신호의 복원된 위상 공간을 이용한 오디오 상황 인지

La The Vinh(1) , Asad Masood Khattak(1), Trinh Van Loan(2), 이승룡(1), 구교호(1)
(1) 유비쿼터스 컴퓨팅 연구실, 컴퓨터 공학과, 경희대학교
(2) Ha Noi University of Technology
e-mail : vinhlt@oslab.khu.ac.kr, asad.masood@oslab.khu.ac.kr, loantv@it-hut.edu.vn, sylee@oslab.khu.ac.kr, yklee@khu.ac.kr

Audio Context Recognition Using Signal's Reconstructed Phase Space

La The Vinh(1), Asad Masood Khattak(1), Trinh Van Loan(2), Sungyoung Lee(1), Young-Ko Lee(1)
(1) Ubiquitous Computing Lab, Computer Engineering Department, Kyung Hee Univerisity, Korea
(2) Ha Noi University of Technology
e-mail : vinhlt@oslab.khu.ac.kr, asad.masood@oslab.khu.ac.kr, loantv@it-hut.edu.vn, sylee@oslab.khu.ac.kr, yklee@khu.ac.kr

So far, many researches have been conducted in the area of audio based context recognition. Nevertheless, most of them are based on existing feature extraction techniques derived from linear signal processing such as Fourier transform, wavelet transform, linear prediction... Meanwhile, environmental audio signal may potentially contains non-linear dynamic properties. Therefore, it is a big potential to utilize non-linear dynamic signal processing techniques in audio based context recognition.

1. Introduction

Nowadays, together with an increasing number of wearable devices an equally increasing number of applications for these devices is available. These applications often require user's attention by various notifications such as incoming telephone calls, short messages, e-mails,...Unfortunately, the notifications, which can happen anywhere at anytime in any situation, sometimes may cause annoyance to the users. For example, there may be a cell phone ring in a meeting or a missed call because of the silent mode when walking around in a noisy shopping center. Clearly, there is a need to handle these events in a smart way depending on particular user's context so that a notification will appear as expected. Therefore, context awareness is becoming an important research area and finding its way to many applications especially in human computer interaction (HCI). Context awareness is concerned with the acquisition of context by sensing hardware, the abstraction and understanding of context through inference algorithms. In terms of the data acquisition, among a variety of input devices such as global positioning system (GPS), motion sensors, environmental sensors... audio recording device seems to be a potential choice [3] since it is available in most wearable or mobile devices for both outdoor or indoor circumstances. Furthermore, audio data provides a valuable source of context-related information, and sound-based context recognition is possible for human to some extent. It is therefore, in this paper we investigate in context recognition using only ambient sound. In terms of processing techniques, most of existing researches used feature extraction algorithms originated from speech signal analysis such as linear prediction, frequency analysis [3], wavelet transform... While these techniques in speech signal processing are base on a well studied linear production model of speech, the nature of ambient sound is much more complicated and potentially contains non-linear dynamic properties. Therefore, the linear processing methods may not extract some important features of the ambient audio signal. In addition, recently, non-linear dynamic signal processing has been proved to be a promised method in time series classification. Nonetheless, to the best of our knowledge, not many researchers focus on audio-based context recognition, especially, the use of non-linear dynamic features of the ambient sound to classify the user context. Therefore, motivated by the remarkable results of the non-linear dynamic processing techniques in some similar areas [1], [2], we decided to take into account these methods in the audio-based context recognition problem introduced above.

2. Methodology

Let us denote $X = x_1 x_2 ... x_T$ as an audio input signal of length T. In our work, the input signal is first embedded in a multidimensional phase space, providing a geometric representation or a phase portrait of the signal. The embedded signal has the form $E = e_1 e_2 ... e_N$, where each

e_i is defined as a point in a m-dimensional space as followings:

$$e_i = [x_i, x_{i+\tau}, x_{i+2\tau}, ..., x_{i+(m-1)\tau}], \text{ i=1,2…T-(m-1)}\tau,$$

where τ and m are the two parameters which we have to estimate from the input signal. Base on chaos theory, τ can be estimated as the first local minimum of auto mutual information function of signal X and m can be calculated by using minimum embedding dimension (Cao) method [4]. From our experiment results, we found that for an audio signal, τ is about 0.008 second, and m is approximately 16. Figure 1 and 2 illustrate the two dimensional phase spaces of an "building site" sound and a "office" sound, respectively. Figure 3 and 4 show the frequency domain representations of the two sounds in that order.

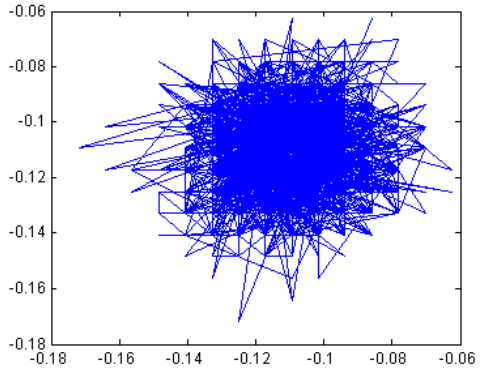

Figure 1. 2-dimensional phase portrait of a "building site" sound

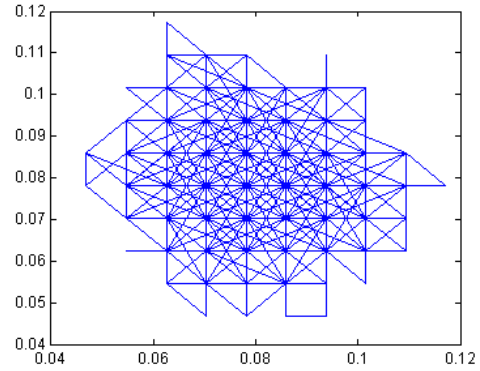

Figure 2. 2-dimensional phase portrait of an "office" sound

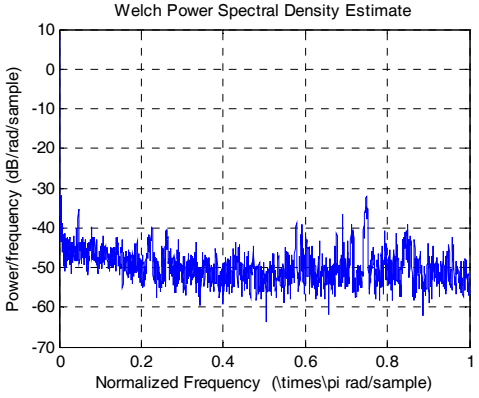

Figure 3. "building site" sound spectral

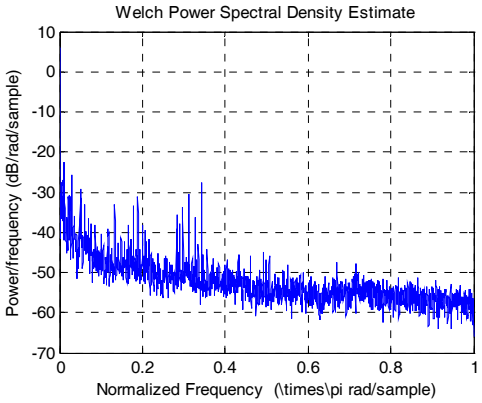

Figure 4. "office" sound spectral

As can be seen in the above figures, the phase portraits expose more distinguishable characteristics than the spectral of the two audio signal. It is therefore, we expect that using the signal's reconstructed phase may produce better result in a recognition system. In a brief experiment, we use Gaussian mixture model with 8 components to learn the distribution of the embedded points in the phase space of the input signal. Then, given a testing sound, we decided the class label of this sound base on the highest likelihood among all trained Gaussian models. The experiment showed an accuracy of over 90% with a dataset containing 12 audio categories (building site, bus, car, car highway, office, train, presentation, shopping centre, street traffic, laundrette).

Acknowledgement

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2009-(C1090-0902-0002)). Also, it was supported by the IT R&D program of MKE/KEIT, [10032105, Development of Realistic Multiverse Game Engine Technology].

References

[1] Guo Feng Wang, Yu Bo Li, Zhi Gao Luo, "Fault classification of rolling bearing based on reconstructed phase space and Gaussian mixture model", Journal of Sound and Vibration 323 (2009), 1077–1089.

[2] Richard J. Povinelli, Andrew C. Lindgren, Jinjin Ye, "Statistical Models of Reconstructed Phase Spaces for Signal Classification", IEEE Transactions on signal processing 54 (2006).

[3] Antti J. Eronen, Vesa T. Peltonen, Juha T. Tuomi, Anssi P. Klapuri, Seppo Fagerlund, Timo Sorsa, Gaëtan Lorho, Jyri Huopaniemi, "Audio-Based Context Recognition", IEEE Transactions on audio, speech, and language processing 14 (2006).

[4] Holger Kantz and Thomas Schreiber, "Nonlinear Time Series Analysis", 2nd Edition, 2004, Cambridge University Press.