CELL PHONE VERIFICATION FROM SPEECH RECORDINGS USING SPARSE REPRESENTATION

Ling Zou, Qianhua He, Xiaohui Feng

School of Electronic and Information Engineering
South China University of Technology, Guangzhou 510640
{eexhfeng, eeqhhe}@scut.edu.cn zou.ling@mail.scut.edu.cn

ABSTRACT

Source recording device recognition is an important emerging research field of digital media forensic. Most of the prior literature focus on the recording device identification problem. In this study we propose a source cell phone verification scheme based on sparse representation. We employed Gaussian supervectors (GSVs) based on Mel-frequency cepstral coefficients (MFCCs) extracted from the speech recordings to characterize the intrinsic fingerprint of the cell phone. For the sparse representation, both exemplar based dictionary and dictionary learned by K-SVD algorithm were examined to this problem. Evaluation experiments were conducted on a corpus consists of speech recording recorded by 14 cell phones. The achieved equal error rate (EER) demonstrated the feasibility of the proposed scheme.

Index Terms— Digital audio forensic, Source cell phone verification, Gaussian supervector, Sparse representation.

1. INTRODUCTION

Reliable recognition of the source device used to acquire a particular speech recording would prove useful in the court for establishing the origin of speech recordings presented as evidence [1, 2]. Source recording device recognition is motivated by the hypothesis that recording device leave behind its intrinsic fingerprint traces in the speech recording [3].

Over the past several years, source recording device recognition has received more attention. Most existing literature related to this problem focus on microphone identification [4-9], telephone handset identification [3, 10-15] and cell phone identification [15-20]. In particular, source cell phone recognition from speech recordings was first pointed out by Hanilçi *et al.* [17]. The authors proposed to identify 14 cell phones from speech recordings using Mel-frequency cepstral coefficients (MFCCs) and support vector machine (SVM). In our recent work [19], a cell phone identification system based on the Gaussian mixture model-universal background model (GMM-UBM) and MFCCs was presented. Kotropoulos *et al.* presented several studies on the telephone handset identification [12-15] and more recently also on cell phones identification [15, 16].

However, most existing studies focus on the source recording device identification (or classification) problem, more specifically, the close-set source recording device identification problem. To our best knowledge, few studies have focused on the source recording device verification problem except that, in a very recent work, a cell phone detection experiment was conducted in [18] based on SVM. Given a speech recording and a claimed recording device, e.g., cell phone, the task of recording device verification is to determine if the speech recording was acquired by the claimed device. This problem is full of significance in the forensic context. Take cell phone as an example, we know that cell phone has become an essential part of our daily life and almost every phone is equipped with the function of voice recording. In the forensic context, the wide availability of cell phones will signify that there will be increasing more recording evidences in the form of cell phone recordings brought to the courts or other law enforcement agencies. Imagine that a person submits a speech recording to the court as evidence and claims that this recording was recorded using his cell phone. Obviously, source cell phone verification from speech recording will aid in justifying the authenticity of this evidence.

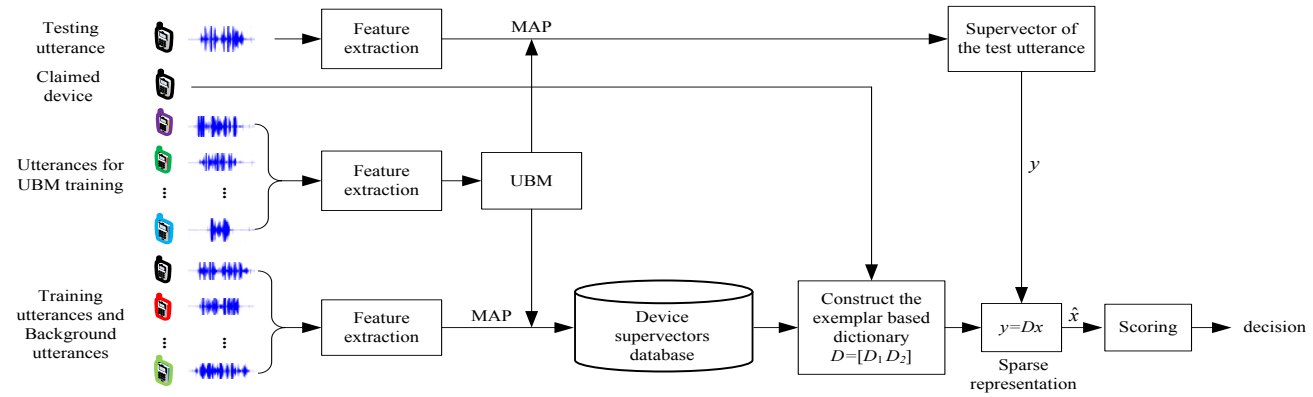
Motivated by the forensic significance of source cell phone verification, partially inspired by the success of sparse representation based speaker verification systems [21-23], in this study, we propose the use of sparse representation for source cell phone verification task. Both the exemplar based dictionary and the learned dictionary are examined. Gaussian supervectors (GSVs) computed from speech recordings have shown to successively represent the intrinsic fingerprint of the recording device [3]. Thus, GSVs are utilized here to construct (or learn) the dictionary. The effects of GMM mean supervector and GMM mean shift supervector to this problem are compared. The performance of the two kinds of dictionaries and various scoring metrics are evaluated and compared on a 14 cell phones verification task.

The remainder of this paper is organized as follows: Section 2 describes the methods of this study. Section 3 details the experimental set up in this paper. The experimental results and discussion are presented in Section 4. Finally, conclusions and future works are summarized in Section 5.

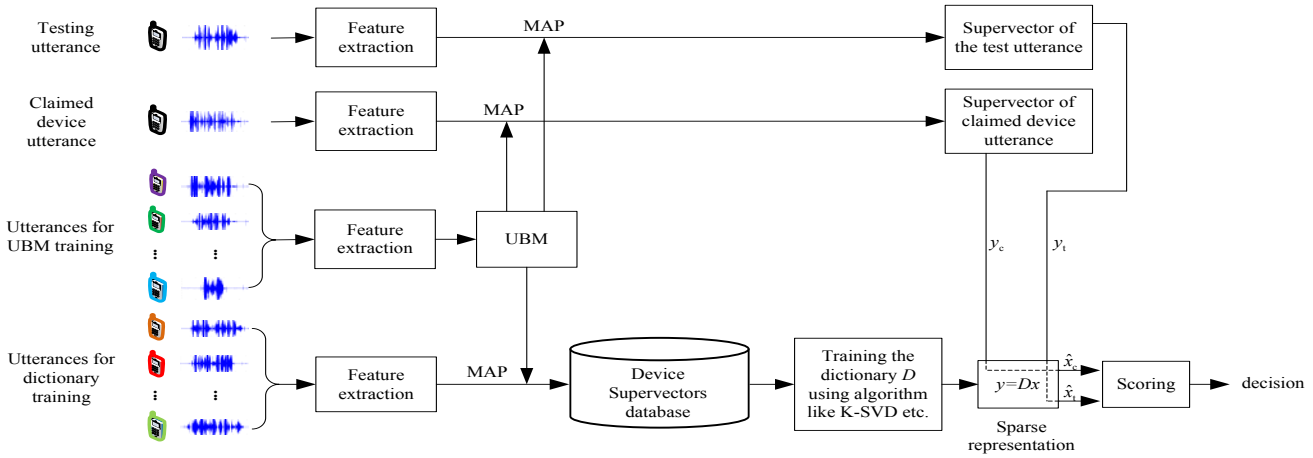
2. METHODS

2.1. Gaussian supervector

The intrinsic fingerprint of the recording device can be effectively represented by the GSVs computed from speech recordings acquired with the device [3]. In this way, given the training data $X = \{x_t\}_{t=1}^{T}$ from an utterance and a diagonal covariance UBM with K mixtures given by $\lambda_{UBM} = \{\omega_i, \mu_i, \Sigma_i\}_{i=1}^{K}$, the means adapted only GMM is updated from The UBM by *maximum a posteriori* (MAP)



(a)



(b)

Fig. 1. Block diagram of source cell phone verification system based on sparse representation when (a) exemplar based dictionary is utilized，or (b) learned dictionary is utilized.

[24, 25]. Suppose that $\lambda_a = \{\omega_i, \mu_i^a, \Sigma_i\}_{i=1}^K$ and $\lambda_b = \{\omega_i, \mu_i^b, \Sigma_i\}_{i=1}^K$ are the means adapted GMMs for two utterances. The Kullback-Leibler (KL) divergence kernel is then defined as the corresponding inner product of the GMM mean supervector which is a concatenation of the weighted GMM mean vectors (For the i^{th} mean vector, the weight is $\sqrt{w_i}\Sigma_i^{-(1/2)}$) [26]:

$$K(\lambda_a, \lambda_b) = \sum_{i=1}^{K} (\sqrt{w_i}\Sigma_i^{-(1/2)}\mu_i^a)^T (\sqrt{w_i}\Sigma_i^{-(1/2)}\mu_i^b). \quad (1)$$

The GMM mean shift supervector [27] for an utterance is defined as

$$y = s - m \quad (2)$$

where s is the GMM mean supervector and m is the device independent UBM mean supervector.

2.2. Source cell phone verification based on exemplar dictionary

In a verification test, for a claimed device, select N_1 object examples from claimed device and N_2 non-target background examples ($N_1 \ll N_2$) as in Figure 1(a). Thus, the exemplar based dictionary [21, 22, 28] is defined by concatenating the examples as

$$D = [D_1 \ D_2] = [a_{11}, a_{12}, \cdots, a_{1N_1}, a_{21}, a_{22}, \cdots, a_{2N_2}] \in \mathbb{R}^{M \times N} \quad (3)$$

here $D_1 = [a_{11}, a_{12}, \cdots, a_{1N_1}] \in \mathbb{R}^{M \times N_1}$, $D_2 = [a_{21}, a_{22}, \cdots, a_{2N_2}] \in \mathbb{R}^{M \times N_2}$, and $N = N_1 + N_2$. Note that $M \ll N$ should be satisfied for constructing an overcomplete dictionary [28]. The atoms in D are normalized to unit ℓ_2 – norm. In our study, each example of the dictionary is a M-dimensional GMM mean (shift) supervector.

For any test vector $y \in \mathbb{R}^M$ with unit ℓ_2 – norm, y can be linearly represented in terms of D as

$$y = Dx = [D_1 \ D_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \quad (4)$$

If y is a valid test, it must lie in D_1, thus $x_1 = [\alpha_1, \alpha_2, \cdots, \alpha_{N_1}]^T$ and $x_2 = [0, \cdots, 0]^T$. Clearly, this representation is sparse. To seek the sparse solution to (4), solving the following optimization problem [28, 29]:

$$\hat{x} = \arg\min \|x\|_1 \ \text{ subject to } \ \|y - Dx\|_2 \le \varepsilon \quad (5)$$

where $\varepsilon > 0$ is a pre-set noise level value. A variant of the problem is also well-known as the unconstrained basis pursuit denoising (BPDN) problem with a scalar weight λ [29]:

$$F(x) \doteq \min_x \tfrac{1}{2}\|y - Dx\|_2^2 + \lambda\|x\|_1. \quad (6)$$

Once the sparse representation $\hat{x}$ are obtained by solving (6), to determine the verification score, we considered the ℓ_1 – norm ratio as scoring metric [22, 23] defined as

$$\|\delta_1(\hat{x})\|_1 / \|\hat{x}\|_1 \quad (7)$$

where $\delta_1(\hat{x})$ denote the entries in $\hat{x}$ which correspond to the claimed device examples (i.e., $\hat{x}_1$). An alternative scoring metric, referred to as the ℓ_2 – norm residual ratio [27], is defined as

$$\|y - D\delta_2(\hat{x})\|_2 / \|y - D\delta_1(\hat{x})\|_2. \quad (8)$$

In addition to model in (4), a more general sparse representation model allow for a error vector [28, 29]. In such condition, the model should be modified as

$$y = Dx + e = [D, I]\begin{bmatrix} x \\ e \end{bmatrix} \doteq Bw \quad (9)$$

where $e \in \mathbb{R}^M$ is an error vector, $B = [D, I] \in \mathbb{R}^{M\times(M+N)}$ and $w = [x; e]$. Similar to x in (6), w can be estimated by solving

$$F(w) \doteq \min_w \tfrac{1}{2}\|y - Bw\|_2^2 + \lambda\|w\|_1. \quad (10)$$

Once the sparse representation $\hat{w} = [\hat{x}; \hat{e}]$ are determined, the aforementioned scoring metrics, ℓ_1 – norm ratio and the ℓ_2 – norm residual ratio, should be redefined as

$$\|\delta_1(\hat{x})\|_1 / \|\hat{x}\|_1 \quad (11)$$

and

$$\|y - \hat{e} - D\delta_2(\hat{x})\|_2 / \|y - \hat{e} - D\delta_1(\hat{x})\|_2 \quad (12)$$

respectively.

2.3 Source cell phone verification based on learned dictionary

Compared to the exemplar based dictionary, the more commonly used dictionary is determined by learning dictionary on a training corpus using a certain algorithm. We considering replacing the exemplar based dictionary with a dictionary $D \in \mathbb{R}^{M\times N}$ learned using K-SVD [30]. The K-SVD algorithm searches for the best possible dictionary for the sparse representation of the training vectors set $Y = \{y_i\}_{i=1}^K$ by solving

$$\min_{D,X}\left\{\|Y - DX\|_2^2\right\} \text{ subject to } \forall i \quad \|\mathrm{x}_i\|_0 \le T_0 \quad (13)$$

where D is the dictionary to be learned, X is the corresponding sparse representation to Y and T_0 is the sparsity constraint. Once the dictionary D is determined, the test vector y can be sparsely represented in terms of D using the orthogonal matching pursuit (OMP) algorithm [31] as

$$\hat{x}_0 = \arg\min\|x\|_0 \text{ subject to } \|y - Dx\|_2 \le \varepsilon \quad (14)$$

where $\hat{x}_0$ is the sparse representation for the test vector. The sparse representation can also be obtained using the basis pursuit (BP) approach [29] by solving:

$$\hat{x}_1 = \arg\min\|x\|_1 \text{ subject to } \|y - Dx\|_2 \le \varepsilon. \quad (15)$$

As there are no class labels associated with the learned dictionary, the scoring metric for the exemplar based dictionary is no longer applicable here. To resolve this problem, we utilized the scoring method as in [23]. The score are determined by comparing the similarity of the sparse representation of the test vector with the sparse representation of the vector of the claimed device using the cosine kernel metric. Then the obtained score will be compared with a threshold for verification purpose as

$$\frac{\langle \hat{x}_c, \hat{x}_t \rangle}{\|\hat{x}_c\|\|\hat{x}_t\|} \gtrless \theta \quad (16)$$

where $\hat{x}_t$ and $\hat{x}_c$ represent the sparse representations of the test supervector y_t and the supervector y_c of the claimed device in terms of the learned dictionary D respectively. We proposed an alternative scoring method which computes the correlation between the two sparse representation as

$$\frac{\langle (\hat{x}_c - \bar{\hat{x}}_c), (\hat{x}_t - \bar{\hat{x}}_t) \rangle}{\|\hat{x}_c - \bar{\hat{x}}_c\|\|\hat{x}_t - \bar{\hat{x}}_t\|} \gtrless \theta. \quad (17)$$

The diagram for source cell phone verification system based on the learned dictionary is illustrated in Fig. 1(b).

Table 1. Brands and models of the 14 cell phones ($\times 2$ denotes two cell phones of the same brand and model).

BRAND	MODEL
SAMSUNG	SAMSUNG E250 ($\times 2$), D900
NOKIA	NOKIA 2730, 6500, 3600 ($\times 2$), 6670
MOTOROLA	MOTOROLA Q
SONY	SONY W880 ($\times 2$), K750I
LG	LG KE970
HP	HP IPAQ514

Table 2. Number of trials for one test.

Experimental Corpus	cell phones	Test utterances	True trials	False trials
LIVE	14	1400	1400	18200
TIMIT	14	1680	1680	21840

3. EXPERIMENTAL SETUP

We evaluated the source cell phone verification system on a corpus consists of speech recordings recorded by 14 cell phones [17-19]. The detail of the cell phones is presented in Table 1. The dataset was collected by two methods and each resulted in a subset. The first subset was constructed by playing a subset (24 speakers are selected) of the TIMIT corpus through all the 14 cell phones in a silent environment using a loudspeaker. (10 sentences for each speaker, approximately 3 seconds per sentence). Thus there are 240 speech recordings for each cell phone. This corpus is referred to as TIMIT subset hereafter. The second subset was collected by a same person speaking into the 14 cell phones a passage in the same room. The length of each recording is approximately of 10 minutes. Then each utterance was evenly divided into 2 utterances (one for training and another for testing). Both the training utterance and the testing utterance were further divided into 100 short utterances each with the length of approximately 3 seconds. This corpus is referred to as LIVE subset hereafter.

When we carried out experiment on LIVE subset, half of TIMIT subset were utilized for training the UBM and vice versa. For the experimental subset, half were randomly selected for constructing (or learning) dictionary and the remaining half were utilized as test trials. Specifically, the number of true trials and fal-

Table 3. EERs for exemplar based dictionary using different GMM supervector, scoring metrics and sparse representation models on two corpus.

SYSTEM	LIVE	TIMIT
Mean shift sv + ℓ_1 – norm ratio (7)	3.5%	5.06%
Mean shift sv + ℓ_2 – norm residual (8)	2.64%	2.80%
Mean shift sv + ℓ_1 – norm ratio (11)	2.43%	3.04%
Mean shift sv + ℓ_2 – norm residual (12)	3.21%	3.39%
Mean sv + ℓ_1 – norm ratio (7)	3.64%	5.18%
Mean sv + ℓ_2 – norm residual (8)	**2.36%**	**2.08%**
Mean sv + ℓ_1 – norm ratio (11)	2.79%	3.69%
Mean sv + ℓ_2 – norm residual (12)	3.07%	4.11%

se trials in a run of test are listed in Table 2, this configuration is identical to that used in [18]. For each speech recording, the whole utterance, including speech segments and non-speech segments, was segmented into frames by a 30 ms Hamming window at 15 ms frame rate. Then 12 MFCCs were computed using 27 triangular filters with c0 excluded. The MFCCs were concatenated with the energy feature as in [19] and resulted in a 13-dimensional feature vector. The number of mixture components in UBM was set to 32, therefore the dimensionality of the GSVs was of 416. For comparing the two kinds of dictionary to this problem, the sizes for the exemplar based dictionary and the learned dictionary were set to be identical (416×1260 for this study, this set also guaranteed that the dictionary is redundant and overcomplete). In particular, for the exemplar based dictionary, we set $N_1 = 90$ and $N_2 = 1170$. Once the test scores were obtained, the equal error rate (EER) was computed as the metric for evaluation.

4. RESULTS AND DISCUSSION

Table 3 shows the EERs when utilizing the exemplar based dictionary. We found that the EERs for the GMM mean supervector and the GMM mean shift supervector are close. We found that when utilizing the ℓ_1 – norm ratio as the scoring metric, the GMM mean shift supervector outperforms the GMM mean supervector slightly in terms of the achieved EERs while the opposite results are observed when utilizing the ℓ_2 – norm residual ratio metric. Table 3 also shows the EERs for the two models defined by equation (4) and (9) respectively. It can be observed that, independent of which type of supervector is utilized, using equation (11) as the ℓ_1 – norm ratio metric outperforms equation (7), however, utilizing equation (8) as the ℓ_2 – norm residual ratio metric outperforms equation (12). Comparing equation (8) and (12), the reason why using equation (8) outperforms using (12) may be due to that the error vector $\hat{e}$ subtracted from y in (12) contains certain information related to the recording device (cell phone here). However, the reason why utilizing equation (11) outperforms (7) is worth further studying. We also found that, in our experiments utilizing the exemplar based dictionary, the lowest EER was achieved when the GMM mean supervector and the ℓ_2 – norm residual ratio metric corresponds to equation (8) are utilized.

Table 4 shows the EERs when utilizing the learned dictionary under various sparse representation methods and scoring metrics. Gaussian mean supervectors are utilized here. We found that, First, the proposed correlation metric outperforms the cosine metric in terms of EER. Second, sparse representation based on BP outperforms OMP for this study. Third, comparing Table 3 and Table 4, we found that, the best EER achieved on the exemplar based dictionary outperforms the K-SVD learned dictionary, the reason might be due to that the size of the training vectors set for learning the dictionary is limited and small in our experiment. In addition, it can be observed from Table 3 and Table 4 that the best EER achieved by the sparse representation based source cell phone verification scheme outperforms the SVM based source cell phone detection system in [18] on the same feature set (GMM mean supervectors with the dimensionality of 416 here).

Table 4. EERs for dictionary learned on GMM mean supervectors using K-SVD when two sparse representation methods (OMP and BP) and two scoring metrics (cosine and correlation) are utilized on two corpus and the EERs for SVM based verification system.

SYSTEM	LIVE	TIMIT
OMP + Cosine metric (16)	5.27%	7.08%
OMP + Correlation metric (17)	5.12%	6.61%
BP + Cosine metric (16)	2.61%	4.32%
BP + Correlation metric (17)	2.57%	4.17%
SVM based system in [18]	4.36%	4.11%

This is a preliminary study of sparse representation based source cell phone verification. We believe the potential of the proposed scheme. It should be noted that this study focus only on the cell phone verification problem, however, it is possible that the proposed scheme could be extended to other types of source recording device verification problem. The size of the experimental corpus is limited and the brand and model of the cell phones are somewhat outdated, source cell phone verification experiments on larger dataset including more current popular brand and model of cell phones deserve attention for future work.

5. CONCLUSIONS

In this paper, for addressing the problem of source cell phone verification from speech recordings, a cell phone verification scheme based on sparse representation is presented. We find that both exemplar based dictionary and dictionary learned by K-SVD are effective to this problem and exemplar based dictionary outperforms dictionary learned by K-SVD in our experiments. The correlation scoring method proposed to be used outperforms the cosine scoring method when utilizing the learned dictionary. The best EER for this study is achieved when exemplar based dictionary is utilized with the ℓ_2 – norm residual ratio as scoring metric and GMM mean supervector as dictionary atoms. To sum up, we propose an alternative cell phone verification scheme in this study. Future work include extending the experimental corpus for further evaluating the reliability of the proposed scheme and applying the method to other source device verification problem such as microphone verification etc.

6. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (No. 61401161 and No. 61301300).

7. REFERENCES

[1] H. Malik and H. Mahmood, "Acoustic environment identification using unsupervised learning," *Security Informatics*, vol. 3, no. 1, pp. 1–17, 2014.

[2] H. Zhao and H. Malik, "Audio recording location identification using acoustic environment signature," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1746–1759, 2013.

[3] D. Garcia-Romero and C. Espy-Wilson, "Automatic acquisition device identification from speech recordings," in *Proc. ICASSP*, 2010, pp. 1806–1809.

[4] C. Kraetzer, A. Oermann, J. Dittmann, and A. Lang, "Digital Audio Forensics: A First Practical Evaluation on Microphone and Environment Classification," in *Proc. 9th Workshop on Multimedia and Security*, Dallas, TX, USA, 2007, pp. 63–74.

[5] C. Kraetzer, M. Schott, and J. Dittmann, "Unweighted fusion in microphone forensics using a decision tree and linear logistic regression models," in *Proc. 11th ACM Multimedia and Security Workshop*, 2009, pp. 49–56.

[6] C. Kraetzer, K. Qian, M. Schott, and J. Dittmann, "A context model for microphone forensics and its application in evaluations," *Media Watermarking, Security, and Forensics III*, vol. 7880, 2011.

[7] R. Buchholz, C. Kraetzer, and J. Dittmann, "Microphone classification using Fourier coefficients," in *Lecture Notes in Comput. Sci*. Berlin/Heidelberg, Germany: Springer, 2010, vol. 5806/2009, pp. 235–246.

[8] H. Malik and J. Miller, "Microphone identification using higher-order statistics," in *Proc. AES 46th Conf. Audio Forensics 2012*, Denver, CO, USA, 2012.

[9] O. Eskidere, "Source microphone identification from speech recordings based on a Gaussian mixture model," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 22, no. 3, pp. 754–767, 2014.

[10] D. A. Reynolds, "HTIMIT and LLHDB: speech corpora for the study of handset transducer effects," in *Proc. ICASSP*, Munich, Germany, 1997, vol. 2, pp. 1535–1538.

[11] D. Garcia-Romero and C. Espy-Wilson, "Speech forensics: Automatic acquisition device identification," *J. Acoust. Soc. Am.*, vol. 127, no. 3, pp. 2044–2044, 2010.

[12] Y. Panagakis and C. Kotropoulos, "Automatic telephone handset identification by sparse representation of random spectral features," in *Proc. 14th ACM Multimedia and Security Workshop*, Coventry, U.K., 2012, pp. 91–96.

[13] Y. Panagakis and C. Kotropoulos, "Telephone handset identification by feature selection and sparse representations" in *Proc. 2012 IEEE Int. Workshop Information Forensics and Security*, Tenerife, Spain, 2012, pp. 73–78.

[14] C. Kotropoulos, "Telephone handset identification using sparse representations of spectral feature sketches," in *Proc. First Int. Workshop Biometrics and Forensics*, Lisbon, Portugal, 2013.

[15] C. Kotropoulos, "Source phone identification using sketches of features," *Biometrics, IET*, vol. 3, no. 2, pp. 75–83, 2014.

[16] C. Kotropoulos and S. Samaras, "Mobile phone identification using recorded speech signals," in *Digital Signal Processing (DSP), 2014 19th International Conference on*, 2014, pp. 586–591.

[17] C. Hanilci, F. Ertas, T. Ertas, and O. Eskidere, "Recognition of brand and models of cell-phones from recorded speech signals," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 625–634, 2012.

[18] C. Hanilçi and T. Kinnunen, "Source cell-phone recognition from recorded speech using non-speech segments," *Digital Signal Processing*, vol. 35, pp. 75–85, 2014.

[19] L. Zou, J. Yang, and T. Huang, "Automatic cell phone recognition from speech recordings," in *Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit & International Conference on*, 2014, pp. 621–625.

[20] M. Jahanirad, A. W. A. Wahab, N. B. Anuar, M. Y. I. Idris, and M. N. Ayub, "Blind source mobile device identification based on recorded call," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 320–331, 2014.

[21] J. Kua, E. Ambikairajah, J. Epps, and R.Togneri, "Speaker verification using sparse representation classification," in *Proc. ICASSP*, May 2011, pp. 4548–4551.

[22] M. Li, X. Zhang, Y. Yan, and S. Narayanan, "Speaker verification using sparse representations on total variability i-vectors," in *Proc. Interspeech*, Aug 2011, pp. 2729–2732.

[23] B. C. Haris, and R. Sinha, "Sparse representation over learned and discriminatively learned dictionaries for speaker verification," in *Proc. ICASSP*, Mar. 2012, pp. 4785–4788.

[24] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models," *Digital Signal Processing*, vol. 10, pp. 19–41, 2000.

[25] G. Liu and J. H. L. Hansen, "An Investigation into Back-end Advancements for Speaker Recognition in Multi-Session and Noisy Enrollment Scenarios," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 22, no. 12, pp. 1978–1992, 2014.

[26] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 308–311, 2006.

[27] M. Li and S. Narayanan, " Robust talking face video verification using joint factor analysis and sparse representation on gmm mean shifted supervectors," in *Proc. ICASSP*, May 2011, pp. 4835–4838.

[28] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 210–227, Feb. 2009.

[29] A. Yang, Z. Zhou, A. Ganesh, S. Sastry, and Y. Ma, "Fast L1-minimization algorithms for robust face recognition," *IEEE Trans. Image Process.*, vol. 22 no. 8, pp. 3234–3246, 2013.

[30] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.

[31] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *Opt. Eng.*, vol. 33, no. 7, pp. 2183–91, 1994.