2024 IEEE 12th International Conference on Computer Science and Network Technology

A Novel Audio Forgery Detection Method Based on Short-time Power Spectral Density

Jiayi Cheng[1,2] ,Xiaolong Li[1,2]

[1] Institute of Information Science,Beijing Jiaotong University,Beijing,China

[2]Beijing Key Laboratory of Advanced Information Science and Network Technology,Beijing,China

Abstract

The popularity of portable recording equipment has made the acquisition of digital audio convenient, and the development of multimedia editing software has made audio editing and modification progressively easier. Therefore, it is essential to ensure the authenticity and integrity of digital media files. Copy-move forgery is a prevalent form of audio tampering, but existing detection methods suffer from low accuracy and poor robustness. To address this issue, we propose a detection method for audio copy-move forgery based on short-time power spectral density (STPSD). Specifically, we first separate the syllables in the voiced part of the speech signal. Then, the STPSD features are extracted from the suspect byte pair. Finally, the content matching algorithm is used for detection. The experimental results show that the proposed algorithm can resist all kinds of common attacks reliably, and can accurately locate copied and moved forged syllables, with high efficiency.

Motivation

- Since the 21st century, digital technology's rapid advance has facilitated faster information dissemination. Digital recordings have become pivotal in evidence collection. However, the manipulation of audio clips using editing tools can alter the original meaning of recordings, which may have adverse societal implications. Consequently, the development of an efficient method to detect digital audio copy-move forgeries is crucial for audio forensics. The copy and paste tamper detection process is shown in Figure 1.
- Most of the existing copy-move detection methods have low detection efficiency and poor audio robustness for post-processing operations.

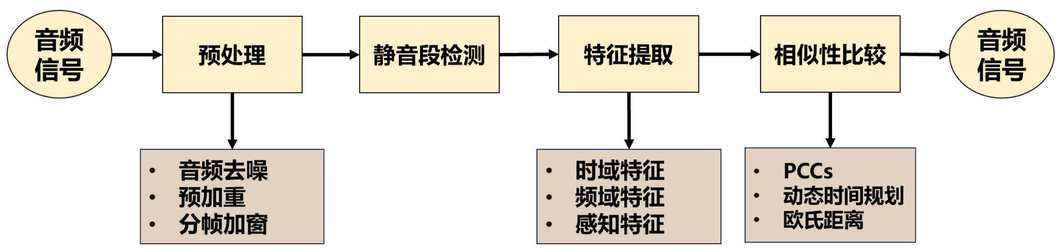


Figure 1 Audio copy-paste tamper detection common process.

Methodology

To improve the accuracy and robustness of audio tamper detection, we focus on extracting common audio features from real audio. As shown in Figure 2, STPSD features were extracted to calculate similarity, and content matching was performed on the sorted byte pairs to build the detection framework.

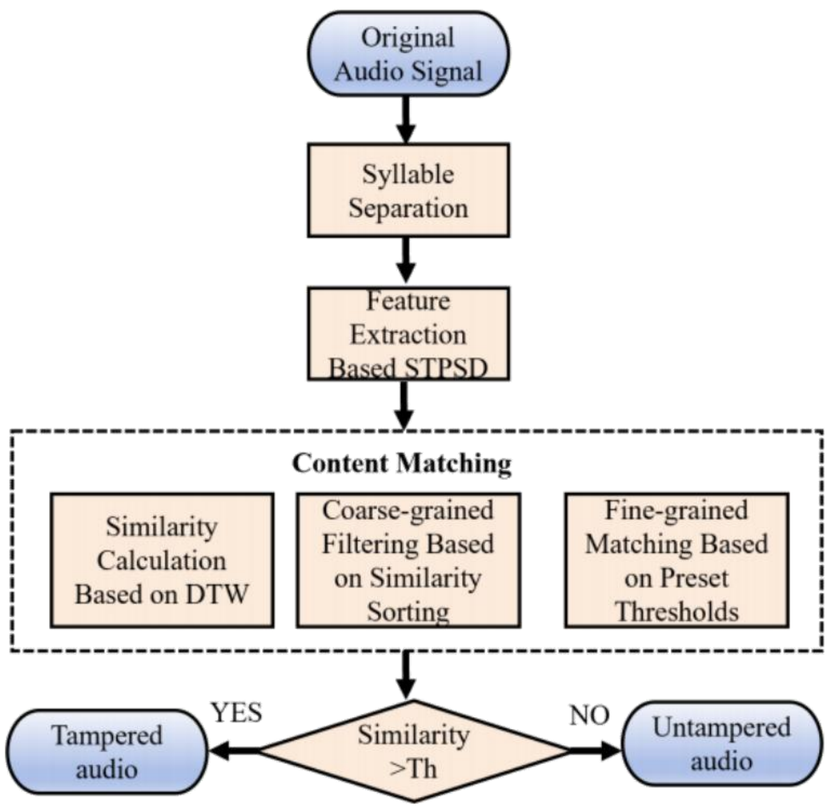
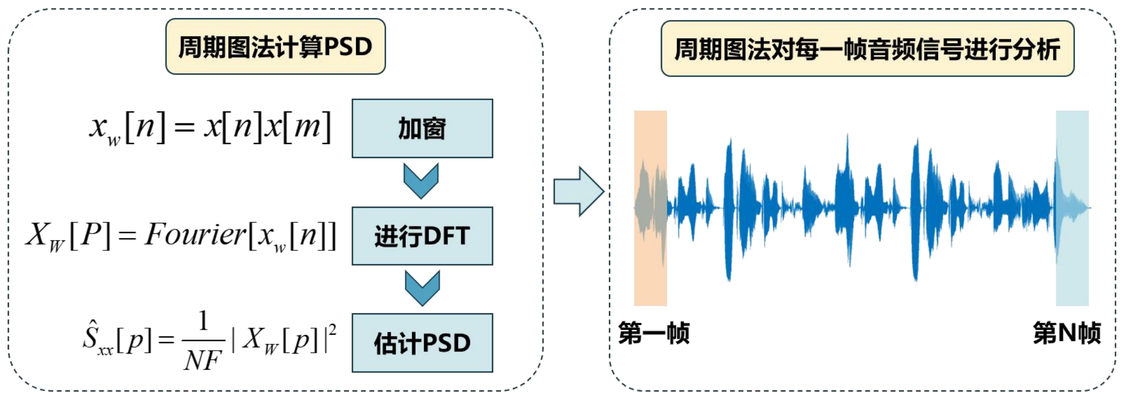
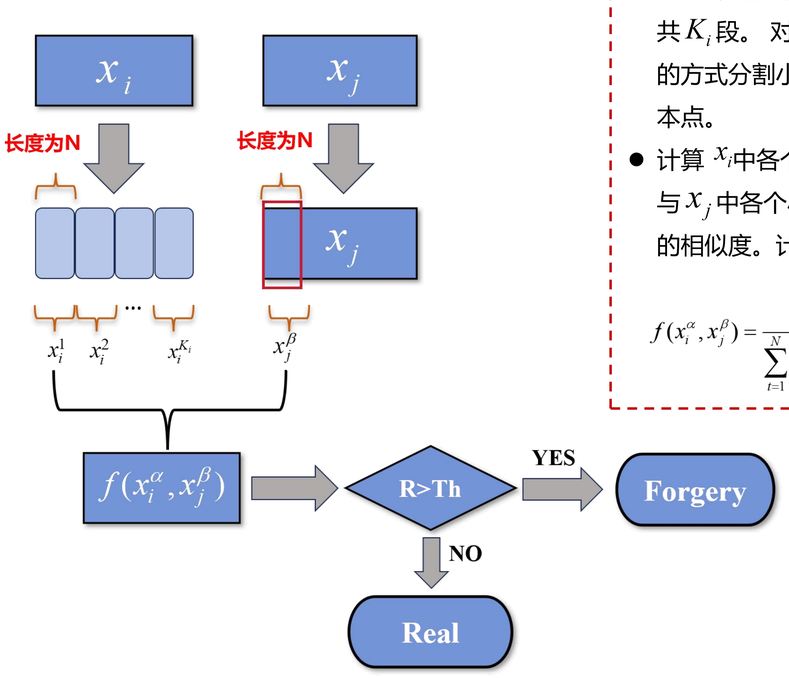


Figure 2 The framework of the proposed copy-move detection method

STPSD Extractor



Content matching



- 对 x_i 平均分割成小片段，每段长度为 N，共 K_i 段。 对 x_j 采取滑动(长度为 N) 窗口的方式分割小片段，每次只滑动一个样本点。
- 计算 x_i 中各个小片段 $x_i^\alpha(\alpha \in (1,2,...K_i))$ 分别与 x_j 中各个小片 $x_j^\beta(\alpha \in (1,2,...K_j - N + 1))$ 段的相似度。计算相似度的公式为：

$$f(x_i^\alpha, x_j^\beta) = \frac{2\sum_{t=1}^{N} x_{i,t}^\alpha x_{j,t}^\beta}{\sum_{t=1}^{N}((x_{i,t}^\alpha)^2 + (x_{j,t}^\beta)^2)}$$

Experiments

Speech	Proposed			Pitch[17]			Formant[18]		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
No attack	**97.1%**	**100.0%**	**96.8%**	78.1%	77.9%	78.5%	70.0%	99.0%	67.0%
Add noise 10dB	**93.6%**	**98.7%**	**87.2%**	75.1%	73.0%	79.9%	68.5%	92.1%	64.2%
Add noise 30dB	**98.2%**	**99.6%**	**96.4%**	79.2%	78.6%	80.1%	90.1%	92.2%	87.7%
Resample 8k	67.7%	90.5%	85.4%	79.5%	78.4%	81.6%	80.1%	90.2%	65.7%
Resample 48k	76.4%	89.3%	75.9%	78.3%	80.9%	74.1%	74.4%	93.2%	75.8%

Table 1 Performance comparison of proposed method with [17] and [18]
Bold values indicate best performance.

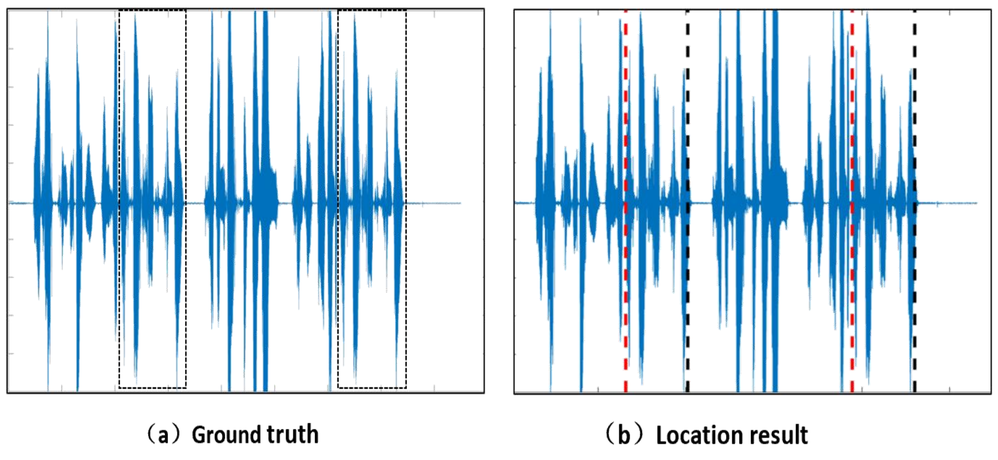


Table 2 The positioning visualization of the copy-move forgery audio.