Media-independent Watermarking Classification and the need for combining digital video and audio watermarking for media authentication

Jana Dittmann, Anirban Mukherjee, Martin Steinebach
GMD - German National Research Center for Information Technology, Institute (IPSI)
{jana.dittmann, anirban.mukherjee, martin.steinebach}@darmstadt.gmd.de

Abstract

In the past five years watermarking has become a major topic to solve authentication problems and copyright protection as major security demands in digital marketplaces. A wide variety of watermarking techniques have been proposed in the literature. Most techniques are developed for still images, currently the research community enforces also approaches for other multimedia data like video, audio and 3D models. In our paper we summarize the main watermarking parameters and introduce a media independent classification scheme. Our classification scheme is based on the application areas. We show the important parameters and possible attacks. Based on our proposed classification the quality of the watermarking techniques can be evaluated.

Furthermore we address the need for combining digital video and audio watermarking for media authentication.

1. Motivation

Digital Watermarking is a powerful technology capable of solving important practical security problems like authentication for copyright protection. Watermarking techniques usually used for digital imagery and now also used for audio and 3D-models are relatively new and are growing at an exponential rate. Well over 90% of all publications in this field have been published in the last 5 years. It is a highly multidisciplinary field that combines image and signal processing with cryptography, communication theory, coding theory, signal compression, and the theory of visual perception. Interest in this field has recently increased because of the wide spectrum of applications it addresses. Today a wide variety of techniques has been proposed but it is quite difficult to classify the approaches and measure their quality. Our intention in this paper is to discuss the main watermarking parameter and to present a media independent classification scheme as a basis for quality evaluation. Our classification scheme takes the application areas into account and shows which parameters and attacks are essential. E.g. in comparison to Pitas [14] we do not take into account the algorithm details like the domain where the watermark is embedded. Our goal is to give the users a scheme where they can find their applications and the main essential parameters of the watermarking techniques. Furthermore we show the need for combining audio and video techniques for a multimedia security solution.

2. Watermarking parameter

The most important properties of digital watermarking techniques are robustness, security, imperceptibility/ transparency, complexity, capacity and possibility of verification.

- **Robustness** describes if the watermark can be reliably detected after media operations. We emphasize that robustness does not include attacks on the embedding scheme that are based on the knowledge of the embedding algorithm or on the availability of the detector function. Robustness means resistance to “blind”, non-targeted modifications, or common media operations. For example the Stirmark or 2Mosaik tool [15] attack the robustness of watermarking algorithms with geometrical distortions.
- **Security** describes if the embedded watermarking information cannot be removed beyond reliable detection by targeted attacks based on a full knowledge of the embedding algorithm and the detector, except the key, and the knowledge of at least one watermarked data. The concept of security includes procedural attacks, such as the IBM attack [3], or attacks based on a partial knowledge of the carrier modifications due to message embedding [9] or embedding of templates [16]. The security aspect also addresses the false positive detection rates.
- **Transparency** is based on the properties of the human visual system or the human auditory system. A transparent watermark causes no artefacts or quality loss.
- **Complexity** describes the effort and time we need for watermark embedding and retrieval like for encoding and decoding of JPEG images or MPEG streams.

This parameter is essential if we have real time applications. Another aspect addresses if we need the original data in the retrieval process or not. Here we distinguish also between non-blind and blind watermarking schemes which influences the complexity.

- **Capacity** decribes how many information bits can be embedded. It addresses also the possibility of embedding multiple watermarks in one document in parallel.
- The **verification** procedure describes if we have a private verification like private key functions or a public verification possibility like the public key algorithms in cryptography.

The optimization of the parameters is mutually competitive and cannot be clearly done at the same time. If we want to embed a large message, we cannot require at the same time large robustness. A reasonable compromise is always a necessity. On the other hand, if robustness to strong distortion is an issue, the message that can be reliably hidden cannot be too long.

3. Media independent classification

Now we propose our media independent classification scheme as a basis for quality evaluation. Our classification scheme is oriented on the application areas where watermarking techniques can be used to meet the needs of the users. Furthermore we show which parameters and possible attacks are essential for each application area. Based on these parameters and attacks the algorithms can be evaluated if a specific algorithm has the adequate properties and can be used for certain application area. Usually existing watermarking techniques can be used in several applications but in each application it is hard to fulfil all quality demands.

3.1. Application based Classification

Altogether we find the following watermarking classes based on application areas for digital watermarking:

- Authentication or Copyright Watermark: Ensuring copyright protection by watermarking the data with an owner or producer identification
- Fingerprint Watermark: Ensuring copyright protection by watermarking the data with customer identifications to track and trace legal or illegal copies
- Copy Control or Broadcast Watermark: Ensuring copyrights with customer rights protocols, for example for copy or receipt control
- Annotation Watermark: Ensuring copyrights by annotations or capturing of the media data, this kind of watermark is also used to embed descriptions of the value or content of the data
- Integrity Watermark: beside the authentication of the author or producer we want to ensure integrity of the date and recognize manipulations

In our classification scheme we do not consider watermarking as information hiding technique to have a secure cover communication.

3.2. Important parameters and attacks

All five watermarks have their own quality parameter and standards. In this section we want to point out the general watermarking parameters described in section 2 for each of the five watermark classes. This can be used as general quality metrics. In addition we show possible attacks which depend on the application area, [2].

For example the Fingerprint Watermark has to deal with the coalition attack. When we watermark the original with different user identifications we produce different copies. Customers could work together by comparing their different copies to find and destroy the Fingerprint Watermark [1,5].

Another problem we recognize is the property of robustness and the recognition of manipulations for the Integrity Watermark. The robustness has to be adapted to content-changing and content preserving manipulation which must be addressed by the watermarking algorithms. A draft classification can be found in the next table.

Content-preserving manipulations	Content-changing manipulations
• Transmission errors- • Noise • Data storage errors • Compression and quantization • Brightness reduction • Resolution reduction • Scaling • Color convertions • γ-distortion • Changes of hue and saturation	• Removing image objects (persons, objects, etc.) • Moving of image elements, changing their positions • Adding new objects • Changes of image characteristics: color, textures, structure, impression, etc. • Changes of the image background (change of the day time or location (forest, ocean)) • Changes of light conditions (shadow manipulations etc.)

Table 1: Content-preserving and content-changing manipulations

The Integrity Watermark has to recognize only content-manipulating changes. Usually, content describing features are extracted and used for watermarking [4,8]. The existing techniques differ in the fragility. Some approaches recognize also content-preserving manipulations mostly used for strong security needs. Depending on the application area the user have to chose the appropriate technique.

Watermark	Parameter	Attacks
Authentication/ Copyright Watermark	-high robustness -high security -non-perceptual -blind methodes are useally more practicable -capacity should fit to the needs for a rightful owner identification -verfication process usually private, public can be also desirable	-Mosaik attack [15] -Stirmark attack [15] -Geometrical attacks [6] -Histogram attacks [13] -Template attacks [16] -Forgery attacks [10] -Rightful ownership attacks (invertability) [3]
Fingerprint Watermark	See Authentication Watermark -also non blind techniques are useful	See Authentication Watermark -additionally the coalition attack [1,5]
Copy Control/ Broadcast Watermark	See Authentication Watermark -also non blind techniques are useful -low complexity required	See Authentication Watermark
Annotation Watermark	-less robustness is in most cases acceptable -security is usually not important -blind methodes are preferable with low complexity -high capacity - verfication process usually private, public can be also desirable	In most cases no interest in attacking the watermark
Integrity Watermark	See Authentication Watermark -but robustness until the semantic of the data is destroyed (semi-fragile, content-fragile)	-Forgery attacks [10] -Rightful ownership attacks (invertability) [3] -attacks on the fragility [4]

Table 2: Important parameters and attacks

4. Combined audio and video authentication

In this section we introduce a new approach to multimedia watermarking. To improve the quality of watermarking technology, we combine watermarking algorithms for different media.

The two main aspects of our security scheme are authenticity and integrity. Authenticity is given when it can be proved that an author is the creator of the marked work . Integrity is given when it can be proved that no changes in the content of the marked work have been made.

Our goal is to provide a solution with robust and fragile aspects to guarantee authenticity and integrity by using secret key watermarks in combination with content information. We combine audio and video watermarking, with the media referring to each other, building a linked structure with mutual information about the content of the linked parts and embed the watermarks with a user key to ensure authenticity. The combination of feature extraction and robust watermarking has been introduced by us as "content-fragile watermarking" in [4].

4.1. Content-fragile watermarking

The concept of a content-fragile watermark can be described as using a robust watermark to embed content information for integrity verification. This information can be compared with the actual content. If changes have been made, content and watermark differ, and a warning message is prompted. Firgure 1 illustrates this process.

Fragility is about losing equality of extracted and embedded content in this case. The idea of content-fragile watermarking is based on the knowledge that we have to handle content-preserving operations, manipulations which do not manipulate the content.

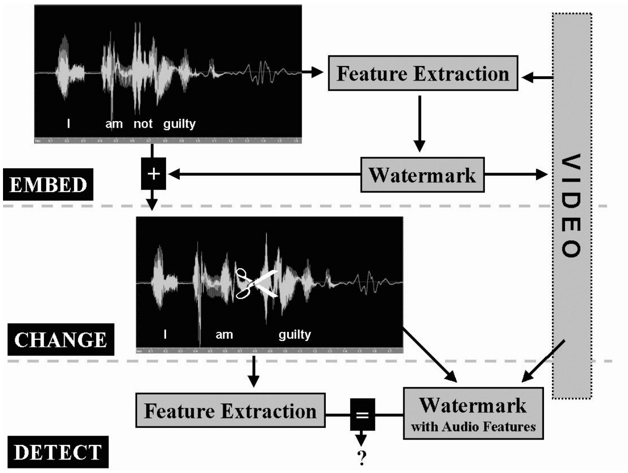

Figure 1: A watermark based on extracted features is embedded, the content is changed and the manipulation is detected with the help of the watermark.

The well-known problem of "friendly attacks" occurs here as in any watermarking scheme: Some signal manipulations must be allowed without breaking the watermark. In our case, every editing process that does not change the content itself is a friendly attack. Compression, dynamics, A/D/D/A-conversion and many other operations that only change the signal but not the content described by the signal should not be detected content manipulation. The idea is to use content information as an indicator for manipulations. The main problem is, which media features are appropriate to distinguish between content-preserving and content-changing manipulations.

4.2. Combined MPEG Watermarking

This section describes an environment for our combined video and audio approach. We use a MPEG system stream (ISO/IEC 11172) that consists of one or more elementary (video, audio, padding and private) streams as an example for combined media. As we apply different types of watermarking alogrithms, we need tools to extract media streams out of the system stream, convert MPEG data to image or PCM data and to rewrite the system stream with the changed data.

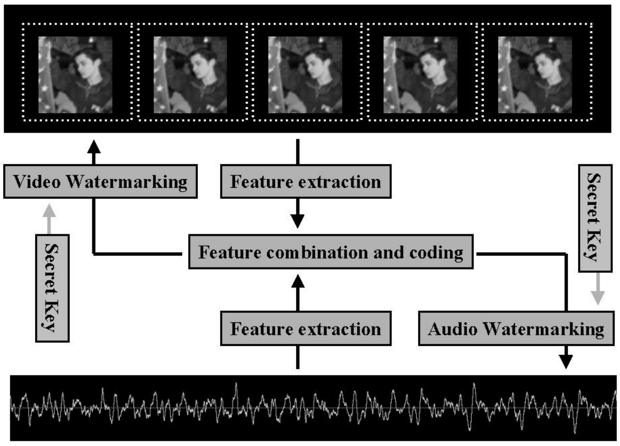

Figure 2: Combined video and audio watermarking based on feature extraction.

We have developed an algorithm for the extraction of audio and video elementary stream from the MPEG system stream and put it back to the original one after embedding the watermark in the elementary streams. During extraction stage, we extract the packets for a particular elementary stream and apply the watermark embedding technique on the decoded version of the extracted elementary stream. During combination, we place these extracted and watermarked data packets in the original system stream.

We have also developed a vision quality model to check the degree of visual quality degradation of the pictures, decoded in ppm format from the MPEG-2 system stream, due to the embedding of watermark. We have calculated “just noticeable difference” between two ppm pictures, one decoded from original video and another from watermarked video.

4.3 Applied algorithms

For our content-fragile watermarking system, both watermarking and feature-extracting algorithms are necessary. Figure 2 shows how the different algorithms interact.

As mentioned above, we use a MPEG system stream as an example for our combined watermarking system. Some of the data is extracted out of the system stream as MPEG data and watermarkes are embedded into the MPEG data while other parts of the system stream are uncompressed, marked and recompressed.

The later aproach is chosen in the video domain. Single images are extracted out of the video stream, marked and compressed again. Our used watermarking method in the video domain is based on overlaying a pattern with its power concentrated mostly in low frequencies. The pattern is created using a pseudo random number generator and a cellular automaton with voting rules. The resulting pattern is applied to the luminance of the image.

In the audio domain we embed the watermark directly into the MPEG data. Given a MPEG-file, an information to embed and a group of three patterns we encode the information into a sequence of patterns and extract the scale factors from the frames of the MPEG-file. Difference patterns based on this scale factors are calculated and the algorithm changes these patterns until a sufficient number matches our desired sequence of patterns. The whole watermark is inserted in this way, if there are more frames than needed the watermark is inserted multiple times. Then the new scale factors are inserted in the source file, overwriting the old ones and so creating a watermarked MPEG-file.

Besides the watermarking algrithms feature extraction algorithms are also necessary for content-fragile watermarking. The concepts described below both work on uncompressed data and therefore the extraction algorithms, introduced in 4.2, have to be applied to the multimedia stream before the features can be extracted.

The main concept of content detection for images is to extract the image characteristics of human perception, called content, which will be used for watermarking. The idea is to determine the edge characteristics of the image or single video frame and transform them into a feature code for the content-fragile digital watermark [4]. The edge characteristics of an image give a very good reflection of the image content, because they allow the identification of object structures and homogeneity of the image. We are using the canny edge detector described as the most efficient edge separator in [8].

Similar to the previous subsection, features for audio content description are needed to ensure the integrity of the audio stream. The extracted feature data has to be robust against all allowed attacks, such as filtering, compression or MP3 encoding, and it has to be coded with significantly less data than the original signal. The latter constraint is important since we want to embed the audio code into the video as a watermark.

To avoid high computational costs our approach is based on a low-level feature. Our idea is to determine signal sequences of equal sign in the audio track of a video and to transform them into a feature code representing the content of the signal. It is obvious that the inversion of the samples of an audio signal does not change the content but is performed easily. Thus, we rather chose the changing of the signs of the sequences as the feature for speech content verification, instead of the sign of each signal sequence in between two consecutive zero crossings.

4.4 Detection / Example

With our mutual watermarking scheme we ensure integrity and synchronization. The video stream watermark describes the video stream and the belonging audio stream and the audio stream describes itself and the video stream. Every time an image or a significant amount of audio information is deleted or changed, the other media will point out this modification. Post production operations that do not change the content of the image will not destroy the robust watermarks. A producer can work with the watermarked material and can transmit it, a receiver can check for integrity with the help of a trust center providing the keys, for example.

A possible application could be: A reporter records an interview with a video camera. He protects the resulting audio and video stream with a content-fragile watermark, then converts it to MPEG. He sends the file to a news distributor. They sell it to a internet news web page. Several post production operations like scaling, cutting and normalization are performed for news production. The edited video placed within other news. A viewer is skeptical about the interview, trusts the reporter, but maybe does not trust the news provider. He sends the received file to a trust center to check if the content has been changed. The trust center sends back a positive or negative result of the content watermarking verification.

5. Future work

The next step of our work is to classify the existing techniques into our proposed scheme for quality evaluation. The biggest problem we have to face is the non-uniform presentation of the existing algorithm and the lack of detailed information about their parameter.

Deleopers of watermarking algorithms must be convinced to classify their work. Without their cooperation the neccessary information can not be obtained. Test results about robustness and transparency can only be the first step for a customers evaluation of watermarking technology.

Our content-fragile watermarking system will be improved and expanded. Feature extraction and media watermarking are both subject to future research. New features to desribe the content with increased robustness against allowed attacks will be evaluated. The robust watermarking schemes to embed the content description will be improved and new algorithms will be researched and implemented.

6. Conclusion

In our paper we summarize the most important properties of digital watermarking techniques: Robustness, security, imperceptibility/ transparency, complexity, capacity and possibility of verification are the most important watermarking parameters. We explain why the optimization of the parameters is mutually competitive.

A media independent classification scheme based on the application areas authentication, fingerprinting, copy control, annotation and integrity is introduced to evaluate the quality of watermarking techniques. We point out the importance of the different discribed paramters for the watermark classes and discuss possible attacks.

An approach to ensure the integrity of multimedia streams using a combination of robust watermarking and content extraction technologies is described: We introduce the "content-fragile watermark" where fragility is not about destroying the watermark but about loosing the identity between extracted and embedded information. A major problem is the robustness of the feature extractions against content-preserving post production operations. The content-fragile watermark must not be destroyed by these operation but has to detect content-changing operations

Aided by the introduced classification scheme the concept of content-fragile watermarking can offer a maximum of security for multimedia data.

References

[1] D. Boneh and J. Shaw, *Collusion-Secure Fingerprinting for Digital Data*, In Proc. CRYPTO'95, Springer LNCS 963, pp. 452-465, 1995.

[2] Cox, Ingemar J, and Linnartz, Jean-Paul M.G.: Public watermarks and resitence to tampering, Proceedings of IEEE Int. Conf. O Image Processing, 1997, available only on CD-ROM

[3] S. Craver, N. Memon, B. Yeo, and M. Yeung: *Can Invisible Watermarks Resolve Rightful Ownerships?* Technical Report RC 20509, IBM Research Division, July 1996.

[4] J. Dittmann, A. Steinmetz, R. Steinmetz: *Content-based Digital Signature for Motion Pictures Authentication and Content-Fragile Watermarking*, Inn Proc. of IEEE Multimedia Systems, Multimedia Computing and Systems, June 7-11, 1999, Florence, Italy, Volume1, pp. 574-579, 1999

[5] J. Dittmann; A. Behr, M. Stabenau, P. Schmitt, J. Schwenk, J. Ueberberg (1999). *Combining digital Watermarks and collision secure Fingerprints for digital Images*, In Proc. of the

SPIE Conference on Electronic Imaging '99, Security and Watermarking of Multimedia Contents, 24-29 January 1999, San Jose USA, Proceedings of SPIE Vol. 3657, [3657-51], pp. 171-182, 1999

[6] Dittmann, Jana, Nack, Frank, Steinmetz, Arnd, Steinmetz, Ralf: *Interactive Watermarking Environments, Proceedings of International Conference on Multimedia Computing and Systems*, Austin, Texas, USA, 1998, pp. 286-294

[7] Dittmann,, Jana, Stabenau, Mark, Steinmetz, Ralf: *Robust MEG Video Watermarking Technologies, Proceedings of ACM Multimedia'98*, The 6th ACM International Multimedia Conference, Bristol, England, pp. 71-80

[8] Fischer, Stephan: *Indikatorenkombination zur Inhaltsanalyse digitaler Filme, D 180 (Diss. Universität Mannheim)*, 1997, Shaker Verlag Aachen.

[9] J. Fridrich: *Methods for data hidung*, Center for Intelligent Systems & Department of Systems Science and Industrial Engineering, SUNY Binghamton, Methods for Data Hiding", working paper, 1997

[10] M. Holliman, N. Memon: *Counterfeitung Attack on Linear Watermarking Schemes*, In Workshop Security Issues in Multimedia Systems, IEEE Multimedia Systems Conference '98, Austin, Texas, 1998

[11] T. Kalker: *System Issues in digital images an video watermarking for copy protection*, In Proc. of IEEE Multimedia Systems, Multimedia Computing and Systems, June 7-11, 1999, Florence, Italy, Volume1, pp. 562-567, 1999

[12] M. Kutter, F. Petitcolas: *Fair Benchmark for Image Watermarking Systems*, In Proc. of the SPIE Conference on Electronic Imaging '99, Security and Watermarking of Multimedia Contents, 24-29 January 1999, San Jose USA, Proceedings of SPIE Vol. 3657, [3657-51], pp. 226-239, 1999

[13] M. J. J. Maes: *Twin peaks: The histogram attack to fixed depth image watermarks*, In Proc. of the Workshop on Information Hiding, Portland, April 1998. Submitted.

[14] N. Nikolaidis, I. Pitas*: Digital Image Watermarking: an overview, In Proc. of IEEE Multimedia Systems, Multimedia Computing and Systems*, June 7-11, 1999, Florence, Italy, Volume1, pp. 1-6, 1999

[15] F. Petitcolas, R. Anderson and M. Kuhn: *Attacks on copyright marking systems*, In Proc of Second International Workshop on Information Hiding '98, 14-17 April, Portland, Oregon, USA; proceedings published by Springer as Lecture Notes in Computer Science v 1525, pp. 219—239, 1998.

[16] T. Pun: *Watermark Attacks*, DFG V3D2 Watermarking Workshop, http://www.lnt.de/~watermarking, 1999

[17] J. J. K. Ó Ruanaidh and T. Pun: *Rotation, scale and translation invariant digital image watermarking*, In Proc. of the ICIP, Santa Barbara, California, Oct, vol. 1, pp. 536-539 1997.

[18] M Swanson, M. Kobayashi, A. Tewfik: *Multimedia Data-Embedding and Watermarking Technologies*, In Proc of the IEEE, vol. 86, no 6, June 1998