

Chapter 4

Daubechies Wavelets Based Robust Audio Fingerprinting for Content-Based Audio Retrieval

Wei Sun

Zhejiang University, China

Zhe-Ming Lu

Zhejiang University, China

Fa-Xin Yu

Zhejiang University, China

Rong-Jun Shen

Zhejiang University, China

ABSTRACT

Audio fingerprinting is the process to obtain a compact content-based signature that summarizes the essence of an audio clip. In general, existing audio fingerprinting schemes based on wavelet transforms are not robust against large linear speed changes. The authors present a novel framework for content-based audio retrieval based on the audio fingerprinting scheme that is robust against large linear speed changes. In the proposed scheme, 8 levels Daubechies wavelet decomposition is adopted for extracting time-frequency features and two fingerprint extraction algorithms are designed. The experimental results from this study are discussed further into the article.

INTRODUCTION

With the development of computer network and multimedia technologies, especially the digital audio compression technology, audio transmission has become more and more convenient and wider and wider. Consequently, the copyright protection and security problems have become more and more urgent. Digital audio watermark-

ing and digital audio fingerprinting provide two effective ways to solve this problem. In fact, an audio fingerprint is a compact content-based signature that summarizes the essence of an audio clip. It has attracted much attention since it can implement audio identification regardless of its format and without meta-data or watermark embedding (Cano et al., 2005). In recent years, many efforts have been made

DOI: 10.4018/978-1-4666-4006-1.ch004

in the field of audio fingerprinting and a good overview can be found in Cano et al. (2005). Existing schemes can be classified into three categories, i.e., time-domain based, transform-domain based and compressed-domain based. For example, a robust audio feature called local energy centroid (LEC) was proposed to represent the energy conglomeration degree of the relative small region in the spectrum (Pan et al., 2011), while a robust audio fingerprinting algorithm in the MP3 compressed domain was proposed with high robustness to time scale modification (Zhou & Zhu, 2011). Among existing transform-based audio fingerprinting schemes, the schemes based on the wavelet transform are very popular, since the wavelet transform or more particularly the discrete wavelet transform is a relatively recent and computationally efficient technique for extracting information about non-stationary signals like audio. Wavelet transform is a local transformation on a signal in time and frequency domains, which can effectively extract information from the signal, and do multi-scale detailed analysis on a function or signal by functions such as scaling and translation, thereby can solve many difficult issues which cannot be solved by the Fourier transform. Therefore, our paper focuses on the wavelet transform based schemes. The existing works based on wavelet transforms can be classified into the following two categories.

The first type of fingerprinting schemes performs the wavelet transform on each audio frame directly to extract time-frequency features for audio fingerprinting. In Lu (2002), the one dimensional continuous Morlet wavelet transform is adopted to extract two fingerprints for authentication and recognition purposes, respectively. In Ghouti and Bouridane (2006), a robust perceptual audio hashing scheme using balanced multiwavelets (BMW) is proposed. They first perform 5 levels wavelet decomposition on each audio frame and divide the 5 decomposition sub-bands' coefficients into 32 different frequency

bands. Then the estimation quantization (EQ) with a window of 5 audio samples is adopted. Finally, 32 bits sub-fingerprinting is extracted according to the relationship between the log variances of each sub-bands' coefficients and the mean of all the log variances for each audio frame. They do several experiments to demonstrate that their scheme is robust to several signal processing attacks and manipulations except for linear speed change.

The other type of fingerprinting schemes introduces the computer vision technique to convert the audio clip into a 2-D spectrogram and then apply the wavelet transform. In Ke et al. (2005), the spectrogram of each audio snippet is viewed as a 2-D image and the wavelet transform is used to extract 860 descriptors for a 10 seconds audio clip. Then apply the pairwise boosting scheme to learn compact, discriminative, local descriptors that are efficient in audio retrieval. This algorithm can finish retrieving quickly and accurately in practical systems with poor recording quality or significant ambient noises. In Baluja and Covell (2006, 2007), the so-called Waveprint, combining of computer vision and data stream processing, was proposed. The Harr wavelet is used for extracting the t top magnitude wavelets for each spectral image. And the selected features are modeled by the Min-Hash technique. In the retrieval step, the locality sensitive hashing (LSH) technique is introduced. This algorithm exhibits an excellent identification rate against content-preserving degradations except for linear speed changes. Furthermore, the tradeoffs between the performance, memory usage, and computation are analyzed through extensive experiments. As an extension, the parameters of the system are analyzed and verified in Baluja and Covell (2008). This system shows superiority in terms of memory usage and computation, while being more accurate when compared with Ke et al. (2005).

12 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/daubechies-wavelets-based-robust-audio/75663

Related Content

A HIPAA Security and Privacy Compliance Audit and Risk Assessment Mitigation Approach

Young B. Choi and Christopher E. Williams (2021). *International Journal of Cyber Research and Education* (pp. 28-45).

www.irma-international.org/article/a-hipaa-security-and-privacy-compliance-audit-and-risk-assessment-mitigation-approach/281681

Visual Gambling

Teng Guo, Feng Liu and Wen Wang (2018). *International Journal of Digital Crime and Forensics* (pp. 43-49).

www.irma-international.org/article/visual-gambling/205522

Lane Detection Algorithm Based on Road Structure and Extended Kalman Filter

Jinsheng Xiao, Wenxin Xiong, Yuan Yao, Liang Li and Reinhard Klette (2020). *International Journal of Digital Crime and Forensics* (pp. 1-20).

www.irma-international.org/article/lane-detection-algorithm-based-on-road-structure-and-extended-kalman-filter/246835

E-Government, Security, and Cyber-Privacy: Individual Rights versus Government Responsibility

Ross Wolfand Ronnie Korosec (2012). *Cyber Crime: Concepts, Methodologies, Tools and Applications* (pp. 1314-1327).

www.irma-international.org/chapter/government-security-cyber-privacy/61011

Information Hiding Model Based on Channel Construction of Orthogonal Basis

Bao Kangsheng (2021). *International Journal of Digital Crime and Forensics* (pp. 1-18).

www.irma-international.org/article/information-hiding-model-based-on-channel-construction-of-orthogonal-basis/277089