



IRM PRESS

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-<http://www.irm-press.com>

ITB11372

This chapter appears in the book, *Managing Multimedia Semantics*, edited by Uma Srinivasan and Surya Nepal
© 2005, Idea Group Inc.

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.

Copyright © 2005, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

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

Misinformation via Tampered Multimedia Content

(2019). *Cross-Media Authentication and Verification: Emerging Research and Opportunities* (pp. 62-86).

www.irma-international.org/chapter/misinformation-via-tampered-multimedia-content/208001

Terminals for the Smart Information Retrieval

Gregor Rozinaj (2009). *Handbook of Research on Mobile Multimedia, Second Edition* (pp. 263-274).

www.irma-international.org/chapter/terminals-smart-information-retrieval/21009

Trends in Managing Multimedia Semantics

Roberto Poli, Achilles D. Kameasand Lambrini Seremeti (2018). *Digital Multimedia: Concepts, Methodologies, Tools, and Applications* (pp. 1690-1702).

www.irma-international.org/chapter/trends-in-managing-multimedia-semantics/189549

Creating Emotions Through Digital Media Art: Building Empathy in Immersive Environments

Paulo Veloso Gomes, Vítor J. Sá, António Marques, João Donga, António Correiaand Javier Pereira Loureiro (2020). *Multidisciplinary Perspectives on New Media Art* (pp. 142-160).

www.irma-international.org/chapter/creating-emotions-through-digital-media-art/260024

WLAN Security Management

Göran Pulkkis (2009). *Encyclopedia of Multimedia Technology and Networking, Second Edition* (pp. 1558-1572).

www.irma-international.org/chapter/wlan-security-management/17585