Issues with Digital Watermarking and Perceptual Hashing

Ton Kalker*, Jaap Haitsma*, Job Oostveen*
Philips Research Eindhoven

Keywords: Robust features, digital watermarking, content recognition, perceptual hashing.

ABSTRACT

Robust identification of audio, still images and video is currently almost always associated with watermarking. Although being a powerful tool, there are some relevant issues with the use of watermarking. In this paper we review these issues, and at the same time propose to reconsider the older technique of robust feature recognition as a serious alternative. Moreover, we argue that not only in the context of content recognition, but also for other applications, a benefit is to be expected from the combination of robust feature recognition and digital watermarking.

1. INTRODUCTION

Imagine the following situation. You're in your car, listening to the radio and suddenly you hear a song that catches your attention. It's the best new song you've heard for a long time. But you missed the announcement and you don't recognize the artist. Still, you would like to know more about this music. What to do? You could call the radio station, but you feel that's too cumbersome. Wouldn't it be nice if you could push a few buttons on your mobile phone, wait a few seconds, and the phone would respond with the name of the artist and the title of the music you are listening too? Even send an email to your default email address with this information?

The above scenario is one of many examples in which the ability to automatically recognize music is a key ingredient. But there are many more. Recognition of audio allows building a reliable filtering mechanism for Napster-like services. Instead of regulating downloading of copyrighted music on the basis of song titles (which has proven to be quite unreliable), music recognition can provide a more reliable filtering mechanism [33]. Moreover, with some additional ingredients, it may at the same time guarantee the quality of the content. Recognition of music also allows building a royalty-tracking infrastructure. Automated monitoring stations that tune into available broadcast channels can create log files of the content that has been aired (which station at what time) [5]. It can furthermore be used to automatically index large music archives, to build copyright aware consumer devices and in general to build intelligent music aware devices [36]. For example, a personal playback device might recognize and remember which music is played most often and use this to build a profile for intelligent random playback. It goes without saying that similar arguments not only hold for music, but also for other audiovisual data types such as still images and moving video.

Until recently digital watermarking has been considered as the most universal method to achieve the above scenarios. Digital watermarking techniques allow the embedding of the content identity into the content itself. This identification method is universal in the sense that it can be used independently of format and quality. Moreover, identification on the basis of watermarks can operate in a stand-alone environment: nothing more than some basic watermark (public or secret) keys are needed to retrieve identity. But recently, an old technique, in this paper referred to as robust or perceptual hashing, has re-emerged as a viable alternative. This has come about due to two reasons. Firstly, there has been considerable progress in boosting performance [3][22][32][36][40]. Secondly, there is the realization that the use of watermarking for content recognition actually amounts to adding redundancy. Theoretically, content is self-identifying and no additional signaling is needed. Most of us are able to identify a song like 'Rolling Stones – Angie' in only a few seconds, and sometimes even less. Not only that, we usually also know whether we are at the beginning, in the middle or at the end of the song. And we as human beings can do this even in a noisy environment. Similarly, watching only a short scene from 'Stanley Kubrick – 2001: A Space Odyssey' will result in positive identification for most of us who grew up in the sixties. And all that without having watermark detectors built into our brains or even having any of the recognized content being imprinted with a watermark. What we **do** have built into our brains is an efficient mechanism to capture (necessarily perceptual) essential

* Philips Research Eindhoven, Prof. Holstlaan 4, 5656 AA, Eindhoven, The Netherlands, tel. 31-40-2743839, fax. 31-40-2744675, email {ton.kalker,jaap.haitsma,job.oostveen}@philips.com.

Multimedia Systems and Applications IV, Andrew G. Tescher, Bhaskaran Vasudev, V. Michael Bove, Jr., Editors, Proceedings of SPIE Vol. 4518 (2001) © 2001 SPIE · 0277-786X/01/$15.00

features, to store these features into our brains (neural archiving) and to efficiently search this neural archive when unidentified content is encountered. This search and retrieval mechanism is amazingly robust and reliable. Only a small and noisy segment will very often lead to a fast, correct and positive identification (low false negative error rate). Also unknown content (i.e. content that has not been encountered before) is usually very reliably identified as such (low false positive rate).

The technique of perceptual (or robust) hashing tries to mimic these neural processes using clever signal processing and database techniques. The former is responsible for extracting essential perceptual features (from now on also referred to as *perceptual hash values* or *hash values* for short), the latter for storing and searching large amounts of pre-computed hash values. In a typical setup (e.g. the opening scene of this introduction), a local client (e.g. a mobile phone) is responsible for capturing the content and transmitting the content (possibly only the hash values if the client is equipped with a feature extractor) to a central database. The central database matches the hash values of the unidentified content with the pre-computed hash values, retrieves the best match and takes appropriate action (e.g. sending an artist name and song title in an SMS message to the requesting client).

The above shows that there are occasions where digital watermarking and perceptual hashing can both serve as the key technical ingredient. It is the purpose of this paper to argue the strong relationship between the two techniques. We do this by giving an overview of the basic qualities of both techniques, by highlighting their differences and similarities and by discussing basic parameters for a choice between or a combination of the two techniques.

2. TECHNOLOGY OVERVIEW

This section introduces the main concepts of digital watermarking and perceptual hashing.

2.1. Digital watermarking

Digital watermarking is a method that provides a communication channel for data associated with multimedia content (audio, still images, moving video) that is embedded *in the content itself*. Digital watermarking achieves this by *actively modifying* content, but in such a way that it causes only *slight and imperceptible* modifications. When properly designed, a digital watermark is able to *resist content degradation*: as long as the quality is sufficiently preserved, the watermark channel provides a reliable communication channel. For example, in the case of audio, MP3 compression at a bitrate of 128Kb/s of a CD audio track should not affect the readability of an embedded watermark. Watermarking allows content creators to embed information in original content that completely identifies the content and is readable in any format as long as quality is sufficiently high.

Digital watermarking is a relatively young branch in the field of signal processing and information theory. Interest started to rise around 1995 [1][31] and has seen since then an enormous growth in the number of scientific publications. A few years later also industrial interest was aroused [13]. A quick scan of the available literature shows that the majority of papers have images as their target audio-visual object [10][31]. The number of publications on audio [38] and video [25] watermarking is growing, but the difference in numbers is significant. What is also apparent from a literature scan is that the main target application is copyright protection or copy protection [20]. This observation explains the great interest in the security aspect of watermarking, and a large number of publications therefore focus on intentional attacks [26][41] and their countermeasures [15]. This silent assumption between security and watermarking is also reflected by the effort put into designing attack tools [29][30] and the title of a number of conferences (e.g. “Security and Watermarking of Multimedia Content”, San Jose). Also the two main industrial applications of watermarking have a focus on security [13][35]. Whether or not this strong perceived relationship between security and watermarking is beneficial to the deployment of watermarking remains to be seen. Recent events around the security of portable audio learn that caution is needed [37]. A more versatile use of watermarking for such benign applications as the embedding of meta-data (as being discussed in MPEG-21) or content recognition (as in the mobile phone application of the introduction) may in the end turn to be more “profitable”.

2.2. Perceptual hashing

Content recognition by comparing features is an art that actually has a longer history than digital watermarking. The oldest source of information is a number of patent applications, some of which go back to the early eighties [21] and others of later date [8][12][14][23]. Scientific publications are of a more recent date, but with numbers increasing. Two somewhat different approaches to content recognition can be distinguished. A first approach consists of extracting and representing

content by *semantically meaningful* features. Examples of such features are the number of beats per minute (audio), the presence or absence of red cars (image) or the average length of scenes (video). The second, more common approach uses features that have no direct semantical interpretation but are nonetheless robust with respect to content quality preserving transformations. In a number of publications, this second approach is used directly in the context of content recognition and archiving [1][3][16][22][27][34][39][40][42]. In a number of other publications this second approach is also used to support digital watermarking with respect to unauthorized embedding attacks (also known as *copy attack* [26]). Examples of this approach can be found in the work of Fridrich [17][18] [19]. In another application, robust features can help out in building a tool for robust authentication by checking for consistency between watermark and feature vector [11].

Already a number of companies, mainly in the field of audio [4][9][36] and video [2][5][6], have realized the business potential of content recognition. Other signs of a growing awareness of the potential of content recognition are a European project on audio recognition [24], the announcement of cooperation between a file-sharing company and content recognition company [33] and a recent Call for Information by the IFPI and the RIAA [24].

In this section we would like to present a new view on content recognition by drawing an analogy with cryptographic hash functions [28]. By pointing out the deficiencies of cryptographic hash functions as a tool for recognition we arrive at the notion of *perceptual hash functions*, sometimes referred to as robust hash functions. This notion hash strong similarities with the notion of a *continuous feature vector* from [42], where, as in our approach, similarity is a continuous notion that needs to be defined with respect to an appropriately chosen threshold. Our *hash* terminology, in analogy with cryptographic hash functions and opposed to the term *continuous feature vector*, stresses the fact that we are concerned with non-semantical features

A cryptographic hash function is a function that maps arbitrary length data to a small and fixed number of bits, usually in the order of 64 to 256. A proper cryptographic hash function **H** allows comparing two large data **X** and **Y**, by comparing their hash values **H(X)** and **H(Y)**. Equality (in the mathematical sense) of the former pair is then equivalent to equality of the latter pair, with only a very small probability of error[1]. Using cryptographic hash functions, an efficient method exists to check whether or not a particular data item **X** is contained in a given and large data set $\mathbf{Y}=\{\mathbf{Y_i}\}$. Instead of storing and comparing with all of the data in **Y**, it is sufficient to store the set of hash values $\{\mathbf{h_i} = \mathbf{H(Y_i)}\}$, and to compare **H(X)** with this set of hash values. This method is more efficient because (i) the storage requirements are usually more relaxed and fewer bits have to be compared. The only caveat is that an initial pre-computation is required to compute the hash values $\{\mathbf{h_i}\}$. Finally we note that good cryptographic hash functions do indeed exist [28].

Given the above arguments, one might suppose that cryptographic hash function are a good tool to identify multimedia content: take a hash function, store hash values for all available content in a large database (costly, but it needs to be done only once) and identify content by hash matching. This method will however fail when using classical cryptographic hash functions. Firstly, for multimedia content we are not interested in *mathematical equality*, but perceptual equality! For example, an original CD quality version of 'The Rolling Stones – Angie' and an MP3 version at 128Kb/s might sound the same to the human auditory system, but their waveforms might be significantly different. Because we do not have mathematical equality, cryptographic hash functions will not, in general, recognize these two versions of 'Angie' as being the same. This problem could be mitigated if cryptographic hash functions would have a continuous behavior, i.e. if perceptually similar content would at least result in mathematically similar hash values. However, cryptographic hash functions typically have rather the opposite property, in the sense that they are bit sensitive: a single bit of difference in the content will result in a completely different hash value (see also footnote 1).

From these observations we conclude that for multimedia identification we need *perceptual hashing* functions, functions that (i) map large multimedia objects to a small number of bits, and (ii) map perceptually similar objects to (mathematically) similar hash values. The observing reader might wonder why instead of (ii) we do not require that perceptual similar objects have *mathematically equal* hash values? The question is valid, but the answer is that such a modeling of perceptual similarity is not possible for reasons of transitivity. To be more precise: it is a known fact that perceptual similarity is not transitive. Perceptual similarity of a pair of objects A and B and of another pair of objects B and C does not necessarily imply the perceptual similarity of objects A and C. However, modeling perceptual similarity by *equality* of perceptual hash functions would lead to a transitive relationship.

[1] One usually also requires that a hash function is one-way: given a hash value **h**, it must be extremely difficult to construct or find a data set **X** such that **H(X)=h**. See also [28]. In this paper we shall not be concerned with this property.

As for watermarking, there is a number of characterizing parameters for a perceptual hashing technology. *Robustness* refers to the ability to positively identify two audio-visual objects O_1 and O_2 as similar, even if O_2 is a heavily degraded version of O_1. As an example and assuming feature extraction is done at the server side, the mobile phone application of the introduction would require a hashing technology which is robust to noise addition and GSM coding! *False positive and negative error rates* refer to errors in matching two objects: either mistakenly declaring O_1 equal to O_2, or mistakenly declaring O_1 unequal to O_2. *Granularity* refers the minimum spatio-temporal segment needed for identification. For example, do we need a complete song to identify, or will three seconds suffice? Are we allowed to crop an image, or do we need the original geometry? Increased granularity usually comes at the cost of *hash bit-sizes*, i.e. the number of bits needed to represent a perceptual hash. Granularity in fact requires that every relevant spatio-temporal segment be treated as a separate audio-visual object with its own perceptual hash, where the hash value of the total object is some kind of union of the hash values of all the relevant segments.

Scalability refers to the ability to efficiently match unidentified objects in large databases of previously computed hash values. Scalability has turned out to be one of the more difficult problems. Recall that perceptual hash functions have a continuous nature, and that therefore perfect matching is not the appropriate way to search a large database. Without a proper structure in a perceptual hash database, searching and retrieving will easily explode into an impractical system. As an example, consider again the mobile phone application of the introduction. To maintain a viable business, one easily needs a hash database for more than a million songs of say 180 seconds each. Moreover, in order to have a satisfactory response time (without annoyed or impatient customers), we require that in principle every 15 seconds of an audio clip is sufficient for identification. Given an allowed synchronization inaccuracy of 1 second, every identification of a fifteen seconds clip would approximately require $10^6 \times 180 = 1.8 \times 10^8$ comparisons in a naïve approach (not even taking time-scale modifications into account). For any reasonable bit-size for the hash values, this would mean impractical database sizes and access times. However, if the space of hash values has some kind of perceptual ordering structure, search complexity can considerably be reduced[2]. In our view research into these ordering structures is essential for practical and large-scale applications of perceptual hashing.

Note that for digital watermarking algorithms, robustness, error rates and granularity are also important parameters. There is obviously no equivalent for scalability in watermarking, as well as there is no equivalent to *capacity* in perceptual hashing. In the following section we make a more detailed analysis of the different aspects of both technologies.

3. DIGITAL WATERMARKING VERSUS PERCEPTUAL HASHING

The purpose of this section is to give a short overview of the main similarities of and differences between digital watermarking and perceptual hashing.

3.1. Robustness

Robustness refers to the ability to positively identify two perceptually similar audio-visual objects as similar. Although this definition seems straightforward, the actual verification of robustness is a difficult issue. This is mainly due to the fact that perceptual similarity is not a well-defined notion. A similar problem exists for digital watermarking where robustness refers to the ability of watermark messages to survive quality-preserving transformations. For both techniques some limited theoretical modeling is possible (for example by restricting the set of allowed transformations and degradations), but in the end only extensive testing can provide full proof. Recent work has shown that perceptual hashing can be made to match digital watermarking in robustness performance [22][32][36].

3.2. Impact

The primary difference between digital watermarking and perceptual hashing is the impact on content. Watermarking is by definition an *active* technique, i.e. it changes the content, however slightly. Perceptual hashing, in contrast, is a *passive* technique, i.e. it does not affect content. This primary difference is one of the reasons that watermarking finds considerable opposition in the ranks of artists, performers and even music labels and movie studios, although they do have a good

[2] For a linear structure and using binary search, complexity is logarithmic in the size of the hash database.

incentive to implement watermarking technology for copy protection purposes. In the case of movie studios, video watermarking is currently being investigated as a tool to prevent widespread illegal copying of copyrighted titles on DVD (or at least to provide a considerable threshold) [13]. In the case of music labels, the Secure Digital Music Initiative (SDMI) is investigating audio watermarking to prevent widespread illegal distribution of music over the Internet [35]. Clearly, the quality of content is such a valued asset, that the cause of copy protection does not provide a clear and unquestioned application of digital watermarking.

3.3. Connected

In most cases, watermark detection is a process that can operate in a *stand-alone* device. In general, if only some small amount of essential information is available, e.g. the (secret) keys that define the watermark embedding/detection process, the message embedded in a watermark can be read. This is one of the main reasons that in stand-alone devices watermark detection can be used for copy protection: without any knowledge of the outside world, the copy protection status of suspected content can be determined. Except for some rare occasion (where the amount of content to be recognized is small), a perceptual hash extractor needs to contact a remote intelligent and usually large database that stores previously extracted hash values. This database also performs the relatively complex searching & matching procedure needed to positively identify the unidentified content. In other words and opposed to watermarking, content identification based on perceptual hashing almost always needs to operate in a *connected* environment.

3.4. Message space

A watermark channel can carry potentially any message, as long as the capacity of the watermarking channel is large enough. This implies that a single original work can be changed into many different, but perceptually similar works, each carrying a *different* message. In case of perceptual hashing, each set of perceptually similar works is associated with a *single* message, viz. the content identity of that set of works. In other words, watermarking allows *multiple* messages, perceptual hashing a *single* message. Of course, in both cases, the 'message' can be linked to additional data structures, that may vary per location (for example, a song title reported in the local language) and in time (the position of the song title in the charts).

3.5. Error characteristics

Similar to cryptographic hash functions, there is a small, but non-zero probability that two perceptually different works have the same hash value. Also for watermarking, there is a small, but non-zero probability that two messages are confused. Well-designed watermarking and hashing technologies quantify these *message error* probabilities and guarantee that they are small enough not to invalidate the intended application. An essential difference lies in modeling and measuring of error rates. Watermarking can be viewed as a communication channel where there is no confusion whether or not the sent message is equal to the received message. No such simple methodology is possible for perceptual hashing: in its essence the perceptual equality of two audio-visual objects remains a subjective decision, that is extremely different to model and may moreover vary in place and time.

Additionally, an important characteristic of any watermarking technology is its *false alarm* probability, the probability that a watermark is detected in unmarked content. A false alarm in the context of a copy protection application has usually serious consequences, and can for example prevent an honest consumer making a perfectly legal copy of his own garage band music. It should be obvious that such an error should be very rare. An analogous error probability, also referred to as false positive probability, exists for perceptual hashing, viz. the probability that a match in a hash database is found, while in reality the hash value for the suspect work is not in the database. These probabilities need ideally to be quantified and be small enough not to invalidate the intended application. Again, as before, this error rate is difficult to model for perceptual hashing. An initial attempt at quantitative statements is presented in [22]

3.6. Legacy content

By definition digital watermarking requires modification of content. This limits watermarking as a solution for applications requiring that *any* work in a similarity set be treated as similar. In particular it *limits applications that need to work on legacy content*, content that already exists today and is evidently not watermarked. An important example of such an application is content monitoring: 'Rolling Stones – Angie' will only be recognized if it contains a watermark, legacy copies (containing no watermark) cannot be recognized. Monitoring of legacy content requires in this case the cooperation of not

only the content owners (permitting modification), but also from content distributors (and these might not have a natural incentive to cooperate). Perceptual hashing, on the other hand, does not require the cooperation of third parties, as it operates directly on the similarity class of work; *whether content is legacy or not is of no significance for perceptual hashing*.

3.7. Security

Security refers to the ability to read, write or modify a message. In case of watermarking, security has proven to be a *controversial* issue [37]. Although some claim that watermarking can be as secure as cryptography (i.e. access to a watermarking channel will necessarily degrade content beyond commercial value), this claim has not been substantiated in practice. In fact, the number of potential hacks for a watermarking system seems to be an order of magnitude larger than the number of watermarking applications. However, this does not mean that watermarking is necessarily a useless security tool. Depending on the application, the security threshold provided by a watermark may have a *large economical value*, although a minority of hackers may still be able to defeat the watermarking system.

On the other hand, the security of hashing system is based upon the correspondence between perceptual similarity and hash values. For most basic perceptual hashing techniques the similarity of hash values is sign of perceptual similarity, but it need not be the other way around. This is a weakness that can be exploited in security applications. For example, a hashing technique in a Napster-like context that is not robust to time-scale modifications will make a very insecure filtering mechanism. Ultimate security can be obtained with a perfect perceptual hashing technique, i.e. a hashing technique where perceptual similarity is completely captured by the perceptual hash function. In this sense security is equivalent to robustness as discussed in Section 3.1

4. APPLICATIONS

We list a few applications where both watermarking and perceptual hashing may serve as a key technology and where a careful consideration is needed in making a choice.

4.1. Connected audio

Connected audio is a general term for consumer applications where music is somehow connected to additional and supporting information. The mobile phone application of the introduction is a typical example. It is a business that is actually pursued [36] and is the lead example of this subsection. It goes without saying that similar applications can be defined for other types of content.

At the core of this application is the ability to recognize a large collection of music even if it heavily degraded. Technically, both digital watermarking and perceptual hashing can function as the key technology. Practically a number of issues exist. On the one hand digital watermarking would require the active cooperation of broadcasters and content owners. Without these parties, no suitably watermarked content will be available. This raises the question of incentive: somehow these parties need to be involved in the value $hain?! And even then the issue of legacy content remains as it difficult to involve every single broadcaster. Perceptual hashing on the other hand would require a large central or distributed database of hash values. Building and maintaining this database is a huge undertaking. Therefore there is a tendency to reuse such a database for a number of other applications. This practice can however easily lead to overload situations. Scalability is therefore a serious concern.

Both techniques have to face serious robustness requirements. Robust watermark detection or feature extraction in a noisy environment and after GSM coding is a considerable challenge for either technology. It is of course possible to mitigate the problem by incorporating the detector or extractor in the mobile phone (avoiding the GSM coding problem), but then again we are faced with a question of incentive. In the end, extensive testing is needed to establish best robustness performance.

For connected audio in stand-alone devices, perceptual hashing is excluded as a viable option as there are no means to connect to the central database.

4.2. Monitoring

Monitoring refers to tracking of radio, television or web broadcasts for, among others, purposes of royalty collection, program verification and people metering. This application is passive in the sense that it has no direct influence on what is being broadcast: the main purpose of the application is to observe and report.

As with the previous application, if legacy content constitutes a significant portion of the programs to be monitored, then digital watermarking alone is not sufficient. On the other hand, in many monitoring applications it is not sufficient to only identify the content, but it is also needed to trace distribution history, a type of information that can only carried by watermarks. In many cases robustness also plays a decisive role: in our practice we have come across situations where the audio to be recognized is 10dB below signal level!

4.3. Filtering

Filtering refers to active intervention in content distribution. The prime example is Napster, where songs are made available for download if and only if the song is not blacklisted (by the content owners and music industry). In a more refined scheme there will also be a need to categorizing songs into free music, different kinds of premium music (accessible to those with the proper subscription) and forbidden music (the blacklist). Depending on the type of the requesting client, access to a music file is either permitted or prohibited.

The main problem with the use of watermarking in this application is the enormous amount of legacy content. Prohibiting the use of legacy content and requiring every song to be watermarked (also known as whitelisting) will make a very unattractive file-sharing network. A proper filtering mechanism therefore needs to incorporate perceptual hashing. That is not to say that digital watermarking has no use: once a music file is identified, the network may modify it by embedding a watermark with the appropriate copyright information. The next time the music file is accessed for download *local watermark detection* is sufficient to establish download permissions.

4.4. Copy protection

Copy protection refers to active intervention in content copying and playback. The main relevant example is DVD-Video, where watermarking is introduced to prevent copying of copyrighted content and to prevent playback of illegal content. As most if not all DVD players and recorders are stand-alone devices, perceptual hashing is not an option for general copy protection[3].

5. CONCLUSIONS

In this paper we have tried to argue that perceptual hashing is a relevant technology that can either replace or support digital watermarking. Both techniques have advantages and disadvantages. Watermarking can operate in a stand-alone environment and has a versatile message set. But is has also a potential impact on perceptual quality and difficulties with legacy content. Perceptual hashing has no impact on quality, is suitable for legacy content but usually needs access to a central database. Depending on the application, a choice for the appropriate mixture of both technologies needs to be made.

ACKNOWLEDGEMENTS

This work has partly been supported by project CERTIMARK (IST -1999 -10987).

REFERENCES

1. M. Abdel-Mottaleb, G. Vaithilingam and S. Krishnamachari, “Signature-Based Image Identification”, in *Proceedings of SPIE 3845, Multimedia Systems and Applications II*, pp. 22-28, Boston, 1999.
2. ACNielsen, http://www.acnielsen.com.

[3] A short-lived exception has been Divix. See also http://www.computerhope.com/help/dvd.htm#05.

3. E. Allamanche, J. Herre, O. Helmuth, B. Fröba and M. Cremer, "AudioID: Towards Content-Based Identification Of Audio Material", in *Proceedings of the 110th Audio Engineering Societ*y, Amsterdam, The Netherlands, May 2000.
4. AudibleMagic, http://www.audiblemagic.com.
5. Arbitron, http://www.arbitron.com.
6. BDSOnline, http://www.bdsonline.com.
7. W. Bender, D. Gruhl and N. Morimoto, "Techniques for Data Hiding", in *Proceedings of the SPIE 2420, Storage and Retrieval for Image and Video Databases III*, 1995, pp. 164-173.
8. Blum et al., Patent US 5918223, "Method And Article Of Manufacture For Content-Based Analysis, Storage, Retrieval, And Segmentation Of Audio Information", July 1997.
9. Cantametrix, http://www.cantametrix.com.
10. I. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia", IEEE *Transactions on Image Processing*, vol.6, no.12, p.1673-87, Dec. 1997.
11. J. Dittmann, A. Steinmetz and R. Steinmetz, "Content-Based Digital Signature For Motion Pictures Authentication And Content-Fragile Watermarking", in *Proceedings International Conference on Multimedia Computing and System*s, vol. 2, pp. 209–213, Florence, 1999.
12. Dhillon et al., Patent Application WO0120483, "Finding Similar Types of Music. Based On Extracting Features And Dividing This Feature Space In Different Regions By Using Human Interaction And Neural Networks", September 1999.
13. DVD-CCA, http://www.dvdcca.org/video.html.
14. Ellis et al., Patent US5572246, "Method And Apparatus For Producing A Signature Characterizing An Interval Of A Video Signal While Compensating For Picture Edge Shift", November 1995.
15. F. Deguillame, G. Csurka and T. Pun, "Countermeasures for Unintentional and Intentional Video Watermarking Attacks", in *Proceedings of SPIE 3971, Security and Watermarking of Multimedia Content II*, vol. 3971, pp. 346 - 357, January 2000.
16. D. Fragoulis, G. Rousopoulos, T. Panagopoulos, C. Alexiou and C. Papaodysseus, "On the Automated Recognition of Seriously Distorted Musical Recordings", IEEE *Transactions on Signal Processing*, vol.49, no.4, p.898-908, April 2001.
17. J. Fridrich, "Robust Bit Extraction From Images", in *Proceedings IEEE ICMCS 99*, vol. 2, pp. 536–540, Florence, 1999.
18. J. Fridrich and M. Goljan, "Robust Hash Functions For Digital Watermarking", in *Proceedings ITCC* 2000, pp. 178-183, Las Vegas, March 2000.
19. J. Fridrich, "Visual Hash For Oblivious Watermarking", in *Proceedings SPIE 3971, Security and Watermarking of Multimedia Content II*, pp. 286–294, San Jose, 2000.
20. "General Requirements and Interoperability", Imprimatur Report on the Watermarking Technology for Copyright Protection, IMP/14062/A, 1998.
21. Greenberg, Patent US4547804, "Method And Apparatus For The Automatic Identification And Verification Of Commercial Broadcast Programs", March 1983.
22. J. Haitsma, T. Kalker, and J. Oostveen, "Robust Audio Hashing For Content Identification", in *Proceedings of the International Workshop on Content-Based Multimedia Indexin*g, Brescia, Italy, 2001 (accepted).
23. A. Hershtik, Patent Application WO070869, "Monitoring System", May 2000.
24. IFPI, "Call For Information On Fingerprinting Technology", http://www.ifpi.org/press/20010615.html.
25. T. Kalker, G. Depovere, J. Haitsma and M. Maes, "A Video Watermarking System for Broadcast Monitoring", in *Proceedings of the SPIE 3657, Security and Watermarking of Multimedia Content,* vol.3657, p.103-12, 1999.
26. M. Kutter, S. Voloshynovskiy and A. Herrigel, "Watermark Copy Attack", in *Proceedings of SPIE 3971, Security and Watermarking of Multimedia Content II*, vol. 3971, pp. 371-380, January 2000.
27. M. Kivanç Mihçak and R. Venkatesan, "A Tool for Robust Audio Information Hiding: A Perceptual Audio Hashing Algorithm", in *Proceedings of the Information Hiding Workshop 2001*, Pittsburgh, USA, 2001.
28. A. Menezes, S. Vanstone and P. van Oorschot, *Handbook of Applied Cryptography*, CRC Press, 1996.
29. S. Pereira et al., http://watermarking.unige.ch/Checkmark/, in *Checkmark*.
30. F. Petitcolas, http://www.cl.cam.ac.uk/fapp2/watermarking/stirmark, in *Stirmark*.
31. I. Pitas and T. Kaskalis, "Applying Signatures on Digital Images", in *IEEE Workshop on Nonlinear Signal Processing*, Thessaloniki, Greece, pp. 460-463, October 1995.
32. RAA, "Recognition and Analysis of Audio", http://raa.joanneum.ac.at.
33. Relatable, http://www.relatable.com.

34. M. Schneider, S.-F. Chang, “A Robust Content Based Digital Signature For Image Authentication”, in *Proceedings of IEEE ICIP 1996,* vol.3, pp.227-230, Lausanne, September 1996.
35. SDMI, http://www.sdmi.org.
36. Shazam, http://www.shazamentertainment.com.
37. J. Stern, “DeSDMI”, http://www.julienstern.org/sdmi.
38. M.D. Swanson, B. Zhu, A.H. Tewfik and L. Boney, “Robust Audio Watermarking using Perceptual Masking”, *Signal Processing*, Vol. 66, pp. 337-355, 1998.
39. R. Venkatesan, S.-M. Koon, M. Jakubowski and P. Moulin, “Robust Image hashing”, in *Proceedings of IEEE ICIP 2000*, vol. 3, pp. 664 – 666, Vancouver, September 2000.
40. R. Venkatesan and M. Jakubowski, “Image Hashing”, in *Proceedings of DIMACS Conference on Intellectual Property Protectio*n, Piscataway, NJ, USA, 2000.
41. S. Voloshynovskiy, S. Pereira, V. Iquise and T. Pun, “Attack modeling: Towards a Second Generation Benchmark”, *Signal Processing, Special Issue: Information Theoretic Issues in Digital Watermarking,* vol. 81, no. 6, pp. 1177-1214, June 2001.
42. H. Wang, F. Guo, D. Feng and J. Jin, “A Signature for Content-Based Image Retrieval using a Geometrical Transform”, in *Proceedings of the Multimedia and Security Workshop at ACM Multimedia*, pp. 229 – 234, Bristol, UK, 1998.