

# 基于稀疏表示的鲁棒性说话人识别系统

于云,周伟栋

(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

**摘要:**基于稀疏表示的说话人识别方法在无噪的环境下已经达到了理想的效果,然而在背景噪声下,此方法的识别性能大幅度下降。为了提高系统的鲁棒性,提出了一种新型的基于稀疏表示的鲁棒性说话人识别系统模型。此系统结合多状态训练和语音增强谱减法,在训练阶段和测试阶段同时利用语音增强技术,然后对增强后的语音进行多状态训练,以便提高训练特征数据集和测试特征数据集之间的匹配度。实验分析和结果表明,所提出的新型模型在所研究的白噪声和有色噪声下达到了很好的抗噪性能,具有很强的鲁棒性。

**关键词:**说话人识别;稀疏表示;多状态训练;谱减法

中图分类号:TN912.3

文献标识码:A

文章编号:1673-629X(2015)12-0041-06

doi:10.3969/j.issn.1673-629X.2015.12.010

## Robust Speaker Recognition System Based on Sparse Representation

YU Yun,ZHOU Wei-dong

(College of Communication and Information Engineering,Nanjing University of Posts and Telecommunications,  
Nanjing 210003,China)

**Abstract:**Robust speaker recognition method based on sparse representation in the absence of noise has reached ideal performance. However,speaker recognition based on sparse representation doesn't perform well where background noise exists. To improve the robustness of this system,describe a new robust speaker recognition system based on sparse representation. The system combines multi-condition training and spectrum subtraction,which is thought to be a preprocessing block not only for the testing stage,but also for the training stage. Then propose to make multi-condition training where various sets of features are extracted,so as to improve the matched degree between training data and testing data. Experimental analysis and results show that the proposed new model under white and colored noises can get the great anti-noise performance,and obviously improve the robustness of the speaker recognition under background noisy environments.

**Key words:**speaker recognition;sparse representation;multi-condition training;spectrum subtraction

## 0 引言

在过去的几十年里,说话人识别技术在无噪的环境下已经发展得相当成熟,但在有背景噪声的环境下识别性能急速下降,所以减小背景噪声的影响成了目前研究的一大热点。由于背景噪声的存在,使得训练特征数据集与测试特征数据集之间的匹配度降低,从而导致识别性能下降<sup>[1]</sup>,这将严重影响说话人识别技术的实际应用。为了解决这个问题,学者们提出了许多方法,比如在识别系统的前端利用语音增强技术:谱减法或者维纳滤波器<sup>[2]</sup>,其他方法有提取鲁棒性特征参数、进行模型自适应等<sup>[3]</sup>。Ji Ming等指出通过训练

多种噪声环境下的语音可以减小训练特征数据集与测试特征数据集之间的失配度,即多状态训练技术<sup>[4-5]</sup>,主要过程是采集多种噪声环境下的语音作为训练数据,为每个说话人建立了多个训练模型,测试时,测试语音与训练模型进行匹配。然而多状态训练也有不足之处:随着添加的噪声种类越来越多或者噪声强度越来越大,多状态训练数据集更加远离测试数据集,导致特征数据集分散,说话人之间产生干扰。因此为了弥补这个缺陷,Kim等提出了一种结合多状态训练和语音增强的方法<sup>[6]</sup>,这样可以解决多状态训练带来的问题。

收稿日期:2015-03-25

修回日期:2015-07-03

网络出版时间:2015-11-19

基金项目:国家自然科学基金资助项目(61271335);国家“973”重点基础研究发展计划项目(2011CB302303);江苏省自然科学基金项目(BK20140891)

作者简介:于云(1990-),女,硕士研究生,研究方向为说话人识别、语音信号处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20151119.1111.060.html>

近几年来,学者们对压缩感知和稀疏表示理论进行了深入研究,为信号处理技术的发展指引了新的方向。稀疏表示理论已经成功地应用于人脸识别中<sup>[7]</sup>,为说话人识别的发展奠定了基础。Imran 等为说话人识别系统提出了一种新型的稀疏表示分类方法<sup>[8]</sup>,即为每个说话人训练一个字典模型,这些字典组合在一起形成系统的大字典,在识别过程中,测试语音被稀疏地分解在大字典上。Xu 等为说话人识别技术提出了一种基于稀疏表示子空间分类的方法<sup>[9-10]</sup>,即为每个人训练一个字典作为字典模型。在纯净语音情况下,这些方法都能得到很好的识别效果,但是在背景噪声存在的环境中,识别性能很不理想。

文中研究的是基于稀疏表示分类方法的说话人识别,解决背景噪声对说话人识别性能的影响,进而提高系统的鲁棒性。文中将多状态训练和谱减法相结合,并且应用到稀疏表示分类系统中。此方法首先在多种噪声语音集训练之前采用谱减法,然后对增强后的语音进行训练,得到每个人的字典模型。识别时,对经过增强后的语音提取特征参数,将测试语音特征数据与训练语音特征数据集进行匹配,根据重构误差对说话人身份进行判决。

## 1 压缩感知理论与稀疏表示思想

近年来,压缩感知理论越来越受到学者们的关注,它是一种新型的数据采集方法,打破了原有的奈奎斯特采样定理。传统的数据采集定理采样率相对较高,所以为了降低采样率,Donoho 等提出了压缩感知理论<sup>[11-12]</sup>。稀疏重构是压缩感知另一种表达形式,也就是文中所用的稀疏表示理论。稀疏表示逐渐成为了国

内外研究的重点,在信号去噪、图像修复、人脸识别中得到了很好的利用价值。它的基本思想是探索信号在字典上的最佳线性表示,通过残余误差进行识别分类。

压缩感知指出如果信号  $x \in R^N$  在某个域中是稀疏的,那么此信号可以表示为  $x = \psi\alpha$ 。其中,  $\psi \in R^{N \times D}$  ( $N < D$ ) 是稀疏基;  $\alpha \in R^{D \times 1}$  是信号的稀疏向量。接下来用观测矩阵  $\varphi \in R^{M \times N}$  对信号进行观测得到观测序列  $y \in R^M$ , 公式如下:

$$y = \varphi x = \varphi \psi \alpha \quad (1)$$

利用线性规划凸优化的方法可以实现信号重构,得到式(2):

$$\min \|\alpha\|_1 \quad \text{s.t.} \quad y = \varphi \psi \alpha \quad (2)$$

文中采用的稀疏表示理论将信号表示为:

$$y = D\alpha \quad (3)$$

其中,  $y$  为测试语音的特征参数;  $D$  为用 K-SVD 方法训练的过完备字典;  $\alpha$  为稀疏系数。

这样测试语音可以由训练语音线性表示,利用式(2)求得重构信号,根据重构误差进行说话人身份的判决。

## 2 系统模型

多状态训练很早就被应用于说话人识别中,具体过程将不同噪声添加到纯净语音上,形成不同的带噪语音,将纯净语音和不同的带噪语音组合在一起作为训练语音集。考虑到多状态训练带来的问题,将多状态训练和谱减法结合在一起,提出了一种新型的系统模型。

图 1 是所提的基于稀疏表示的鲁棒性说话人识别系统框图。

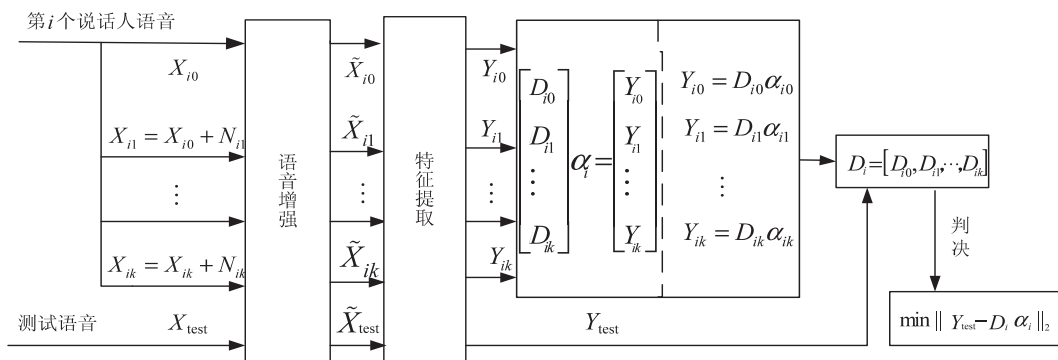


图 1 新型的鲁棒性说话人识别系统模型

在训练过程中,首先为每个说话人收集受到不同信噪比噪声影响的语音,经过谱减法系统之后得到已增强的语音,对增强后的语音进行特征提取,然后利用 K-SVD 算法为每个说话人训练出用于识别的过完备字典<sup>[13-14]</sup>,即说话人系统的子空间模型。这里要注意的是:

(1) 语音增强不仅降低了噪声,而且也改变了原

始纯净语音特性,所以需要在纯净语音之前也利用语音增强技术;

(2) 提出两种训练说话人字典结构的方法,即联合字典训练和独立字典训练。

为了保持训练与测试的一致,在训练和测试过程中同时利用语音增强方法。识别时,提取增强后的测试语音的特征参数,分别投影到每个人的字典上,根据

重构误差进行身份的判决。

### 3 字典训练模型和识别过程

#### 3.1 特征提取

对于说话人识别系统来说,特征参数有着不可替代的作用。目前普遍使用的特征参数有线性预测倒谱系数(LPCC)、Mel 倒谱系数(MFCC)等。MFCC 特征参数考虑了人的听觉特性,它是基于 Mel 三角滤波器组提取出的倒谱系数,能够有效地提高识别系统的抗噪性能。文中针对 Mel 倒谱系数做了一些研究,在 Mel 倒谱特征参数提取的过程中,最后一步 DCT 主要目的是聚集特征和降低特征参数的维度,许多学者们指出 DCT 这一步可以舍去<sup>[15]</sup>。Xu 等通过实验论证了舍去 DCT 步骤的 Mel log 能量特征可以作为特征参数,简化了特征提取流程,并且得到了理想的识别效果。Mel log 能量特征参数提取过程如图 2 所示。

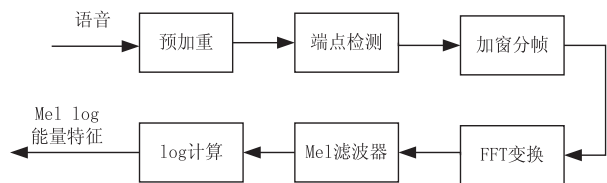


图 2 Mel log 能量特征参数的提取过程

#### 3.2 字典结构

说话人识别系统性能的好坏取决于两个关键因素:一个因素是特征参数的提取;另一个关键因素是训练模型的建立。因此,字典结构作为系统训练模型,对识别性能的影响至关重要,所以需要设计一套合适的字典训练结构方案。

##### (1) 独立字典训练。

首先设计的是独立字典训练方法。 $X_{i0}$  表示第  $i$  个说话人的原始纯净语音,  $k$  种互不相同信噪比的噪声  $\{N_{i1}, N_{i2}, \dots, N_{ik}\}$  分别增加到原始纯净语音上,从而为第  $i$  个人建立了不同噪声语音数据集  $\{X_{i0} + N_{i1}, X_{i0} + N_{i2}, \dots, X_{i0} + N_{ik}\} = \{X_{i1}, X_{i2}, \dots, X_{ik}\}$ 。对第  $i$  个人的语音数据集  $\{X_{i0}, X_{i1}, \dots, X_{ik}\}$  分别提取 Mel log 能量特征参数集  $\{Y_{i0}, Y_{i1}, \dots, Y_{ik}\}$ , 然后利用 K-SVD 算法对特征参数集中的每段参数进行单独训练,训练出一个小字典,将小字典联合在一起形成第  $i$  个人的训练大字典  $\{D_{i0}, D_{i1}, \dots, D_{ik}\}$ 。具体过程见式(4)。

$$Y_{i0} = D_{i0} \alpha_{i0}, Y_{i1} = D_{i1} \alpha_{i1}, \dots, Y_{ik} = D_{ik} \alpha_{ik} \quad (4)$$

字典训练包含了噪声语音的多种状态,即所谓的多状态训练。将带噪的测试语音特征与每个说话人的

训练特征集进行特征匹配,这样测试语音特征数据集与训练语音特征数据集的匹配程度将会明显提高,解决了背景噪声引起的系统不匹配问题。显而易见,字典训练过程是独立进行的,每个人大字典中的每段参数互不相关,所以称此过程是独立字典训练过程。

这样的字典训练过程有两个明显的好处。首先,它可以使得此字典具有更强的区分能力;其次,此字典训练过程简单易懂,建立字典模型比较方便,而且 K-SVD 算法使得字典原子数目明显减小,在降低复杂度的同时也提高了系统的鲁棒性。

##### (2) 联合字典训练。

然而,以上提出的独立字典训练过程存在一定的不足之处。当对原始纯净语音增加噪声进行字典训练的时候,随着添加噪声强度越来越大,噪声语音特征越来越远离纯净语音特征,大字典训练集会变得很分散、不集中,这样可能会造成一个说话人训练特征集与其他说话人训练特征集交叠,从而增加了识别错误率。

为了解决独立字典训练带来的问题,提出了联合字典训练方法。联合字典训练不同于独立字典训练方法,它将第  $i$  个说话人的原始纯净语音和添加不同信噪比的噪声语音特征融合在一起进行训练,若  $Y_{i0}$  表示原始纯净语音的 Mel log 能量特征参数,  $Y_{i1}, Y_{i2}, \dots, Y_{ik}$  分别表示添加不同信噪比的噪声语音特征参数,则利用 K-SVD 算法对  $\{Y_{i0}, Y_{i1}, \dots, Y_{ik}\}$  集合进行训练,见式(5)。

$$[Y_{i0}, Y_{i1}, \dots, Y_{ik}]^T = [D_{i0}, D_{i1}, \dots, D_{ik}]^T \alpha_i \quad (5)$$

其中,  $\{D_{i0}, D_{i1}, \dots, D_{ik}\}$  表示过完备字典;  $\alpha_i$  为第  $i$  个说话人的稀疏系数矩阵。

联合字典训练也是为每个说话人训练一个字典,仅仅是在字典训练的过程中将每种状态的特征参数集融合在一起进行训练,这样使得每个说话人的训练集更加集中稳定,带来的好处是一个说话人训练数据集与另一个人的训练数据集交叠几率减小,识别时造成的误差也会变小,进而可以提高说话人识别系统的性能。

##### (3) 多状态训练与谱减法的结合模型。

在背景噪声环境下,为了提高说话人识别系统的鲁棒性,语音增强技术被普遍地运用到语音前端处理中。语音增强方法有很多,其中谱减法是基于语音幅度谱估计的算法,具有较低的计算复杂度、较高的去噪效果等特点,广泛地应用于语音前端处理中。文中选取的谱减法是由 Berouti 改进后的谱减法,公式如下:

$$|\hat{S}_w(k)|^2 = \begin{cases} |Y_w(k)|^2 - \alpha |\hat{N}_w(k)|^2 & |Y_w(k)|^2 - \alpha |\hat{N}_w(k)|^2 > \beta |\hat{N}_w(k)|^2 \\ \beta |\hat{N}_w(k)|^2 & \text{其他} \end{cases} \quad (6)$$



其中,  $|S_w(k)|^2$  为估计语音谱;  $|Y_w(k)|^2$  为带噪语音谱;  $|N_w(k)|^2$  为估计噪声谱;  $\alpha$  为谱减因子;  $\beta$  为调节因子。

在利用纯净语音训练模型时,训练数据集与测试数据集不匹配,导致识别性能不理想。但是在多状态训练的过程中,当添加到第  $i$  个人的噪声  $\{N_{i0}, N_{i1}, \dots, N_{ik}\}$  种类越来越多,或者噪声强度越来越大的时候,第  $i$  个人的训练模型将会变得很分散,这会导致第  $i$  个说话人的训练模型与另一个说话人交叠。因此,为了解决此问题,结合多状态训练和谱减法,在训练语音前利用谱减法,得到增强后的语音  $\{\tilde{X}_{i0}, \tilde{X}_{i1}, \dots, \tilde{X}_{ik}\}$ ,如图 1 所示,为了保持训练与测试的一致,对增强后的测试语音提取特征参数。此方法使得训练特征更加集中化,每个说话人之间的干扰程度降低。

3.3 识别阶段

根据以上字典训练方法,为每一个说话人训练字典模型  $D_i = [D_{i0}, D_{i1}, \dots, D_{ik}]$ ,由于使用 K-SVD 算法训练冗余字典,这种方法能够降低复杂度,有效地减少计算时间,满足实时性的要求。提取测试语音特征  $y$ ,利用正交匹配追踪稀疏分解方法 (OMP) 求解,表示为:

$$\min \|\alpha_i\|_1 \quad \text{s.t.} \quad y = D_i \alpha_i \tag{7}$$

以测试语音特征投影到每个说话人的字典上的重构误差作为判定身份的条件,判定结果是:

$$\min \|y - D_i \alpha_i\|_2 \tag{8}$$

以重构误差最小对应的  $i$  就是此说话人的身份。

4 实验设计与结果分析

4.1 实验设计

文中采用的语音库是笔者团队在消音室里录制的,包含 210 个说话人的语音,每个人 180 条语句,采样率是 16 kHz。实验中,选择 28 个说话人,每个人的 10 条语句用于训练字典,20 条语句用于测试。

在特征提取过程中,文中提取的 Mel log 能量特征参数是 26 维的。对每个人的特征参数训练出一个冗余字典,字典的大小是 26x52 维。

在现实环境中,测试环境的噪声信息是未知的,所以多状态训练时添加的噪声频谱应该尽可能地包含大部分测试噪声频谱。白噪声是一种宽带噪声,其功率谱密度覆盖到整个频谱范围并在整个频谱范围内为一常数,白噪声可以近似地表示测试噪声谱,添加白噪声进行字典训练可以粗略地对有噪测试语音起到补偿作用。因此,训练语音是由添加不同信噪比的白噪声组成的,分别添加了 0 dB,10 dB,20 dB 的白噪声。噪声

来源于 noisex-92 数据库,将四种噪声 white, m109, f16, factory1 以 0 dB,5 dB,10 dB,15 dB,20 dB,25 dB,30 dB 分别添加到测试语音上,形成不同的背景噪声环境。

4.2 不同说话人识别系统的识别率对比

首先利用纯净语音进行字典模型训练,识别效果见表 1。

表 1 纯净语音训练识别率							%
	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	30 dB
white	4.64	5.89	9.11	23.04	47.86	76.61	89.82
m109	11.79	32.14	61.61	86.07	94.64	95.89	96.07
f16	3.04	10.36	25.36	56.96	82.50	94.11	95.89
factory1	5.18	16.07	41.07	75.36	92.32	95.00	96.61

由于测试语音含有噪声,在信噪比较低的条件下,噪声改变了测试语音的特征结构,测试语音特征数据集与训练语音特征数据集之间的匹配程度急速下降,重构误差变得很大,从而导致系统的识别性能不理想。甚至在 20 dB 白噪声环境下,识别率还未达到 60%;小于 15 dB 的 m109 噪声下,识别率相对也很低;f16 和 factory1 噪声下,识别率不容乐观。因此提高较低信噪比噪声环境下的识别率至关重要,也成为了提高识别系统性能的当务之急。

表 2 比较了四种说话人识别系统性能。前两种系统表示多状态训练稀疏表示分类方法, SRC-1 指的是不加谱减法的独立字典多状态训练, SRC-2 指的是不加谱减法的联合字典多状态训练。后两种系统表示结合多状态训练和谱减法的稀疏表示分类方法, SRC-3 表示增加谱减法的独立字典训练, SRC-4 表示增加谱减法的联合字典训练,系统如图 1 所示。

从表 2 可以看出,已提出的四种说话人识别系统性能比表 1 识别性能好得多,尤其在较低信噪比下,识别率得到了很大提高。在 5 dB 白噪声环境下,识别率将近达到了 60%;在 10 dB 的 m109 噪声环境下,识别率可以达到 90% 以上; f16 识别率也得到了一定的提升;15 dB 的 factory1 噪声识别率达到了 90%,比纯净语音训练提高了将近 15 个百分点。而且正如前面分析的结果,联合字典训练的识别率比独立字典训练的识别率有了一定的提高,结合多状态训练和谱减法的识别率比仅仅利用多状态训练的识别率也得到了进一步的提升。

4.3 识别结果的分析

从表 2 看出,在利用多状态训练时,当添加到测试语音中的噪声是 f16 时,系统识别率提升的程度不是很高,而对于其他三种噪声来说,识别率提升的程度相对大一些。当将多状态训练与语音增强谱减法相结合的时候,对于 f16 和 factory1 来说,识别率得到了明显

		表 2 四种说话人识别系统识别率							%
		0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	30 dB	
white	SRC-1	42.14	57.85	73.39	83.39	87.14	88.92	91.61	
	SRC-2	48.04	61.96	76.61	85.35	87.68	88.92	90.71	
	SRC-3	52.68	58.92	76.61	86.79	92.32	93.75	95.71	
	SRC-4	54.46	61.96	80.71	89.46	89.82	93.04	94.64	
m109	SRC-1	46.25	76.96	82.50	91.43	95.90	96.61	96.96	
	SRC-2	61.96	74.82	80.90	92.14	94.82	96.07	96.79	
	SRC-3	76.61	88.93	93.04	93.93	94.46	95.00	96.96	
	SRC-4	80.71	90.19	94.64	95.36	97.32	96.96	97.14	
f16	SRC-1	5.00	15.54	31.25	64.82	84.64	94.82	96.43	
	SRC-2	10.36	17.32	35.89	67.50	84.10	93.92	96.61	
	SRC-3	14.64	40.18	66.61	85.36	94.46	95.71	96.43	
	SRC-4	21.61	41.07	67.32	83.57	93.04	95.53	95.71	
factory1	SRC-1	12.68	47.32	80.19	90.10	94.29	96.79	97.50	
	SRC-2	15.54	50.00	76.68	90.54	94.82	96.61	96.78	
	SRC-3	54.46	82.50	92.14	95.54	96.07	96.61	96.96	
	SRC-4	44.46	75.36	86.07	91.43	94.29	96.07	96.96	

的提高,这说明谱减法对噪声也起到了补偿作用。为了探究其中的原因,分析了四种噪声的谱结构。图 3 表示的是噪声频谱图,m109,factory1,f16 和白噪声频谱结构大部分地方都很相似,只是 f16 噪声频谱结构部分地方出现了毛刺,它和白噪声频谱结构相差最大。因此与白噪声越相似的噪声,其识别率提高的空间越大。

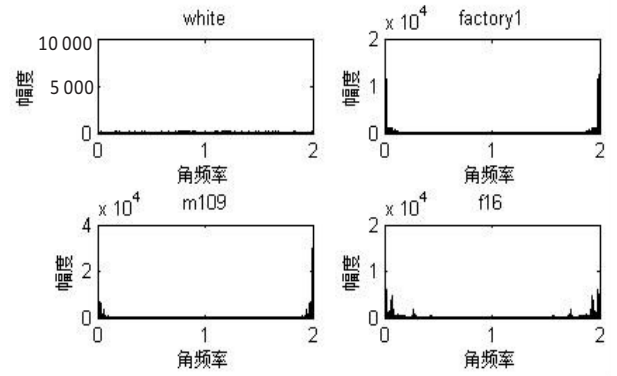


图 3 噪声频谱图

4.4 添加不同噪声训练字典的性能对比

图 4 表示的是分别添加 white 白噪声和 f16 有色噪声训练字典的识别率对比图。采用 SRC-1 系统进行训练,发现利用 white 白噪声训练字典在大多数情况下识别率较高,尤其在较低信噪比下,识别性能高于有色噪声训练情况,更加适于未知噪声信息的环境中。因此,利用白噪声训练字典比用有色噪声识别效果更

理想。

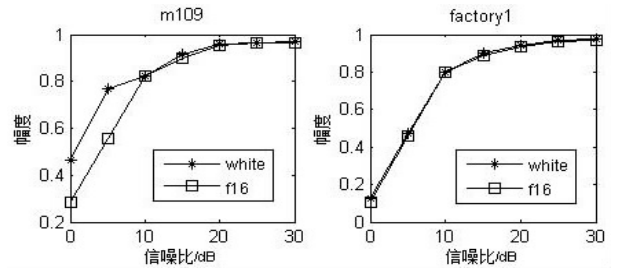


图 4 添加 white 和 f16 噪声训练字典的识别率对比图

5 结束语

文中基于稀疏表示分类方法,设计出了一种新型的鲁棒性说话人识别系统模型。利用多状态训练技术,在已增强的 Mel log 能量特征域提取特征参数。多状态训练扩大了每个人的训练字典集,每个人的训练集包含不同的状态,使得训练模型与测试语音集的匹配程度大幅度提高,这样会尽可能地寻找到最佳原子来线性表示测试语音。然后分别提出了两种字典训练结构,即独立字典训练结构和联合字典训练结构,并且利用 K-SVD 方法训练学习字典,代替原本的样本字典的方法,减小了计算复杂度,同时也保证了语音在字典上的稀疏性。实验结果和分析表明:文中所提出的系统在噪声环境下依然保持着良好的性能。所提出的系统具有很强的鲁棒性,对说话人识别技术研究具有现实的指导意义。然而该系统只是利用

白噪声进行多状态训练,而测试环境可能包含了各种噪声,将来的工作会在训练阶段添加不同种类的噪声,以符合实际需要。当然该系统计算复杂度依然很高,因此有必要减小计算复杂度,提高系统的运行效率。

#### 参考文献:

- [1] 何勇军,孙广路,付茂国,等. 基于稀疏编码的鲁棒说话人识别[J]. 数据采集与处理,2014,29(2):198-203.
  - [2] Ortega-Garcia J, Gonzalez-Rodriguez L. Overview of speaker enhancement techniques for automatic speaker recognition [C]//Proc of ICSLP'96. Philadelphia: [s. n.], 1996:929-932.
  - [3] Teunen R, Shahshahani B, Heck L P. A model-based transformational approach to robust speaker recognition [C]//Proc of ICSLP'00. Beijing, China: [s. n.], 2000:495-498.
  - [4] Ji Ming, Hazen T J, Glass J R, et al. Robust speaker recognition in noisy conditions [J]. IEEE Trans on Audio, Speech, and Language Processing, 2007, 15(5):1711-1723.
  - [5] Ji Ming, Stewaryt D, Vaseghi S. Speaker identification in unknown noisy conditions - a universal compensation approach [C]//Proc of ICASSP. Philadelphia: [s. n.], 2005:617-620.
  - [6] Kim K, Kim M Y. Robust speaker recognition against background noise in an enhanced multi-condition domain [J]. IEEE Trans on Consumer Electron, 2010, 56(3):1684-1688.
  - [7] Wright J, Yang A Y. Robust face recognition via sparse representation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2009, 31(2):210-227.
  - [8] Imran N, Roberto T, Mohammed B. Sparse representation for speaker identification [C]//Proc of ICRP. Istanbul, Turkey: [s. n.], 2010:4460-4463.
  - [9] Xu L T, Yang Z. Speaker identification based on sparse subspace model [C]//Proc of Asia-Pacific conference on communications. Bali, Indonesia: [s. n.], 2013:37-41.
  - [10] Saon G, Padmanabhan M, Gopinath R. Maximum likelihood discriminant feature spaces [C]//Proc of ICASSP. [s. l.]: [s. n.], 2000:1129-1132.
  - [11] Donoho D L. Compressed sensing [J]. IEEE Trans on Information Theory, 2006, 52(4):5406-5425.
  - [12] 石光明,刘丹华,高大化,等. 压缩感知理论及其研究进展 [J]. 电子学报, 2009, 37(5):1070-1081.
  - [13] Haris B C, Sinha R. Speaker verification using sparse representation over KSVD learned dictionary [C]//Proc of 18th national conference on communications. Kharagpur: [s. n.], 2012:1-5.
  - [14] 马 振,张雄伟,杨吉斌. 一种基于 K-SVD 的说话人识别方法 [J]. 计算机工程与应用, 2012, 48(34):112-115.
  - [15] Yu H, Waibel A. Streaming the front-end of a speech recognizer [C]//Proc of 6th international conference on spoken language processing. Beijing, China: [s. n.], 2000:353-356.
- 
- (上接第 40 页)
- [2] 魏 岩,涂铮铮,郑爱华,等. 结合 RGB 颜色特征和纹理特征的消影算法 [J]. 计算机技术与发展, 2013, 23(10):72-74.
  - [3] Thirion J P. Realistic 3d simulation of shapes and shadows for image processing [J]. CVGIP: Graphical Models and Image Processing, 1992, 54(1):82-90.
  - [4] Koller D, Daniilidis K, Nagel H H. Model-based object tracking in monocular image sequences of road traffic scenes [J]. International Journal of Computer Vision, 1993, 10(3):257-281.
  - [5] 胡园园,王让定. 基于局部纹理不变性的运动阴影去除算法 [J]. 计算机应用, 2008, 28(12):3141-3143.
  - [6] 付 萍,方 帅,徐心和,等. 视频监控系统中运动目标检测的阴影去除方法 [J]. 计算机工程, 2007, 33(10):22-24.
  - [7] Cucchiara R, Grana C, Piccardi M, et al. Detecting moving objects, ghosts, and shadows in video streams [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(10):1337-1342.
  - [8] 张红颖,李 鸿,孙毅刚. 基于混合高斯模型的阴影去除算法 [J]. 计算机应用, 2013, 33(1):31-34.
  - [9] 代科学,李国辉,涂 丹,等. 监控视频运动目标检测减背景技术的研究现状和展望 [J]. 中国图象图形学报, 2006, 11(7):919-927.
  - [10] 王 栓,艾海舟. 基于差分图象的多运动目标的检测与跟踪 [J]. 中国图象图形学报, 1999, 4(6):470-475.
  - [11] Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction [C]//Proc of ECCV 2000. Berlin: Springer, 2000:751-767.
  - [12] 李鹏飞,陈朝武,李晓峰. 智能视频算法评估综述 [J]. 计算机辅助设计与图形学学报, 2010, 22(2):352-360.
  - [13] Prati A, Mikic I, Trivedi M M, et al. Detecting moving shadows: algorithms and evaluation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(7):918-923.
  - [14] 邓亚丽,毋立芳,李云腾. 一种有效的图像阴影自动去除算法 [J]. 信号处理, 2011, 27(11):1724-1728.

基于稀疏表示的鲁棒性说话人识别系统

作者：[于云](#)，[周伟栋](#)，[YU Yun](#)，[ZHOU Wei-dong](#)  
作者单位：[南京邮电大学 通信与信息工程学院, 江苏 南京, 210003](#)  
刊名：[计算机技术与发展](#)[ISTIC](#)  
英文刊名：[Computer Technology and Development](#)  
年，卷(期)：2015, 25(12)

引用本文格式：[于云](#). [周伟栋](#). [YU Yun](#). [ZHOU Wei-dong](#) [基于稀疏表示的鲁棒性说话人识别系统](#)[期刊论文]-[计算机技术与发展](#) 2015(12)