CONTENT IDENTIFICATION IN CONSUMER APPLICATIONS

Ching-Wei Chen, Randall Cook, Markus Cremer, Peter DiMaria

Gracenote, 2000 Powell Street, Emeryville, CA, USA 94608

ABSTRACT

Much public attention has recently been drawn to the use of content identification technologies to filter or block the unauthorized publication of copyrighted material on Internet sites, as part of the content industry’s effort to recover ground in the fight against Internet piracy. However, the underlying fingerprinting algorithms that are broadly used for content identification have actually been in existence for some time, and have been successfully deployed in numerous applications that enrich and enhance consumers’ experiences of all types of multimedia content. This publication will present an in-depth view into the state-of-the-art of these applications, and discuss the corresponding requirements that said fingerprint technologies have to fulfill. Furthermore, future scenarios are investigated where fingerprinting will play a crucial role in consumer electronics, stretching from home and car entertainment devices to the world of mobile and portable applications.

Index Terms— Fingerprinting, Content Identification, Multimedia Applications

1. INTRODUCTION

The advent of the digital media age has posed new challenges to consumers with regard to browsing and finding content to their liking. While in the past century people would browse by the packaging of physical media, listen to an entire record at a time, or manually assemble a playlist on an audio tape, compact digital devices today hold thousands of albums and videos. An important design factor in modern media players thusly resides in the user interface, to enable users to find the media item of interest efficiently, and to enhance the enjoyment of multimedia content in the same way that cover art and booklet information enriched the user’s experience 20 years ago. On top of this, new distribution channels provided by the omnipresent connectivity of modern day devices offer novel ways to peruse and discover content in an unprecedented fashion. We are still beginning to explore the possibilities opened up by networked Blu-ray disk players, TV sets, and mobile phones that are increasingly replacing traditional unconnected portable media players.

A key component for all applications that provide either media browsing capabilities or content enrichment lies in the associated metadata, i.e. the descriptive data for each media content item. However, for a large amount of media content currently in the field, metadata is insufficient, inaccurate, or not present at all. This makes it an unreliable source for content identification, which is needed to enable the delivery of rich data to empower interesting applications that are easy and convenient to use [1].

In the following chapters, the authors discuss several emerging applications in the digital media space that can be driven by content identification, as well as the need for rich metadata associated with content. From this, we derive a number of requirements for the underlying identification technologies. Subsequently, a flexible system architecture is proposed that is able to deliver the rich media experience of the dawning twenty-first century. The paper concludes with a summary and outlook for the future.

2. CONSUMER APPLICATIONS USING IDENTIFICATION TECHNOLOGIES

Figure 1 below highlights the main classes of applications that rely on content identification to enable and enhance their usefulness.

Figure 1: Applications Driven by Identification

2.1 File Naming, Tagging and Library Normalization

Probably the most familiar use of content identification technology in consumer applications is the automatic naming and tagging of digital files as part of a file encoding process. Users of iTunes, WinAMP, Nero, RealPlayer, Windows Media Player, and many others [2] all rely on content identification technology to identify the compact...
disks that they rip and encode into digital files. Because there is actually no human-readable information embedded in most audio CDs, the ability to automatically identify CDs and tag files has greatly fueled the MP3 revolution.

Because consumers often have assembled collections of digital media files from a variety of sources, and there has been no common standard for embedding accurate and consistent metadata in such files, the consumer is burdened with a collection of music that is difficult to navigate and manage due to inconsistent, inaccurate, or incomplete metadata [1]. Fingerprinting technology makes it possible to automatically re-label a library with a consistent, accurate, and complete set of metadata (e.g. the TuneUp plug-in for iTunes) [4].

Fingerprinting can also identify duplicate recordings, regardless of their associated tag data or file names, thereby removing redundant files. It can also perform the opposite: revising identically tagged files that in fact contain different content (e.g. live and studio versions of a song).

2.2 Media Navigation, Playlisting and Recommendation

Using content identification technology to accurately associate not only information such as song names, but also other descriptive attributes such as genre and mood, makes it possible to add highly functional navigation features to media applications and devices. These include multi-level browsing hierarchies, faceted navigation trees, and enhanced search that allow the user to very quickly and easily find the media content they seek.

Once a media item has been recognized using content identification technology, a variety of other descriptive, statistical, and relational data may be associated with it that may then be used to automatically generate a playlist comprised of similar songs from the user’s own library (e.g. Apple’s iTunes “Genius” feature [5]). Alternatively, following content identification, descriptive metadata such as granular genres, mood, and tempo attributes may be used to compute similarity and derive a playlist.

Again, by definitively identifying all the recordings in the library, media recommendation systems can provide suggestions for new content not present in the consumer’s media library. Alternatively, an application could offer to automatically retrieve a set of higher quality files to upgrade their current library, regardless of where the files were originally acquired. Lastly, content identification allows applications to offer to “complete the collection” if the user currently owns some, but not all, of the works by a particular artist.

2.3 Taste Profiling and Related Content Association

Thanks to content identification, once the composition of a user’s personal collection is known, a taste profile may be generated by examining the genres, time periods, moods, artists, etc. present in their collection. With this profile, one may use collaborative filtering or other statistical techniques to make further recommendations for new content or new personal associations. This technique applies equally well to audio and video collections.

Content identification technology is used extensively to associate related content and other information with media items, and importantly, to do so in a dynamic manner. When the identity of media is definitively known, additional content such as cover art, lyrics, photos, biographies, and reviews may be dynamically associated and updated. Furthermore, dynamic content such as new releases, concert listings, and merchandise may also be maintained.

2.4 Data Mining

Content identification technology makes it possible to not only efficiently create aggregate logs of the media browsing, collecting, purchasing and listening activity of millions of consumers, but also to then mine these logs to produce useful and interesting information for the enjoyment of consumers. Examples of this type of analysis include the creation of various digital media “Top Ten” charts that show the most popular music or video content being enjoyed, or can focus on activity in specific genres or regions.

2.5 Streaming Media and Ambient Content Identification

Whereas traditional consumer content identification technology had been focused on just a single identifier for a complete media item, in recent years, streaming media identification technology has made it possible to make instantaneous identifications throughout the duration of an arbitrarily long piece of content, or even a broadcast stream. This makes it possible to dynamically provide, in real-time, additional information, content, and commerce options that are associated with the media items recognized in the stream without requiring that anything be embedded into the broadcast stream beyond the content itself.

A very popular extension of streaming media identification is the recognition of sounds in the user’s surroundings. Consumer services such as TrackID [6] and Shazam [7] allow any mobile phone owner to simply use their handset to identify and purchase, in real-time, any piece of music audible in a club, bar, restaurant, house, or other location. Full-song waveform fingerprint technology allows the music to be identified regardless of what portion of the recording is currently being played, and can do so even in cases where the audio is highly degraded or masked by ambient noise.
3. REQUIREMENTS FOR CONTENT IDENTIFICATION SYSTEMS

Content identification systems must fulfill certain requirements in order to support the applications discussed above. The core technology must be suitably robust to produce usable results, and must scale to support large installations and high-volume activity. Depending on the application, the technology may need to provide simple, one-shot item identification, or continuous, stream-based identification.

3.1. Robustness Requirements

Robustness is critical to the usefulness of a content identification system. Systems that are too loose in their matching generate excessive numbers of false positive matches, which can cause high-level rules like content-blocking to be applied inappropriately. Systems that are too exacting generate excessive numbers of false negative matches, which make the system unusable due to lack of results.

With audio recorded with a microphone, background noise, weak signals, and poor-quality microphones can degrade the signal. With audio transmitted by radio, equalization, dynamic range compression, and playback speedup relative to the reference audio must be overcome to produce good match results. With audio extracted from digital files, different pressings of the source CD or complete remasterings lead to subtly different versions that all must match to each other.

Equivalently, with video, signal degradations involve adding or removing information from each frame, changing the frame rate, aspect ratio, or colors, and rotating the image or otherwise changing its geometry. Information that is added to the frame includes logos, promotions, news crawls, and letterbox borders.

3.2. Scalability Requirements

The scalability of a content identification system has three main aspects. First, it must provide fast recognition so that large volumes of queries can be economically handled in a given time period. Second, it must support essentially an open-ended reference collection size. Third, it must allow for fast addition and removal of reference content independent of the size of the collection.

3.3. Item-Based and Stream-Based Identification

For some applications, like identification of digital music files, it is reasonable to assume that the unidentified content represents a single cohesive work. In this case the system can stop when it finds at least one match for a single small subsequence of the unidentified content. This type of matching is obviously unable to detect the presence of additional pieces of content outside the short range it checks.

Since only a fixed portion of each piece of reference content needs to be indexed, the capacity of an item-based recognition server can be quite large. In some cases an essentially comprehensive index for several million reference items can be stored on a single server machine, or even reduced somewhat and stored on a portable device [8].

Stream-based identification is a more general form of content identification in which identification is attempted at some number of points along the duration of a piece of unidentified content, depending on how thorough the identification needs to be. At each point of identification, a list of matches is returned. The union of all these lists forms a complete identification report for a given piece of content [9, 10].

The scalability requirements for stream-based identification are harder to achieve than for item-based identification. This is largely because the entire duration of each piece of reference content needs to be indexed, not just a fixed subsequence. In commercially-scaled systems, the size of the index is generally much larger, and so must be distributed across several machines.

4. PROPOSED SYSTEM ARCHITECTURE

The success of a media enrichment and enhancement service hinges upon having an authoritative back-end database that catalogs every piece of content, and keeps track of the myriad relationships between the media, its identifiers (e.g. fingerprints), and all of its associated metadata and other enhancement data. There must be mechanisms to add content and related metadata to the database, as well as to associate each piece of content with a reference fingerprint or identifier for lookups. The system must also allow efficient, reliable, and secure access to its data from large numbers of clients through a scalable and high-performance service.

4.1. Database Construction

At the core of the system is a comprehensive database representing every piece of media in the desired domain. The most efficient way to populate the database with authoritative data is to obtain direct data feeds from large content holders such as media publishers and content aggregators, although user-submitted metadata is also sometimes appropriate. Such feeds include both human-readable metadata and automatically generated fingerprints and other attributes. Each individual piece of content should be assigned a unique identifier which serves as the authoritative link between all information related to that piece of content. As with any large database with multiple
sources of data, normalization and de-duplication are required to provide a solid foundation for applications.

4.3. Application Servicing

In order to provide metadata and media enrichment services for several million pieces of content to millions of users worldwide, it is best to decouple the underlying database from the application service module by means of a separate system optimized for interfacing with large numbers of users at once and servicing queries as quickly and reliably as possible.

In addition to performing identification and metadata retrieval, the application service may also be used to collect statistical data about query events. By logging (preferably non-personal) information about each lookup request, one can generate interesting statistics, such as which content is most popular (by any number of criteria), which pieces of content commonly occur in the same users’ collections, etc.

4.4. Client Application

As detailed above many application scenarios are conceivable. Memory and CPU of the receiving device, network bandwidth, and targeted application determine the requirements and demands to the overall system. The more varied the supported services and client architectures are, the more flexible the application servers have to interact with the client. However, experience and statistics from one application can be used to fuel and optimize another, which benefits the overall quality of the rendered services.

5. SUMMARY AND OUTLOOK

In this paper we have illustrated how identification of media content is a vital part in today’s and tomorrow’s consumer applications. From browsing and navigation of locally stored audio and video files via commercial recommendations and personal radio program services to large databases hosting user generated content, content identification is the key element that empowers a great user experience. From this it becomes evident that going forward, identification will become a compulsory requirement in consumer oriented media products. Moreover, the cross-linking between media sources as well as media types will allow for a more immersive media experience that consumers will come to expect in their product of choice.

6. REFERENCES


