A NEW CONCEPT FOR ROBUST VIDEO MARKING

Niels Thorwirth
Verimatrix, Inc.

Abstract

Digital watermarking is the concept of permanent and imperceptible tagging of video data. User-specific watermarking embeds tracking information that accurately identifies the last legal recipient of a video. This type of watermark does not enforce content use restrictions but it enables identification of sources of content abuse, acting as a deterrent and encouraging responsible consumer behavior. These types of techniques represent a way to maintain control over video content while permitting convenient content consumption that can compete with free but inconvenient and illegal content distribution.

Required is a secure, robust and imperceptible marking technology that does not degrade consumer experience and yet establishes a reliable trace that survives attacks or passages through the analog hole. This paper presents our implementation approach, optimized to create a commercially significant capability incorporating new concepts tailored to digital video.

A key element of the approach may appear to be counter intuitive, in that the embedded information is not designed to be machine readable but it can be extracted into a human readable form. While it requires human interaction, it allows use of well honed human perception processing to aid the extraction task - which is simply superior to machine recognition and thereby solves the problem of content synchronization. It overcomes the challenges of content misalignment that fools machine readable approaches, which are often unable to read the mark after small geometric content transformations.

This paper outlines the research results that enable human readout and the challenges to make the mark robust yet invisible. The result allows for wide content distribution, ensures the mark is secure and can establish strong evidence linking the content to a consumer while maintaining privacy.

DIGITAL WATERMARKING 101

Over the last few years, the representation, storage and distribution of digital video have grown massively in diversity and popularity, and are now approaching the level of digital music in its ease of use and ubiquity. The advantages of this type of distribution and consumption, just as with music, also fuel the growth of large scale piracy. The rightful owners of video content are deprived of their revenues and could be discouraged from investing in the creation of content in the future.

Several approaches have been applied to secure digital media:

Digital encryption technology is effective to enable secure delivery. However, once decrypted and presented in a human visible format, it can be re-recorded to obtain and distribute an illegal, unsecured copy. No secure technology currently exists to prevent re-recording using a camcorder.

Amongst other applications, the marking of media can help investigation to identify individuals that are responsible for abuse,

either by embedding recipient information in the media or ownership information that indicates copy restriction.

One way of marking media is done by adding information to the digital media file that is ignored during normal playback, but can be extracted and used by more specialized tools. Apple iTunes is widely assumed to be using this method to tag DRM-free file distribution (1). This kind of tagging enables identification of the unmodified file, but are easily removed and destroyed when the file is re-recorded or converted to another format.

For more robust marking, visible text or dots have been used to carry identifying information in video, and while this information does survive re-recording, it can easily be identified and removed in order to disable the ability to track the content. In order to maintain robustness, these markings are also visible, a compromise that degrades the consumer experience.

Digital watermarking is another marking approach that has been suggested in many variations. Common digital watermarking schemes embed information, by introducing manipulations at a certain positions in space or time. These watermarks are detected by specialist software (2). In other words, it is a process for modifying media content that embeds a machine-readable code into the data content.

When interpreting the manipulations during machine processing of the content, processing can be greatly reduced through pre-knowledge of the insertion positioning. When the positions are modified, i.e. misplaced or weakened, the readout may become difficult or impossible. Such modifications do routinely occur during simple media treatment such as cropping, rotation or conversion to another file format, including lossy compression, during which perceptually insignificant information is eliminated to reduce the size of a digital media file.

Critical transformations that occur during piracy re-recording include:

- Camcorder capture
- Digital-to-Analogy (DA) conversions to S-video or VHS
- Re-compression to low bitrates using popular formats like H.264, DivX, MPEG2
- Geometric distortions like heavy scaling, rotation, AR change, random bending, cropping
- Color conversion, e.g. grayscale
- Filtering like blur, sharpen, contrast modification, noise reduction and de-flicker
- Frame rate conversion

Relative misplacement of the mark and underlying content can also be created intentionally by imperceptible, slight or combined modifications, such as shifts, rations and time jitter. Publicly-available tools (3, 4) apply these modifications, also called attacks, in an automated fashion. Since current image processing algorithms are not very successful in recognizing misplacements in distorted content (a process also called registration) these modifications render these types of machine-readable digital watermarks ineffective.

Digital still images have been the early focus of watermarking research and commercial exploitation (5). Video watermark approaches often have been based on the application of the image watermark applied to video frames. While this is the natural progression and allows the embedding of much data, this approach does

not efficiently use the time domain for gathering embedded information because detection is only successful if some information in individual frames can be recovered. This approach fails when none of the watermarks can be read at least in part due to a failure in registration or destruction of relevant areas.

Digital watermarking typically involves a complex transformation of the original image and of the message to be embedded in order to allow invisible embedding. In the forensic, user specific application, watermarks are used to embed information about an individual playback device to allow tracing of the last legal recipient. It this case, it acts as a serial number or license plate and requires execution of the embedding in a playback device, which typically has little processing power to spare for additional functionality like digital watermarking.

HUMAN-READABLE WATERMARKING

To overcome the challenges outlined above, a new approach is required that takes the specific requirements of forensic watermarking into account. The approach starts with the realization that the recognition of distorted content is a task that humans can still perform better than machine. This fact is used for example in so called CAPTCHA images (6) that blocks Web site robots and identify human users. So instead of embedding machine-readable information, a human recognizable image is invisible embedded in the video, distributed over time. During extraction, that information is aggregated in order to derive a human-readable image containing the embedded information.

The result is a robust mark that exists with the actual media and survives transformations thereof, unlike bitstream manipulation and encryption. Unlike visible marking it is unnoticeably hidden in the media. The difference to previous watermarking approaches is the possibility of using the human perceptual system for actual recognition, which is superior to current machine reading capabilities, in particular for degraded, noisy and transformed content.

The message is emphasized by a computer, but the actual interpretation of the mark can be performed by a human, making it possible to detect the mark in degraded content and independent of registration. While it is very difficult for a watermark technology to interpret a misplaced mark, a message can be easily read by a human even if it is rotated, made smaller, stretched and on a noisy background.

During the embedding process, only some areas are modified, while all video data is used for detection. The areas that are modified are chosen by a perceptual model that takes into account where in the content the modification is able to stay just below any noticeable difference and thereby make the modification invisible to the consumer. Another component of user-specific forensic watermarking is that it ensures security by distributing the embedding locations in a random fashion. Consequently, the precise locations of the alterations can not be observed or removed, even if the mark can be recognized.

The mark is spread over the entire frame area and combined with the media using the basic content, such that the embedding portion can not be generally removed without destroying the video.

Additionally the mark is spread over time, and while it generally can not be

recovered from a single frame, each frame contributes to the detection result. The detection process accumulates results from several frames over time, contributing to invisibility and robustness of the mark.

The basic principle of combining frames in time is that during the process the actual movie content of the frames will average out to be a more or less smooth surface with pixel values close to the mean –with a faster convergence for high-motion content. The information embedded in the content in contrast is at constant locations and therefore will increase in relative signal strength as the frame content diminishes. A simple averaging though is not sufficient to expose a readable mark. To actually archive robust detection, additional filtering is applied to enhance the mark in every frame. The filter is aimed to reduce the effect of underlying content to be significant in the combined image while at the same time improve the effect of individual frames. Over the last years of researching this approach, we have evaluated a variety of different filters and often found surprising, counter intuitive results on their individual effectiveness. The best achievement so far to highlight the remaining bits of information is a combination of filters applied in sequence.

Our initial research however already revealed that the method of combining frames over time shows remarkable robustness, even in the presence of modifications that would suggest a removal of the minute, unnoticeable variations. These degradations include compression of the color domain or analog transformation. The reason for the approach being effective in these scenarios is that a fraction of the modification will survive any transformation that leaves the content in a reasonable quality.

The specifics of the technology do not allow for automatic measurement of the actual extracted signal strength as it is part of the design that the information is read out by a human. While the mark can be measured for verification by correlation to the known embedded information or OCR applied for reading, this will not be adequate to determine the level of human readability. The detection is therefore somewhat asymmetric. For one, the readout is done by a human and the embedding by a machine. In addition the extraction filter is mostly independent of the embedding and does not follow the specifics used during embedding.

For instance, the embedding locations are spread in a pseudo random fashion that does not have to be known during extraction. Both levels of asymmetric embedding and reading increase the security against attacks that aim to understand the embedding algorithm and to invert them. It also provides robustness against the so called oracle attack that degrades the watermarked content in an automated loop to find the least amount of degradation that causes the detector to fail to recognize the mark.

Unlike machine-readable digital watermarking, the detection process displays an obvious and unambiguous human understandable outcome, e.g. a serial number. When uncovering the mark, the embedded graphic slowly appears from the marked video, showing that the mark is derived from the content. The extraction can be interpreted or read by a layman. The result is easy to understand and can be used as persuasive evidence of wrongdoing during an investigation into content misuse.

HOW TO USE ROBUST WATERMARKING

User-specific digital watermarking is ideally suited for forensic applications that aim to identify the source of piracy. It acts as a deterrent because it registers the individual video to its owner. The embedding task is performed in each set-top box (STB) or DVR, allowing individually-marked copies, without the requirement to process individual videos for each receiver at the head-end. The extraction is performed on copies that are in public distribution and originate from a single piracy source. To identify the embedded information, the copies are read in a central location avoiding the distribution of the extraction service preventing a possible attacker to verify the success of modification applied to the video with the intention to remove the mark.

CONCLUSION

With the obvious need for better protection and infrastructure in the growing business of digital video distribution, a layered protection approach that incorporates digital watermarking can help alleviate the revenue loss content owners are facing through current piracy. Digital video delivered through an STB provides a unique environment for content protection using user-specific marking, potentially providing a superior distribution channel and a preferred form of distribution by movie studios that are looking to secure their sensitive early release windows. As a business stimulus, such early release windows translate into more demand for pay-TV operators and they are a crucial component to increase the momentum and customer base of this distribution channel.

Transparency of consumption enabled by watermarking may be an additional important factor for the bottom line - maintaining the consumer motivation to actually *pay*-per-view for premium content compared to trawling for "free" peer to peer downloads. Effective content protection can make the difference between few subscribers that distribute movies to a large Internet community and a growing consumer base that values premium content.

References

1.http://www.theinquirer.net/inquirer/news/357/1050357/drm-free-itunes-files-have-your-number (retrieved on 1/29/09)

2. Jian Zhao, "Look, It's Not There", in: BYTE Magazine, January, 1997

3. Fabien A. P. Petitcolas, "Watermarking schemes evaluation", I.E.E.E. Signal Processing, vol. 17, no. 5, pp. 58–64, September 2000

4. Fabien A. P. Petitcolas, Ross J. Anderson, Markus G. Kuhn, "Attacks on copyright marking systems," in David Aucsmith (Ed), Information Hiding, Second International Workshop, IH'98,, Proceedings, LNCS 1525, April 1998

5. Saraju P. Mohanty, "Digital Watermarking: A Tutorial Review," Dept. of Comp Science and Eng, University of South Florida, Tampa, 1999

6. Luis von Ahn, Manuel Blum and John Langford, "Telling Humans and Computers Apart Automatically," In Communications of the ACM, vol. 47, pp. 56 – 60, February 2004

Niels Thorwirth
Director Advanced Technology Implementations
Verimatrix, Inc.
6825 Flanders Drive
San Diego, CA 92121
+1 (858) 677-7800 x3003
+1 (858) 677-7804 Fax
+1 (858) 357-1529 Cell
nthorwirth@verimatrix.com