View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Explore Bristol Research

Winne, D. A., Knowles, H. D., Bull, D. R., & Canagarajah, C. N. (2002). Spatial digital watermark for MPEG-2 video authentication and tamper detection. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'02) Orlando, FL, USA. (Vol. 4, pp. IV-3457 - IV-3460). Institute of Electrical and Electronics Engineers (IEEE). 10.1109/ICASSP.2002.1004656

Link to published version (if available):
10.1109/ICASSP.2002.1004656

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms.html

Take down policy

Explore Bristol Research is a digital archive and the intention is that deposited content should not be removed. However, if you believe that this version of the work breaches copyright law please contact open-access@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline of the nature of the complaint

On receipt of your message the Open Access Team will immediately investigate your claim, make an initial judgement of the validity of the claim and, where appropriate, withdraw the item in question from public view.

SPATIAL DIGITAL WATERMARK FOR MPEG-2 VIDEO AUTHENTICATION AND TAMPER DETECTION

Dominique A. Winne, Henry D. Knowles, David R. Bull, C. Nishan Canagarajah

Image Communications Group, Centre for Communications Research
University of Bristol
Woodland Road, Bristol, BS8 1UB, U.K.
e-mail: {D.Winne,Henry.Knowles,Dave.Bull,Nishan.Canagarajah}@bristol.ac.uk

ABSTRACT

The widespread adoption of digital video techniques has generated a requirement for authenticity verification in applications such as criminal evidence, insurance claims and commercial databases. This work addresses problems that arise from a spatial digital watermarking technique developed to detect frame reordering and dropping scenarios. It discusses the differences between mutual frame types at different bit-rates. Many papers consider detection after MPEG-2 decoding as a naïve approach. However, this approach does offer significant advantages for a slight increase of computation load. This paper also establishes a link between the detector performance and the sequence content. The uniqueness of this work is the comparison of the test results using 18 different standard MPEG test-sequences. The functionality of the algorithm is demonstrated with a simulated attack.

1. INTRODUCTION

Authentication has always been an important issue throughout history [1]. Now, in the digital world of imagery and video, the aid of digital watermarks is useful to establish authenticity.

The attacks on video signals fall into two categories, [2]: The first type covers the attacks that tamper with the intensity patterns of video, e.g. compression, noise, etc.. While the second type, time-base tampering, disrupts the frame sequencing, e.g. frame cuts, swapping, deletion or foreign frame insertion. Digital watermarking becomes a useful technique to stamp imperceptible labels to the monitoring content, allowing detection of time-base tampering.

There are several ways to embed and detect a robust watermark in a compressed video sequence. Previously reported work embeds a watermark in the pre-compressed MPEG-2 bitstream and utilizes a drift compensation function to neutralize the introduced distortion [3]. Mobasseri *et al.* [2] embed a spread-spectrum watermark in the video stream at bitplane level and Kalker *et al.* [4] embed a spread-spectrum watermark in the spatial domain and detect its presence in the Displaced Frame Difference (DFD) signal. Jordan *et al.* [5] propose a video watermarking method that embeds information in the motion vectors. This method is not robust to re-encoding. Linnartz *et al.* [6] embed information by modifying the picture type of a GOP and so creating irregular GOP structures. This method is not useful for time-base attacks as it is not robust to re-encoding and it contradicts efforts to improve coding efficiency. Deguillaume *et al.* [7] propose a method to embed a spread-spectrum watermark by employing a 3-D DFT. By using a template, their scheme can invert effects of frame-rate, aspect-ratio changes and frame re-scaling.

Spatial watermarks are easy and fast to embed but are generally considered fragile [5]. In this work, we show that it is possible to recover spatial watermarks from compressed video at different locations and different bit-rates. In section 2, we briefly introduce a spatial watermarking scheme where the detection is based on a similarity check between the original watermark and the watermarked image in question. In section 3, we extend the scheme to MPEG-2 compressed video and discuss the drawbacks of the embedding methods. In section 4, we compare and discuss the different detection locations and examine the differences between frame types and sequence contents. We present our conclusions in section 5.

2. EMBEDDING AND EXTRACTION OF A SPATIAL DIGITAL WATERMARK

The information of a digital image can be seen as a group of two-dimensional samples ordered in grid formation represented as $I(x,y) \in \mathcal{N}$. x and y are bounded by the row and column size of the image I. These samples represent the captured and mapped three-dimensional information. The watermark W consists of a string of pseudo-random Gaussian distributed numbers with zero mean and unit variance. $W = \{w_1, w_2, \ldots, w_n\}$ with $\forall w \in \mathcal{R}$.

0-7803-7402-9/02/$17.00 ©2002 IEEE

Authorized licensed use limited to: UNIVERSITY OF BRISTOL. Downloaded on February 4, 2009 at 07:04 from IEEE Xplore. Restrictions apply.

2.1 Watermark embedding algorithm

The watermark is embedded by signal adaptive addition (see Eq. 2.1) [8] $I^{'}$ represents the watermarked image and α symbolizes the strength factor.

$$I_i^{'} = I_i + \alpha_i \cdot W_i \qquad (2.1)$$

The log of the variance of neighboring pixels within a window controls the strength factor α. Because the eye responds logarithmically to changes of intensity, see [9]. ε $(\in \Re_0^+, \varepsilon << 1)$ limits β within practical range.

$$\beta_i = \log(\mathrm{var}(window(I_i, n)) + \varepsilon) \qquad (2.2)$$

$$\alpha_i = C \cdot \frac{|\beta_i|}{\max(\forall \beta_i)} \qquad (2.3)$$

The width, n, of the square window, affects the distortion level of edges and small objects. The wider the window size, the more blurry the edges. (Eq. 2.3) normalizes the output of the variance and C represents a user-defined constant. This constant is higher for video watermarking as the human visual system decreases its sensitivity with increasing motion [10], typically C = 10.

2.2 Watermark detection

The detection algorithm is a cross-correlation process. From the possibly corrupted image $I^{'*}$, the spatial pixel information is correlated with each one of the possible embedded watermarks using the following rule [11]:

$$corrcoef = \frac{\sum_{i=1}^{N} I_i^{'*} \cdot W_i}{N} \qquad (2.4)$$

The watermark W for which the detector response is the highest, is assumed to be the embedded mark. To be more resilient to errors, the watermark is embedded twice: once in its normal state and once in its inverted state. Equation. 2.1 then becomes:

$$I_i^{'} = I_i + \alpha_i \cdot W_i \text{ for i<L} \qquad (2.5)$$

$$I_i^{'} = I_i - \alpha_i \cdot W_{i-L} \text{ for i≥L} \qquad (2.6)$$

L represents the mid pixel number. The summation of the two detector responses creates a larger SNR between the cross-correlation of the actual embedded W and I'_w and the cross-correlation of the set of random watermarks and I'_w.

3. SPATIAL DIGITAL WATERMARK FOR MPEG-2

Each frame of a video sequence is equally important for authenticity verification applications. Watermarking each frame with a different watermark can prevent watermark leakage [12]. The algorithm described in section 2 is adapted to enable video watermarking as follows. The pseudo-random watermark seed number represents the frame number. The comparison between the extracted seed number (using equation 2.4) and the frame position in the video sequence allows detection of time-base tampering.

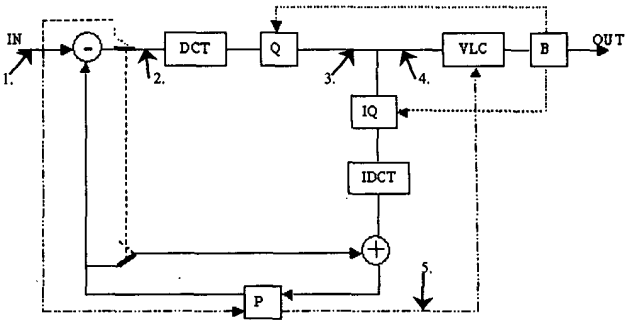

Figure 1. Video coder block diagram with different watermark embedding locations

The round-robin embedding scheme of the labels is a trade off between computation time (the larger the modulus base the more cross-correlation values need to be generated) and detection accuracy. The latter depends on the motion entropy of the sequence. Normally the removal of one frame in an active video sequence is noticeable. However, security camera footages are mostly non-active so that the removal of a few frames can become unnoticeable. To speed up the detection process, the number of checked watermarks has to be reduced to a random subset of watermarks including the possible embedded watermark.

3.1 Watermark embedding location

There are three distinct locations to embed the watermark: before, in or after the motion compensation loop.

Embedding of the watermark in the motion compensation loop (see Figure 1, location 2), using the DFD signal is possible as the DFD signal hardly ever becomes exactly zero. There is always observation noise, occlusion or variation of scene illumination between adjacent frames. Embedding in the same domain as compression (location 3) has proven to be an advantage [1]. The coefficients that become small due to the quantization process can be avoided by the watermarking system in order not to distort the Variable Length Coding (VLC). Embedding of the watermark in the motion vector information (by adding a small change, (location 5)) is not robust to the re-encoding process and hence not favorable. Every watermark embedding system after the motion compensation loop needs a drift-compensation system to neutralize the propagation of errors. (Figure 1, location 4)

Embedding of a watermark in the DFD signal is less robust to compression as there is less 'space' available to host the watermark. In addition, the rate/buffer control system adjusts the quantizer step size, which results in a larger deterioration of the reference frame at lower bit-rates. This results in a low auto-correlation of the DFD

Authorized licensed use limited to: UNIVERSITY OF BRISTOL. Downloaded on February 4, 2009 at 07:04 from IEEE Xplore. Restrictions apply.

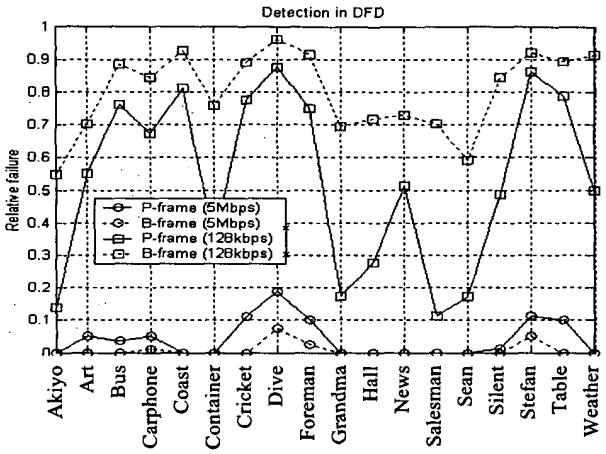

Figure 2. Correlation detector in DFD signal for different bit-rates and test sequences

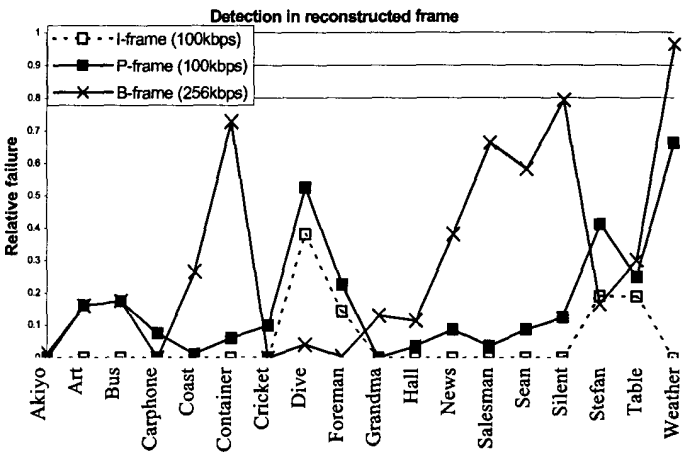

Figure 3. Correlation detector in reconstructed frame for different bit-rates and test sequences

signal at different bit-rates as the motion vectors are pointing to a different location in the reference frame. Hence, embedding a watermark in the DFD signal is less robust to bit-rate changes. Inserting a watermark before the motion compensation loop (see Figure 1, location 1) is more favorable than watermarking the poorly correlated DFD signal.

4. IMPLEMENTATION AND SIMULATION RESULTS

The tag embedding experiment has been carried out on 18 QCIF test sequences (30 frames/sec) compressed with MPEG-2 at different bit-rates: 5Mbps, 2Mbps, 1Mbps, 512kbps, 448kbps, 384kbps, 320kbps, 256kbps, 192kbps, 128kbps and 100kbps. A 300-frame sequence generates 21 I-frames, 80 P-frames and 199 B-frames. The strength factor α is frame type independent. The detection of the embedded label happens in two different locations: once in the DFD and once in the reconstructed signal.

4.1 Detection in DFD signal

The detection performance is proportional to the watermark/image ratio. This number describes the ratio of watermark versus host signal energy, the higher this ratio, the lower the detection failure and visa versa. Reducing the image energy can increase the watermark/image ratio by detecting the watermark in the DFD signal (where the predicted frame is removed from the current frame). The DFD signal of P and B-frames also contain the propagated watermark from the reference frame flipped by 180°. This phase shift results in a negative cross-correlation and does not influence the performance of the detector. Non-active video sequences (akiyo, grandma and salesman) perform better in comparison with active video sequences (cricket, dive and foreman). Because the motion estimation is very effective and the motion compensation function reduces almost all the image content of the frame, this results in a high watermark/image ratio, hence better performance. Detection in the DFD signal is illustrated in Figure 2. It describes the relative failure, the ratio of frames from which the watermark could not be detected versus the total number of frames, in function of the tested sequences. Each graph depicts the performance of different frame types at different bit-rate. The detector is more accurate for B-frames at high compression ratios than P-frames because B-frames are better predicted. (Their prediction is bi-directional.) This is illustrated in Figure 2 for 5Mbps. However, P-frames have a higher detection accuracy at lower bit-rates due to the IPB quantizer mode. B-frames are more heavily quantized than P-frames resulting in a poor watermark/image ratio. This is shown in Figure 2 for 128kbps. The detection system in the DFD signal is only reliable for high bit-rates, e.g. 5Mbps.

4.2 Detection in the reconstructed signal

Prior to detection of the embedded frame number, the local mean is removed from the reconstructed frame. This process can be modeled as a high-pass filter and reduces the interference (cross-talk) between the watermark and the host signal, which is more concentrated in the low frequencies. Equation (2.4) then becomes:

$$corrcoef = \frac{\sum_{i=1}^{N} (I_i'^{*} - \bar{I}_i'^{*}) \cdot W_i}{N} \qquad (2.7)$$

$\bar{I}_i'^{*}$ represents the local mean from a 3x3 neighborhood of pixels. The size of this neighborhood represents the cut-off frequency of the high-pass filter.

The detection reliability in the reconstructed frame depends on the frame type, sequence content and bit-rate. Figure 3 describes the relative failure of different test-sequences for I, P and B-frames compressed at

Authorized licensed use limited to: UNIVERSITY OF BRISTOL. Downloaded on February 4, 2009 at 07:04 from IEEE Xplore. Restrictions apply.

100kbps and 256kbps respectively. The performance of the detector when extracting the frame number in the reconstructed signal depends on the correct-watermark/image ratio. This correct-watermark level is reduced by the level of compression or by the propagation of the reference frame watermark.

The loss of the watermark in I-frames is only due to compression. Complex sequences (highly textured, camera panning, ...), which are more prone to quantization, have a higher relative failure rate. This can be observed in Figure 3 for the dive, stefan, foreman and table test sequence. (dotted line)

The loss of the watermark in P and B-frames are due to compression or propagation of the reference frame watermark. One can distinguish 3 different cases. Firstly, the non-active test-sequences with a high textured background are susceptible to the propagation. Due to the small motion vectors, the negligible shift of the propagated reference watermark reduces the correct-watermark/image ratio, which result in a high relative failure rate. One can observe this in the container, salesman, sean, silent and weather test-sequences of Figure 3.

Secondly, the non-active test-sequences with a low textured background are also prone to propagation. However, the low textured regions of these sequences do not contain a strong presence of the reference frame watermark, hence does not significantly influence the correct-watermark/image ratio. These sequences result in a low relative failure at a bit-rate of 256kbps for B-frames and 128kbps for P-frames. Examples are: akiyo, carphone, grandma and.

Finally, the active test-sequences with a high textured background are more susceptible to compression than to propagation, because the large motion vectors de-synchronize the propagated reference frame watermark. This result in a low relative failure at a bit-rate of 192kbps. Examples are cricket, dive, foreman and stefan.

Figure 4. depicts the frame order as a function of the extracted label for a simulated time-base tampering scenario of the 'Cricket' test sequence compressed at 192 kbps. Ten frames were re-ordered at the beginning and dropped at the end. In addition, two frames were flipped in the middle of the sequence. The round-robin modulus base was set to 100. The tampering can be reversed to a certain extent from the extracted information of Figure 4. Ideally, a straight line is desirable.

5. CONCLUSIONS

This paper establishes a link between detection failure and sequence content. It explains the differences in detection performance at different bit-rates between mutual frame types and detector locations. The proposed method has successfully been tested on 18 different test sequences, allowing detection of time-base tampering in video authentication.

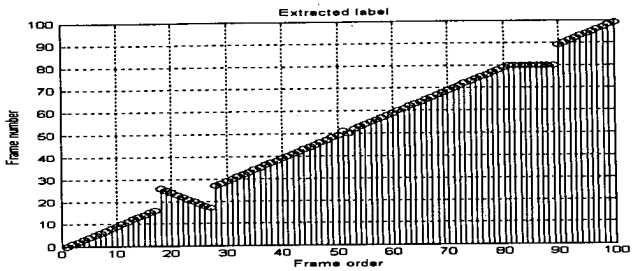

Figure 4. Detector response of a time-base attack: reordering (frame 25), flipping (frame 51) and dropping attack (frame 80)

5. ACKNOWLEDGEMENT

This work was supported by Motorola Research Laboratory UK and EPSRC under grant number GR/M81885.

REFERENCES

[1] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *IEEE Proceedings*, Vol. 87, No. 7, pp. 1076-1107 July 1999.

[2] B. C. Mobasseri, M. J. Sieffert and R. J. Simard, "Content authentication and tamper detection in digital video," *IEEE Int'l Conf. on Image Proc.*, Vancouver, paper ID:1973, Sept. 2000.

[3] F. Hartung and B. Girod, "Digital watermarking of MPEG-2 coded video in the bitstream domain," *Proc. Int'l Conf. on Ac., Sp., & S. P.*, Munich, Vol. 4, pp. 2621-2624, April 1997.

[4] T. Kalker and J. Haitsma, "Efficient detection of a spatial spread-spectrum watermark in MPEG video streams," *IEEE Int'l Conf. on Image Proc.*, Vancouver, paper ID: 1365, Sept. 2000.

[5] F. Jordan, M. Kutter and T. Ebrahimi, "Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video", ISO/IEC Doc. JTC1/SC29/WG11 MPEG97/M2281, July 1997

[6] J.P.M.G. Linnartz and J. C. Talstra, "MPEG PTY-marks: cheap detection of embedded copyright data in DVD-video," *5th ESORICS 98,* Belgium, Vol. 1485, pp. 221-240, 1998

[7] F. Deguillaume, G. Csurka, J. O'Ruanaidh and T. Pun, "Robust 3D DFT video watermarking," *In IS&T/SPIE 11th Elect. Im., Sec. & Watrm. of Mult. Cont.*, San Jose, Vol. 3657-13, pp. 113-124, Jan. 1999.

[8] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing*, Vol. 6, nr. 12, pp. 1673-1687, 1997

[9] V. Bruce, P. R. Green and M. A. Georgeson, *Visual Perception, Physiology, Psyhcology, and Ecology*, Psychology Press, UK, pp. 25-27, 1997.

[10]S. Suthaharan, S-W. Kim, S. Sathananthan, H-K. Lee and K.R. Rao, "Perceptually tuned video watermarking scheme using motion entropy masking", *IEEE Proceedings of TENCON*, Korea, pp. 182-185, 1999

[11] M. Barni, F. Bartolini, V. Cappellini and A. Piva, "Copyright protection of digital images by embedded unperceivable marks," *Image and Vision Computing*, Vol. 16, pp. 897-906, February 1998.

[12] M. D. Swanson, B. Zhu and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE Journal in communications*, Vol. 16, No. 4, May 1998.

Authorized licensed use limited to: UNIVERSITY OF BRISTOL. Downloaded on February 4, 2009 at 07:04 from IEEE Xplore. Restrictions apply.