A Content-Sensitive Approach to Search in Shared File Storages

Gábor Richly
Budapest University of Technology and Economics, Hungary

Gábor Hosszú
Budapest University of Technology and Economics, Hungary

Ferenc Kovács
Budapest University of Technology and Economics, Hungary

INTRODUCTION

The article presents a novel approach to search in shared audio file storages such as P2P-based systems. The proposed method enables the recognition of specific patterns in the audio contents, in such a way it extends the searching possibility from the description-based model to the content-based model. The targeted shared file storages seam to change contents rather unexpectedly. This volatile nature led our development to use real-time capable methods for the search process.

The importance of the real-time **pattern recognition** algorithms that are used on audio data for content-sensitive searching in stream media has been growing over a decade (Liu, Wang, & Chen, 1998). The main problem of many algorithms is the optimal selection of the reference patterns (*soundprints* in our approach) used in the recognition procedure. This proposed method is based on distance maximization and is able to choose the pattern that later will be used as reference by the pattern recognition algorithms quickly (Richly, Kozma, Kovács & Hosszú, 2001).

The presented method called ***EMESE (Experimental MEdia-Stream rEcognizer)*** is an important part of a light-weight content-searching method, which is suitable for the investigation of the network-wide shared file storages. This method was initially applied for real-time monitoring of the occurrence of known sound materials in broadcast audio. The experimental measurement data showed in the article demonstrate the efficiency of the procedure that was the reason for using it in shared audio database environment.

BACKGROUND

From the development of the Napster (Parker, 2004), the Internet-based communication is developing toward the **application level network**s (**ALN**). On the more and more powerful hosts, various collaborative applications run and create virtual (logical) connections with each others (Hosszú, 2005). They establish virtual **overlay**, and oppositely to the older **client/server model** they use the **peer-to-peer** (**P2P**) communication. The majority of such systems deal with file sharing (Adar, Huberman, 2000), that is why their important task is to search in large, distributed shared file storages.

A large amount of effort is dedicated for improving their searching (Yang & Garcia-Molina, 2002) and downloading capability (Cohen, 2003; Qiu & Srikant, 2004), however, the searching is quite traditional, it is based on the descriptive metadata of the media contents, as the file name, content description, and so forth. Such method has an inherent limitation, since the real content remains covered and even in case of the mistake of the file description this search can fail. Oppositely to the widely used description-based seeking, the content-based searching has been the topic of the research, only. Its main reason is that the content and its coded representation have a huge variety that is why a comprehensive method has not developed yet.

The novel system, EMESE, is dedicated for solving a special problem, where a small but significant pattern should be found in a large voice stream or bulk voice data file in order to identify known sections of audio. The developed method is light-weight, meaning that its design goals were the fast operation and the relatively small computing power. In order to reach these goals, the length of the pattern to be recognized should be very limited, and the total score is not required.

This article deals mainly with the heart of the EMESE system, the pattern recognition algorithm, especially with the creation of the reference pattern, called *reference selection*.

THE PROBLEM OF THE PATTERN RECOGNITION

In the field of sound recognition there are many different methods and applications for specific tasks (Coen, 1995; Kondoz, 1994).

The demand for working efficiently with streaming media on the Internet increases rapidly. These audio streams may contain artificial sound effects besides the mix of music and human speech. These effects furthermore may contain signal fragments that are not audible by the ear. As a consequence, processing of this kind of **audio signal** is rather different from the already developed methods, as for example, the short-term predictability of the signal is not applicable.

The representation of digital audio signal as individual sample values lacks any semantic structure to help automatic identification. For this reason, the audio signal is transformed into several different orthogonal or quasi-orthogonal bases that enable detecting certain properties.

Already, there are solutions for classifying the type of broadcast on radio or television using the audio signal. The solution in Akihito, Hamada, and Tonomura (1998) makes basically a speech/music decision by examining the spectrum for harmonic content, and the temporal behavior of the spectral-peek distribution. Although it was applied successfully to that decision problem, it cannot be used for generic recognition purposes. The paper (Liu et al., 1998) also describes a scheme classifying method where the extracted features are based on the short-time spectral distribution represented by a bandwidth and a central frequency value. Several other features, for example, the volume distribution and the pitch contour along the sound clip, are also calculated. The main difficulty with this method is its high computation-time demand, so real-time monitoring is hardly possible, when taking the great number of references to be monitored into account.

A similar monitoring problem was introduced in Lourens (1990) and the used feature, a section of the energy envelope of the record signal (*reference*) was correlated with the input (*test*) signal. The demand on real-time execution drove the development of the recognition scheme introduced in Richly, Varga, Hosszú, and Kovács (2000), and Richly, Kozma, Kovács, and Hosszú (2001) that is capable of recognizing a pattern of transformed audio signal in an input stream, even in the presence of level-limited noise. This algorithm first selects a short segment of the signal from each record in the set of records to be monitored.

THE SOUND IDENTIFICATION IN THE EMESE

The reference selection algorithm needs a well understanding of the recognition method. The audio signal, sampled at f_s=*16kHz* is transformed into a spectral description because

Figure 1. The sound representation in the recognition system