2.3.4 Content Based Identification (Fingerprinting)

Jürgen Herre

I Introduction

Ever since the broad availability of efficient source coding methods (data reduction) and digital distribution channels (including the Internet), consumers have seamless access to an enormous amount of multimedia data. This includes audio material and still and moving pictures within a wide range of quality, ranging from “pre-view” (e.g. Internet radio) to broadcast quality. As a result, efficient handling of this considerable amount of data has become a challenge of its own (e.g. “how can I find desired material efficiently?”). This has led to the definition of a number of so-called metadata standards. Examples for such specifications include the Dublin Core initiative\textsuperscript{225}, the SMPTE/EBU Dynamic Metadata Dictionary\textsuperscript{226}, the P/Meta project of the European Broadcasting Union (EBU)\textsuperscript{227} and, more recently, the MPEG-7 standard\textsuperscript{228}. The general idea behind these standards is to define data formats which provide a comprehensive description of the actual multimedia content in an interoperable way. Such meta-data (i.e. “data about data”) structures may include a wide range of descriptions of the origin and identity of the content, its structure, usage rules, and various perceptual or semantic aspects.

Among the many conceivable ways of characterizing a piece of audiovisual content, the unique description of the content identity based on its signal representation (so-called “content–based identification”) is of great importance. This functionality is frequently also referred to as fingerprinting\textsuperscript{229} and enables automatic identification (including title, author and other description of the works) of content which has been registered previously in an internal database of reference data. The topic of fingerprinting has received much attention recently in both research and commercial deployment and current technological development has shown that, depending on the underlying technology, reliable and efficient identification can still be achieved even for distorted input signals and large databases of multimedia material.

This article discusses the concept of content–based identification and the underlying technological challenges as well as some of its many attractive applications.

\textsuperscript{224} Fraunhofer Institut für Integrierte Schaltungen, Erlangen, Germany.

\textsuperscript{225} Web site of the Dublin Core Metadata Initiative: http://dublincore.org/

\textsuperscript{226} See: SMPTE (2001).

\textsuperscript{227} See: Hopper (2000).

\textsuperscript{228} See: MPEG-7 Introduction (2001).

\textsuperscript{229} As a note to the reader it should be mentioned that the term fingerprinting is occasionally also used in the literature in the context of digital watermarking where the idea is to enable unambiguous identification of the content by imprinting a unique mark into the signal (rather than deriving a fingerprint from it). Unfortunately, this use of terminology may lead to considerable confusion and is, therefore, not endorsed by the author.

\R{E. Becker et al. (Eds.).:} Digital Rights Management, LNCS 2770, pp. 93–100, 2003.
\copyright Springer-Verlag Berlin Heidelberg 2003
in the multimedia area. Owing to the underlying idea, the fingerprinting approach is very different in its nature from (and in fact in a sense complementary to) the concept of watermarking. Thus, the article is concluded by contrasting both approaches with respect to their use cases.

II The Concept of Fingerprinting

During the recent years, a number of technologies for fingerprinting of multimedia material were developed. In contrast to the identification of content based on embedded digital watermarks, fingerprinting is a “non-invasive” approach which does not require any modification of the original multimedia signal. The underlying idea consists of identifying the audio/image/video content directly by examining the characteristics of its signal representation using a pattern recognition process. As usual within the framework of pattern recognition, a training phase is required so that the characteristics of the items to be recognised are introduced into the system. This leads to a two-stage process (see Figure 1):

- During the training phase, characteristic features are extracted from a set of known reference items such that the extracted feature data forms a unique combination which allows for the unambiguous distinction of a particular item from all other entries. Such feature representations can be made extremely compact (e.g. several orders of magnitude smaller than MP3-compressed audio) and are frequently called fingerprints, signatures or robust hashes. For each item which should be recognised later by the system, such a fingerprint is generated and stored in a reference database together with some of its descriptive metadata. This metadata may be just enough for the identification of the individual item in terms of bibliographic reference (e.g. title name, artist) or may contain richer descriptions of the content.

- During the recognition phase, the signal to be identified (query item) is presented to the system and used for the extraction of a fingerprint in a way similar to that of the training phase. The actual recognition process is based on comparing this query fingerprint with the fingerprints that are stored in the reference database. The most “similar” fingerprint found in the database corresponds to the best matching (and most likely) reference item. As a result of this comparison process, the system delivers an indication of whether the presented signal has been successfully identified and, if this is the case, the database ID of the identified item together with a measure of the achieved recognition confidence. Furthermore, the metadata associated with this database item may be returned by the system.