Chapter 5

Content-Based Music Summarization and Classification

Changsheng Xu, Institute for Infocomm Research, Singapore

Xi Shao, Institute for Infocomm Research, Singapore

Namunu C. Maddage, Institute for Infocomm Research, Singapore

Jesse S. Jin, The University of Newcastle, Australia

Qi Tian, Institute for Infocomm Research, Singapore

ABSTRACT

This chapter aims to provide a comprehensive survey of the technical achievements in the area of content-based music summarization and classification and to present our recent achievements. In order to give a full picture of the current status, the chapter covers the aspects of music summarization in compressed domain and uncompressed domain, music video summarization, music genre classification, and semantic region detection in acoustical music signals. By reviewing the current technologies and the demands from practical applications in music summarization and classification, the chapter identifies the directions for future research.

INTRODUCTION

Recent advances in computing, networking, and multimedia technologies have resulted in a tremendous growth of music-related data and accelerated the need to analyse and understand the music content. Music representation is multidimensional and time-dependent. How to effectively organize and process such large variety and quantity of music information to allow efficient browsing, searching and retrieving is an active research area in recent years. Audio content analysis, especially music content understanding, posts a big challenge for those who need to organize and structure music data. The difficulty arises in converting the featureless collections of raw music data to suitable forms that would allow tools to automatically segment, classify, summarize, search and retrieve large databases. The research community is now at the point where the limitations and properties of developed methods are well understood and used to provide and create more advanced techniques tailored to user needs and able to better bridge the semantic gap between the current audio/music technologies and the semantic needs of interactive media applications

The aim of this chapter is to provide a comprehensive survey of the technical achievements in the area of content-based music summarization and classification and to present our recent achievements. The next section introduces music representation and feature extraction. Music summarization and music genre classification are presented in details in the two sections, respectively. Semantic region detection in acoustical music signals is described in the fifth section. Finally, the last section gives the concluding remarks and discusses future research directions.

MUSIC REPRESENTATION
AND FEATURE EXTRACTION

Feature extraction is the first step of content-based music analysis. There are many features that can be used to characterize the music signal. Generally speaking, these features can be divided into three categories: timbral textural features, rhythmic content features and pitch content features.

Timbral Textural Features

Timbral textural features are used to differentiate mixture of sounds that may have the same or similar rhythmic and pitch contents. The use of these features originates from music-speech discrimination and speech recognition. The calculated features are based on the short-time Fourier transform (STFT) and are calculated for each frame.

Amplitude Envelope

The amplitude envelope describes the energy change of the signal in the time domain and is generally equivalent to the so-called ADSR (attack, decay, sustain and release) of a music song. The envelope of the signal is computed with a frame-by-frame root mean square (RMS) and a third-order Butterworth low-pass filter (Eiilis, 1994). RMS is a perceptually relevant measure and has been shown to correspond closely to the way we hear *Loudness*. The length of the RMS frame determines the time resolution of the

34 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/content-based-music-summarization-classification/25970

Related Content

Web Search Results Discovery by Multi-granular Graphs
Gloria Bordogna, Alessandro Campi, Giuseppe Psailaand Stefania Ronchi (2012). *Quantitative Semantics and Soft Computing Methods for the Web: Perspectives and Applications (pp. 118-136).*
www.irma-international.org/chapter/web-search-results-discovery-multi/60118

Contiguous Placement on Hierarchical Storage Systems
Phillip K.C. Tse (2008). *Multimedia Information Storage and Retrieval: Techniques and Technologies (pp. 156-160).*
www.irma-international.org/chapter/contiguous-placement-hierarchical-storage-systems/27010

On the Applicability of Speaker Diarization to Audio Indexing of Non-Speech and Mixed Non-Speech/Speech Video Soundtracks
Robert Mertens, Po-Sen Huang, Luke Gottlieb, Gerald Friedland, Ajay Divakaranand Mark Hasegawa-Johnson (2012). *International Journal of Multimedia Data Engineering and Management (pp. 1-19).*
www.irma-international.org/article/applicability-speaker-diarization-audio-indexing/72890

A Comparative Study of Graph Kernels and Clustering Algorithms
Riju Bhattacharya, Naresh Kumar Nagwaniand Sarsij Tripathi (2021). *International Journal of Multimedia Data Engineering and Management (pp. 33-48).*
www.irma-international.org/article/a-comparative-study-of-graph-kernels-and-clustering-algorithms/271432

Generating Window of Sign Languages on ITU J.200-Based Middlewares
Felipe Lacet Silva Ferreira, Tiago Maritan Ugulino de Araújo, Felipe Hermínio Lemos, Gutenberg Pessoa Botelho Neto, José Ivan Bezerra Vilarouca Filhoand Guido Lemos de Souza Filho (2012). *International Journal of Multimedia Data Engineering and Management (pp. 20-40).*
www.irma-international.org/article/generating-window-sign-languages-itu/69519