A Review of Video Watermarking and a Benchmarking Framework

V. Moutselakis, S. Tsekeridou
Democritus Univ. of Thrace, Dept. of Electrical & Computer Engineering,
67100 Xanthi, Greece
{emoutsel,tsekerid}@ee.duth.gr

Abstract. The rapid use of Internet has led to the investigation of digital watermarking as a complementary technology to traditional protection mechanisms. Significant research efforts and review works presenting unifying characteristics of different methods have been reported for audio and image watermarking. In the context of video watermarking, though, there is a great deal of non-uniformity in presented approaches. The objective in this paper is to give an in-depth overview of different video watermarking techniques in order to single out the particularities of that field. Furthermore, the paper presents a benchmarking framework for objective video watermarking performance evaluation. We conclude that novel techniques need to be implemented and unexplored video-driven approaches have to be investigated.

1 Introduction

The rapid growth of Internet and networked multimedia systems in the past decade has raised concerns from the content designers, since multimedia data nowadays can be flawlessly copied and rapidly disseminated at large scale. Encryption and steganography were proved to be insufficient for digital media protection and thus digital watermarking emerged, aiming at embedding auxiliary information into a host digital signal by imposing secure, imperceptible signal changes (with the employment of a special constructed signal, called watermark that is embedded into original content such as image, video, or audio, producing a watermarked signal). Digital watermarking allows the user to manipulate the content.

We focus on digital video watermarking, where time enhances the flexibility of the solution space. Available data are greater than image data, a fact that during watermark design is useful both for the designer and the attacker as it supports reliable embedding of auxiliary data using sophisticated temporal masking, but also, allows the attacker to make greater use of correlators that lead to more effective watermark estimation and removal attacks.

There is a great academic and industrial interest on the design of a copyright protection system for MPEG-2 coded video distributed on Digital Versatile Disk (DVDs), employing the

Please use the following format when citing this chapter:
Moutselakis, Vangelis, Tsekeridou, Sofia, 2006, in IFIP International Federation for Information Processing, Volume 204, Artificial Intelligence Applications and Innovations, eds. Maglogiannis, I., Karpouzis, K., Bramer, M., (Boston: Springer), pp. 665–672

digital video watermarking technology [25]. A video watermarking system has also been designed by the Galaxy Group to complement the existing content scrambling system (CSS) that is part of the DVD standard; the technology is now called WaterCast and is being applied in the automatic monitoring of digital video broadcasts [9].

In Fig. 1 a general model is provided presenting the entire video watermarking process.

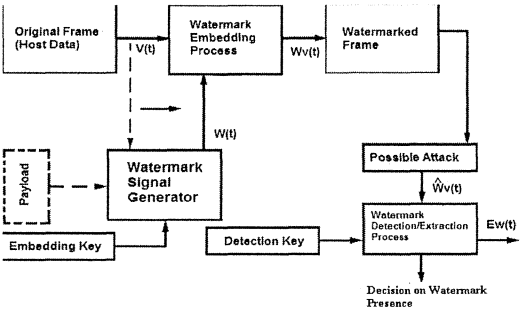

Fig. 1. Video watermarking process model

The video watermarking process consists of two main stages: Watermark Generation and Embedding and Watermark Detection and/or Extraction. At first, the watermark signal generator creates the watermark signal and is provided with an embedding key (the use of a secret such key, to create and embed the watermark is often required for security reasons) and possibly a payload (auxiliary information), and produces $W(t)$, the watermark, to be inserted into the video. Some watermarking techniques further use the original video frame sequence $V(t)$ to achieve more effective watermark embedding. Once the watermark is constructed, it is inserted into the original video frame to produce the watermarked video frame. The specific methods by which the watermark is constructed and embedded is dependent on the watermarking technique. The output of the embedder is the watermarked video $W_v(t)$.

In Fig. 1, $\widehat{W}_v(t)$ denotes the watermarked video that is possibly attacked and is provided to the detector. If the video has not been attacked, then $W_v(t)$ is identical to $\widehat{W}_v(t)$ for all t. The watermark detector examines the received video and determines if the watermark is present. In Fig. 1, $E_w(t)$ denotes the extracted watermark. The detector is also provided with a detection key necessary for the detection of the watermark. A symmetric (private key) watermark uses identical embedding and detection keys, whereas asymmetric (public key) watermarks use distinct but related such keys, similar in concept to public key cryptography.

In the sequel, a review of existing video watermarking techniques is given and a video watermarking benchmark framework is proposed.

2 Types of Attacks

An attack is any processing that aims at impairing watermark detection or communication of information conveyed by it [5]. An attack causes watermarked video to be altered, intending to remove the embedded watermark or make detection more difficult (*intentional attacks*).

Watermarked data on the other hand is often processed in some way prior to detection. This may include compression, signal enhancement, or digital-to-analog (D-A) and analog-to-digital (A-D) conversion. Thus, we should take into account the case that an embedded watermark is unintentionally impaired by such processing (*non-intentional attacks*).
In this section, we concentrate mainly on *intentional attacks*:

Simple or noise/waveform attacks: attempt to modify both host data and watermark without intending to trace and remove the watermark. Linear/non-linear, temporal/spatio-temporal filtering, waveform-based compression, noise addition are included in this category.

Geometric attacks (or *synchronization attacks*): are accomplished by geometrically transforming the data. For video data, this means frame spatial shift, frame rotating and temporal filtering attacks. The watermark is not ultimately removed by the data (as the goal of these attacks is to force the detector to confront a more difficult synchronization problem), so it is possible to successfully detect and recover it. Temporal synchronization attacks in video include frame dropping, insertion, transposition, averaging (temporal interpolation or scaling).

Removal attacks: are focused on detecting the watermark, isolating it from the host data and eventually removing it, without breaking the security of the watermarking algorithm (e.g., without the key used during watermark embedding, as in [4]). This category includes *denoising*, *quantization* (e.g., for compression), *remodulation*, and *collusion attacks* (these occur when an attacker obtains collections of video frames that are analyzed or combined with the purpose of producing a non watermarked copy of the original).

Forging attacks: attempt to sabotage the owner's watermark, that is, the attacker wants to forge the original watermark.

Statistical attacks: try to detect the embedded watermark by comparing and finding similarities among a number of watermarked signals that belong to the same owner (whereas collusion attacks involve many copies of a given data set, each signed with a different key).

Protocol attacks: attempt to subvert the security of the watermark, hence attack the entire concept of the watermarking application. They do not directly impact watermark detection.

Ambiguity attacks are based on the concept of invertible watermarks. The malicious forger knows that the data are watermarked. He tries to subtract his own watermark from the watermarked data to later claim to own them and therefore cause uncertainty regarding their true owner. It is essential for copyright protection applications to employ non-invertible watermarks to eliminate the possibility of ambiguity attacks.

Another attack in this category is the *copy attack*: it aims at estimating a watermark from the watermarked data and copies it to some other "target" data without ultimately destroying the watermark or hindering its detection [5].

3 Video Watermark Embedding Methodologies

The embedding process of a watermark into multimedia signals is divided in three categories regarding the entry domain:

1) The watermarks that are constructed in the *spatial/temporal domain*, commonly named as *spatial watermarks*.

[1] models a multi-stage watermarking process. The amount of watermarking imposed on a specific stage counterbalances the quality of the final result. Each selected stage is watermarked by selecting a set of "constraints" (that indicate the presence of the author's

signature), then using preprocessing of the stage's input and post processing of the stage's output to ensure that a disproportionate number of these constraints are satisfied.

In [2], the embedding process employs meaningful information bits in the luminance mean values of each frame. To deal frame removal attacks, synchronization bits are also integrated alternating with the watermark information bits (in both cases a pseudo random sequence (PRS) generator of different length is used). The watermark PRS values are per frame embedded by modifying the mean luminance value of individual frames.

In [10], a state machine key generator is used to produce time-invariant, time-independent, and time-periodic key schedules, to support temporal synchronization for blind video watermarking. The design of the watermark and its key schedule affect the ease of synchronization. The use of a feature vector allows the key sequence produced to be video-dependent. A video-dependent key schedule can increase the difficulty of inverting the watermark and make it more robust against ownership [26] and copy [27] attacks but may cause temporal synchronization loss due to attacks changing the feature vectors [10].

2) The watermarks incorporated into the *frequency/transform domain*, commonly named as *spectral (or transform-based) watermarks.*

They are integrated within the related transform coefficients. In particular, they involve use of DCT, DWT and DFT or FFT within the embedding process. In video watermarking, significant research efforts are reported to employ 3D DCT, 3D DWT, 3D DFT [24], 3D TWT [12]. The *Temporal Wavelet Transform* (TWT) has scalable temporal resolution.

In [4] the Integer-to-Integer DWT (IIDWT) is used so that both the input and output data to DWT are characterized in integer values. The watermark data is embedded in high frequency regions [4], [6] to improve the watermark effectiveness. Embedding is done in only those coefficients whose norm is greater than a specified threshold in order to achieve perceptual invisibility and robustness against MPEG encoding and re-encoding.

3) The watermarks inserted at the *compressed domain.*

The process of partial or full decompression of video files is skipped thus avoiding quality loss and extra computational cost. The watermarking process can be executed in real time.

In [8] the watermark is embedded directly in an MPEG-2 compressed bit stream by intentionally forcing bit errors. Thus, the error recovery option of a bidirectionally decodable packet (initially used to handle channel errors in [22]) is exploited so as to embed and retrieve the watermark. Reversible VLCs (RVLC) exhibiting error resiliency are implemented due to their two-way decoding directions capabilities. The watermark is encrypted prior to insertion to make it indistinguishable from randomly extracted bits.

4 Video Watermark Detection/Extraction Methodologies

A prevalent classification of watermark detection is based on whether the original data are used or not. Specifically, if the watermark detector does not require access to the original signal, the watermarking technique is called *blind*. Otherwise, it is known as *non-blind.*

A cryptographic system used in [1] based on public key encryption prevents the forger from discovering a set of constraints that match the original signature. A single metric, P_c is used, showing the probability of how many of the selected constraints (used to map an author's signature) are satisfied. Basically, P_c is the probability of a non-watermarked solution carrying the watermark. If the value of P_c is very low, the more effective the watermark scheme is. P_c is calculated as a sum of binomials, as shown in [1].

Table 4.1. Performance of method discussed in [1] under various attacks

Attack	*Performance*
Ambiguity Attacks	Brute-force attacks become computationally infeasible if the proof of authorship threshold is set sufficiently low (e.g., $P_c \leq 2^{-56}$)
Removal Attacks	Possible for an attacker to use tampering methods to remove a signature known to him, or to add an entirely new signature.
Forging Attacks	Successfully prevented when using a key encryption system

In [2], the detection process is based on the cross correlation of the embedded PRSs and the video frame mean luminance sequence. Each video frame has different luminance, so the use of an amplitude limiting filter followed by a whitening filter prior to correlation is proposed, in order to improve the detector performance.

Table 4.2. Performance of method discussed in [2] towards different kinds of attacks

Attack	*Performance*
Geometric Attacks: (frame *spatial shift* and frame *rotating* attacks)	Good performance against frame spatial shift attacks and frame rotating attacks.
Frame Removal Attacks and Temporal Filtering Attacks	Successfully prevented. Detection based on application of a low pass-filter on the luminance values and observation of the corresponding cross-correlation

In [4], the whole watermark extraction process occurs in the decoded video per frame based on a detection key. The averaged watermark obtained is compared to the embedded original watermark in order to ensure that it is exactly the same. Each frame is randomized prior to embedding of the watermark according to the value of a pair of keys, derived from the detection key, to successfully deal collusion and statistical attacks. In [4], it is proved that low values of PSNR (<34db) are obtained with quite big watermarks (> 25 bits).

Table 4.3. Performance of method in [4] under MPEG encoding and re-encoding

Attack	*Performance*
MPEG Encoding	Average BER ranges from 20-23% and increases steadily when the watermark length is 24 bits and over. Compared to the common DWT technique, the IIDWT one has a 4-8% better performance whatever the watermark length.
MPEG Re-encoding (2 MPEG encoding iterations)	Overall BER increased about 2-3% compared to single MPEG encoding. More robust than DWT, by an 8% difference in BER rate

In [8] the watermarked VLC must be identified the moment it is decoded. Therefore, the inserted watermark must immediately cause decoding failure in order to trigger reverse decoding that begins from the end-of-packet. To ensure forward detection failure right at the edge of a watermarked VLC, the decoded watermarked bit stream must begin with a sequence of so called *flag bits*, guaranteeing detection failure. If the packet length is known to the decoder, the last VLC to be recovered on reverse decoding is the same VLC that failed detection on forward decoding. At the end of this process, the watermark bits are extracted, whereas the stream is restored to its initial state. If the packet length is unknown to the decoder, another flag is used, a *reverse flag*, that causes detection failure on reverse decoding. The watermarking process of compressed media in the VLC domain is inherently fragile since the watermark is vulnerable to re-compression or transcoding. Errors during the detection process-when an incorrect watermark was decoded- are significantly low (they range from 0 - 0.15%) and they are not proportionally affected by the file size.

In [10] a model for symmetric blind video watermark detection is described using a detection key. The watermark detector applies a spatial de-correlating filter to reduce the host-signal interference, followed by a correlation detector and comparison with a threshold.

Table 4.4. Performance of method in [10] towards various synchronization attacks

Attack	*Performance*
Frame Dropping	Poor performance in cases of little temporal redundancy.
Frame Transposition	Performance similar to the frame dropping one.
Frame Insertion	Does not affect watermark detection. Method achieves a detection rate of 100%.

5 Performance Evaluation

One of the metrics widely used to evaluate watermarking schemes is the *False acceptance rate (FAR).*FAR states the probability that an unknown individual will be falsely 'recognized' as the rightful owner of the reference video data upon presentation of his or her verification data. FAR is dependent on the selected tolerance limit within which the verification and reference data must match for there to be a successful authentication: the lower the tolerance limit, the lower the FAR and the higher the probability of FRR errors.

The *False rejection rate (FRR)* metric states the probability that the rightful owner of the reference data will be wrongly rejected. FRR is dependent on the tolerance limit within which the verification and reference data must match for there to be a successful authentication: the higher the tolerance limit, the lower the FRR and the higher the probability of FAR errors.

A quite simple metric used for evaluation is the *Bit Error Rate (BER)* that denotes the number of error bits divided by the watermark length (BER is calculated per frame).

6 Benchmarking Framework for Video Watermarking

Since the complete theoretical analysis of a watermarking algorithm performance with respect to different attacks is rather complicated, the developers of watermarking algorithms refer to the results of experimental testing performed in the scope of some benchmark. The benchmark combines the possible attacks into a common framework and weights the resulted performances depending on the possible application of the watermarking technology.

In image watermarking, several benchmarking tools have been developed to evaluate different methodologies, such as **stirmark**, **checkmark** and **optimark**. *Stirmark* is a generic tool provided with a watermarked input image, that generates a number of modified images used to verify watermark existence after a number of attacks. Stirmark proposes combination of different detection results and computation of an overall score. Stirmark has limited potentials for sophisticated image watermarking schemes as it does not properly model the watermarking process. *Optimark* is a benchmarking tool [32] that supports various attacks and employs differentiated performance metrics depending on the type of the detector used (and the output it produces) as well as on the characteristics of the watermarking algorithm.
Likewise, the main design challenges for a video benchmark framework are listed below:

- Detection performance evaluation using multiple trials employing different sets of data
  - For watermarking schemes that follow frame-by-frame approaches where a different watermark is inserted in each video frame, the chosen set of data must include all the different watermarks used in order to evaluate their robustness,
  - For watermarking schemes that also follow frame-by-frame approaches but embed the same watermark in all video frames, a much smaller set of frames need to be chosen,
  - For more sophisticated watermarking methodologies based on a compression standard or embed a watermark in a three-dimensional (3-D) transform, the chosen set of data must be carefully chosen in order to ensure that all the possible watermarks used are evaluated and also the range of the testing data excludes the possibility of estimation errors (i.e. the case where the entire set of data is unwatermarked),
- Evaluation of the following detection/decoding performance metrics:,
  - Bit error rate,
  - Signal to Noise Ratio (SNR), and Peak Signal to Noise Ratio (PSNR) (as indirect measures for watermarked video quality estimation),
  - False acceptance rate,
  - False rejection rate,
- Evaluation of the mean embedding and detection time,
- Interface to input watermarking schemes and deploy watermark embedding and detection processes, thus weighing the outputs for certain attacks based on the target application
- Option for the user to choose any combination of attacks based on the target application.

Types of attack that could be included in such a benchmarking tool are: Copy attacks, Geometric attacks, Simple Waveform attacks (i.e. MPEG compression), Removal attacks (such as denoising, frame removal, frame linear transformations).

7 Conclusions and Future Prospects

Video watermarking is a recent area of exploration of digital watermarking. The increasing concern of multimedia owners for copyright protection motivates further research here. In this paper, we have reviewed a number of existing video watermarking schemes that cover a wide range of applications, varying from frame-based watermarking to more sophisticated video specific watermarking in a three-dimensional space. Furthermore, we compared a number of existing video watermarking techniques performance against attacks, and found that there is indeed room for improvement since all attacks cannot be completely dealt with. We further need to define detailed constraints based on the targeted application. We have also proposed a benchmarking framework for video watermarking presenting the main requirements to be met in order to objectively evaluate and rate a wide range of video watermarking methodologies.

Finally, we observe that there are only few video watermarking algorithms that meet the real-time or the three-dimensional constraint. These technical challenges remain unexplored and future research on these will play a decisive role in digital video watermarking.

References

1. B. Kahng et al., Watermarking Techniques for Intellectual Property Protection, *35th IEEE Design Automation Conf.*, San Francisco, USA, 1998, pp. 776 - 781
2. Yao Zhao, Reginald L. Lagendijk, Video Watermarking Scheme Resistant to Geometric Attacks, *IEEE Int. Conf. on Image Processing ICIP 2002*, 2002, pp. II-145- II-148 vol.2
3. A. Kejariwal, Watermarking, *Potentials IEEE Volume 22, Issue 4*, Oct-Nov 2003, pp. 37-40

4. S.N. Merchant et al., Watermarking of Video Data Using Integer-to-Integer Discrete Wavelet Transform, *Conf. on Convergent Technologies for Asia-Pacific Region TENCON 2003*, pp. 939- 943
5. S. Voloshynovskiy et al., Attacks on Digital Watermarks: Classification, Estimation-Based Attacks, and Benchmarks, *IEEE Communications Magazine*, August 2001, pp. 119-126
6. M. Kutter, Digital Image Watermarking: Hiding Information in Images, *PhD thesis*, EFPL, Lausanne, Switzerland, 1999
7. C. I. Podilchuk, E. J. Delp, Digital Watermarking: Algorithms and Applications, *IEEE Signal Processing Magazine*, July 2001, pp. 33-46
8. B. G. Mobasseri, D. Cinalli, Lossless watermarking of compressed media using reversibly decodable packets, *Elsevier Signal Processing, Article In Press, Corrected Proof*, Sept. 2005
9. K. Su; D. Kundur, D. Hatzinakos, Statistical invisibility for collusion-resistant digital video watermarking, *IEEE Trans. on Multimedia, Vol. 7, Issue 1,* Feb. 2005, pp. 43 - 51
10. E.T. Lin, E.J. Delp, Temporal synchronization in video watermarking, *IEEE Trans. on Signal Processing, Vol. 52, Issue 10, Part 2*, Oct. 2004, pp. 3007 - 3022
11. F. Deguillaume, G. Csurka, and T. Pun, Countermeasures for unintentional and intentional video watermarking attacks, *in Proc. SPIE, vol. 3971*, Jan. 2000, pp. 346–357
12. M. D. Swanson, B. Zhu, and A. T. Tewfik, Multiresolution scene-based video watermarking using perceptual models, *IEEE J. Select. Areas Commun., vol. 16, no. 4*, May 1998, pp. 540–550
13. C.-S. Lu, J.-R. Chen and K.-C. Fan, Real-time frame-dependent video watermarking in VLC domain, *Elsevier Signal Processing: Image Communication, Vol. 20, Issue 7,* Aug. 2005, pp. 624-642
14. S. H. Kwok, C. C. Yang, K. Y. Tam and Jason S. W. Wong, SDMI-based rights management systems, *Elsevier Decision Support Systems, Vol. 38, Issue 1,* Oct. 2004, pp. 33-46
15. X. Kong, Y. Liu, H. Liu and D. Yang, Object watermarks for digital images and video, *Elsevier Image and Vision Computing, Vol. 22, Issue 8,* Aug. 2004, pp. 583-595
16. G. Doërr and J.-L. Dugelay, A guide tour of video watermarking, *Elsevier Signal Processing: Image Communication, Vol. 18, Issue 4,* Apr. 2003, pp. 263-282
17. P. Judge and M. Ammar, WHIM: watermarking multicast video with a hierarchy of intermediaries, *Elsevier Computer Networks, Vol. 39, Issue 6, 21* Aug. 2002, pp. 699-712
18. M.P. Queluz, Authentication of digital images and video: Generic models and a new contribution, *Elsevier Signal Processing: Image Communication, Vol. 16, Issue 5,* Jan. 2001, pp. 461-475
19. F. Hartung and B. Girod, Watermarking of uncompressed and compressed video, *Elsevier Signal Processing, Vol. 66, Issue 3, 28 May 1998,* pp. 283-301
20. B.G. Mobasseri, M.P. Marcinak, Watermarking of MPEG-2 video in compressed domain using VLC mapping, *7th ACM Wor. on Multimedia and Security MM&Sec '05,* Aug. 2005
21. X. Zhang, S. Wang, A new watermarking scheme against inserter-based attacks suitable for digital media with moderate size, *3rd Int. ACM Conf. on Information Security InfoSecu '04,* Nov. 2004
22. M. Kutter and F. A. P. Petitcolas, A fair benchmark for image watermarking systems, *Security and Watermarking of Multimedia Contents*, http://citeseer.ist.psu.edu/kutter99fair.html, pp. 1-14
23. B. Girod, "Bidirectionally decodable streams of prefix ode-words", *IEEE Comm. Lett. Vol. 3, Issue 8*, Aug. 1999, pp. 245–247
24. F. Deguillaume et al., "Robust 3D DFT video watermarking,", *SPIE Conf. on Security and Watermarking of Multimedia Contents I, vol. 3657*, San Jose, USA, Jan. 25–27, 1999, pp. 113–124
25. J. A. Bloom et al., "Copy protection for DVD video", *Proc. of IEEE, vol. 87*, Jul. 1999, pp. 1267–1276
26. S. Craver et al., "Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks, and implications," *IEEE J. Select. Areas Commun., vol. 16,* May 1998, pp. 573–586
27. M. Kutter, S. Voloshynovskiy, and A. Herrigel, The watermark copy attack, *SPIE Security and Watermarking of Multimedia Contents II, vol. 3971, San Jose, CA*, Jan. 24–26, 2000, pp. 371–380
28. 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*, Oct. 2004 pp. 2955-2964
29. S. Pereira et al., Second generation benchmarking and application oriented evaluation, *Information Hiding Workshop III,* Pittsburgh, USA, April 2001
30. S. Voloshynovskiye et al., Attack modeling: Towards a second generation benchmark, Signal Processing, *Special Issue: Information Theoretic Issues in Digital Watermarking*, May, 2001
31. F. Petitcolas, R. Anderson, M. Kuhn, Attacks on copyright marking systems, *2nd Int. Wor. On Information Hiding*, IH'98, Portland, U.S.A., April 15-17, 1998
32. V. Solachidis, A. Tefas, N. Nikolaidis, S. Tsekeridou, A. Nikolaidis, I.Pitas, A benchmarking protocol for watermarking methods, *IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, 7-10 Oct., 2001, pp. 1023-1026