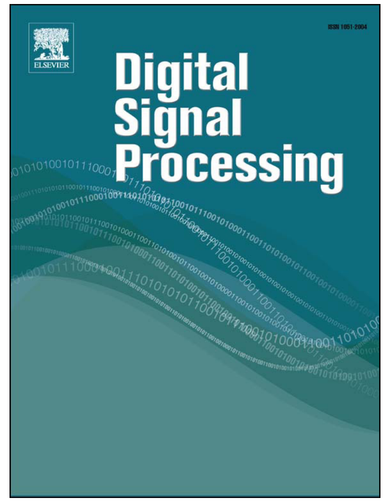
Accepted Manuscript



Source cell phone verification from speech recordings using sparse representation

Ling Zou, Qianhua He, Junfeng Wu

PII: S1051-2004(16)30186-5
DOI: http://dx.doi.org/10.1016/j.dsp.2016.10.017
Reference: YDSPR 2041

To appear in: *Digital Signal Processing*

Please cite this article in press as: L. Zou et al., Source cell phone verification from speech recordings using sparse representation, *Digit. Signal Process.* (2016), http://dx.doi.org/10.1016/j.dsp.2016.10.017

Source cell phone verification from speech recordings using sparse representation

Ling Zou, Qianhua He*, Junfeng Wu

School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China

ABSTRACT

Source recording device recognition is an important emerging research field in digital media forensics. The literature has mainly focused on the source recording device identification problem, whereas few studies have focused on the source recording device verification problem. Sparse representation based classification methods have shown promise for many applications. This paper proposes a source cell phone verification scheme based on sparse representation. It can be further divided into three schemes which utilize exemplar dictionary, unsupervised learned dictionary and supervised learned dictionary respectively. Specifically, the discriminative dictionary learned by supervised learning algorithm, which consider the representational and discriminative power simultaneously compared to the unsupervised learning algorithm, is utilized to further improve the performances of verification systems based on sparse representation. Gaussian supervectors (GSVs) based on MFCCs, which have shown to be effective in capturing the intrinsic characteristics of recording devices, are utilized for constructing and learning dictionary. SCUTPHONE, which is a corpus of speech recordings from 15 cell phones, is presented. Evaluation experiments are conducted on three corpora of speech recordings from cell phones and demonstrate the effectiveness of the proposed methods for cell phone verification. In addition, the influences of number of target examples in the exemplar dictionary and size of the unsupervised learned dictionary on source cell phone verification performance are also analyzed.

Keywords: Digital audio forensics, Source cell phone verification, Gaussian supervectors, Sparse representation, Supervised dictionary learning.

1. Introduction

Today, with the wide availability of various easy-to-use and portable recording devices, lots of forensic evidences in the form of digital speech recordings are presented in courts. The availability of low-cost, powerful digital multimedia editing software has made manipulating and editing digital speech recording easier. Hence the authenticity and integrity of digital speech recordings presented as forensic clues or evidences in courts or other law enforcement agencies must be established. This is a challenging task especially when no extrinsically imposed security measures such as digital watermarks are embedded into the recording.

Over the last decade, techniques for detecting traces of tampering and forgeries in digital audio file have been developed utilizing various forensic fingerprints like the *electric network frequency* (ENF) [1-7], acoustic reverberation [8-14], background noise [14, 15], compression artifacts [16-18], etc. Specifically, recording device-specific fingerprints or signatures, which are left by the acquisition device during the recording procedure, are also utilized for digital audio forensic applications (e.g., identifying the source device of the speech recordings).

Source acquisition device recognition has received considerable attention during the past decade in the forensic community, especially in digital image forensics. There exists rich literature on imaging device (e.g., digital camera, scanner, cell phone, etc.) identification from the image file [19]. For example, the well-known photo-response nonuniformity (PRNU) fingerprints [20-24], which make use of the noise introduced by imperfections in the sensor of digital camera, have been proven to be effective in source camera recognition etc. There also exists many works which focus on source recording device recognition from the recording. The majority of these works utilize a set of representative features and powerful tools from machine learning and pattern recognition such as support vector machine (SVM) [25]. However, compared to the considerable works on extraction of intrinsic fingerprint of imaging device from images and applications to digital image forensics, extracting the intrinsic fingerprint of recording device from audio recordings and the applications to digital audio forensics are still relatively less studied.

* Corresponding author.
E-mail addresses: zou.ling@mail.scut.edu.cn (L. Zou), eeqhhe@scut.edu.cn (Q. H. He).