beach, CA, 1998

6 SWANSON, M., ZHU, B., CHAU, B., and TEWFIK, A.: 'Object-based transparent video watermarking'. Proceedings of IEEE Signal Processing Society, 1997 Workshop on *Multimedia signal processing*, Princeton, NJ, 23–25 June 1997, pp. 369–374

7 WOLFGANG, R., PODILCHUK, C., and DELP, E.: 'Perceptual watermarks for digital images and video', *Proc. IEEE*, 1999, **87**, (7), pp. 1108–1126

8 KIM, S., SUTHAHARAN, S., LEE, H., and RAO, K.: 'Images watermarking scheme using visual model and BN distribution', *Electron. Lett.*, 1999, **35**, (3), pp. 212–213

9 ISO/IEC 13818-2. Information Technology—Generic coding of moving pictures and associated audio information

10 COVER, T., and THOMAS, J.: 'Elements of information theory' (Wiley, 1991)

11 EGGERS, J., SU, J., and Girod, B.: 'A blind watermarking scheme based on structured codebooks'. Proceedings of IEEE Conference on *Secure images and image authentication*, April 2000, London, UK, pp. 1–6

12 PODILCHUCK, C., and ZENG, W.: 'Image-adaptive watermarking using visual models', *IEEE J. Sel. Areas Commun.*, 1998, **16**, (4), pp. 525–538

13 HERNANDEZ, J., DELAIGLE, J., and MACQ, B.: 'Improved data hiding by using convolutional codes and soft-decision decoding', *Proc. SPIE-Int. Soc. Opt. Eng.*, 2000, **3971**, pp. 24–47

14 DIVSALAR, D., and POLLARA, F.: 'Multiple turbo codes for deep space communications'. TDA Progress Report 42–121, May 1995, pp. 66–77

15 BENEDETTO, S., GARELLO, R., and MONTORSI, G.: 'A search for good convolutional codes to be used in the construction of turbo codes', *IEEE Trans. Commun.*, 1998, **46**, (9), pp. 1101–1105

16 HARTUNG, F., SU, J., and GIROD, B.: 'Spread spectrum watermarking: malicious attacks and counterattacks', *Proc. SPIE-Int. Soc. Opt. Eng.*, 1999, **3657**, pp. 147–158