BLIND MOTION-COMPENSATED VIDEO WATERMARKING

Peter Meerwald and Andreas Uhl

Department of Computer Sciences,
University of Salzburg, Jakob-Haringer-Str. 2, A-5020 Salzburg, Austria
Email: {pmeerw, uhl}@cosy.sbg.ac.at

ABSTRACT

The temporal correlation between adjacent video frames poses a severe challenges for video watermarking applications. Motion-coherent watermarking has been recognized as a strategy to embed watermark information in video frames, resistant to collusion attacks. The motion-compensated temporal wavelet transform (MC-TWT) provides an efficient tool to separate static and dynamic components of a video scene and enables motion-coherent watermarking.

In this paper, we extend a MC-TWT domain watermarking scheme with blind detection, i.e. motion estimation and watermark detection is performed without reference to the unwatermarked video. Our results show that motion-coherent watermarking can be combined with a blind detector, widening the applicability of MC-TWT domain watermarking beyond forensics (where the unwatermarked content is assumed to be available).

Index Terms— motion-coherent, blind watermarking

1. INTRODUCTION

Watermarking has been proposed as a technology to ensure copyright protection by embedding a signal in digital multimedia content such as video [1]. Direct application of image watermarking schemes on the individual video frame gives rise to inter-frame attacks [2]. Adjacent video frames are typically highly correlated along the temporal axis. This fact can be exploited by averaging frames in case of an uncorrelated watermark or by performing perceptual remodulation of the averaged per-frame watermark estimate (WER attack [3]). To counter above attacks, the embedded watermark should exhibit correlation similar to the host signal frames [4], i.e. the watermark should be motion-coherent [5].

Frame registration and temporal transforms employing motion-compensation (MC) have been proposed as tools to align components of a video scene [6]. While the temporal transform approach uses block-based motion estimation (ME) to track motion of background and foreground objects, the frame registration technique merely separates and aligns the background. Motion-compensated frame prediction and evaluation of the local variance statistics of the residual frame has been proposed to assess the motion-coherency of a video watermarking scheme [7].

In this paper, we propose a blind video watermarking scheme based on a motion-compensated temporal wavelet transform. It extends the work of Pankajakshan et al. [6] by employing blind ME and blind watermarking detection, i.e. without reference to the unwatermarked content.

Supported by Austrian Science Fund project FWF-P19159-N13.

In section 2 we review motion-compensated watermarking and propose our novel blind detection scheme. Experimental results are presented in section 3, followed by concluding remarks in section 4.

2. MOTION-COHERENT WATERMARKING

Early video watermarking schemes simply adopted image watermarking techniques on a per-frame basis. Two prototypical key schedules, repetitive and independent watermarking, can be distinguished, i.e. the same key is used for all frames or a different key is used to generate the watermark signal for each frame. In case of independent frame watermarking, flickering may become noticeable even when the watermark is imperceptible for each frame.

Furthermore, the redundancy between video frames permits to drop or swap frames to hinder synchronization, but also gives rise to powerful watermark estimation and collusion attacks which threaten the security of the watermarking scheme by revealing information on the secret watermark signal. Only recently, the notion of watermark security has been established alongside watermark robustness. In this paper we do not consider synchronization or inter-video attacks but concentrate on inter-frame attacks.

A repetitive video watermark can be attacked by estimating and remodulating the watermark's high-frequency components in each frame (e.g. via Wiener filtering [3]). The watermark estimate can be refined by combining estimates derived from dissimilar frames thus exploiting the redundancy of the watermark signal.

An independent video watermark is susceptible to the frame temporal filtering (FTF) or collusion attack: representing adjacent video frames by their temporal low-pass approximation averages out the uncorrelated watermark in the high frequency components. This attack's effectiveness can be greatly increased by employing MC-FTF [6] or FTF after frame registration [5].

Watermarking schemes aim to cope with the redundancy between the host frames using temporal transforms: Swanson et al. [8] apply temporal wavelet filtering to separately mark static (low-pass approximation) and dynamic (detail subbands) components of the video. 3D DCT [9] and DFT [10] transforms have also been proposed. Recently, watermarking schemes have been presented which explicitly take video motion into account to resist MC-FTF attacks. Kundur et al. [4] depend on anchor points to embed a correlated watermark in similar host video components, Doërr et al. use frame registration to align the video's background component before watermarking. Pankajakshan et al. [6] embed the watermark in the low-pass approximation obtained by a motion-compensated temporal wavelet transform (MC-TWT) [11]. Figure 1 shows the temporal low-pass frame with and without MC of the first 16 Foreman sequence frames.

(a) Frame 1 (b) Frame 8 (c) Frame 16 (d) Temporal lowpass (e) MC temporal lowpass

Fig. 1. Frames of the CIF Foreman sequence, (a) to (c). Temporal low-pass (average) of the first 16 frames with (d) and without (e) MC.

For detection of a motion-coherent watermark, the video motion information is required. In case the detector has access to the original host video, i.e. non-blind or private detection, an accurate motion model is available. Clearly, this requirement restricts the range of possible application scenarios, e.g. to forensic watermarking. For blind watermark detection, i.e. detection without reference to the original host, the approximate motion model has to be estimated from the watermarked – and potentially further altered – video. Robustness of the more versatile blind detector therefore also depends on the robustness of the motion model.

In the next section, we review the scheme of Pankajakshan et al. and then extend it to blind watermark detection. The MC-TWT offers the advantage of an efficient, fine-grained motion model based on block-based ME to track both foreground as well as background video components and is compatible with potential future video coding standards [6, 11, 12].

2.1. MC-TWT watermarking

The MC-TWT can be efficiently computed via lifting steps. Here, we follow the notation of [6] and restrict ourselves to the Haar wavelet and a motion model, M, with integer pixel accuracy. Extensions to the $5/3$ wavelet for bidirectional filtering and sub-pixel accuracy motion can be found in [11].

A video sequence is split into scenes of N frames, $\{X_k[n],\ k = 0, 1 \ldots, N-1\}$, which are recursively decomposed in low-pass, l_k^i, and high-pass, h_k^i, temporal frames of decomposition level i,

$$h_k^i[n] = l_{2k+1}^{i-1}[n] - l_{2k}^{i-1}[M_{2k\cdot 2^{i-1}\to(2k+1)\cdot 2^{i-1}}(n)] \quad (1)$$

$$l_k^i[n] = l_{2k}^{i-1}[n] - \frac{1}{2}h_k^i[M_{(2k+1)\cdot 2^{i-1}\to 2k\cdot 2^{i-1}}(n)] \quad (2)$$

where $k = 0, 1, \ldots, {}^{N}/_{2^i} - 1$ and $l_k^0[n] = X_k[n]$. A spread-spectrum watermark, $W[n]$, is then added to the temporal low-pass frame

$$\hat{l}_0^\iota[n] = l_0^\iota[n] + W[n] \quad (3)$$

of maximum level ι. The marked video sequence is obtained by the reconstruction steps given by

$$l_{2k}^i[n] = l_k^{i+1}[n] - \frac{1}{2}h_k^{i+1}[M_{(2k+1)\cdot 2^i\to 2k\cdot 2^i}(n)] \quad (4)$$

$$l_{2k+1}^i[n] = h_k^{i+1}[n] + l_{2k}^i[M_{2k\cdot 2^i\to(2k+1)2^i}(n)]. \quad (5)$$

After embedding the watermark in the low-pass temporal frames at decomposition level ι, the resulting reconstructed, watermarked frames $\hat{X}_k$ carry the same watermark sample in different frames along the motion trajectories (assuming composition and invertibility of the motion vectors):

$$\hat{X}_k[n] = \begin{cases} X_k[n] + W[n] & k = 0 \\ X_k[n] + W[M_{0\to k}(n)] & k = 1, \ldots, {}^{N}/_{2^\iota} - 1. \end{cases} \quad (6)$$

Watermark detection can be performed by computing the normalized correlation,

$$NC(\tilde{W}, W) = \frac{< \tilde{W}, W >}{\|\tilde{W}\| \cdot \|W\|}, \quad (7)$$

between the embedded watermark, $W[n]$, and the extracted watermark, $\tilde{W}[n]$, from a potentially altered frame $\tilde{X}_k$,

$$nc_k = NC(\tilde{X}_k[n] - X_k[n], W[M_{0\to k}(n)]). \quad (8)$$

and comparing nc_k against a detection threshold $T_{NC}(P_{fa})$ designed to yield a probability of false-alarm, P_{fa}, suitable for a given application,

$$nc_k \overset{?}{\geqq} T_{NC}(P_{fa}). \quad (9)$$

This non-blind detector, designed for Gaussian noise interference, subtracts the original frames in eq. 8 to suppress the non-Gaussian interference due to the host signal.

2.2. Blind detection

When the original host signal is not available to the watermark detector, the watermark has to be correlated directly with received video frames, $\tilde{X}_k[n]$, instead of the extracted watermark. The host signal acts as noise and interferes with watermark detection. By applying a block-wise 8×8 DCT transform on the temporal low-pass frames and adding the watermark only to the mid-frequency bands of the transform blocks, substantial energy of the host signal can be rejected. It is well known that the mid-frequency coefficients of the 8×8 DCT can be modeled by a generalized Gaussian distribution for which an optimal detector has been derived [13].

For our blind MC-TWT video watermarking we select 18 frequency bands, band 3 to 21 in zig-zag scan order, from the 8×8 DCT blocks of the temporal low-pass frame l_0^ι. We construct a frequency domain bipolar watermark, $W'[\eta]$, where only the coefficients in the selected bands are non-zero. The marked temporal low-pass frame is then obtained by

$$\hat{l}_0^\iota[n] = DCT_{8\times 8}^{-1}(DCT_{8\times 8}(l_0^\iota[n]) + W'[\eta]). \quad (10)$$

Applying the inverse 8×8 DCT on $W'[\eta]$ yields the spatial domain watermark

$$W[n] = DCT_{8\times 8}^{-1}(W'[\eta]). \quad (11)$$

Due to the linearity of the DCT, it follows that

$$\hat{l}_0^\iota[n] = l_0^\iota[n] + W[n]. \quad (12)$$

For blind watermark detection we construct the vectors v and w from the selected frequency bands of $DCT_{8\times 8}(\tilde{X}_k[n])$ and $DCT_{8\times 8}(W_k[n])$, $W_k[n] = W[\tilde{M}_{0\to k}(n)]$, respectively, and compute the generalized Gaussian detection statistic

$$GGd_k(v, w) = \sum_j \beta(\left|v_j\right|^c - \left|v_j - w_j\right|^c), \quad (13)$$

Sequence	Non blind	Blind ME			
		SR 16, L 4	SR 32, L 4	SR 16, L 3	SR 32, L 3
Foreman	1.00	0.80	0.79	0.90	0.89
Coastguard	1.00	0.48	0.45	0.63	0.60
Akiyo	1.00	0.98	0.98	0.99	0.99
Mobile	1.00	0.34	0.29	0.45	0.38
Stefan	1.00	0.47	0.47	0.64	0.61

Table 1. Normalized Correlation (nc) results for watermark detection with non-blind and blind motion estimation (ME) for different search ranges (SR) and temporal decomposition levels (L).

where the shape parameter of the distribution, c, and β are computed using maximum-likelihood estimation on v. Note that for blind detection, an approximate motion model, $\tilde{M}$, has to be estimated from the received video frames $\tilde{X}$.

The detection statistic is again compared against a decision threshold, $T_{GGd}(P_{fa})$, to decide upon the watermark presence. We can turn the above motion-coherent watermarking into a repetitive or independent watermarking scheme by setting $W_k[n] = W[n]$ or generating uncorrelated $W_k[n]$, respectively.

3. EXPERIMENTAL RESULTS

We have implemented the reference non-blind scheme [6] for comparison and the proposed blind watermarking schemes with per-frame repetitive and independent watermarking as well as motion-coherent watermarking. The MC-TWT is performed with Haar wavelet lifting with a decomposition level of 4 and integer pixel accuracy.

The same binary bipolar watermark has been embedded in the luminance component of all sequences with no perceptual masking applied. The embedding strength has been adjusted for all algorithms so that the average PSNR of the watermarked video is around 38 dB.

For ME, a simple hierarchical variable size block matching (HVSBM) technique with minimal block-size 8×8 pixels and integer-pixel accuracy is adopted. We have chosen the first 64 frames of the widely available video sequences Foreman, Coastguard, Akiyo, Mobile and Stefan in CIF format, 352×288 pixels. A characterization of the motion of these video sequences can be found in [7]. The reported results were obtained by averaging the per-frame results over 5 test runs.

3.1. Evaluation of ME robustness

First we evaluate the impact of blind ME, see table 1. Given the original motion information, the detector can perfectly recover the embedded watermark. However, when ME has to be performed on the watermarked video, the detection performance degrades, strongly depending on the video content. The detection improves when constraining the search range or decomposition level.

Next, we assess the robustness of the non-blind detector under H.264 and MC-EZBC [12] compression attack with bit rates ranging from 3000 to 250 kbit/s and contrast the performance with (simulated) blind ME. Figure 2 presents the plots for the Foreman and Coastguard sequence. The lack of accurate motion information decreases the detector response, but the NC result stays well above the detection threshold of 0.02 for a $P_{fa} = 1e^{-6}$.

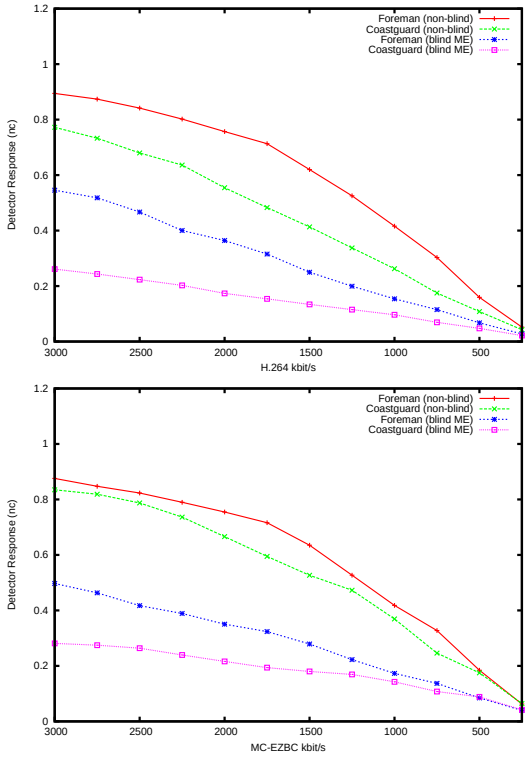

Fig. 2. Robustness of the non-blind MC-TWT watermark detector against H.264 and MC-EZBC compression and impact of blind ME.

3.2. Robustness of the blind detector

The blind MC-TWT watermarking scheme's robustness against H.264 and MC-EZBC compression is illustrated in figure 3 using the Foreman and Coastguard sequence. We plot the ratio $d = (GGd - T_{GGd}(1e^{-6}))/\sigma^2$, where T_{GGd} is the detection threshold and σ^2 the estimated variance of the generalized Gaussian detection statistic. The decrease in detection performance due to inaccurate ME is less pronounced compared to the non-blind detector. Only for bit rates less than 250 kbit/s the watermark cannot be detected reliably.

3.3. FTF attack on the blind detector

We test our proposed blind watermarking scheme with repetitive, independent and motion-coherent watermarking with FTF and MC-FTF, i.e. inter-frame collusion attacks. For the FTF attack we confine the investigation to a collusion window size of 3 as higher values lead to very noticeable motion blur, compare with figure 1 (d) and (e). MC-FTF is performed with window size 7, nevertheless PSNR is consistently higher.

As expected, FTF is ineffective against the repetitive watermark. The motion-coherent watermark is more resistant against MC-FTF than the repetitive or independent watermark.

4. CONCLUSION

We have extended MC-TWT domain watermarking with blind detection. Although the inaccurate motion information derived by the blind detector impairs robustness, the motion-coherent watermark

Sequence	FTF Attack (window size 3)						MC-FTF Attack (window size 7)					
	Repetitive WM		Independent WM		Motion-coherent WM		Repetitive WM		Independent WM		Motion-coherent WM	
	PSNR (dB)	d	PSNR (dB)	d	PSNR (dB)	d	PSNR (dB)	d	PSNR (dB)	d	PSNR (dB)	d
Foreman	33.04	1.35	34.02	0.45	33.75	0.35	36.96	0.79	37.96	0.60	36.92	0.86
Coastguard	30.16	1.33	31.01	0.42	30.94	0.29	33.47	0.51	33.84	0.38	33.12	0.68
Akiyo	38.18	1.35	41.66	0.49	38.49	0.86	38.25	1.12	41.33	0.75	38.36	0.96
Mobile	26.62	1.33	27.69	0.41	27.46	0.36	28.53	0.69	29.01	0.52	28.58	0.76
Stefan	26.06	1.56	27.07	0.52	26.63	0.47	30.06	0.63	31.02	0.53	30.26	0.79

Table 2. PSNR and detector response results for the FTF and MC-FTF attack with collusion window 3 and 7, respectively.

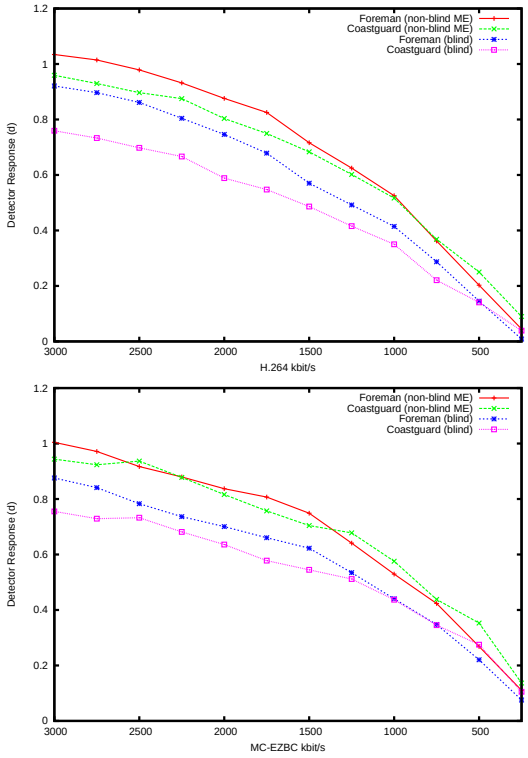

Fig. 3. Robustness of the blind MC-TWT watermark detector against H.264 and MC-EZBC compression, contrasted with (simulated) non-blind ME.

remains detectable even under severe compression. There clearly is a trade-off to be made between robustness and watermark security.

The motion-coherent watermark can either be detected in the temporal low-pass frame, permitting progressive, blind detection integrated in MC-TWT based video codecs such as MC-EZBC [12], or in the decoded frames. Further research will evaluate watermark estimation attacks and assess the robustness against explicit tampering with block-based ME.

5. REFERENCES

[1] G. Doërr and J.-L. Dugelay, "A guide tour of video watermarking," *Signal Processing: Image Communication*, vol. 18, no. 4, pp. 263–282, Apr. 2003.

[2] G. Doërr and J.-L. Dugelay, "Security pitfalls of frame-by-frame approaches to video watermarking," *IEEE Trans. on Signal Processing*, vol. 52, no. 10, pp. 2955–2964, 2004.

[3] S. Voloshynovskiy, S. Pereira, A. Herrigel, N. Baumgärtner, and T. Pun, "Generalized watermark attack based on watermark estimation and perceptual remodulation," in *Proceedings of IS&T/SPIE, Security and Watermarking of Multimedia Content II*, San Jose, CA, USA, Jan. 2000, vol. 3971.

[4] K. Su, D. Kundur, and D. Hatzinakos, "Statistical invisibility for collusion-resistant digital video watermarking," *IEEE Trans. on Multimedia*, vol. 7, no. 1, pp. 43–51, Feb. 2005.

[5] G. Doërr, J.-L. Dugelay, and D. Kirovski, "On the need for signal-coherent watermarks," *IEEE Trans. on Multimedia*, vol. 8, no. 5, pp. 896–904, May 2006.

[6] V. Pankajakshan and P. K. Bora, "Motion-compensated interframe collusion attack on video watermarking and a countermeasure," *IEE Proceedings on Information Security*, vol. 153, no. 2, pp. 61–73, June 2006.

[7] V. Pankajakshan, G. Doërr, and P. K. Bora, "Assessing motion coherency in video watermarking," in *Proceedings of the ACM Multimedia and Security Workshop, MMSEC '06*, Geneva, Switzerland, Sept. 2006, pp. 114–119, ACM.

[8] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 540–550, Apr. 1998.

[9] H. Park, S. H. Lee, and Y. S. Moon, "Adaptive video watermarking utilizing video characteristics in 3D-DCT domain," in *Proceedings on the 5th International Workshop on Digital Watermarking, IWDW '06*, Korea, Nov. 2006, vol. 4283 of *Lecture Notes in Computer Science*, pp. 397–406, Springer.

[10] Y. Y. Lee, H. S. Jung, and S. U. Lee, "Multi-bit video watermarking based on 3D DFT using perceptual models," in *Proceedings of the 2nd International Workshop on Digital Watermarking, IWDW '03*, Seoul, Korea, Oct. 2003, vol. 2939 of *Lecture Notes in Computer Science*, pp. 301–315, Springer.

[11] J.-R. Ohm, M. van der Schaar, and J. W. Woods, "Interframe wavelet coding – motion picture representation for universal scalability," *Signal Processing: Image Communication*, vol. 19, no. 9, pp. 877–908, Oct. 2004.

[12] S.-T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," *Signal Processing: Image Communication*, vol. 16, no. 8, pp. 705–724, May 2001.

[13] J. R. Hernández, M. Amado, and F. Pérez-González, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 55–68, Jan. 2000.