# Content-Based Multimedia Identification: A New Approach for Digital Rights Management

Tiejun Huang, Yonghong Tian, Wen Gao
(Peking University, Beijing, China)

**Abstract**: *Content-based Multimedia Identification* (CBMI) is a *passive* approach for Digital Rights Management that identifies a media content item by the automatically-extracted descriptors herein named as *mediaprints*.

--------------------------✦ -------------------------

The Internet is revolutionizing multimedia content distribution, shifting the way content producers and users approach digital rights, especially along with the rapid increase in popularity of online content-sharing sites such as Flickr and YouTube. These sites offer immense opportunities for users to upload and share digital images, audio and video content. However, the ability for anyone to make perfect copies and the ease by which those copies can be freely distributed also facilitate misuse, illegal copying and distribution ("piracy"), plagiarism, and misappropriation [1]. In early 2006, a short video titled "*The Bloody Case that Started from a Steam Bun*" became popular very quickly on the Internet in China. This video was re-made by a blogger from a hit movie "*The Promise*," thereby raising a wide-range feud about the online video copyright protection in China. Another well-known example is Viacom's $1 billion lawsuit against YouTube in March 2007 for "massive intentional copyright infringement." Since the year of 2000, the explosive growth of the unlicensed distribution and sharing of digital content on the Internet has raised a large amount of copyright issues, consequently producing a serious impact on the development of media industry. Therefore, how to manage the copyright of multimedia content on the Internet has become an important issue of global concern.

To address the illegal copying and distribution of multimedia content, digital rights management (DRM) has been widely studied. DRM is generally taken to refer to the technologies or systems that protect and enforce the rights associated with the use of digital content [1]. Encryption and watermarking are the two major DRM approaches in the past two decades [2], by either *proactively* encrypting multimedia content or digital watermarking/fingerprinting for *posterior* authentication. However, the technical challenges for securing media content on the Internet are formidable. The contents released by millions of Internet users cannot be canned again into "bottles" being locked by encryption or affixed with watermarks. In this article, we describe a new approach, namely *Content-based Multimedia Identification* (CBMI), to offer a *passive* but reproducible and reliable *DRM* measure.

## Current DRM Approaches: Encryption and Watermarking

Firstly, this section gives a brief review of two current DRM approaches — encryption and watermarking, and then explains why they cannot successfully solve the problem of managing the copyrights of multimedia content on the Internet.

As one of the most fundamental technologies of information security, encryption is the process for controlling access to confidential data, known as *plaintext*, by scrambling the data into an unintelligible form (i.e., *ciphertext*) with knowledge of an encryption key [1]. The inverse process, known as decryption, is very easy to perform with knowledge of the decryption key and very difficult to perform without it. The security of an effective encryption technique lies in the secrecy of the decryption keys. Compared with normal data, encrypting multimedia content needs to take some special application requirements into account. For example, some errors in multimedia bit stream may not crash the usage of the content. However, even the multimedia content encryption approach is enough perfect, it still faces several more fundamental problems when applied to content on the Internet:

- ***Analog hole***: The encrypted multimedia content, once decrypted and played back, can be illegally copied or recorded by digital/analogue devices, consequently getting rid of the control of the encryption systems. The so-called *analog hole* is one of the main reasons why piracy is rampant on the Internet.
- ***Protection cost and complexity***: Encryption-based DRM systems can be implemented in a constrained environment, but is almost impossible in the scale of the Internet since it relies on the full deploy-

ment of a global information security infrastructure which needs huge cost and has high complexity. For example, all user devices must be equipped with an *interoperable* decryption module, say, which is able to decrypt multimedia content packaged by different encryption techniques. Moreover, once the encryption system is cracked, to fix the damage or upgrade the system needs more costs.

- ***Fair use and public availability after copyright expiration***: From the legal point of view, using encryption techniques to protect multimedia content would cause unnecessary troubles for the fair use. Moreover, encryption also hampers the public availability of multimedia content after copyright expiration
- ***Conflict with the intrinsic value of media content:*** Different with the security of general confidential information, the intrinsic value of multimedia content in general depends on their visibility and accessibility to the public. In most cases, encryption runs in the opposite direction.

Another DRM approach that has been widely studied is digital watermarking. A watermark is a signal embedded within the multimedia content. In addition to being perceptually invisible or inaudible to humans, watermarks should also be statistically undetectable and resistant to any malicious attempts to remove them. The embedded watermarks may be detected by a watermark detector. There are two types of digital watermarks, i.e., robust watermarks and (semi-)fragile watermarks. Robust watermarks are able to resist a designated class of transformations in copyright applications (to carry ownership or forensic information) and copy protection applications (to carry copy and access control information). Fragile and semi-fragile watermarks are commonly used to provide authenticity or signal/content integrity verification. Among them, fragile watermarks fail to be detected after the slightest modification; while semi-fragile watermarks resist benign transformations but fails detection after malignant transformations. As a derivative from digital watermarking technology, digital fingerprinting is to embed a distinct set of marks into a given host signal to produce a set of fingerprinted signals that "appear" identical for use, but have a slightly different bit representation from one another.

However, there still exist several problems for either robust or (semi-)fragile watermarks:

- ***Inevitable degradation of quality***: It is certain that the quality of multimedia content will be degraded more or less when watermarks are embedded. In general it is easy to create robust watermarks *or* imperceptible watermarks, but the creation of robust *and* imperceptible watermarks has proven to be quite challenging [3].
- ***Cannot cover the multimedia content having been spread out:*** As a DRM approach featured by *proactively* embedding and *posteriorly* detecting mechanism, digital watermarking approach can be used to newly-released multimedia content, but not to those that have been spread out on the Internet.
- ***Failure to solely resolve the ownership authentication issue:*** In practice, everyone can embed his/her watermarks into multimedia content. Thus to authenticate the content ownership, there must be an independent registration authority who neutrally executes the functions of registration and authentication. However, with the presence of such authority, the ownership of the registered content can be directly authenticated even without embedded watermarks.

Over the past decade, there are many practices that attempt to exploit encryption-based or watermarking-based techniques to reduce the piracy problem on the Internet. As a direct response to the widespread success of MP3, Secure Digital Music Initiative (SDMI) was formed with the purpose of developing specifications that protected the playing, storing and distributing of digital music. However, SDMI audio watermarking scheme was cracked at its open challenge phase. Apple's FairPlay might be one of the most successful commercial encryption-based DRM systems. However, Steve Jobs, CEO of Apple, claimed "(Encryption-based) DRMs haven't worked, and may never work, to halt music piracy" in Feb 2007 [7].

## Concepts Related to Mediaprint

*Perceptual hash* (or *robust hash*, *soft hash*), as the extensions to the conventional *hash*, is a unique binary string or code for multimedia authentication. Different from the conventional hash function which generates different hash values for different inputs, the perceptual hash value is expected to change only when the input is perceptually different. Similar to cryptographic hash function, a good perceptual hash function should be robust, unique and unpredictability. Unpredictability means it is very hard to find (forge) perceptually different inputs having similar hash values. For authentication, unpredictability is the security foundation. But for CBMI, it is not essential in most cases and very difficult to design such hash function. Some existing works on perceptual hashing were really concerned about unpredictability and some others didn't care for at all. In fact, the perceptual hash may be generated by hashing the unique descriptor of CBMI with conventional hashing functions.

*Fingerprinting*, which was first introduced in 1983 [1], is the process of embedding a distinct set of marks into a given host signal to produce a set of fingerprinted signals which "appear" identical for use, but have a slightly different bit representation from one another. The fingerprinting technique used in CBMI in last several years is totally different with the original meaning of fingerprinting.

Another similar concept is *audio signature* and *visual signature* from MPEG. Since one meaning of signature is "a distinctive mark, characteristic, or sound indicating identity", audio and video signature seems to be a good choice for CBMI. The weakness of this choice is possible confusion with well-known digital signature or electronic signature which is attached by data being authenticated. The audio signature is used in MPEG-7 [2] as a descriptor to identify an audio item in 2001. At July 2007, MPEG started to define visual signature as a MPEG-7 visual descriptor to uniquely identify individual image and video items [3].

To avoid the possible confusion, this article proposes to use mediaprint (similarly, audioprint, videoprint and so on) to denote a robust and unique descriptor for multimedia content.

***References***

[1]N. R. Wagner, Fingerprinting. In: *Proceedings of IEEE Symp. Security and Privacy*, 1983, 18-22.

[2]ISO/IEC 15938-4: *Information Technology Multimedia Content Description Interface-Part 4: Audio*, 2002.

[3] ISO/IEC MPEG W9216. Call for proposals on image & video signature Tools. July 2007, Lausanne, CH.

## Content-Based Multimedia Identification

The difficulties that both encryption-based and watermarking-based approaches faced bring forward the emergence of a new approach, which attempts to *passively* protect copyrights by identifying multimedia content items and monitoring whether they are illegally distributed and shared on the Internet. In general, there are three ways to identify media items. The first way is to use *manually-assigned* identifier which is independent on the data representation (e.g., bit stream) and content of an item, such as ISRC numbers for music works or UUID for any item. These identifiers like an ID card number to a citizen. The second way is to use *data-relevant* methods to identify media items. For example, a hash function can be used on the bit representation of a media item to generate a digest value as its identifier. In this way, only the association at the bit level exists between a media item and its identifier. However, one obvious disadvantage of the two ways is that they are irrelevant to media content, making them still unable to effectively find the modified copies. As an analogy, the ID card number alone is not sufficient to identify and authenticate a terrorist, and the face, fingerprint or even DNA should also be used. The third way is thus to used *content-relevant* methods to identify media items. In this article, we refer to it as *Content-based Multimedia Identification* (CBMI for short).

The key idea of CBMI is to automatically generate a new descriptor from the content as the identifier of a media item. The new descriptor should be robust (unchanging) across a wide range of transformations (e.g., editing operations), but would be sufficiently different for every "original" content item to identify it uniquely and reliably. However, there is a wide range of disagreement about what the descriptor should be referred to. Some existing examples of its designation include *fingerprint*, *perceptual hash*, *audio and visual signatures* from MPEG or *media DNA* (e.g., *video DNA*) in industry. These designations either are not able to reflect the intrinsic properties of the new descriptor, or may produce some confusion with other technologies that are basically different with CBMI (The "Concepts Related to Mediaprint" sidebar gives a short discussion for these designations). In this article, we refer to the descriptor as *mediaprint* by following the word-formation of fingerprint and voiceprint. It is expected that different types of mediaprints may be needed for different types of media, which are referred to *imageprint*, *audioprint*, *videoprint* respectively for image, audio and video, or *visualprint* and *auralprint* in more general context. This concept can also be naturally extended to *docprints* for documents and *softwareprints* for software source codes. Correspondingly, *mediaprinting* is used to denote the process of extracting mediaprints from the media content for CBMI.

Basically, mediaprint should have at least the following two intrinsic properties:

- **Robustness:** Mediaprint shall be identical for an "original" content item and its copies modified by using a wide range of modifications (e.g., editing operations) or transformations (e.g., transcoding, analog VCR recapturing to digital, movie camcording);
- **Uniqueness**: Mediaprints extracted from different "original" media items which are not modified copies of one another shall be significantly different. In other words, mediaprint shall have the strong ability to identify a media item.

Moreover, mediaprints shall be based on intrinsic measurements from the media content, rather than extrinsically affixed labels like the watermarks. In addition, MPEG proposed some other common requirements for its visual signature [4], including fast matching, fast extraction, compactness, non-alteration, self-containment, coding agnosticism. For example, the non-alteration property reveals the fact that visual signature should be extracted and measured without having to alter the content. These requirements can be used to further describe and constrain the extraction, expression and matching of mediaprints.

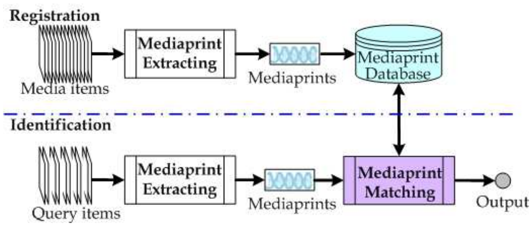


**Figure 1. Processes of Mediaprint-based CBMI.**

Mediaprints provide an effective means to identify media items based on the content. As shown in Figure 1, mediaprint-based CBMI typically includes two processes. In the registration process, mediaprints of the copyrighted media items are extracted and stored in the mediaprint database. Then for a query item, its mediaprint is also extracted and then compared with all mediaprints in the database to verify whether it matches a registered item. We can see that no extrinsic labels are needed in the identification process.

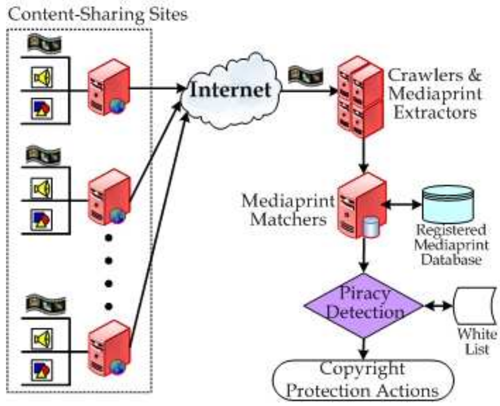


**Figure 2. Mediaprint-based CBMI for piracy detection on the Internet.**

By mediaprints, CBMI provides a passive approach to protect the copyrights of multimedia content on the Internet, without proactively altering multimedia content by encryption or embedding extrinsic labels such as watermarks or digital fingerprints. The paradigm shown in Figure 2 illustrates how mediaprint-based CBMI is used for piracy detection on the Internet. Firstly, media content in copyright are registered with their mediaprints and copyright information (e.g., expiration). Secondly, the system uses crawlers to discover and download the monitored media items from the content-sharing sites or P2P systems. Then mediaprints are extracted from these media items and compared with all mediaprints in the registration database. Finally, piracy judgement is carried out to determine whether copyright protection actions should be taken. It should be noted that the system can be used by the intellectual property authorities to discover and reduce piracy, plagiarism and misappropriation, or by content-sharing operators to prevent their users from uploading media items that might potentially cause copyright infringements. It also can help the end users to reduce misuse by online verifying the mediaprints of media content to be played.

For the DRM purpose, mediaprint-based CBMI is significantly different with encryption, watermarking and digital fingerprinting. Figure 3 compares encryption-based and mediaprint-based DRM approaches from the costs for legal and illegal consumption, in which legal and illegal usages of media content are respectively denoted by white and black circles. Since encryption-based approach employs encryption and authorization techniques to actively protect the media content, legal users must pay the additional cost for copyright protection while illegal users do not pay any cost to consume the cracked content. Clearly, this is not a rational and normal case. On the contrary, by using mediaprint-based approach, legal users can freely consume media content without any additional cost for copyright protection; while illegal users, if they want to distribute copyrighted media content on the Internet, must pay the additional cost to elude the copyright monitoring of mediaprint-based DRM systems by using the encryption or alike techniques to transform the media content into scrambled or quality-loss versions. Moreover, this approach can effectively cope with the problems of the analog hole, fair use and public availability after copyright expiration. It can also be easily implemented in the scale of the Internet by deploying servers at content-sharing sites or Internet gateways in a scalable way, say, according to the values of the content or the distribution of piracy activities.

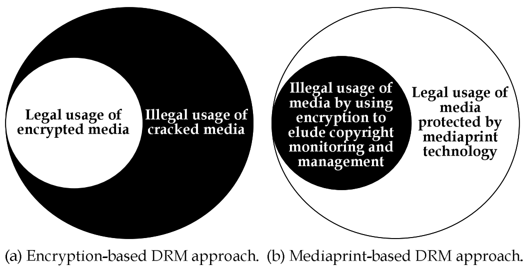


**Figure 3. Comparison of encryption-based and mediaprint-based DRM approaches from the costs for legal and illegal consumption.**

We can also compare watermark-based and mediaprint-based DRM approaches from the perspective of content quality loss in the cases of legal and illegal usage, as shown in Figure 4. In watermark/digital fingerprint based approach, legal users consume quality-loss content embedded with watermark or digital fingerprints, but illegal users can consume lossless content if they crack the watermarking or digital fingerprinting algorithms, or misappropriate the original copy. The situation is different in mediaprint-based approach, where legal users can freely consume lossless content while illegal users can only consume lossy content since they must transform the media content in order to elude copyright monitoring. In addition, this approach is able to protect copyrights of multimedia content that have been spread out on the Internet, and can also prevent the copyright disputation even pirates embed their watermarks into multimedia content.

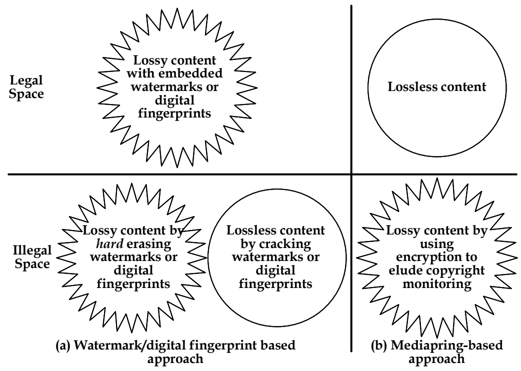


**Figure 4. Comparison of watermark/digital fingerprint based and mediaprint-based DRM approaches from the content quality loss.**

However, the application coverage of the mediaprint-based DRM approach is limited in the public space (e.g., on the Internet). That is, this approach cannot be used to protect the copyrights of media content that are consumed in the users' private spaces (e.g., in private PCs or MP3 players), unless these private spaces are permitted to be accessed by the third-part mediaprint-based DRM systems or the end users actively install a mediaprint verification client software.

## Key Issues of Mediaprinting

To make mediaprin-based CBMI applicable for DRM on the Internet, there are at least three key issues should be addressed:

- **Content distortion modeling:** Although dozens of modifications and transformations are listed out by some existing works (e.g., [4], [8]), there is no a general content distortion model. In general, content distortions that should be addressed by CBMI must catch up with the quality change limitation acceptable by human. As one of most important properties of human auditory/vision system, sparseness plays a crucial role in aural-visual perception and cognition. For example, human being can easily tell whether the video has been ever watched or not, even it was played by different devices (e.g., cinematograph or TV) or in different environmental conditions (e.g., illumination). In these cases, the sparse representation of the visual information may be the same, or at least similar, which as a consequence can be used as the ideal "mediaprint."

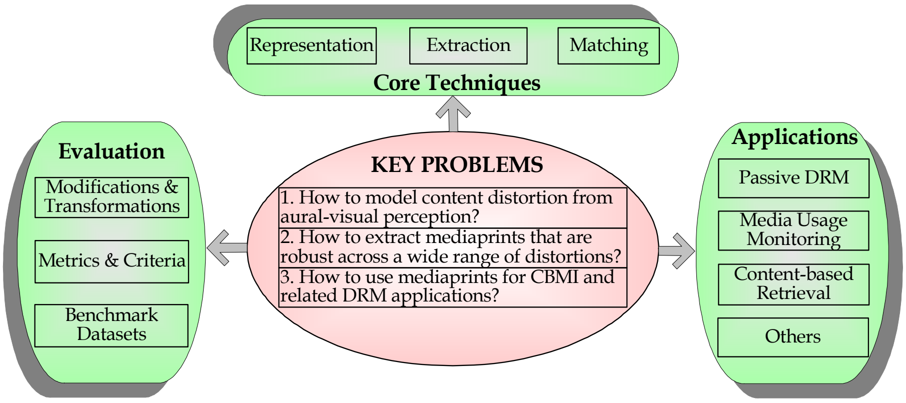


**Figure 5. The view of the various facets of mediaprinting technology.**

Naturally, understanding the similarities and differences between an "original" content item and its "modified" copies from the perspective of human aural-visual perception plays a fundamental role for mediaprinting. This can further boil down to the content distortion models, which aims to understand and characterize the variant and invariant components of the media content when they are submitted to a wide range of modifications and transformations. Ultimately, this model will provide the theoretical basis for mediaprinting.

- **Mediaprint extraction:** For various CBMI applications, how to extract unique mediaprint from the content that are robust across a wide range of distortions is the core task. Intuitively, mediaprint extraction is similar to feature extraction in content-based multimedia retrieval in that they all aim at describing media content according to discrete dimensions, such as color distribution, texture and the shape and motion of objects. However, what makes mediaprint extraction more difficult is mediaprints should describe images/audios/videos as unique entities.
- **Mediaprint usage:** The third issue is how to use mediaprints for CBMI and related DRM applications. This involves at least two aspects: how to fast search and compare mediaprints to allow large volumes of media content to be matched rapidly; how to evaluate the performance of mediaprinting (e.g., verification capabilities and speed). Moreover, the development of industry standards should also be taken into account for deploying mediaprint-based interoperable DRM systems.

## Mediaprinting: Core Techniques and Evaluation

To address above issues, a possible organization of the various facets of mediaprinting technology is shown in Figure 5. Roughly speaking, there are three aspects related to mediaprinting: core techniques, evaluation and applications. In this section, the first two aspects are addressed.

***Core Techniques: Representation, Extraction and Matching***

Expressive representation models, robust extraction and efficient matching algorithms are three core techniques in mediaprinting. From the design perspective, the representation and extraction of mediaprints cannot be totally separated. The representation model of mediaprints determines to a large extent the realm for extraction techniques. Figure 6 shows a possible view of mediaprint representations at different levels and their corresponding extraction procedures. At the bit level, media content can be treated as general data, and hereby data hashing functions can be used to generate a fixed-length string as its mediaprint. Similar to content-based multimedia retrieval which finds similar content by audio-visual features at the signal level, a more sophisticated approach is to extract invariant features in spatial/temporal/frequency domains and then convert them into mediaprints. Then the task transfers to how to find such invariant features. To further explore

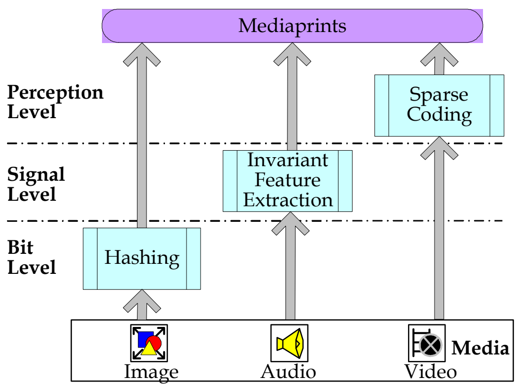


**Figure 6. The mediaprint representations at different levels.**

the *intrinsic* differences between an "original" content item and its "modified" copies, content distortion models at the perception level should be investigated. Thus the ideal approach is to simulate human visual/aural system which generates a compact expression (i.e., mental image) for a media item by physiological sparse coding. One example is that you can easily recall a song only based on a weak melody flying from the street corner. Another example is that even on a black-white TV, you can almost immediately recognize a movie being watched many years ago in a cinema. These cases show that human brain stores multimedia content in the long-term memory by a very compact or sparse way. However, how to generate such high-level mediaprints remains an open problem.

Roughly speaking, the mediaprint extraction approaches can be further classified into two categories: *feature-based* and *process-based*. Feature-based approach generates a mediaprint by extracting physically meaningful features that can reflect the uniqueness of the media content from some certain facets. Typically, the features that are selected for generating mediaprints should be robust to content distortion models — in contrast to features in pattern recognition applications, for example, which must be robust to categorization models. Note that these features can be directly extracted from the content, but also can be obtained after transforms such as dimension reduction. In general, human auditory/vision system only extracts some transformation-invariant salient features in identification tasks. Thus we need to investigate which salient features can be used to generate mediaprints.

Process-based approach generates mediaprints from media content directly by a linear or non-linear mapping function. One typical mapping function is artificial neural networks. Neural networks identify the watched/listened content by functionally simulating human auditory/vision process (e.g., sparse coding). Compared with feature-based approach, process-based approach is in essence a non-analytic procedure in which mediaprints do not rely on the explicit representation such as the feature set, but are hidden in mapping functions. In practice, a combination of the two approaches may obtain the better CBMI performance. An example is to map a feature-based large-size mediaprint to a compact one by process-based approach.

Efficient matching is the third core technique in mediaprinting. In general, there are two different query scenarios when mediaprints are used in CBMI and related DRM applications [4]: direct matching and partial matching. Their difference lies in whether *the whole mediaprint* or its *certain segment* of the query item matches with a certain segment of one or more mediaprints in the database. As a consequent, there are different requirements for the matching algorithms. For example, the direct matching algorithm is required to output the start point and the end point of the matched segment in a registered content item; while the partial matching algorithm is required to output the start position and duration in both the registered and the query items. Moreover, to enable fast mediaprint matching in a very large database, efficient indexing models should also be elaborately designed. For example, some researchers (e.g., [6]) proposed an invert indexing model of mediaprints, which quantizes a mediaprint as a coded string and then creates an invert list of all possible segments of codes for efficiently indexing.

In recent years, mediaprinting and related technologies have attracted a wide range of research interest. Please refer to the "Related Work on Mediaprinting" sidebar for more details.

### *Evaluation*

With the numerous mediaprinting techniques and systems proposed and in operation, evaluation becomes a critical issue, allowing us to choose from many different proposed ideas and to test new approaches against older ones. For any media print-based CBMI system, a strategy for evaluation involves determining the following aspects: various types of modifications and transformations that simulate various content distortions, appropriate metrics and criteria for evaluating competing approaches, and benchmark datasets.

In general, the content modifications and transformations can be divided into three categories:

- Coding format changes, such as coding in different standards, change from trancoding such as coding formats, bitrates, and frame sizes;

## Related Work on Mediaprinting

Most of current efforts on mediaprinting have focused on technologies that deal with how to extract invariant features to generate mediaprints. Some representatives of such developments for different types of media content are reviewed as follows.

### *Imageprinting*

According to the way in which features are extracted from images, it is possible to distinguish three extraction approaches for imageprints. In the first approach (e.g., [1]), features extracted from the whole image are used to generate imageprints. Even though this approach performed well in many cases, it can not keep robust to local modifications such as cropping, embedding, combining. Thus an alternative approach is to extract local features to generate imageprints for the image. One case is to partition an image into several blocks (or regions) and then to extract features of these blocks [2, 3]. Recently, a keypoint-based approach is attracting more and more research interest. For example, Monga *et al.* [4] proposed an image perceptual hashing algorithm using visually significant feature points.

### *Audioprinting*

There are many proposals for audioprinting, differentiated by the features used for generating audioprints and the matching algorithms. Some audio features originally designed to content-based audio retrieval are used to generate audioprints, such as mean energy [5], normalized spectral sub-band moments [6], audio spectrum flatness (ASF) for MPEG audio signature. One example addressing high-dimensional audioprint matching is the approximate nearest neighbor search algorithm for binary audioprint vectors proposed by Miller *et al.* [7].

### *Videoprinting*

A video clip can be treated as a sequence of images or a 3D data stream. Thus videoprinting approaches can be divided into two categories. The first approach employs 3D data transforms (e.g., spatio-temporal DCT [8]) to extract a global descriptor of a video clip. However, this approach is difficult to be applied to partial content matching, namely, to identify whether only one segment of a query clip matches with a certain segment of a registered one. Instead, by summarizing one video clip with a set of sampled frames or keyframes, another approach (e.g., [9], [10]) is to employ imageprinting methods on these frames, and then assemble the corresponding frameprints together to form the videoprint. To further improve the performance, the temporal and spatial information such as the difference or correlation of neighboring frames (e.g., [11], [12]) can be utilized to generate videoprints that would be more robust and have stronger ability to identify video clips.

### *References*


[1] S. S. Jin, H. Jaap, K. Ton, and D. Y. Chang, A robust image fingerprinting system using the Radon transform. *Signal Processing: Image Communication*, 19: 325-339, 2004.

[2] W.-G. Oh, A. Cho, I.-H. Cho, W.-K. Yang, J.-K. Jin, J.-W. Lee, and D.-S. Jeong, Concentric circle partition-based image signature. *MPEG Doc. No.M14956*, Oct. 2007.

[3] C.-Y. Lin and S.-F. Chang, A robust image authentication method distinguishing JPEG compression from malicious manipulation. *IEEE Trans. Circuits and Systems for Video Technology*, 11(2): 153-168, Feb 2001.

[4] V. Monga. and B. L. Evans, Perceptual image hashing via feature points: performance evaluation and tradeoffs. *IEEE Trans. Image Processing*, 15(11): 3452- 3465, Nov 2006.

[5] V. Venkatachalam, L. Cazzanti, N. Dhillon and M. Wells. Automatic identification of sound recordings. *IEEE Signal Processing Magazine*, 21(2): 92-99, 2004.

[6] J. S. Seo, M. Jin and S. Lee. Audio fingerprinting based on normalized spectral subband moments. *IEEE Signal Processing Letters*, 13(4): 209-212, Apr 2006.

[7] M. L. Miller, M. A. Rodriguez and I. J. Cox, Audio fingerprinting: Nearest neighbor search in high dimensional binary spaces. *Proc. IEEE Workshop on Multimedia Signal Processing*, 182-185, 2002.

[8] B. Coskun, B. Sankur, N. Memon, Spatio-temporal transform based video hashing. *IEEE Trans. on Multimedia*, 8(6), 1190-1208, Dec. 2006.

[9] S.-C. S. Cheung and A. Zakhor. Efficient video similarity measurement with video signature. *IEEE Trans. Circuits and Systems for Video Technology*, 13(1): 59-74, Jan 2003.

[10] S. Lee and C. D. Yoo, Robust video fingerprinting for content-based video identification. *IEEE Trans. On Circuits and System for Video Technology*, 18(7), Jul 2008.

[11] J. Oostveen, T. Kalker, and J. Haitsma. Feature extraction and a database strategy for video fingerprint. *Visual Information System Architectures*, 117-128, 2002.

[12] X. Zhou, M. Schmucker, and C. L. Brown. Video perceptual hashing using interframe similarity. *GI Sicherheit 2006*.


- Editing operations, such as deletion or insertion of frames, insertion of text or logo, image processing (e.g., brightness change, rotation, and scaling, flip, crop, blur, skew, perspective, aspect ratio change);
- Quality change, such as addition of noise, analog VCR recording & recapturing, movie camcording etc.

Recently, MPEG launched a series of robustness tests to various types of modifications, including 11 basic and 6 complex modifications for images [4], and 9 modifications for videos [5]. These modifications are also subjected to different levels (e.g., light, medium, and heavy). In the evaluation experiments, they are used to generate the database of modified media items from the original content.

To test the robustness of different media-printing techniques, detection capabilities are often evaluated in the presence of various modifications. The evaluation criteria are relatively simple for mediaprinting algorithms of non-temporal media such as images. In this case, the comparison of the mediaprints of the original and modified images is used to determine whether the modified ones match their corresponding originals, and the number of detected matches is counted to measure the detection capabilities of the algorithms. The evaluation criteria are more complex for media-printing algorithms of temporal media such as audio and video, mostly due to the complexity how the match is considered as a success. The evaluation criteria of successful match are given respectively for direct matching and partial matching in [5]: For direct matching, more than 50% overlap between the estimated position and the ground-truth position is required. While for partial matching, difference between the durations of the ground truth segment and the (estimated) matched segment must be shorter than 2 seconds. More importantly, the overlap between the ground-truth segments respectively in the original and query media items must be at least half of the duration of the ground-truth segment. Given such criteria, two metrics can be used to evaluate the detection capabilities: *false alarm rate* (FAR) and *miss alarm rate* (MAR). False alarm denotes that two media items or segments with distinct content are misjudged into a match by the mediaprinting algorithm. Thus FAR is used to measure the proportion of mismatches to the total number of query items. The success ratio (SR) proposed in [5] can be viewed as a variation of FAR where SR = 1 - FAR. Similarly, miss alarm denotes that two media items or segments with the same content are misjudged into a non-match by the mediaprinting algorithm. Thus MAR is used to measure the probability that miss alarms take place. Clearly, MAR = 1 – recall, where *recall* is a statistical measure that reflects the fraction of correct matches in all homologous pairs. TRECVID also proposed an overall measure, minimal normalized detection cost rate (DCR), to evaluate the detection effectiveness for content-based copy detection, a new task firstly appeared in 2008 [8]. In addition, the *match number per second* (*MPS*) is used in MPEG to measure the matching efficiency [5].

With respect to benchmark datasets, three well-known evaluations related to media-prints collected testing datasets and offered to participants. For the anti-piracy movie fingerprinting test organized by MPAA (Motion Picture Association of America), MovieLabs prepared amount of movie clips and about fifty transformations to test a dozen candidate systems in 2007. In content-based copy detection task at TRECVID 2008, the reference dataset consists of approximately 200 hours of AV data in 438 reference files [8]. For visual signature competition, MPEG released a dataset of 135,609 images selected from the CD-ROMs "Art Explosion 800000" in Jul 2007 [4], and released a dataset containing 1,900 video clips of >3 minutes for test in Oct. 2008 [5]. These images and videos are then transformed into modified items, creating a large database for evaluating competing visual signature algorithms. Larger-scale datasets publicly available are expected for mediaprinting research and development, especially for audioprinting and videoprinting.

## The Potential Applications

This section discusses the potential applications of mediaprint-based CBMI. Although initially inspired by DRM requirements, mediaprinting technology can extend its reach to non-DRM applications such as media usage monitoring, and content-based retrieval.

***Passive DRM***

Mediaprint-based CBMI offers a passive but reproducible and reliable DRM measure. This *passive DRM* approach can be widely used in different application scenarios such as piracy detection, royalty collection and brand management. In Figure 2, we have shown a mediaprint-based CBMI system for piracy detection on the Internet. The system can be easily deployed in the scale of the Internet either by the intellectual property authorities or content-sharing providers. The mediaprint-based CBMI approach can also be easily used to royalty collection. Recently, KTV operators in China are required to pay royalty fees by China Audio-Video Copyright Association (CAVCA) on behalf of singers and composers. In this case, mediaprint-based CBMI system can help the proprietors of content in copyright know how and where their works are being consumed. Similarly, imageprinting technology can be used to effectively detect the plagiarism and misappropriation of registered brands. In the converse situation, an author who has acquired a content item or a consumer who has downloaded a media item from P2P systems can employ this technology to check the provenance and copyright information.

***Media usage monitoring***

Mediaprint technology provides an effective way to media usage monitoring, namely, monitoring how the media content is being used. For streaming media such as video and audio, we can even exploit mediaprinting technology to track not only which content that is being played, but also which part of the content is being played by using partial mediaprint matching strategy. The ability to track media usage automatically and precisely provides content owners with valuable data regarding media content usage. For example, an advertising agency may confirm that advertisements have been broadcasted correctly. The mediaprints of advertisements is useful for automatically linking the users to the corresponding sale sites. Mediaprint technology can also be used to prevent illegal insertion to broadcasting programs. Here the mediaprints of broadcasting programs can be extracted in real-time and then used to check whether they remain unchanged when being played.

***Content-based retrieval***

Mediaprint technology provides a feasible solution to identify near-duplicates or similar content, consequently beneficial to many multimedia retrieval applications, such as finding the content item with the slightly different appearances on different sites, clustering similar media content, presenting uncluttered search results, and providing additional clues to the retrieval of low-quality media content. For example, with the rapidly increased popularity of online video-sharing sites such as MySpace and YouTube, a huge amount of small-size, low-quality user-generated videos (exactly for this reason they are often called as *microvideos*) are distributed online. These microvideos are often much harder to analyze semantic meanings than professionally produced videos. A possible approach is to link these microvideos with high-quality videos of the same content, and then use the semantic clues extracted from these high-quality videos for microvideo annotation and retrieval. It can be implemented by matching their videoprints which keep robust to quality changes.

***Others***

A lot of products and applications related to CBMI are emerging in the last several years. Some companies focusing on mediaprint-based CBMI have rushed in the market — Gracenote, Audible Magic and Shazam for audio, Advestigo, iPharro, MotionDSP, Vobile and Zeitera for video.

As a new topic which is not enough attended previously, mediaprint-based CBMI influences various multimedia systems. A simple example is to simplify the personal photo-collection management. A more ambitious possibility is to reconstruct the Web space by mediaprints rather than manually-assigned URIs.

## Conclusion

This article introduces the new approach for multimedia security and copyright management, i.e., content-based multimedia identification (CBMI). The key idea is to automatically generate new descriptors, called mediaprints, to identify the media content. Compared with the active protection provide by encryption and authentication by watermarking, CBMI offers a passive but reliable DRM measure. It is expected that mediaprinting technology will be explored in a wider range of applications.

## References


[1] E. T. Lin, A. M. Eskicioglu, R. L. Lagendijk, and E. J. Delp, Advances in digital video content protection. *Proc. of the IEEE*, 93(1): 171-183, 2005.

[2] W. Zeng, H. Yu and C. Lin (Editors). *Multimedia Security Technologies for Digital Rights Management*, Elsevier, July 2006.

[3] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich and T. Kalker, *Digital Watermarking and Steganography* (2nd Ed.), Morgan Kaufmann, 2008.

[4] ISO/IEC MPEG W9216. Call for proposals on image & video signature Tools. Lausanne,CH, Jul 2007.

[5] ISO/IEC MPEG W10155. Call for proposals on video signature tools. Busan, Korea, Oct 2008.

[6] J. Oostveen, T. Kalker, and J. Haitsma. Feature extraction and a database strategy for video fingerprinting. *Proc. Int'l Conf. Recent Advances in Visual Information Systems*, 117-128, 2002.

[7] S. Jobs. Thoughts on Music. http://www.apple.com/honews/thoughtsonmusic,Feb6,2007.

[8] W. Kraaij, P. Over, J. Fiscus, A. Joly. Final CBCD Evaluation Plan TRECVID 2008 (v1.3). Jun 3, 2008. http://www-nlpir.nist.gov/projects/tv2008/Evaluation-cbcd-v1.3.htm.