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 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.3.0 [**Information Storage and Retrieval**]: General; 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, media security

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. Also, strong limitations in the use of DRM-protected data hinder a broad acceptance among consumers. Thus, digital forensic watermarking as an anti-piracy tool has recently gained increased attention.

Digital watermarking is a technique of embedding additional 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 still read the embedded information afterwards. Using such a watermark, copyrighted content can be tracked to determine where and when illegal distribution occurred, without limiting legal use. 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') in each individual 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 massively distributed in file-sharing networks or even sold, its watermark is extracted and the identifier can be looked up in a tracking-database.

In order to serve the purpose of tracking, a watermarking scheme has to fulfill certain requirements. Naturally, the information embedded has to be secure against unauthorized extraction, embedding 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 (which guarantees the identification of alterations), robust watermarks should survive distortions (e.g., rotation, scaling and translation in case of images [5]) 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 any attack, the watermark is supposed to survive common

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

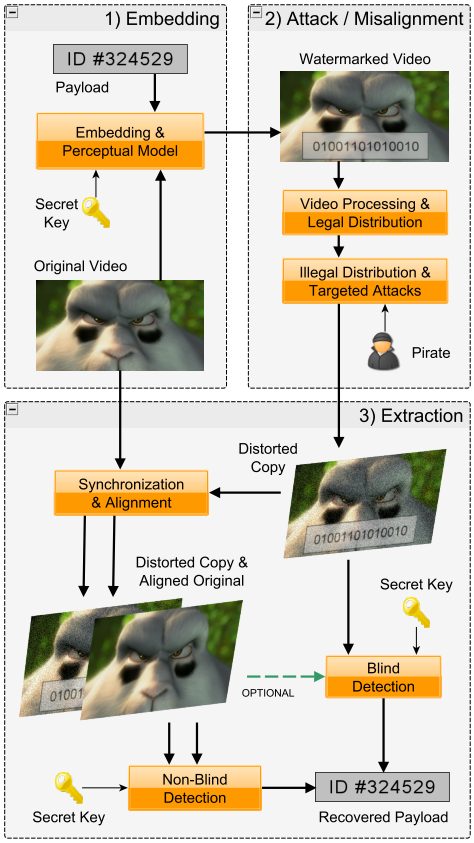

Figure 1: Watermarking overview

signal processing operations as well as non-hostile modifications, and remain in the media throughout the complete (legal and illegal) distribution chain. This poses significant challenges to a watermark design. 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 into a host signal is called *embedding*, while the reading out is most often referred to as *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 requires the unmodified content for extraction. 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 steps.

In this paper, we present a novel approach of watermarking videos using 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 watermarking scheme is described in detail in Sections 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 (about one to three seconds). In the following, we describe how to encode one bit into frames of the corresponding shot.

Luminance-based: The luminance based approach employs a method by Arno van Leest et al. [1]. 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 unset. 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 Chang-Hsing Lee and Yeuan-Kuen Lee [2]. Again, frames are modified only if the bit to encode is set. In this case, a frame is divided into multiple blocks of the size 4x4. A pattern bitmap is now 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 bitmap can be reconstructed. This is compared to the original pattern through image distance functions for all frames of the shot. The average of these values is thresholded to determine the state of the bit encoded.

Frequency-based: Frequency-based approaches try to imperceptibly mark frames by performing modifications in the frequency domain. In [3], the Discrete Wavelet-Transformation (DWT) is proposed to decompose a frame into high, low and middle frequency bands using Haar-Wavelets. The implemented way of marking a frame 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 DWT-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.

3. ROBUST WATERMARKING BASED ON GEOMETRIC TRANSFORMATIONS

The basic idea of our proposed watermarking scheme is to deliberately apply small (affine) geometric transformations to frames. The transformation corresponds to the encoding of a bit. Although human observers easily notice sudden changes to the geometric alignment in videos, a slight trans-

formation alone is usually very hard to notice. In order to hide the changes, 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 watermarking payload using these transformation, there are several possibilities. We chose a robust approach, encoding only one bit per interval/shot: Applying any transformation encodes a bit set, no modification encodes an unset 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 used for encoding data. Both is done using our video registration toolkit application and algorithms developed and presented in previous work [4].

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 (cuts), the length of the intervals is basically determined by these.

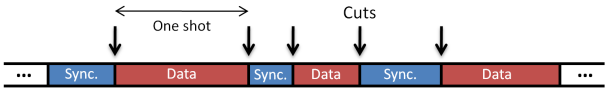

Figure 2: Synchronization and embedding intervals

The geometric synchronization is based on finding corresponding feature points as described in [4], 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.

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 set bit) or not (encoding an unset 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:

$$b_i = \begin{cases} 1, & \alpha_\tau \leq \overline{\alpha} \leq \alpha_max \\ 0, & \alpha_min \leq \overline{\alpha} \leq \alpha_\tau \end{cases} \quad (2)$$

The thresholding value α_τ is calculated by using a thresholding 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 $\gamma = 0.25$. 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 exposure adjustment.
- **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 changed.
- **Capture:** Whereas all previous attacks can be simulated, the use of a camcorder may induce a combination of different distortions like changes in color and luminance distribution, a modification of the resolution, geometric transformations, and noisy or blurred pixels. Because these effects are most critical, we used two camcorders – in PAL and HD resolution – to validate the reliability of the different watermarking algorithms.

Preliminary test were performed with a larger collection of test videos based on PAL resolution. The parameters of the watermarking algorithms were chosen based on these tests. The main evaluation was done with the new video "Big Buck Bunny" [1]. A watermark was added to the video, and one of the attacks above was performed. In the next step, our software tool for video registration (see Section 3.2) is used to re-align frames and to correct temporal and spatial misalignments. Although we analyzed the precision and recall of all combinations of watermarking algorithms and attacks, we will only focus on the most relevant results in the following.

The watermarking algorithm based on *luminance* handles the attacks *scaling*, *rotation*, *noise*, and *blur* very well and no bit errors did occur. Slightly worse but still within reasonable thresholds are the attacks *luminance*, *crop*, and *capture*. Especially if the luminance of a shot changes, the automatic gain control of camcorders might reduce the reliability of the watermark extraction. By focusing on the luminance, users could recognize slight modifications in two static and uneventful shots. In case of high motion, modifications could not be recognized.

The *contrast-based watermarking* algorithm handles most attacks very well. An attack based on the luminance caused several errors, because the contrast of the blue color channel was not significant enough in several shots. It is nearly impossible for users to recognize whether a watermark is embedded or not.

The algorithm based on manipulating the *frequency domain* is very reliable in case of most attacks, except *luminance* and *capture*. Similar to the contrast-based technique, it is all but impossible to recognize image modifications caused by the embedded watermark.

The different geometric transformations are considered separately in the following. In case of *translation*, the position of pixels is shifted by 4, and a threshold $f_\tau = 0.2$ was chosen. The watermark could be extracted in all cases. Nevertheless, the parameters of the translation change significantly over time, especially in case of camcorded videos. This is caused by the analog merging of adjacent frames which reduces the precision of the camera parameter estimation algorithm. If changes occur only at shot boundaries, users cannot recognize that some columns or lines are missing.

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

Watermarking based on *zooming* is robust in most cases. Significant error rates occur only in case of Gaussian smoothing. This is caused by the fact that many feature points do not match, and that camera parameters become less precise. In some cases, this is also true for *cropping*, especially if image regions with many feature points are no longer visible. The watermark of all captured videos is fully recovered, but the results are not very robust due to a high standard deviation of the parameters. Users could not recognize modifications in the embedded video.

To summarize the results, all watermarking algorithms are highly robust against most attacks, and only in some cases the embedded data may be destroyed, e.g. if the *contrast-based* or the *zoom-based* watermarking algorithm is used. Watermarks may become clearly visible (e.g., the *luminance-based* algorithm). For other watermarking techniques like frequency-based or contrast-based algorithm, even a direct comparison of original and watermarked frames did not show visible differences. Comparing our novel watermarking algorithm based on geometric transformations to the other algorithms, it is one of the most reliable techniques and the visual quality of the watermarked videos is also very high. Another major advantage of our algorithms is the fact, that it can be combined with other techniques and more than one geometric transformation can be applied at the same time.

5. CONCLUSIONS AND FUTURE WORK

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

Still, there are a lot of open issues we want to consider in the future. A major goal is to define a perceptual model to decide whether the embedding of the watermark might be perceived as such or not. Based on this model, the algorithm can select the type and strength of the geometric transformation. For example, a rotation is unsuited if straight lines are visible, while a zoom or translation might not be noticed in this case.

6. REFERENCES

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

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

[3] M. B. D. Taskovski, S. Bogdanova. Digital watermarking in wavelet domain. Technical report.

[4] 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, New York, NY, USA, 2010. ACM.

[5] D. Zheng, Y. Liu, J. Zhao, and A. E. Saddik. A survey of RST invariant image watermarking algorithms. In