INVISIBLE WATERMARKING FOR DIGITAL VIDEO - APPLICATIONS AND CHALLENGES

Anthony Huggett and Clive Stubbings*

Abstract

Digital Television offers us many potential benefits in picture quality. Included amongst these is the ability to store and copy material with no loss of quality or fidelity. It may seem that in this environment the invisible watermark has many applications including audit trails, copyright protection, piracy prevention, copy control, revenue collection and archive tagging. However, digital television systems also offer many potential challenges to the usefulness of a watermark.

Digital Television

Digital television is now a reality in many countries around the world. In order to get the programme content from the studio to the viewer, it must first undergo a process of lossy compression encoding in order to reduce the amount of data by a factor of 40 or so. This is followed by a quasi error free transmission chain in which errors occur less than once per hour. The viewer then uses a digital receiver/decoder (currently a set top box but integrated digital televisions will follow) in order to produce video images and sound.

One advantage of digital television is the quality of the picture available to the viewer – it is as good in the viewer's lounge as it was when it left the studio at the output of the compression encoder. This has led to concerns from content providers and broadcasters over the protection of their rights to the content; copyright protection, copy prevention and copy detection have all been suggested as potential applications for watermarking technology.

Watermarking can also be applied to content detection. It may be easier to detect a watermark in a particular piece of content than to detect that content itself. This has applications in archiving, advert detection (revenue payment) and revenue collection.

Compression

The MPEG 2 standard has achieved world wide dominance as the compression standard for broadcast digital video, whatever the delivery medium (cable, terrestrial or satellite). If we are to use a watermarking scheme in this arena, we need to be aware of the characteristics of MPEG 2 compression

- It is a lossy compression scheme.
- The aim of the compression algorithm is to discard detail that is imperceptible to the eye but which nevertheless exists in real images. This detail is exactly what an invisible watermark should be. If the MPEG 2 algorithm were perfect in its ability to discard imperceptible data, then all invisible watermarks would be removed by the compression process! No two compression engines perform the same and the designers are constantly striving to improve the ability of the systems to remove imperceptible detail.
- Noise, particularly noise of high spatial frequency on the video data badly degrades the performance of the compression engine. Many bits are taken up in transmitting the noise, which leaves fewer bits to transmit the image.

* Anthony Huggett and Clive Stubbings are employed by Tandberg Television.

Invisible Watermark applications in Digital Broadcast Systems

Within the digital broadcast chain there are a number of applications for which the use of invisible digital watermarks has been suggested. Here we select and describe five, and later we will look at some of the potential issues that arise:

1 Copyright protection: A content owner places an invisible watermark on the content. This then proves the ownership. One example where this could be applied would be proving the ownership of news footage, where multiple news agencies were at the scene of the event. The "really good" picture has significant value but a number of different organisations could all claim ownership.

2 Piracy Prevention: Broadcasters are worried that content, particularly feature films that are transmitted by them may be digitally copied and hence digitally perfect illegal copies could be made. Using watermarks to counter this would involve a system in which the content is watermarked by the equipment through which it passes, including the receiver. Detection of the watermark in a recovered copy of the content would then identify the receiver that had been used, and hence the owner could be identified and prosecuted.

3 Copy Control: The invisible watermark would be used as part of a scheme to enable or disable copying in the receiver/recorder [1].

4 Advert checking: A large proportion of some television broadcasters' revenue comes from advertising. To the advertiser, there are a number of important parameters related to the screening of his advert and these significantly effect the price he pays. Factors include time of day, location within a commercial break and location with respect to other commercials. A watermark could be placed on the advert and every time the advert was transmitted the watermark could be detected and the playing of the advert logged. An automated system could then be used to check that the broadcaster was fulfilling its contractual obligations.

5 Archive tagging. Watermarks can be used to identify images or video segments in an archive. This is effectively using the watermark as a means of attaching metadata to the video sequence in a way that ensures that the data and video do not become separated.

Inherent Watermarking Challenges

There are a number of requirements for an invisible watermark for all digital video applications. In some senses, an invisible watermark is "incompatible" with a compression system and careful design is required by the watermark designer.

As we said above, video compression works by removing detail that is invisible to the eye. With a perfect compression system, it would be impossible to add invisible detail and preserve this through the compression chain. Fortunately, with the current implementations of the MPEG 2 compression system it is possible to design the watermark in such a way as to be preserved by the compression process.

Part of the MPEG 2 approach to image compression is to discard details with high spatial frequency.
Experiments have shown that watermarks with low spatial frequency exhibit good resilience to MPEG 2 compression chains. The popular spread-spectrum technique of embedding the watermark in the low frequency DCT coefficients is one such method.

There is a problem in the opposite direction too. Additive noise is potentially a problem in compression systems, and many systems have extensive and complex pre-filters to remove noise. Additive noise can be shown to hurt the compression performance by "stealing" bits – bits are required to transmit the noise and the quality of the image suffers. There are 2 consequences of this:

- The watermark must be designed in such a way that the watermark signal, when added to the picture, does not degrade the compression performance requiring more bitrate for the same image quality.

- The watermark must also be resistant to any noise filtering which may be carried prior to compression.

Again, low spatial frequency watermarks of practical amplitude do not appear to adversely affect the compression chain.

For many applications, the watermarking algorithm may not be secret. Where security is required, the watermark must then be identified by some kind of secret key. If the key is changed on every frame then we need to know which frame we are on before we can detect the watermark. If we use only one key for the whole content (e.g. a feature film), then the watermark may be the same on every frame. Both of these approaches have implications for the security and manageability of the watermarking scheme.
For most watermarking applications we would like to be able to detect the watermark without having to check all the frames. The mark must therefore be resistant to temporal cropping.
Furthermore, in the case of pan & scanned films (where films are converted from cinema to TV aspect ratio by selecting the portion of the image of "most" interest), we have to be resistant to constantly varying spatial cropping.

Deliberate Attacks

In addition to resilience to the compressed digital environment, watermarking schemes must often be resistant to deliberate attacks. We consider two classes of deliberate attack on a watermarking scheme for a particular application.

- Direct attacks attempt to remove, attenuate, obscure or otherwise render the watermark undetectable in the content.
- Indirect attacks leave the watermark undamaged, but seek to undermine the validity of the scheme that uses the watermark as its basis – the watermark is still there, but it doesn't help.

There is not room here to fully discuss the resilience of various watermarking methods to direct attacks. We will assume that a method may be found which is sufficiently resilient to signal processing attacks (cropping, scaling, rotating, quantizing, filtering) so that it cannot be removed without severely damaging the appearance of the content.

However, there are other direct attacks that are especially relevant in a broadcast environment.

- Collusion. If multiple copies of the same content are available to the attacker with different watermarks, then the content may be combined in order to attenuate the watermark. A modest PC now has processing power and hard drive storage to perform frame-by-frame averaging in reasonable time. Averaging 4 copies of the source together that are differently watermarked must have the effect of attenuating each mark by a factor of 4, or more if the watermarks are not mutually orthogonal.
- Known frames. Any watermark that occurs on a known frame e.g. a black frame, a white frame, a title screen etc. may be isolated by subtraction of the known frame. Clearly, the known frame in the video can be replaced by a frame which looks identical but which has no watermark. More importantly, if the mark and the watermarking scheme are known to the attacker (the secrecy of the parameters of the watermarking scheme being its only defence) then obtaining the mark on its own might provide a very good starting point from which to search for or estimate the parameters. Great care must be taken so that either known frames are left without a watermark or so that obtaining an image of the watermark alone by comparison with a known frame does not help a hostile party in guessing the watermarks on the unknown frames.
- Averaging. If a hostile party knows that a number of video frames are marked with the same watermark, it is possible that these frames could be averaged together to give a good estimate of the watermark. This could then be used to attenuate the watermark in all the frames with that watermark.

All of the above attacks emphasise the need for very careful control and usage of the parameters (keys) used to create the watermark pattern. If the same parameters are used throughout the video so that the same watermark is produced on every frame, then averaging attacks become easier. If we use a different set of parameters for every frame then the resultant database of keys will be very large; we will require some way of finding the correct key in order to detect the watermark for every frame. If we are providing near video on demand by broadcasting the same content at staggered intervals, then we must make sure that the same parameter set is used for the same frame of every staggered broadcast in order to prevent collusion attacks.

We now consider indirect attacks:

- Inversion. For the purposes of establishing copyright, the watermarking algorithm must be in the public domain. Otherwise, anyone can claim that their watermark is in the image by choosing a scheme and finding some parameters that cause their mark to be found in the image. Any method of copyright protection that relies on detection of the mark by comparison of the public (watermarked) copy with an original (unwatermarked) image is fundamentally flawed, since all that it proves is that the watermark exists in the difference image. Consider the following situation: the legitimate copyright owner places his content C in a safe, and only distributes a single watermarked copy C+W, where W is a watermark that may only be detected by direct comparison with C (for instance by subtracting one image from the other). Now suppose that a pirate takes the public content C+W, subtracts his own watermark V, and places his image C+W-V in a safe. Consider trying to use the watermark to prove ownership. The legitimate owner detects his watermark on the public copy by performing (C+W)-C=W, and on the pirate's safe copy by performing (C+W-V)-C = W-V. However the pirate may similarly compare the public copy with his private copy yielding just his watermark, (C+W)-(C+W-V)=V, and the legitimate party's copy with his private copy yielding the difference of the two (legitimate and private) watermarks C-(C+W-V)=V-W. Thus the pirate's evidence is as good as the legitimate owners evidence, unless we knew a priori which of the 2 safe images is the original, but that is what the watermark is supposed to be telling us! This is an example of an inversion attack as proposed in [2]. More generally, a scheme is invertible if it is possible for a hostile party to choose parameters of the scheme so as to cause their watermark to be detected in both the public copy and the real owner's private copy (to which the hostile party has no access!) In a later paper, the authors of [2] showed that it was a necessary but not sufficient condition for the establishment of copyright ownership that the watermarking scheme be non-invertible [3]. It has recently been suggested [4] that it may be very difficult, if not impossible, to prove mathematically that any particular watermarking scheme is non-invertible, although it may be possible to make estimates of the computational time required to invert a particular scheme. Furthermore, it may not be necessary for the pirate to actually perform the inversion. The suggestion that it is possible that the legitimate party had in fact inverted the watermark may be sufficient to discredit the validity of the scheme.
- Circumvention. This attack does not degrade the watermark itself in any way. The effect is to limit the usefulness of the watermark by discrediting or otherwise rendering useless the information that the system gives us. If set top boxes were to uniquely watermark the content that they decoded we might think that this would allow us to identify persons who might be illegally copying the content. However, this system is most useful if the pirate uses a correctly purchased and registered set top box. Similarly, one might implement a copy control system using watermarks that enable or disable copying

within the recorder. This requires that all recorders implement the inhibition feature. "Professional" recorders would doubtless become available which have the inhibition feature removed (as happened with digital audio tape in the early 1990s.)

We will now take each of our applications and discuss the likely attacks to which they are vulnerable.

Copyright protection

With copyright protection, there is only one copy of the content in the public domain; this copy contains an invisible watermark, which (in theory) unambiguously identifies its owner. Because there is only one copy in the public domain, collusion attacks are not possible. A suitable key management strategy must be used. Furthermore, we must ensure that watermarks do not occur on known frames e.g. black frames, or that if they do occur, then this tells the pirate nothing about the watermark in the rest of the content.

For this application, we are potentially vulnerable to an inversion attack. In our news example above, both parties could make credible claims to have been the original owner of the material. It would appear that this challenge, of making a logically watertight case that the existence of an invisible watermark in a piece of content proves its rightful owner, is the most difficult.

Piracy prevention

The aim here is not to use the watermark to prove that a particular individual or company is committing piracy. The evidence for this would be the seizure by the police of equipment etc. Rather, we are using the watermark to assist in identifying the pirate in order to assist with the discovery of the equipment.

Where we are applying the watermark in the receiver to uniquely identify the pirate, there will be many differently watermarked versions of the same source within a given locality. Collusion attacks are thus a definite possibility – a pirate would need to obtain several different copies of the same video source and use a PC to perform frame by frame averaging to attenuate the watermarks.

Circumvention of the scheme is even simpler. Assuming that our system is targeting the professional pirate, who is already engaging in a criminal act of theft, it is unreasonable to suppose that he or she will be using a legitimately purchased and registered decoder once he knows that this might incriminate him or her. It is more likely that the decoder would be obtained from the black market and used for a short period before another was obtained. Thus, the watermark found on a pirated video might not point to the pirate.

Watermarking one stage up the broadcast chain, i.e. at the transmitter, may not be helpful. For example, a satellite footprint may cover several countries, and terrestrial single frequency networks may, in theory, span continents.

Copy Control

In copy control applications, we use a watermark as part of a system to either watermark the video to prevent copying, or watermark it to enable copying. The copy inhibition is provided by the recorder within the receiver. This requires the watermark to be secure against removal (or faking). This may be difficult to achieve since the algorithms used in the recorder are likely to get into the public domain. That aside, there are no special weaknesses in the scheme. Only one copy of the video need be in the public domain, preventing collusion attacks. Inversion is not an issue in this case.

The problem with this application is preventing the manufacture of "professional" recorders that do not contain the copy control mechanism. This was the problem with Digital Audio Tape (DAT) recorders and it is very likely that the industry would be unable to prevent such devices reaching the market if this were adopted as a solution for digital video.

Advertising Monitoring/Revenue Collection

With this application, a watermark is used to identify a particular piece of content; it being easier to automatically identify a known watermark than to identify a particular piece of content. In both advertising monitoring and revenue collection there is a measure of professional trust between the two parties. Nevertheless, it is desirable that the broadcaster would not be able to fake the watermark for a particular advert and hence fraudulently gain extra revenue from the advertiser. For revenue collection where a broadcaster is paying to broadcast a piece of video, it is desirable that the watermark could not be removed and hence the broadcaster could avoid paying.

Archive Tagging

With this application, there is no hostile party; hence, there are no deliberate attacks.

Conclusions

In writing this paper, we have sought to highlight the challenges that are posed in designing digital watermarks and their associated key management strategies for particular video applications.

We can see several applications where invisible watermarks may be useful for digital video. However, as watermarking solutions begin to emerge into the marketplace for all these applications, it is important that the purchasers be made aware of any limitations in the technology that they are buying.

The design of invisible watermarking schemes for digital video applications rather than still images imposes a number of additional constraints upon the watermarking scheme. Not only must the watermark be resilient to compression, but we must also be very careful in our key management strategy so that the resistance to direct attack is maximised. Indirect attacks upon the credibility and usefulness of a watermarking scheme are as important as direct attacks upon the watermark itself. For the purposes of copyright protection, pirate identification, and copy control, they may be the most important challenge to be overcome.

For less hostile environments watermarking technology is already capable of providing good solutions.

References

[1] DAVIC, "DAVIC's Seventh Call for Proposals", *The Digital Audio-Visual Council Fifteenth DAVIC Meeting,* Hong Kong, 9-13 December 1996.

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

[3] Craver S., Memon N., Yeo B.L., and Yeung M., "On the Invertibility of Invisible Watermarking Techniques", in *Proc. Int. Conf. Image Processing* 1997, vol.1, pp. 540-543.

[4] Zeng W. and Liu B. "A Statistical Watermark Detection Technique Without Using Original Images for Resolving Rightful Ownerships of Digital Images", *IEEE Transactions on Image Processing,* vol. 8, no.11, November 1999, pp1534-1548.