Robust Digital Watermarking in Videos Based on Geometric Transformations

Philipp Schaber, Stephan Kopf, Fabian Bauer, Wolfgang Effelsberg
Department of Computer Science IV, University of Mannheim
Mannheim, Germany
{schaber|kopf|effelsberg}@informatik.uni-mannheim.de
fbauer@rumms.uni-mannheim.de

ABSTRACT

In the efforts to fight piracy of high-valued media content, forensic digital watermarking as a passive content security scheme is a potential alternative to current, restrictive approaches like DRM. In this paper, we present a novel watermarking scheme for videos based on affine geometric transformations. Frames can be modified in an imperceptible manner by applying a small, global rotation, translation, or zooming, which can be detected later on by comparison with the originals. To compensate geometric distortions that have been introduced while a video travels down legal as well as illegal distribution chains, a spatio-temporal synchronization is performed using our video registration toolkit application. To evaluate our approach, we compare it with several other schemes regarding the robustness against common attacks, including camcorder capture.

Categories and Subject Descriptors

H.1.1 [**Coding and Information Theory**]: Information Theory; I.4.5 [**Image Processing and Computer Vision**]: Reconstruction

General Terms

Security, Verification, Algorithms

Keywords

Digital watermarking, geometric transformations

1. INTRODUCTION

A major concern with the digital distribution of high-valued content such as movies is theft by piracy. Organizations like the *Motion Picture Association of America* (MPAA) calculate very high losses to the studios from movie piracy every year. Apart from the question whether these numbers are trustworthy or not, it is apparent that current approaches to deter content theft, such as Digital Rights Management systems (DRM), have little control over video piracy. Thus, digital forensic watermarking as an anti-piracy tool has recently gained increased attention.

Digital watermarking is a technique of embedding information in host data, most often into media data such as pictures, audio or video. Contrary to meta data where information is stored alongside the host data, watermarks store the information in the content itself by modifying it. Besides visible watermarking (such as station logos), invisible watermarking tries to introduce modifications that are imperceptible to human observers. Nevertheless, an appropriate watermark detector can read the embedded information. Using such a watermark, copyrighted content can be tracked to determine where and when illegal distribution occurred. In contrast to systems like DRM, which actively try to hinder any form of distribution, this is a passive content security scheme. The idea is to embed a unique, traceable identifier as watermark data ('payload') that is different for each legally distributed copy. For client side-watermarking, this is done as soon as the media leaves the (DRM-)protected domain, e.g., in a set-top box receiving an encrypted video-on-demand stream. If an illegal copy is then distributed or even sold, its watermark is extracted, and the identifier can be looked up in a tracking database.

In order to support tracking, a watermarking scheme has to fulfill certain requirements. Naturally, the information embedded has to be secure against unauthorized extraction or modification. This can be achieved using encryption, checksums and the like, but is beyond the scope of this work. Next, the watermark's modifications shall not alter the quality of the marked content, i.e., they must be imperceptible to human observers. Last, but most important, the watermarking scheme has to be classified as *robust*. While fragile watermarks are intended to immediately degrade when any modification is performed to the host content, robust watermarks should survive distortions and remain extractable even after severe degradations. This is an important issue for two main reasons: 1) Removing the watermark or making it undetectable is the primary goal of targeted attacks. 2) Even if a copy is not the target of an attack, the watermark is supposed to survive common signal processing operations as well as non-hostile modifications, and remain in the media throughout the complete (legal and illegal) distribution chain. Another important aspect is that robust watermarks should remain in the host content even if the digital domain is left, e.g., when a movie is captured using a camcorder.

The process of inserting a watermark signal is called *embedding* while the reading out is most often referred to as

MM'10, October 25–29, 2010, Firenze, Italy.

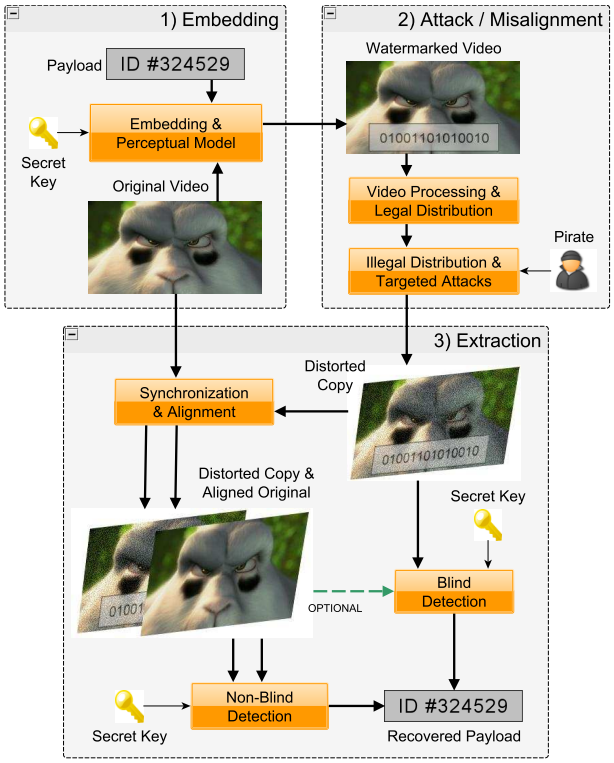

Figure 1: Watermarking overview

extraction. The need for the original data during extraction categorizes watermarking schemes: With *blind extraction* watermarking does not need the unmarked, original host data to retrieve the watermark (although it typically will profit significantly from its availability). On the other hand, *non-blind* watermarking also requires the unmodified content. Since the marked copy might have undergone temporal, spatial, as well as other distortions, original and copy need to be *synchronized* beforehand. Figure 1 gives an overview over all the steps.

In this paper, we present a novel approach of watermarking videos using global geometric transformations. While the modifications are imperceptible to human observers, the robustness and reliability of the extraction is very high. The paper is structured as follows: In Section 2, related work is discussed. Next, our proposed marking scheme is described in detail in Section 3. After the presentation of evaluation results in Section 4, the last section concludes the paper.

2. RELATED WORK

Regarding watermarking for videos, numerous techniques and methods have been published. We classify existing approaches into *luminance-based*, *contrast-based* and *frequency-based*. For each group, we selected a representative candidate and implemented the proposed scheme to be able to compare it against our geometric approach. For a fair comparison, all algorithms were adjusted to encode the same number of bits per time unit, which is one bit per shot of the video (to fit our scheme). Of course, all schemes described here can encoded the payload with a higher density in practice. However, this is always a trade-off between the amount of bits to encode and the encoding's robustness. As our goal is to compare the robustness, it is feasible to focus on encoding one bit per time unit. In the following, we describe how this can be achieved for each algorithm.

Luminance-based: The luminance-based approach employs a method by Arno van Leest et al. [5]. Information is encoded by adjusting the mean luminance of all frames in the shot. If the bit to encode is set, the luminance is increased by a given value whereas the frame is not manipulated at all in case the bit is zero. To extract the mark, the mean luminance of the copy's frames within one shot is compared to the mean luminance of the original frames.

Contrast-based: The second method implemented adjusts frames' contrast values as described by Lee [1]. Again, frames are modified only if the bit to encode is set. In this case, a frame is divided into blocks of the size 4x4. A pattern bitmap is used to alter pixel values in some blocks and leave others unchanged. The strength of the alteration is dynamically derived from the contrast value of these blocks.For extraction, the frames of the marked copy and the original are again decomposed into 4x4 pixel blocks. By comparing corresponding blocks, the embedded pattern can be reconstructed. It is matched to the original pattern through image distance functions for all frames of the shot. The average of these values is thresholded to determine the bit encoded.

Frequency-based: This approach tries to mark frames by performing modifications in the frequency domain. In [4], the Discrete Wavelet-Transformation (DWT) is proposed to decompose a frame into frequency bands using Haar-Wavelets. The implemented way of marking is similar to the contrast-based approach as again a pattern bitmap is used. The lowest frequency DWT coefficients are additively modified by the values of the pattern. The extraction is achieved accordingly by decomposing the frames of the marked and original video using DWT. The coefficients of the lowest frequency are compared (original frame vs. marked copy), so the pattern can be reconstructed by analyzing the differences. Again, values are averaged over all encoded frames, and a thresholded image comparison will yield whether a pattern was embedded (encoding a set bit) or not.

In 1998, Maes and van Overveld already proposed to modify geometric features instead of color components [2]. However, they applied *local* changes by geometric warping, which has several disadvantages: First of all, the modifications are much more likely to be perceived by viewers unless the amount of the modification is relatively small. This, on the other hand, limits the robustness of the embedding. Our solution addresses these issues.

3. ROBUST WATERMARKING BASED ON GEOMETRIC TRANSFORMATIONS

The basic idea of our proposed watermarking scheme is to *globally* apply small (affine) geometric transformations to frames. Although human observers easily notice sudden changes to the geometric alignment in videos, a slight, global transformation alone is usually very hard to notice. Considering this, we apply the same transformation to all frames of a shot and only change it at shot boundaries (cuts). For extraction, we need to compare the marked frames with the original, unmarked ones to analyze the transformation that was introduced during embedding. As this is a non-blind extraction, an additional synchronization is required.

3.1 Embedding

For embedding, a geometric transformation is applied to each frame within a shot. For our watermark, we consider the three affine transformations *rotation* by an angle α, *translation* by an offset of t pixels, and *scaling* (*zooming in*) by a zoom factor s. Using homogeneous coordinates, all transformations can be expressed using a single 3×3 matrix T, so every pixel $(x, y, w)^\top$ is transformed to its destination $(x', y', w')^\top$ using

$$\begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} = \begin{bmatrix} s\cos\alpha & s\sin\alpha & t_x \\ -s\sin\alpha & s\cos\alpha & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ w \end{pmatrix} \quad (1)$$

To avoid undefined border areas after applying a rotation, frames additionally have to be zoomed in to cut off these areas. The same applies to translations (alternatively, the frame size could just be reduced). For zooming-in, no additional transformations have to be performed.

To actually encode a bit-sequence as a watermarking payload using these transformation, there are several possibilities. We chose a robust approach, encoding only one bit per interval/shot: Applying a transformation encodes a 1 bit, no modification encodes a 0 bit.

3.2 Spatio-temporal synchronization

As mentioned in the introduction, a watermarked video will most likely undergo signal processing operations as well as format adaptations to different (legal or illegal) distribution channels or target devices. These may include changes of the frame rate (*temporal misalignment*) and also changes of the video's resolution (resizing, cropping) or its aspect ratio. Also, if the copy is acquired using a camcorder, perspective distortions might occur. In any case, *geometric misalignment* is the result. However, as our watermarking scheme is non-blind, it relies on a comparison of the marked (misaligned) and unmarked (original) frames for extraction. To allow this, corresponding frames have to be determined first (*temporal synchronization*). Also, the geometric misalignment resulting from distortions has to be compensated (*spatial synchronization*) in order to be able to detect the transformations. Both are done using our video registration toolkit application and algorithms developed and presented in previous work [3].

To only compensate the distortions and *not* the transformations that have been applied by our embedding, synchronization intervals have to be inserted with no intentional modifications. Although the geometric distortion (e.g., resulting from screen size adaptations) is usually constant, we recommend to have multiple synchronization intervals, depending on the required degree of robustness. Figure 2 shows an alternating scheme, having a synchronization interval ('Sync.') preceding each encoding interval ('Data'). As we temporally align our transformations to shot boundaries, the length of the intervals is determined by these.

The geometric synchronization is based on finding corresponding feature points as described in [3], and is performed once on each synchronization interval. For all frames of the following one or more encoding intervals, the temporally corresponding frames of the original, unmarked video are aligned to the distorted copy frames so that the only geometric differences should be those introduced by the watermark encoding.

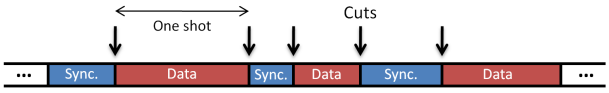

Figure 2: Synchronization and embedding intervals

3.3 Extraction

The algorithm for extraction has to analyze all frames of each encoding interval and has to decide whether a specific transformation was applied by the embedder (encoding a 1 bit) or not (encoding an 0 bit). To come to that decision, the transform matrix T between each original frame and the corresponding counterpart of the copy is estimated using corresponding feature points, similar to the way the geometric synchronization is done but without aligning any frames. Instead, the parameter of the chosen transformation is derived (rotation angle α, translation offset t, and zoom factor s), and its average values $\overline{\alpha}, \overline{t}, \overline{s}$ are calculated over all frames of the encoding interval.

In order to rank this value, the overall minimum and maximum regarding all frames of the video also has to be determined. With this information, a bit value b_i can then be calculated from the average value for each encoding interval i, e.g., in case of rotation:

$$b_i = \begin{cases} 1, & \alpha_\tau \le \overline{\alpha} \le \alpha_{max} \\ 0, & \alpha_{min} \le \overline{\alpha} \le \alpha_\tau \end{cases} \quad (2)$$

The thresholding value α_τ is calculated by using a factor f_τ in the following way: $\alpha_\tau = \alpha_{min} + f_\tau(\alpha_{max} - \alpha_{min})$.

4. EVALUATION

We evaluate the quality of the watermarking algorithms based on *luminance*, *contrast*, *frequencies*, and our new *geometric approach*. Several attacks are used, and the reliability of the watermark extraction is analyzed for each algorithm. We consider the following attacks:

- **Scaling:** Using bi-linear interpolation, the frames are scaled by a factor of $F = 0.5$. This attack represents the re-encoding of a video using a different resolution.
- **Rotation:** Each frame is rotated by a small fixed angle of $\gamma = 0.25$ degrees. This represents a typical transformation when videos are captured by camcorders.
- **Luminance:** The luminance of each pixel is increased by $L = 8$. This attack represents the effect that the average luminance changes even if a camcorder does not use automatic gain control.
- **Noise:** Additional noise is added to each frame based on a Gaussian distribution ($\sigma = 16$, $\mu = 0$). This attack simulates noise that is caused by the capture and processing of the video.
- **Blur:** A Gaussian filter with a mask size of $G = 9$ is used to smooth the frames. Blurred images may always occur in case of camcorded videos.
- **Crop:** A fixed border of the video is removed ($B = 25\%$). This attack is typical if wide-screen movies are captured with standard camcorders, or the aspect ratio is different from the original video.
- **Capture:** Whereas all previous attacks can be simulated, the use of a camcorder induces a combination of different distortions like histogram changes, a modification of the resolution, geometric transformations,

and noisy or blurred pixels. Because this attack is most critical, we used two camcorders – in PAL and HD resolution – to validate the reliability of the different watermarking algorithms. In both cases, the camcorder was positioned at a small angle and rotation. Also, the test video was captured twice: Once zoomed out, with the surroundings of the screen still visible, and once zoomed in, cropping parts of the captured video frames.

Preliminary tests were performed with a larger collection of test videos with PAL resolution. The parameters of the watermarking algorithms were chosen based on these tests. The main evaluation was done with a PAL video and the HD video "Big Buck Bunny"[1]. A watermark was added, and one of the attacks above was performed. Our software tool for video registration (see Section 3.2) was used to re-align frames and to correct temporal and spatial misalignments, before the mark was read out again. Although we analyzed all combinations of watermarking algorithms and attacks, we can only summarize the most relevant results in the following.

The watermarking algorithm based on *luminance* handles the attacks *scaling*, *rotation*, *noise*, and *blur* very well (no bit errors). Slightly worse but still within reasonable thresholds are the attacks *luminance*, *crop*, and *capture*. Especially the automatic gain control of camcorders reduces the reliability of the extraction. Also, users could recognize slight modifications in two static and uneventful shots. The *contrast-based* watermarking and the algorithm based on manipulating the *frequency domain* both handle all attacks very well, except changes to the *luminance* and *capturing*. These attacks caused several errors, for example, because the contrast of the blue color channel was not significant enough any more for the contrast-based extraction. However, the modifications are impossible for users to recognize.

The different geometric transformations are considered separately in the following. In case of *translation*, the position of pixels was shifted by 4, and a threshold $f_\tau = 0.2$ was chosen. If changes occur only at shot boundaries, users cannot recognize that some columns or lines are missing. The watermark could be extracted in all cases, which was also true for *rotation*, where we chose a rotation angle of $\alpha = 0.5$ degrees. Nevertheless, the extracted parameters of the translation and rotation changed significantly over time in case of camcorded videos. This is caused by the analog merging of adjacent frames which reduces the precision of the feature point algorithms. By averaging and thresholding the extracted parameters as proposed in our algorithm, the correct result can still always be extracted.

Watermarking based on *zooming* is robust in most cases; however, the current implementation smoothes the image, which results in feature points becoming less precise. With additional Gaussian smoothing or capturing as an attack, significant error rates occurred. The watermark of all videos is fully recovered, but the results are not very robust due to the high standard deviation of the parameters.

Regarding the embedding capacity of our watermark based on transformations, the *average shot length* (ASL) of a video is the determining factor. This value not only depends on a movie's genre, but also varies from year to year. However, as the intended application scenario for our watermark is tracking, it is usually sufficient to embed a unique identifier, for example of 32 bit length. Roughly assuming an ASL between two and 15 seconds as boundaries according to the CineMetrics database[2], the payload is distributed over about one to eight minutes. This is based on encoding only one bit per shot; however, more than that can be embedded with a single geometric transform, if not only the existence but also parameters of a transformation are used for embedding.

To summarize the results, all watermarking algorithms are robust against most attacks. Changes to the luminance as well as camcorder capture is problematic for contrast- and frequency-based algorithms. On the other hand, even a direct comparison of original and watermarked frames did not show visible differences, while other watermarks may become noticeable (e.g., the *luminance-based* algorithm). Comparing our novel algorithm based on geometric transformations to the others, it is one of the most reliable techniques, and the visual quality of the watermarked videos is also very high. Also, it can be combined with other techniques, and since any combination of affine transformations can be expressed with a single matrix, even more than one transformation can be applied at quite fast or even in real time on accelerated hardware.

5. CONCLUSIONS AND FUTURE WORK

We presented our new watermarking algorithm that uses global geometric transformations to encode data into videos. A comparison with luminance-, contrast-, and frequency-based watermarking algorithms indicates a high quality with regard to invisibility and robustness of the watermark. Seven attack scenarios were considered in the evaluation: the capture by camcorder was most challenging due to temporal, spatial, and color-based distortions.

Still, there are many open issues we want to consider in the future. A major goal is to define a perceptual model to automatically decide on an appropriate type and strength of the geometric transformation for each shot. To overcome the limitation of shot boundaries, a dynamically changing transformation could be considered, avoiding noticeable changes.

6. REFERENCES

[1] C.-H. Lee and Y.-K. Lee. An adaptive digital image watermarking technique for copyright protection. *IEEE Transactions on Consumer Electronics*, 45(4):1005–1015, 1999.

[2] M. Maes and C. van Overveld. Digital watermarking by geometric warping. In *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, volume 2, pages 424 –426 vol.2, 4-7 1998.

[3] P. Schaber, S. Kopf, W. Effelsberg, and N. Thorwirth. Semi-automatic registration of videos for improved watermark detection. In *Proceedings of the first annual ACM SIGMM conference on Multimedia systems*, MMSys '10, pages 23–34, 2010.

[4] D. Taskovski, S. Bogdanova, and M. Bogdanov. Digital watermarking in wavelet domain. First IEEE Balkan Conference On Signal Processing, Communications, Circuits, And Systems, 2000.

[5] A. van Leest, J. Haitsma, and T. Kalker. On digital cinema and watermarking. *Proceedings of SPIE-IS&T Electronic Imaging*, 5020:526–535, 2003.

[1] www.bigbuckbunny.org, (c) Blender Foundation

[2] http://www.cinemetrics.lv