**This paper was presented at the Electronic Imaging Symposium, on January 24, 2003, in Santa Clara, California. Please use the following format to cite this paper:**

**Jeffrey Lubin, Jeffrey A. Bloom, and Hui Cheng, "Robust, Content-Dependent, High-Fidelity Watermark for Tracking in Digital Cinema", in *Security and Watermarking of Multimedia Contents V*, Edward J. Delp III, Ping Wah Wong, Editors, Proceedings of SPIE Vol. 5020, (2003).**

# Robust, Content-Dependent, High-Fidelity Watermark for Tracking in Digital Cinema

Jeffrey Lubin, Jeffrey A. Bloom, Hui Cheng

Sarnoff Corporation, 201 Washington Road, Princeton, NJ 08540

## ABSTRACT

Forensic digital watermarking is a promising tool in the fight against piracy of copyrighted motion imagery content, but to be effective it must be (1) imperceptibly embedded in high-definition motion picture source, (2) reliably retrieved, even from degraded copies as might result from camcorder capture and subsequent very-low-bitrate compression and distribution on the Internet, and (3) secure against unauthorized removal. No existing watermarking technology has yet to meet these three simultaneous requirements of fidelity, robustness, and security. We describe here a forensic watermarking approach that meets all three requirements. It is based on the inherent robustness and imperceptibility of very low spatiotemporal frequency watermark carriers, and on a watermark placement technique that renders jamming attacks too costly in picture quality, even if the attacker has complete knowledge of the embedding algorithm. The algorithm has been tested on HD Cinemascope source material exhibited in a digital cinema viewing room. The watermark is imperceptible, yet recoverable after exhibition capture with camcorders, and after the introduction of other distortions such as low-pass filtering, noise addition, geometric shifts, and the manipulation of brightness and contrast.

Keywords: Forensic watermark, human visual perception, masking model, digital cinema, piracy.

## 1. THE NEED FOR A FORENSIC WATERMARK

A major barrier to the development and deployment of digital distribution channels for motion imagery content (e.g., digital cinema and other distribution paths such as video download) is the concern of content providers that their copyrighted material may be copied and then subsequently distributed without appropriate authorization. Encryption and data hiding (the covert incorporation of metadata into a digital bitstream) are important components of a Digital Rights Management (DRM) approach to controlling access to the content, but cannot prevent all instances of theft.

As illustrated in Figure 1, piracy can occur at various points in content distribution, in particular at the end of a given distribution pathway, where the content is decrypted, decoded, and displayed. The final stage of the distribution chain, exhibition, has been identified by members of the Motion Picture Association (MPA) as the most susceptible and damaging source of theft and unauthorized distribution of motion picture data [1]. Sophisticated pirates may tamper with hardware or software to capture decrypted data files, while less sophisticated pirates can use screen capture software to capture decompressed data from video buffers, or camcorders to capture exhibited data from the screen. We call this last process *exhibition capture*.

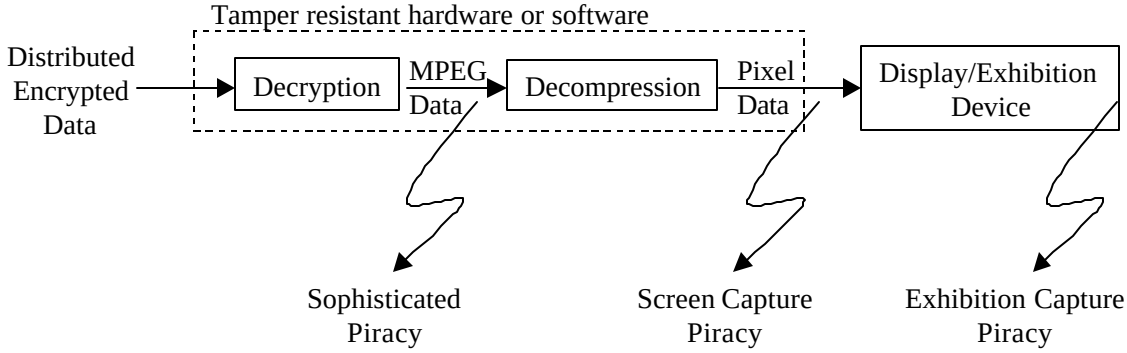


**Figure 1. Piracy holes at the end of a digital motion imagery content distribution path**

This potential leakiness at the end point of content distribution is exacerbated by the sheer number of such end points in some distribution pathways. Thus, while the problem of digital cinema theft from any of the multiple screenings at multiple theaters is large, the problem of content theft over the Internet from the millions of potential download customers is huge. The potential "Napsterization" of Hollywood content is thought by many to be a deal-breaker for the deployment of a digital video distribution infrastructure on the Internet.

Content owners are aware of current cinema piracy because illicit copies of first-run movies are available "on the street." It has become common for such illicit copies to be available far in advance of the scheduled video release and, in some cases, within days or even prior to first theatrical release. Exhibition theft is not unique to digital cinema; it is common with today's film-based distribution and display technologies. However, the move to digital distribution and display will provide an opportunity to employ the same digital solution to both the video download and digital cinema piracy problems.

Given these potential leaks, a content owner needs forensic tools that enable the tracking of unauthorized copies back to the party who licensed the use of the content, and who was responsible for preventing its further distribution. The ability of the content owners to identify the exact distribution point at which material was stolen can be used as a tool to identify the responsible parties and can act as a deterrent to such theft. A digital watermark uniquely identifying the licensee of that copy of the content can serve this purpose. This tracking watermark will give content owners a powerful forensic tool against piracy, because it allows them to trace pirated copies to the individual customers (e.g., for video download), or to a specific post-production house, or to the time and location (e.g., for digital cinema) at which theft occurred.

Consistent with those stated by SMPTE DC28.4 for their *Download Watermark* and *Exhibition Watermark* [2], a forensic watermark used for purchaser identification must have the following properties.

- It must satisfy the high fidelity requirements of the content owners.
- Exhibition watermarks must be robust to the combination of exhibition capture and compression.
- Exhibition watermarks must be secure against unauthorized removal and unauthorized embedding.
- Embedding must fit into the process chain without adding undue delay. While latency may be acceptable, the embedding process should be as fast as the preceding process.

Additionally, it has been suggested that for Digital Cinema, the watermark must carry 34 bits of information in order to uniquely identify the theater at which the motion picture was shown as well as a time and date stamp indicating the particular performance. For video download, the number of bits needed is the log of the number of unique downloads of the content. The same 34 bits can represent approximately 16 billion individual copies.

In addition to these requirements, we note the following features that a forensic watermark can have:

- Informed Detection: Detection of these watermarks can rely on the use of a reference. This is called *informed detection* [3]. The reference may be an unmarked video or a data vector derived from an unmarked video.
- Detection can be an expensive process as there will be few detectors, few applications of the detectors, and the nature of detection is such that it need not be done in real-time.

## 2. PREVIOUS WORK

Of the four requirements listed above, we consider the first three – fidelity, robustness, and security – critical to the work proposed here[1], as no existing watermarking algorithm has yet demonstrated acceptable performance in all three. Spread spectrum techniques, can be made quite robust, secure, and imperceptible, however most do not meet the level of robustness required by camcorder capture. Below we highlight two promising approaches presented specifically for this application and one addition work of particular interest.

[1] We do not address computational complexity here, but claim without proof that requirements will be easy to meet.

The work of Honsinger and Rabbani [4] is a spread spectrum watermark that has been demonstrated to survive camcorder capture, but at inadequate fidelity. Their watermark detection algorithm is blind (no information from the original is required) and a tiling pattern is used for automatic registration. This approach reduces the security against unauthorized removal by allowing the adversary to identify the tile size.

Haitsma & Kalker [5] also embed a "spread spectrum" watermark, but theirs is spatially DC and biased toward temporal low frequencies. The mean luminance from each frame is extracted to obtain a 1D temporal signal. This is then effectively subsampled (e.g. by a factor of 5) and a spread spectrum watermark is embedded in the resulting 1D signal. The watermark is embedded in each of the phases. In our implementation of this work (the adaptive embedding algorithm for improved fidelity) the watermark could made to survive camcorder capture, but not without introducing visible flicker in the exhibited imagery.

The trend to low frequencies as the key to extreme robustness was also pursued briefly by Fridrich [6], who creates a low-frequency spatial watermark by processing (cellular automata followed by low-pass filtering) a white noise pattern. This watermark has the potential to be extremely robust, but will be quite a challenge for imperceptible embedding since the low-frequency watermark is not tailored to the host imagery.

## 3. APPLICATION REQUIREMENTS

***Fidelity*** refers to the perceivable difference in quality between the original, unwatermarked content a watermarked copy; i.e., a high fidelity copy is one that is visually indistinguishable from the original. Fidelity is critical for the intended initial beneficiaries of this technology – entertainment content providers – who will not accept processing that detracts from the intended perceptual experience. Our target for fidelity is < 1 JND of distortion, as measured in Sarnoff's JNDmetrix[2] units, and/or with subjective rating by experts and trained naïve observers.

***Robustness*** can be described as the ability of the watermark to be recovered after probable distortions. For this application, the set of probable distortions includes all combinations of exhibition capture, low bitrate compression, D/A/D conversion, VHS recording, noise addition, noise reduction, spatial and temporal filtering, changes in brightness or contrast, cropping, and geometric distortions (translation, rotation, scaling, perspective distortions). Without a high level of robustness, a watermark is much less attractive to content owners, since its lack implies that untraceable but marketable pirated content can get out. We assume that as the level of distortion increases, the recovery rate (an thus the achieved capacity) declines. Thus, robustness for a given application can be quantified in terms of minimum data rate for watermark recovery over an agreed upon epoch.

***Security*** is the ability of the watermark to withstand specific attempts by an adversary to interfere with its intended purpose. For a forensic tracking watermark, such interference might include unauthorized embedding (i.e. forgery) or unauthorized removal. Many such "attacks" have been presented in the literature, as summarized in [3]. Without security, a forensic watermark is worthless against skilled pirates. Security against unauthorized embedding can be accomplished with the use of encryption and digital signature technologies. Security against unauthorized removal can be measured in three complementary ways:

1. In JND units, the level of distortion required for successful watermark removal (*jamming*) in a cryptographically secure setting. Here, a distortion above 5 JNDs is the target, estimated to produce unusable pirated content.
2. The probability of breaking the security by chance. This probability must be low with respect to the maximum computing time that even the most dedicated band of pirates would endure.
3. The probability that world-leading watermark security experts could break the security, expressed here in the number of person-hours spent trying to crack our security by a local Princeton University team now famous for a recent watermark-breaking success.

[2] http://www.jndmetrix.com

*False positive probability* is the likelihood that a watermark is detected in unwatermarked source. For many applications, the false positive probability is critical and must be provably low (given some assumptions about the distribution of source material.) However, in this application, we do not expect to be applying the detector to unwatermarked material since we assume all exhibited material contains a tracking watermark. We are more concerned with bit errors as we do not want to accuse the wrong customer of piracy.

This application has asymmetric *computational costs* for the various parts. As will be seen in the following section, our embedding process involves two major steps: a preprocessing step in which the motion imagery is analyzed and an insertion step in which the content is actually modified. Preprocessing need only be applied one time per movie and can therefore be more costly, but insertion must be applied a number of times per copy. The last application of the insertion process occurs during the actual showing (addition of the time stamp) and will likely be performed by a projector, so this insertion must be a real-time and have low computational cost.

Detection, on the other hand, has no real-time restrictions. Detection is only performed occasionally (when the content owner obtains pirated content), is performed by, or at the behest of, the content owner, and can therefore make use of the original imagery during the detection process. While there is some urgency in discovering the source of the piracy, there are no strict computational restrictions.

The *payload* requirements for this application are not well defined. As mentioned earlier, some have suggested that the watermark must carry 34 bits of data [7]. Those authors also suggest that these 34 bits must be recoverable from a single frame of a pirated movie. While higher bit rates may always be better, a more modest target may be acceptable for this tracking application. For example, others have suggested that the payload must be 100 bits recoverable from a more significant portion of the movie [8]. If a "significant portion" is defined as 2-5 minutes[3], then an embedded bit rate of approximately 0.5 to 1 bit per second is sufficient.

## 4. THE APPROACH

In this section, we describe a forensic watermarking approach that simultaneously maintains the fidelity, robustness, and security needed for digital cinema and other digital distribution applications. In summary, we achieve the required level of robustness and fidelity by restricting the watermark pattern to be very low frequency in both space and time. In the following, we elaborate on the approach by first providing motivation for use of low spatial and temporal frequencies. We then discuss techniques implemented to help ensure security against unauthorized removal. Finally, we describe the architecture of both the embedding and detection processes.

### 4.1 Motivation for use of very low spatio-temporal frequencies

Sensitivity to low spatiotemporal frequency distortions is much higher for physical measurement devices (e.g., photometer or digital camera/scanner) than it is for humans. For example, photometric measurement of CRT luminance non-uniformity often shows up to 25% center-to-corner drops[4], without any noticeable effect to the human viewer. More general corroboration for this claim derives from inspection of the human visual contrast sensitivity function, such as that reproduced in Figure 2 below. Here, a substantial dip in visual sensitivity is evident at the low spatial, low temporal frequency corner of the plot, a region in which the sensitivity of physical measuring devices is remaining roughly constant. Any disparity of sensitivities such as suggested here for low frequencies is useful in discovering watermark carriers that can be robustly measured by physical devices, but are invisible to humans.

Corroboration of this analysis is offered by van Hateren [9], who modeled the spatiotemporal filters of mammalian vision, based on the spatial and temporal statistics of natural images, and on an optimization assumption that maximized the flow of information through noisy channels of limited dynamic range. In his analyses, and as reproduced in Figure 3 below, van Hateren noted that information is maximal at the lowest spatial and temporal frequencies, but that the optimal

[3] Personal communication with studio executives.

[4] Based on Sarnoff collective wisdom from years in the CRT design business

filter reduces sensitivity dramatically to these frequencies, since these frequencies are so pervasive in natural images that they tend to occupy too much of the dynamic range of the channel, at the expense of other frequencies. From our perspective, this discrepancy between information and sensitivity provides strong justification for the modulation of low frequencies in watermark embedding, since it implies that there is a great deal of information availability in low frequencies, even though humans are relatively insensitive to it.

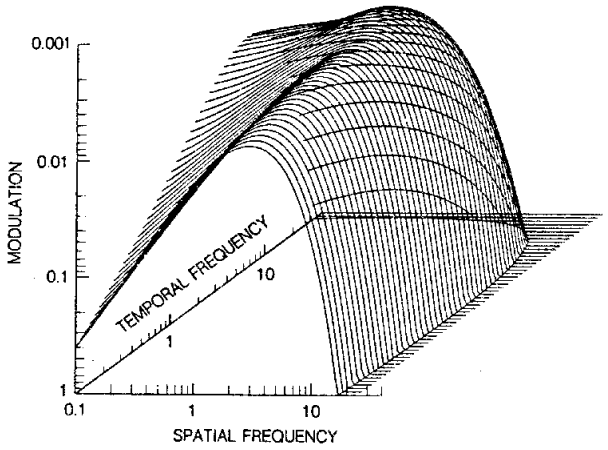


**Figure 2. Spatiotemporal contrast sensitivity plot, reproduced from Kelly [10]. Note the dip in sensitivity at the low spatial, low temporal frequency corner of the plot.**

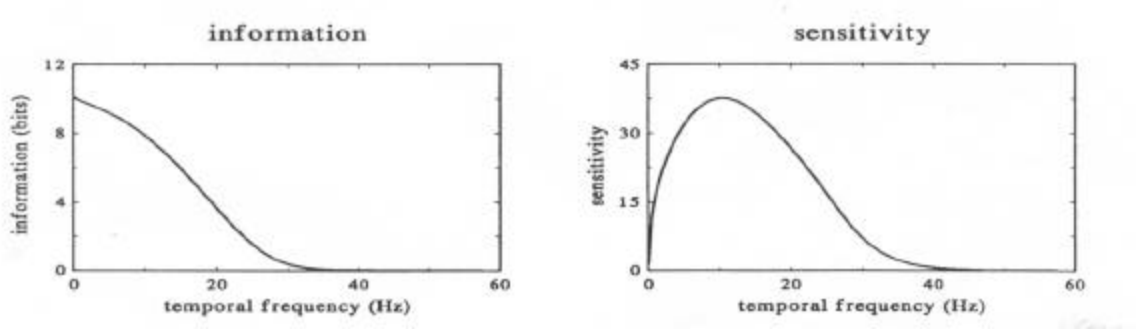


**Figure 3. Information in natural image sequences vs. visual sensitivity as a function of temporal frequency, reproduced from van Hateren [9]. Note the discrepancy between information and sensitivity at the low frequency end of the two plots.**

From this discussion, we conclude that Cox et al. (e.g., [11]) were correct that watermarks should be hidden in perceptually salient components of the image. High frequencies should be avoided because the reduced sensitivity of the HVS at high frequencies allows these components to be distorted by processing or attacked by adversaries without significant degradation to the fidelity. Thus, watermark data in these components can be damaged. One might conclude, from the reduced HVS sensitivity at the low spatio-temporal corner of the CSF of Figure 2, that low frequencies should be avoided for the same reason. However, the high degree of information in the low frequency components makes them difficult to distort without degrading the fidelity. Most optical and computational processes that are applied to moving imagery and result in "watchable" quality tend to reproduce these low frequency / high information components with high fidelity. For example, camcorder piracy, which often degrades middle and high frequencies to the extent that typical spread-spectrum watermarks are significantly damaged, still generally produces a video stream from which a viewer could described in detail what is happening in each scene.

Very low frequencies have, until now, been notoriously difficult for watermarking. It is common to read in the watermarking literature, "mid frequencies are used because the high frequencies are not robust and the low frequencies cannot be imperceptibly modified[5]." We address this fidelity challenge by composing the watermark pattern from a number of local, low frequency carrier functions, and by using a masking model to analyze the original content and determine where, in space and time, each of the carriers can be imperceptibly added. The watermark can be described parametrically as a list of carrier descriptions (shape and size, spatial and temporal location, amplitude, etc.). The design of the carrier functions (shape, size, amplitude) controls the robustness of the watermark and, along with the masking model (defining allowable spatial and temporal locations), controls the fidelity as well.

### 4.2 Security

There are two security concerns of interest: security against unauthorized embedding and security against unauthorized removal. The former is necessary to prevent an adversary from "framing" another party by embedding a different theater identifier into a pirated movie. This security is provided by standard cryptographic techniques; prior to error coding, the message is encrypted and a digital signature is appended. Security against unauthorized embedding is enhanced with the use of a secret watermark key as described below.

The natural security of high dimensionality spread-spectrum watermarks, due to the statistical orthogonality of any two such high dimensional patterns, is largely lost when watermarks are constrained to low frequencies, given the severe dimensionality reductions imposed by this constraint. Therefore, alternate means of security against unauthorized removal are needed. To achieve this security, we algorithmically select a set of possible watermark carriers for each spatio-temporal region based on the ability of the underlying cover work to mask the carriers. Then, using a secure random number generator, we choose only one of these for actual inclusion. The logic behind this Kerckhoff's assumption-based approach is that an adversary, even knowing the details of the selection algorithm, will not be able to know which of the possible carriers was used for any given region. He or she will therefore be forced to jam across all of these carriers. (By jamming, we mean addition of noise to a carrier at an amplitude similar to that of the modulation of the watermark itself.)

Such a system achieves good security to the extent that jamming across all of the potential carriers results in unacceptable visual distortion, even if any one of the carriers could be jammed individually without serious visual loss. But how do we develop a set of carriers with the required joint properties that (1) any one of them could be inserted invisibly at a magnitude sufficient to ensure recovery, but also that (2) the insertion of all of them, each at a magnitude below its own threshold of visibility, results in a strongly noticeable degradation?

An answer can be outlined with reference to the psychophysics literature of sub-threshold summation. Experiments such as those of Graham and Nachmias [12], Wilson and Bergen [13], and Watson [14] based their evidence for the existence and characterization of separate visual channels (e.g., sensitive to different spatial frequency ranges) on the notion that combinations of sub-threshold signals that stimulate a single channel will result in a detectable pattern, whereas similar magnitudes of signals across different channels will not. Using this same logic, we design our multiple carriers for a given region so that they all tend to stimulate a single visual channel or a small number of visual channels. In this way, carriers that are individually below threshold will tend to produce distortions that rise above threshold when multiple such carriers are jointly present.

### 4.3 Embedding Architecture

Figure 4 below shows the architecture for embedding the watermark into a stream of motion imagery. An original sequence enters the process and is first subjected to a vision model-based masking computation that determines spatiotemporal regions of the sequence that could invisibly support the addition of our carriers. The resulting masking map is scalar, and can be used to modulate watermark amplitude as a function of spatiotemporal location. Alternatively, it can be thresholded, with regions above threshold indicating allowable locations for fixed amplitude carriers. We have chosen the latter approach in our current implementation.

---

[5] We state this "quote" without reference because it is so common in the literature.

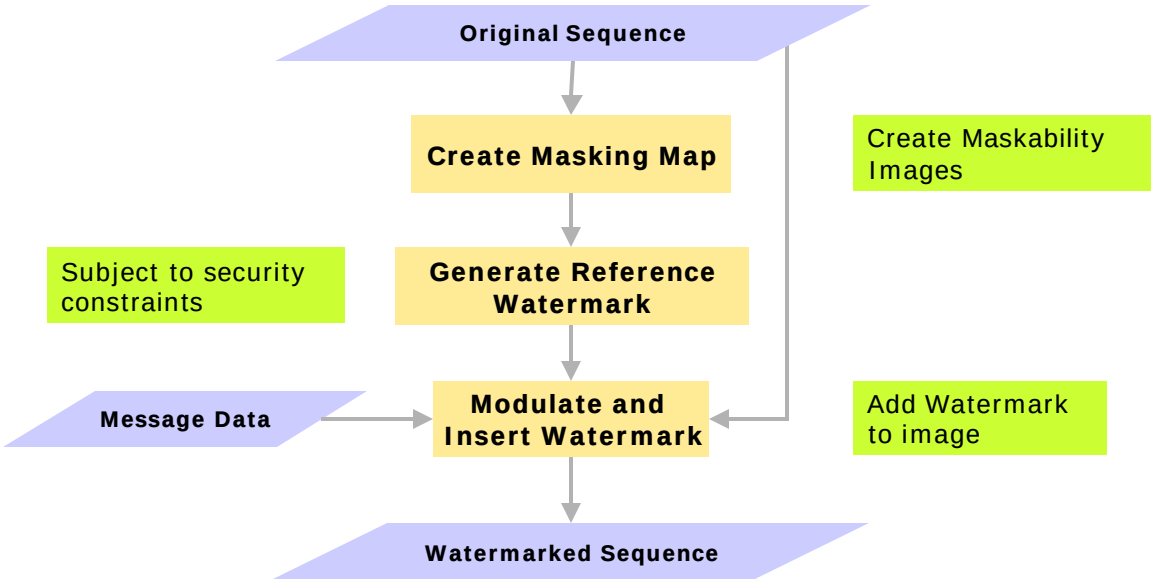


**Figure 4. Embedding architecture**

The reference watermark pattern is then constructed by sparse sampling of the allowable carriers, according to a key-based pseudo-random number generator, as suggested in the security discussion above. A message, after appropriate encryption and error coding, is represented by mapping bits to the sign of each carrier in the reference pattern. This modulated watermark is then inserted into the sequence.

### 4.4 Recovery Architecture

Figure 5 below shows the architecture by which the watermark is recovered from a captured video sequence.

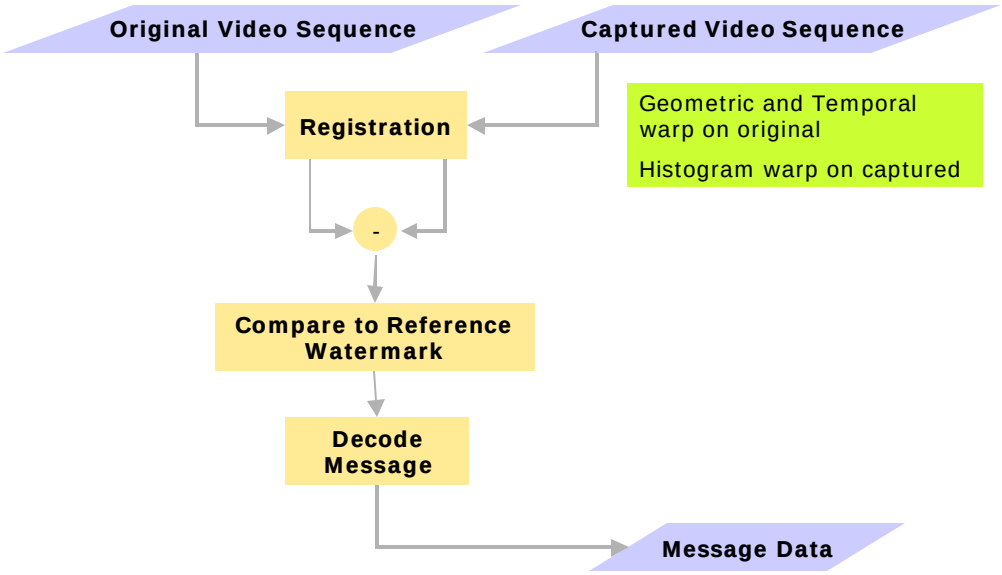


**Figure 5. Detection architecture**

As indicated in the figure, a captured video sequence is first registered with the original sequence. This registration is performed in space and time, as well as grayscale. According to the method developed by Cheng [15], the original is warped so as to be aligned temporally and spatially with the capture. Histogram warping is applied to the grayscale values of the capture to align the luminance with that of the original. A difference sequence between these two input sequences is computed, and then locally compared, using standard correlation techniques, to the carriers comprising the reference pattern. Correlation signs are interpreted as coded bit values and the message is then decoded.

## 5. EXPERIMENTAL RESULTS

We performed experiments to determine both fidelity and robustness of the watermarking algorithm. Colleagues at Princeton University performed security analysis and testing. In this section, we summarize the results of all three testing tasks.

### 5.1 Fidelity testing

Three two-minute HD Cinemascope clips of digital cinema content were obtained and subjected to the watermark insertion process. These three clips, labeled here "House", "Buddha", and "Love", were chosen from available digital content to span a wide range of cinematographic style and settings. All three clips were 10-bit in each of three color channels, 1280x1024 (non-square) pixel spatial resolution[6], and 24 frames per second temporal resolution. These clips were displayed using a TI DLP-based Christie projector, onto twenty-foot wide screen in a cinema viewing room at Sarnoff. Peak luminance of the projector was approximately 24 cd/m$^2$, and the ambient light was below 5 lux, chosen to recreate typical cinema ambient lighting. Observers viewed the screen from approximately two picture heights.

Five expert observers, familiar with the details of the watermark and adept at visual detection tasks, participated in a two alternative forced choice experiment in which each trial consisted of two presentations of the same clip, once with and once without the watermark present. Observers were required to indicate which of the two clips contained the watermark. Each source clip was used in four such trials, with position (first or second) of the watermarked version varied randomly from trial to trial. An experiment therefore consisted of twelve trials, with each trial lasting approximately four minutes.

The results are simple to report: no observer was able to determine reliably the identity of the watermarked sequence in any case. These results matched debriefing comments by the observers, all of who reported a complete inability to determine watermark presence or absence.

### 5.2 Robustness testing

Robustness of the watermark recovery process on these same clips was tested after camcorder capture, and after numerous synthetic distortions. Table 1 below shows the robustness results after camcorder capture. Here a DV camcorder was mounted on a tripod and placed approximately 3 picture heights away from the screen. The camera was centered horizontally and directed at the center of the screen at an upward angle of approximately 10 degrees. Automatic focus, white balance, and gain were disabled and the lens was zoomed to capture the full width of the projected imagery. Due to the difference in aspect ratio, this "letterbox" format capture resulted the capture of broad "stripes" at the top and bottom of the picture. Of the 486 captured lines, the motion picture imagery is approximately 300 lines in the center.

[6] In cinemascope format, the imagery is intended to be viewed through an anamorphic lens with horizontal : vertical scale ratio of 1.91:1.

**Table 1. Watermark Recovery After Camcorder Capture**

| Source sequence | Number of bits embedded | Number of bits correctly recovered | Recovery percentage |
|---|---|---|---|
| Love | 28 | 28 | 100% |
| House | 80 | 79 | 99% |
| Buddha | 56 | 53 | 95% |

The totals can be summarized as follows: over six minutes of source material, at a bit-rate of 0.46 bits/second, the watermark was successfully recovered after camcorder capture with a 97.5% probability. Although we have not yet implemented any error correction coding, the bit error rate obtained here is obviously amenable to such operations.

Note also, that we envision that the entire motion picture will be marked at this bitrate. Assuming an average bitrate of 0.46 bits/second, a typical 100-minute movie will hold 2760 bits. If a 34-bit message is encrypted and error coded it could easily expand to 100 bits. This coded message could be represented 27 times in a 100-minute movie. One period will be recoverable from any 3.6 minute clip and, with longer clips, corroboration of decoded message from additional periods will improve the robustness (bit errors not corrected by the ECC), increase the security against unauthorized removal (an adversary will have to successfully attack all copies of the embedded message), and decrease the probability of a false accusation (likelihood that the same error is introduced in multiple periods of the message is vanishingly small.)

The sequence Love was also subjected to a number of different synthetic distortions, including:

1. Horizontal shifts of 1, 2, 5, 8, and 16 pixels
2. Spatial low-pass filtering with standard deviation of 1 and 3 pixels
3. Addition of uniform white noise with peak values of 3.125%, 6.25%, and 12.5% of the maximum representable amplitude
4. $1/f^2$ noise with peak value of 6.25% of the maximum representable amplitude
5. Quantization to 4 and 6 bits (from 10 bits)
6. Gamma (exponentiation) of 0.8 and 1.2

In all of these cases, the 28 marked bits were recovered with 100% reliability, even when the introduced distortions were glaringly noticeable. Of special interest is the $1/f^2$ noise, which has power concentrated at the low frequency end of the spectrum, and is thus more likely than other distortions to interfere with low frequency watermark modulations. Even here, with highly disruptive noise, the watermark was recovered perfectly.

### 5.3 Security testing

There are a number of techniques that an adversary could use to try to thwart watermark detection. One approach is to modify the low spatio-temporal frequencies of the marked imagery (perhaps guided by the same visual masking model used for embedding) with the hope of colliding with the embedded carriers. The probability of success can be determined analytically by considering the sparseness of the carrier selection with respect to the maximum packing possible without introducing the subthreshold summation artifacts discussed in Section 4. A second approach is to introduce distortions that make registration difficult. Note however that while the registration process employed in this work was completely automatic (after a single manually generated seed for the spatio-temporal registration), the process of forensic analysis could make use of manual intervention and improvements in the registration tools. A third category of attack would be attempts on behalf of an adversary to estimate the location, shape, and size of the embedded carriers.

Our security analysis and testing was performed by colleagues at Princeton University. They were selected for this task in part due to their success in a similar analysis and testing of the SDMI, Phase-II watermarking proposals [16]. A summary of their conclusions is that they were able to increase bit error rates to approximately 20%, with some significant loss in visual quality, by spatially non-uniform manipulations of luminance and contrast. These attacks were designed to confuse the registration algorithm and so could be at least partially thwarted by improvements in registration, for example by performing more local registration operations or, because forensic operations are not strongly limited in time or compute power, by adding manual adjustments to the registration approach. However, even if

these registration refinements are not performed, the Princeton team concludes that these bit error rates are still well within the range of error-correcting codes, and therefore that the algorithm described here is quite suitable for forensic applications.

## 6. CONCLUSIONS

We have presented a watermark approach that simultaneously maintains excellent fidelity as well as robustness to camcorder capture and other highly visible distortions. Current results on the security of the approach are also encouraging. Together, these results indicate that the watermark presented here is quite viable for forensic application to digital cinema and other digital distribution of motion imagery content.

## ACKNOWLEDGEMENTS

This work was funded in part under the U.S. Department of Commerce, National Institute of Standards and Technology, Advanced Technology Program, Cooperative Agreement Number 70NANB1H3036. The authors also wish to thank our colleague Nurit Binenbaum, who did most of the actual work to generate the results presented here. We also thank Professor Bede Liu and Scott Craver of Princeton University for their security analysis and testing.

## REFERENCES

1. B. Hunt, "The Role of Watermarking in Confronting Digital Cinema Piracy", SMPTE DC28.4 Watermarking Workshop, March 15, 2001.
2. M. Watson, "Conditional Access Study Group Watermark Recommendations", SMPTE DC28.4 Watermarking Workshop, March 15, 2001.
3. I. J. Cox, M. L. Miller, and J. .A. Bloom, *Digital Watermarking*, Morgan Kaufmann Publishers, 2001.
4. C. Honsinger and M. Rabbani, "Data embedding using phase dispersion," International Conference on Information Technology: Coding and Computing (Invited Paper), 2000.
5. J. Haitsma and T. Kalker, "A Watermarking Scheme for Digital Cinema", International Conference on Image Processing, 2001.
6. J. Fridrich, "Digital Watermarking by Adding Random, Smooth Patterns", *United States Patent* 6,101,602, 2000.
7. C. Honsinger and P. Jones, "Challenges in Presentation Watermarking for Digital Cinema", SMPTE DC28.4 Watermarking Workshop, March 15, 2001.
8. A. Bell, Private Communication, August 2002.
9. J. H. van Hateren, "Spatiotemporal Contrast Sensitivity of Early Vision", Vision Research, 33, 257-267, 1992.
10. D. H. Kelly, "Motion and Vision II: Stabilized Spatio-Temporal Threshold Surface", Journal of the Optical Society of America, 69(10): 1340-1349, 1979.
11. I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Images, Audio, and Video", Proceedings of the International Conference on Image Processing (ICIP '96), vol. III, pp. 243-246, 1996.
12. N. Graham and J. Nachmias, "Detection of grating patterns containing two spatial frequencies: A test of single-channel and multiple-channels models", Vision Research, 11, 251-259, 1971.
13. H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision", Vision Research, 19, pp. 19-32, 1979.
14. A. B. Watson, "Summation of grating patches indicates many types of detector at one retinal location", Vision Research, 22, 17-25, 1982.
15. H. Cheng, "Spatial-Temporal and Histogram Registration of Captured Video Sequences", submitted to the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003.
16. S. A. Craver, M. Wu, B. Liu, A. Stubblefield, B. Swartzlander, D. Wallach, D. Dean, and E. Felton "Reading Between the Lines: Lessons from the SDMI Challenge", Proc. of the 10th USENIX Security Symposium, August 2001.