Recording Device Identification Based on Ceptral Mixed Feature

Wei Zhong, Xiangwei Kong, Xin-gang You, Bo wang

Dalian University of Technology

Introduction

The recording equipment classification is the latest audio forensics research hotspot[1]. In the course of the audio evidence provided, somebody claims that he used a device to record audio evidence, but there is no effective way to verify it, hence, people carry out researches in this area. In 2006, Lukas studied on the effects of the sensor output noise on VCR recognition. Since 2007, Dirik, who studied the impact of dust characteristics of the sensor to VCR recognition, achieved valuable results. Tsai and Li *et al*. [2-3] had a in-depth study cellular phone recognition. Cemal *et al*. [4] extracted cell. phone's characteristics from the cell phone recording signal, and using the MFCC parameters as feature parameters and SVM as a recognition model, a high recognition rate of 96% is achieved for 14 different phones.

Cemal had studied and analyzed the characteristic parameters and recognition model of recording equipment, its characteristic parameters and recognition model are based on the existing speaker recognition features and models, either Fourier Transformation parameters or MFCC is not for a special recording device identification parameters. Characteristic parameters that specifically for recording equipment are still very few. In terms of the MFCC, low-dimensional parameters generally reflect the speaker's semantic features, and high-dimensional parameters generally reflect the speaker's personality traits. The MFCC will definitely affect recognition of recording devices accuracy rate, when it is used as a characteristic parameter of recording equipment. Therefore, we must find or construct characteristic parameters consistent with the characteristics of the recording device.

From the recording equipment itself, taking into account copyright and other reasons, there may be a difference in terms of recording circuit and chip, sampling rate, the number of quantization bits and the compression algorithm, where we can find the recording equipment personality characteristics. Also, recording equipment parameters are not only mixed in semantic features bands but also mixed in speaker feature parameters.

Hence, considering the lack of special characteristic parameters of recording equipment, we study and propose a number of characteristic parameters characterizing feature of recording equipment firstly, and then combining with existing audio feature consist of mixed feature of recording equipment.

Methods

The proposed method adopts the following mixing 92-dimensional feature mixing parameters as the characteristic parameter of recording equipment. Table 1 shows the details.
The MFCC and DCT minimum amplitude proportion features based on frequency domain have been proved to be effective in prior works. In last two subsection, we have demonstrated that the effect of quantization step of difference devices. The feature vectors in spacial domain are sensitive to the effect. A reasonable approach can be obtained to combine the time-domain and frequency-domain features to construct a better classifier. Base on this, we mix 44-dimensional MFCC features, 10-dimentional DCT minimum amplitude proportion features, 20-dimensional time domain minimum amplitude proportion features and 20-dimensional time-domain low proportion roughness features for the feature vector.

Table 1 Time-frequency mixing characteristic parameters of recording equipment

Mixed characteristic parameters	Description
MFCC1-10，33-64	Using 64-dimensional MFCC parameters' low-dimensional and high-dimensional parts
10-dimensional DCT minimum amplitude proportion	Frequency domain features, after DCT transform, calculate the number of the minimum value of 10 points in the proportion of all point values.
20-dimensional minimum amplitude proportion	Time-domain characteristics, calculate the number of the minimum value of 20 points in the proportion of all point values. Definition is shown in Equation (3).
20-dimensional time-domain low proportion roughness	Definition is shown in Equation (6).

Recognition model uses SVM classifier.

Results

The recording device used in the experiment are five recording device:(1) Sony PCM-M10, (2) Tong Fang TF-A20, (3) Jing Hua DVR-818, (4) Modern HYM-3698, (5) Sanyo ICR-PS004M (each type of equipment is two). Recording subjects were 60 persons consist of 30 young men and 30 women. Everyone speaks 10 different Mandarin, and every word is about 10 seconds, generating 6000 wav audio data. The sampling frequency is 44.1KHz, quantization bits are all 16-bit, frame length is 2048 points, a frame shift of 50%.Take a word each person and each device as training audio, the other as a test audio.

The proposed method with hybrid characteristic parameters is compared with a baseline proposed in [4]. Experimental results are listed in Table 2.

Table 2 gives a comparison of recognition rate between hybrid characteristic parameters and the baseline system. Recognition mode uses the text-independent manner. From the table, recognition rate of hybrid characteristic parameters increases by more than 6% compared with baseline system. The most obvious improvement is Sony, which improve by 9.5%. For a variety of devices, recognition rate of Sony is highest, Sanyo and modern secondly, between 75% to 83%. Tong Fang and Jing Hua are poor, around 70%. An average accuracy of 80.0% is achieved, compared with that of 73.4% obtained by the baseline. The results shows that combination of the proportion of low time-domain roughness and MFCC can improve the performance of the device identification

Table 3 shows the result of picking up characters from characteristic parameters of base line and mixing characteristic parameters through the way of orthogonal projection operator. From the table, it is obvious that adopting the orthogonal projection operator improves the recognition rate of system. For example, equipments like Sony, Sanyo and Modern get a significant improvement of 3% to 5% approximately. However, the improvement of property seems not very obvious for Tong Fang and Jing Hua, whose improvements are approximately below 1%.

Table 2 Time-frequency mixing characteristic parameters of recording equipment

	Sony	Sanyo	Modern	Tong Fang	Jing Hua	AVGERAGE
Baseline proposed in [4]	82.2%	74.6%	76.9%	65.1%	68.4%	73.4%
Proposed method	91.7%	78.5%	81.4%	73.0%	75.5%	80.0%

Table 3 Identify performance comparison of no projection of hybrid feature parameters

	Sony	Sanyo	Modern	Tong Fang	Jing Hua	AVGERAGE
[4] with orthogonal projection operator	86.3%	77.9%	80.6%	66.7%	69.2%	76.1%
Proposed method with orthogonal projection operator	93.1%	83.2%	84.0%	74.4%	75.9%	82.1%

Conclusions

The original-evidence research mainly consists of obtaining evidence with the recording equipment, recognizing the time and place of recording and so on. The progress of recognizing recording time and place achieve less among home and abroad. The judge mainly depends on the relevance of other evidence during the actual operation. But research of obtaining evidence of recording equipment is still the hot issue among domestic and overseas in terms of speech single processing , which remains in the technology trigger and has not raised or analyzed the special characteristic parameter of recording evidence. The article goes deep into the characteristic parameter of recording evidence, raises the time-domain low proportion roughness and other two characteristic parameters of recording evidence, which constitutes 92-dimensional feature mixing parameters combined with the modified MFCC characteristic parameters . The experiment demonstrates that the mixed characteristic parameters are able to represent the feature of recording evidence effectively, by collecting sixty youth that ten different speech each of them and two speech of the same model with five different brand of recording evidence, whose recognition rate raises up by 10.4 percent comparing with the ordinary parameters of cepstrum.

Bibliography

1. **Panagakis, Y., Kotropoulos, C.. Automatic telephone handset identification by sparse representation of random spectral features[C]. MM and Sec'12 - Proceedings of the 14th ACM Multimedia and Security Workshop, ACM, USA. pp, 91-95(2012).**
2. **Tsai M. J., Lai C. L., Liu J.. Camera/mobile phone source identification for digital forensics[C]. Proceeding of IEEE International Conference on Acoustics, Speech Signal Processing, IEEE, USA. pp, II-221 - II-224(2007).**
3. **Celiktutan O., Sankur B., Avcibas I.. Blind identification of source cell phone model[J]. IEEE Transaction on Information Forensics and Security, 3(3): 553–566(2008).**
4. **Hanilci, C., Ertas, F.. Recognition of Brand and Models of Cell-Phones From Recorded Speech Signals[J]. IEEE Transaction on Information Forensics and Security, 7(2): 625-634(,2012).**