

Watermark-based Media Annotation For Blu-ray Disks

Ron van Schyndel

School of Computer Science and Information Technology

RMIT University

ronvs@rmit.edu.au

ABSTRACT

This short paper is an interim report on a project to embed watermarks in multimedia where the watermark provides a meta-content layer, and can be used to embed meta-content, so that a compatible viewer can use this layer to: add content value in the form of annotation; link with existing data-bases; and inform accessible devices.

We propose the meta-content layer to be added to media intended for wide-scale distribution, in which the distributor or content-creator intends to add considerable value in the form of meta-data or annotations.

The watermark then serves to provide a benefit to the customer instead of only to the content-creator for the price of including a watermark. Removing the watermark removes this additional benefit. Retaining it allows the watermark to remain a part of the DRM that could be used to ensure compliance.

Author Keywords

Media indexing, Metadata, Media Ontologies, Digital Watermarks

ACM Classification Keywords

H5.1. Multimedia Information Systems: video, K.4.2 Social Issues: Assistive technologies for persons with disabilities

INTRODUCTION AND BACKGROUND

Gotuit (Gotuit, 2009) describes meta-content to be the *currency of the internet*. Images, video and audio can now easily be obtained by the common person, then uploaded and shared on the internet.

So the above quote is an apt description, given the social value usually attributed to such video or audio on places like YouTube or Facebook. However, for such media to be found and useful beyond its social value, the value-added content much be rendered in a form that can easily be searched and indexed.

To a limited degree, such meta-data is already available in the form of format-specific tags. For example, the PNG, TIFF and JPEG/JFIF (JPEG, 1992) image formats all have proprietary tags describing characteristics of the image recording itself (such as image resolution, height, width, colour format, compression type, etc) that is essential for display.

These image files also optionally include simple tags describing the image content. For JPEG images, the most common form of these tags is the EXIF (EXIF, 2002) standard. As technology develops, we can expect a greater level of automated content description generation.

For example, nowadays cameras can GPS-label the shot location, as well as be time-stamped onto their images.

Similarly audio and video files possess tags describing the recording format, and other information important for play-back. And these also include optional content-level descriptions. For example, for MPEG-based MP3 audio files, the most common tag formats are the MP3 ID3 audio tags (MP3 ID3, 1996)

These tags consist of: essential tags, without which a player cannot process the file (eg frame height, width); optional content tags, which provide extra information to the viewer but may have little effect on the display (such as capture date, location, owner, etc); and extension or linkage tags, which relate this file to others, or other parts of itself, but which require appropriate infrastructure to support.

While these tag formats are proprietary or semi-open, they are widely used and conversion from these to a platform-independent XML schema is straightforward and easily available.

Tag use is widespread, yes. But is it accurate?

Maintaining the *accuracy* of such tags is vital, as programs increasingly come to rely on such information, despite the fact that much of this information must be supplied by human-level intelligence. Anyone with a significant digital music or image collection can attest to the importance of accurate meta-information, and also how easily it is for this meta-information to get mismatched to the file or otherwise corrupted, as well as the oft-avoided tedium of correcting these.

Automated systems have existed for many decades that attempt to mitigate this information inaccuracy. An early form of this was CDDB from *Gracenote*, introduced in 1995 (Gracenote, 1995), which attempted to exactly match a sequence of byte values at the start of every CD track against a database of such values which properly identify the track, artist and album. Since the data was digital but uncompressed, exact matches were straightforward. With the rise and frequent use of compressed, recompressed and resampled data, such exact matching is no longer practical.

Media Fingerprinting Technology

Nowadays, exact matching is no longer practical, a form of perceptually significant markers can be used as a fingerprint for the media, which is relatively independent of the sampling artefacts caused by recompression.

For example, *Civolution* (Civolution, 2008), uses media fingerprint technology that is robust to significant noise

including compression artefacts. It was derived from research done by Philips around 2005.

There are many other web vendors touting the use of such fingerprints (for example see Cano, 2005 or Cox et al, 2007), and even some open-source projects such as Foosic (Foosic, 2009) with various degrees of success.

While we expect increased use of such *data-cleansing* tools over the years to come, the first requirement is simply to identify the media. Once properly identified, an online data base could be used to fill in the missing metadata for that medium.

Indeed, it is not unreasonable to expect a significant market to come into existence to fill this need for clean meta-data. However, one significant factor against this market development is the customer-perceived close tie and suspicion between online content-identification processes and associated and implied privacy concerns, with the processes of copyright enforcement. I believe that ultimately, content ID will become important enough to overcome this fear.

Dangers of Using Open Tags

Adding content value using open tags (those free or easily readable) adds some risks where a Digital Rights Management (DRM) scheme is used to ensure user-compliance (typically against unauthorised copying or distribution).

A DRM is essentially a mechanism to automatically enforce the license arrangement under which the media had been purchased. This may include limiting the number of copies of the media (or whether copying is allowed at all), limiting the ability to share the media with others, limiting the ability to excerpt parts of the media (eg removing inserted advertisements), or being forced to watch parts in sequence. Customer resentment at some of these enforcements, has resulted in a significant market for pirates to remove or compromise the DRM.

After illegal DRM-stripping by a pirate, the value-added component may be separated from the video/audio stream, allowing the existence of low-quality versions of the media without the value-added parts to proliferate. This could hurt the commercial viability of adding such content.

Market damage due to piracy is particularly bad if the business model requires the use of online data access per view. If pirates can extract the necessary information once, they can abstract this into a form that does not require online access.

Unfortunately, much of the additional information needed to make audio or video *accessible* falls into this category of value-added metadata. For example, a text-extractable subtitle or audio-description track on a movie could easily double as a movie script.

Blu-ray is a more accessible DVD

The next-generation DVD-like vehicle for movie playing is the Blu-Ray™ disk (BD) format.

Nowadays BD come in two profiles: Disk-based, and Disk and Network-based (so-called BD-Live™). While the initial offerings were essentially a recoding of the DVD format to HD resolutions, the latest batch includes games and increased use of value-adding. Figure 1 shows a recent movie, which includes Gracenote/Sony's MovieIQ menuing system, allowing "online access to up to date information".

This extra content is normally downloaded to the device's local storage using the disk itself as a security access key. The advantage of this model is that only 'authorised' people will have access to the extended content. However this local storage cannot be backed up, so in practise it limits the amount of extra material that can be downloaded. Also, although this extra content is only downloaded once, the disk needs to be present and inserted when this extra content is accessed.



Figure 1: An examples of extra content provided via an online service exclusive to owners of this particular disk title - "Angels and Demons" (from Gracenote, 2009 Press release).

We believe that although there will be a market for this approach, we also believe that ultimately most people will prefer all content to be on the disk.

So Is It Accessible?

As with DVD, there appears to be no attempt to address the accessibility of the BD interface. However, unlike DVD, this may be less of a problem.

There is no technical reason for the lack of accessibility as there was with DVD due to its limited and proprietary interface. BD uses an optimised and assisted Java 1.3 level virtual engine and a Java Mobile (JME/PBP) structure for the menuing and playing system. Thus potentially much of the java code written for mobile devices can be utilised.

The BD menus can be created using Java code and XML. Through the use of separate sound files describing each of the options, it would be relatively straightforward to programmatically add spoken options to the menu system. Additionally, making button clicks and icon

shape/colour toggles will maximise the accessibility of the unenhanced hardware platform.

The content is normally accessed using a TV-style remote control device. By its nature, this reduces the sensitivity of the interface to the problems experienced by the motor-challenged such as those with Parkinson's disease – as long as the auto-repeat on any of the buttons can be switched off, or ignored.

THE CASE FOR WATERMARKS

The importance of metadata has already been argued. We now argue that placing that metadata indelibly within the media serves three major players:

- the content-provider that will assure license compliance;
- the customer in providing the additional content;
- the disabled in providing that additional content in a form that can be programmatically enhanced to serve their particular needs.

We propose the meta-content layer to be added to media intended for *wide-scale distribution*. Since the distributor or content-creator intends to add considerable value in the form of this meta-data, sales will need to be significant to recoup the costs of doing so.

Ordinarily, such a layer may be added using embedded text streams within the video, and typically these are in XML using a proprietary schema. MPEG, Quicktime, Ogg Vorbis/Theora and Matroska (Matroska, 2002) are all *container* file formats that can handle such text streams. Indeed, during development, we will be pursuing this line.

However, in end usage, such added value layers, could easily be removed by pirates and used separately, perhaps with a DRM-stripped version of the video. This risks separation of the two components when further distributed.

If this layer was inserted as an audio and/or video watermark within the existing streams, then the added content cannot be easily separated from the host data.

Contrary to many currently used implementations, which only seem to serve the content-provider, this watermark serves to also provide a benefit to the customer instead of only to the content-creator for the price of including / retaining the watermark. Removing the watermark will remove this additional benefit. Retaining it, allows the watermark to remain a part of the DRM that could be used to ensure compliance.

The design of our system is universal in nature, so that accessibility is naturally covered, but not specifically so.

This is an important point, since specific adaptation effort may often be seen as an additional cost, and instead, a separate *accessible* version be released subsequently.

Experience (including with the web) suggests that with a much smaller market, many such subsequent releases never actually happen.

ARCHITECTURE

The descriptive layer is built using timed text streams within a *Matroska* media container (Matroska, 2002), and an appropriate player and editing workbench have been built in-house.

In parallel, we are building the watermark technology, described in outline below. The process, shown in figure 2, combines a robust and fragile watermark. An overview of watermarking can be seen in (Cox et al, 2007)

A robust watermark is one that will resist any attempt to remove or destroy it. A successful attack on a robust watermark is one where the watermark is removed from the host video and the resulting host video is still 'good enough' that it retains some market value. Clearly this is an imprecise definition, but one should consider that in some places even a mobile-video of a movie taken in a cinema has market value.

A fragile watermark is one that is extremely sensitive to changes in the media (but may allow changes caused only by resampling or requantisation). It acts as a kind of visible checksum, and depending on use may identify places in the video that have been damaged.

Tag entry

The process begins by capturing most of the annotation usually created in the making of a film, and lost by the time the movie is sold.

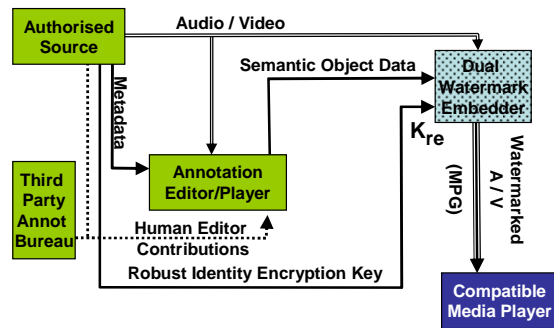


Figure 2: Overview of the encoding process

This includes such details as camera positions, continuity information (which gives us vital clues as to what semantic elements are significant enough to warrant continuity checks), script and stage directions, titling (especially that shown during the film).

In addition, any special effects visible to the viewer (such as multiple camera angles, break-out into separate story lines, or links to online games and other websites).

With the assistance of a third-party annotator who sifts through all this technical annotation, we produce the final annotation stream via a video editing console.



Figure 3: An example of a manual semantic object classification process. An editor ‘lasso’s the relevant objects in a scene (top-left, bottom-left). The lasso is *tightened* around the objects (top-right) and they are classified as the four people, and ‘dining table’. One might classify some of the items on the table, or simply text-summarise them.

At the same time, the authorised source produces the DRM technology (if any), and robust identification key to the film itself. Any customer-unique information will be embedded later by the distributor.

A player must first decode the robust watermark using a decryption key provided in order to extract from it a decryption key for the fragile watermark. This key is then used to extract the semantic content.

The logic here is that the robust watermark is required to extract the object data from the fragile watermark, but if there is some damage to the video due to resampling or requantisation, this object data is fatally corrupted.

The customer is thus encouraged to leave the watermark alone, lest they lose the object data. While the system can still be compromised by pirates who might offer duplicate disks, they cannot pass these off as originals, as the player code would need to be significantly altered accordingly.

Semantic Objects: Description and Activities

The techniques described here allow a player to appear to become ‘aware’ of the content of what it is playing. With suitable world-view ontology, it is possible to customise the BD player to relate elements in the currently playing movie with other elements previously ‘discovered’ in other movies.

This project is ongoing and will result in a better experience for the consumer interested in detailed information about a movie.

At the same time, some of the meta-information can be consolidated into libraries enabling more sophisticated lookups.

What is needed

Of primary concern is the need for a unified ontology for movies that goes beyond the kind of production details provided in websites such as IMDB (IMDB, 1990), but also includes story elements.



Figure 4: Entering activity attributes for these objects

REFERENCES

- Cano, P. Batlle, E., Gómez, E., Gomes, C.T. and Bonnet, M., Audio Fingerprinting: Concepts And Applications in Computational Intelligence for Modelling and Prediction, ISBN 978-3-540-26071-4, Pp 233-245, Springer Heidelberg, 2005
- Civolution 2008, a spinoff from Philips in 2008 that deals with fingerprinting technology. Details at <http://www.civolution.com/> accessed Sept 2009
- Cox, I, Miller, M., Bloom, J., Fridrich, J. and Kalker, T., Digital Watermarking and Steganography, 2nd edition, Morgan Kaufman, 2007, ISBN-13: 978-0123725851
- EXIF, Exchangeable Image File Format for Digital Still Cameras (JEITA CP-3451). 2002, at <http://www.exif.org/Exif2-2.PDF> access Sept 2009
- Foosic, open source software dated 2006, at <http://www.foosic.org>, accessed Sept2009
- Gotuit, “The Currency of Internet Video”, White Paper accessed Sept 2009 and available at <http://www.gotuit.com/about/WhitePapers.html>,
- Gracenote, CDDB (1996), at <http://www.gracenote.com> accessed Sept 2009.
- GraceNote 2009, Press Release “Sony Pictures Home Entertainment Partners with Gracenote To Deliver First Live, In-Movie Film Information with movieIQ” at http://www.gracenote.com/company_info/press/06/18/2009
- ID3, <http://www.id3.org> and its accessibility addendum at <http://www.id3.org/id3v2-accessibility-1.0> accessed Sept, 2009
- IMDB, started in 1990, at <http://www.imdb.com> accessed Sept 2009
- JPEG/JFIF, 1992, accessed Sept 2009 and available at <http://www.w3.org/Graphics/JPEG/jfif3.pdf>
- Matroska Media Container, and Open Standard, started in 2002, at <http://www.matroska.org/>, accessed Sept 2009