Journal of Computer Applications
计算机应用,2013,33(7):1935 - 1937,1968
ISSN 1001-9081
CODEN JYIIDU
2013-07-01
http://www.joca.cn

文章编号:1001-9081(2013)07-1935-03
doi:10.11772/j.issn.1001-9081.2013.07.1935

基于段级特征主成分分析的说话人识别算法

储　雯[1,2*],李银国[2],徐　洋[2],孟祥涛[1,2]
(1.重庆邮电大学 计算机科学与技术学院,重庆 400065;　2.重庆邮电大学 汽车电子与嵌入式系统工程研究中心,重庆 400065)
(*通信作者电子邮箱 rukia_chu@163.com)

摘　要:为了提高说话人识别(SR)系统的运算速度,增强其鲁棒性,以现有的帧级语音特征为基础,提出了一种基于段级特征主成分分析的说话人识别算法。该算法在训练和识别阶段以段级特征代替帧级特征,然后用主成分分析方法对段级特征进行降维、去相关。实验结果表明,该算法的系统训练时间、测试时间分别为基线系统的47.8%、40.0%,同时识别率略有提高,抑制了噪声对说话人识别系统的影响。该结果验证了基于段级特征主成分分析的说话人识别算法在识别率有所提高的情况下取得了较快的识别速度,同时在不同噪声环境下的不同信噪比情况下均可以提高系统识别率。

关键词:说话人识别;非线性分段;主成分分析;说话人识别系统
中图分类号:TP18　　**文献标志码**:A

Speaker recognition method based on utterance level principal component analysis

CHU Wen[1,2*], LI Yinguo[2], XU Yang[2], MENG Xiangtao[1,2]
(1. *College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*;
2. *Research Center of Automotive Electronics and Embedded System Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: To improve the calculation speed and robustness of the Speaker Recognition (SR) system, the authors proposed a speaker recognition algorithm method based on utterance level Principal Component Analysis (PCA), which was derived from the frame level features. Instead of frame level features, this algorithm used the utterance level features in both training and recognition. What's more, the PCA method was also used for dimension reduction and redundancy removing. The experimental results show that this algorithm not only gets a little higher recognition rate, but also suppresses the effect of the noise on speaker recognition system. It verifies that the algorithm based on utterance level features PCA can get faster recognition speed and higher system recognition rate, and it enhances system recognition rate in different noise environments under different Signal-to-Noise Ratio (SNR) conditions.

Key words: Speaker Recognition (SR); non-linear partition; Principal Component Analysis (PCA); speaker recognition system

0　引言

语音识别是指计算机对人类语音进行正确响应的技术[1]。广义的语音识别技术具体包括:语音识别、说话人识别、语种识别、语音评分[2]。说话人识别(Speaker Recognition,SR)技术是一项根据语音中反映说话人生理和行为特征的语音参数自动识别说话人身份的技术,其关键问题之一是提取反映说话人个性的语音特征参数。说话人识别系统常用的语音特征参数主要有梅尔倒谱系数(Mel-Frequency Cepstrum Coefficient, MFCC)[3-6]、线性预测系数(Linear Prediction Coefficient,LPCC)[7-9]以及它们的变体。为了提高特征的可识别性,往往会对特征进行二次处理,包括差分、组合等,这导致特征参数变得庞大,增加了存储量和计算量。

本文主要从特征域入手,希望在减少模型训练、识别阶段运算量的同时提高系统鲁棒性。说话人识别系统在训练和识别过程中需要对输入的语音进行逐帧计算,时间开销很大。鉴于语音识别和说话人识别之间的紧密联系以及许多语音识别领域中的技术也在说话人识别中得以成功应用的实现[10],引入了孤立词识别领域中的非线性分段(Non-Linear Partition, NLP)技术。然而,在实验中发现,NLP无法从原理上保证分段的准确性,极容易受到微小干扰的影响,分段稳定性较差[11]。文献[11]中提出了一种改进的NLP方法——基于马氏距离的分段规则(Mahalanobis Distance Non-Linear Partition,MDNLP),通过对分段规则进行改进,提高了分段的合理性,对语音中的微小干扰具有良好的鲁棒性,并且基于这种新规则的说话人识别系统在干净语音条件下不但获得了较少的时间开销,也取得了更好的识别效果。文献[12]中提出了一种新的基于主成分分析(Principal Component Analysis, PCA)的说话人识别方法,该方法在降维时选取特征值贡献率低且累计贡献率达与100%以上的特征值。这样提取出来的是所有样本中的共有的个性特征信息,这种语音特征参数高度聚敛了说话人的个性特征,而且计算简单、模型规模小。本文结合了两者的优点,提出了一种基于段级特征主成分分析的说话人识别算法。该算法特点如下:

1)在训练和识别阶段,以段级特征代替帧级特征,有效地压缩计算量,从而提高识别速度;

收稿日期:2013-01-18;**修回日期**:2013-02-18。　**基金项目**:重庆市科委自然科学基金资助项目(cstc2012jjA60002)。
作者简介:储雯(1985 -),女(土家族),重庆人,硕士研究生,主要研究方向:语音识别;　李银国(1955 -),男,湖北黄梅人,教授,博士生导师,博士,主要研究方向:模式识别、人工智能、系统辨识与智能控制;　徐洋(1977 -),男,重庆人,副教授,博士研究生,主要研究方向:仪器仪表、嵌入式数字系统。

2)利用新 PCA 方法对段级特征进行降维、去相关,抑制噪声对说话人识别系统的影响,提高系统鲁棒性,同时提高运算速度。

实验结果表明,该算法在识别率有所提高的情况下取得了较快的识别速度,同时在不同噪声环境下的不同信噪比情况下均可以提高系统识别率。

1 说话人识别基线系统

笔者研究的基线系统是基于动态时间规整(Dynamic Time Warping,DTW)的说话人自动识别系统。首先,对用户语音进行 8 kHz 采样和 16 bit 的量化,接着用凯塞窗设计的 20 阶有限长单位冲激响应(Finite Impulse Response, FIR)带通滤波器进行预滤波,然后用一阶数字滤波器进行预加重,再用汉明窗截取语音段,将信号分成 32 ms 等长帧,相邻帧间隔为 16 ms。在语音进行处理前,用短时平均幅度进行端点检测,采用 24 维的 MFCC 特征。在进行识别匹配的时候将语音特征序列按照 DTW 方法与模板序列进行匹配,并计算匹配得分,根据得分和阈值的相对大小来确定识别结果。

2 基于段级特征主成分分析的说话人识别算法

2.1 NLP 分段

非线性分段是将语音依据非线性的方法划分为段的技术。根据语音在时间上的变化情况,NLP 将其划分为长度不等的 N 段,对每一段中的所有帧,认为它们是相似的。在训练和识别阶段,将以段的形式进行对待,而非以帧的形式进行处理。在基于 NLP 的说话人识别系统中,保证分段的合理性是极为关键的一环。分段的合理性包含两个方面的内容:一是段内距离小;二是段间距离大。如果假定某一帧为段与段的分界点,则该帧前后帧之间的距离一定较大,根据这个假设改进的分段规则 MDNLP 将某一帧作为分界点的分数公式进行了重新定义,提高了分段准确性,本文所选取的非线性分段规则为 MDNLP。

假设一段语音有 M 帧,每一帧所对应的特征为 $\boldsymbol{x}_i(1 \leqslant i \leqslant M)$,则对应特征序列 $\boldsymbol{X} = (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M)$,定义语音特征变化信息为 $\boldsymbol{x}_t$ 和 $\boldsymbol{x}_{t+1}$ 之间的距离:

$$d_t = d_{cep}(\boldsymbol{x}_t,\boldsymbol{x}_{t+1}) = \sum_{k=1}^{K}[W_k(\boldsymbol{x}_t^k - \boldsymbol{x}_{t+1}^k)]^2;\ 1 \leqslant t \leqslant T-1 \tag{1}$$

其中 W_k 指定了特征向量第 k 维的权重,是一个实验经验数据。本文中采用近似的下标权重 $W_k = k$。为了找出段分界点,定义某帧 i 作为分界点的分数 s_i 如下:

$$s_i = \begin{cases} (s'_i - \overline{s'})^2/\sigma, & (s'_i - \overline{s'}) > 0 \\ 0, & \text{其他} \end{cases} \tag{2}$$

其中:$s' = \frac{1}{M}\sum_{k=1}^{M} d_{i-k,i+k}$,$d_{i,j}$ 表示帧 i 与帧 j 之间的距离,M 为以帧 i 为中心选取的窗的宽度,$\overline{s'}$ 和 σ 则代表所有 s' 的均值和方差。s_i 越大,则表明第 i 帧前后帧的差异越大,则该帧越有可能是一段分界点。

2.2 PCA 降维

PCA 是一种在均方误差最小意义上的最优降维方法[13]。PCA 假定具有大变化的方向的数据比很少变化方向上的数据携带有更多的信息,因而它寻找具有最大方差的那些称之为主轴的方向来表征原始数据。通过把原始特征向量向更小的子空间投影,PCA 达到了降维和去除冗余的效果。因此,经过 PCA 降维,损失的特征信息最少,在保证识别性能的同时,后续阶段的计算开销将会大大减少。另外,对于受某些噪声影响的语音,在其上面提取出特征,经过 PCA 转换,噪声感染部分作为次要因素,往往被作为高维信息去除了,因此,PCA 转换还有降噪的功能。但是,传统的 PCA 算法中,选取较大特征值对应的特征矢量作为基向量对数据进行降维,而没有考虑到不同特征参数的累计贡献率的问题。这里采用 PCA 新方法,在选取特征值时取特征值贡献率低且累积贡献率大于 100% 以上的特征值。

记 $\boldsymbol{x}(i) = [\boldsymbol{\alpha}_{i1}^{\mathrm{T}},\boldsymbol{\alpha}_{i2}^{\mathrm{T}},\cdots,\boldsymbol{\alpha}_{iK}^{\mathrm{T}}]^{\mathrm{T}}$ 为 K 帧语音特征向量组成的段级特征,其中,$\boldsymbol{\alpha}_i$ 为帧级特征。$\boldsymbol{X} = \boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N$ 为整个语音的段级特征集,N 代表段数。具体的基于段级特征主成分分析的说话人识别算法步骤如下:

1)计算样本均值向量 $\boldsymbol{\mu}_j$:

$$\boldsymbol{\mu}_j = \sum_{i=1}^{N}\boldsymbol{X}(i)_j;\ j = 1,2,\cdots,d \tag{3}$$

其中 $\boldsymbol{\mu}_j(j = 1,2,\cdots,d)$ 为某一维的均值向量。

2)计算段级特征集 $\boldsymbol{X}$ 的协方差矩阵 $\boldsymbol{S}$:

$$\boldsymbol{S} = \frac{1}{n-1}(\boldsymbol{X}(i) - \boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{X}(i) - \boldsymbol{\mu}) \tag{4}$$

计算矩阵 $\boldsymbol{S}$ 的特征值 λ 并将其从大到小排序。设 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_d$,求出其对应的特征向量 $\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\cdots,\boldsymbol{\omega}_d$。

3)按累计贡献率准则提取 k 个主成分。主成分贡献率如式(5)所示:

$$l_i = \lambda_i / \sum_{j=1}^{d}\lambda_j;\quad i = 1,2,\cdots,d \tag{5}$$

其中 l_i 就是第 i 个主成分的贡献率。累计贡献率如式(6)所示:

$$H = \left(\sum_{i=1}^{j}\lambda_i\right)/\left(\sum_{i=1}^{N}\lambda_i\right) \tag{6}$$

其中 H 就是第 i 个主成分的累积贡献率。

4)取累计贡献率之和大于 100% 的主成分作为新的特征矢量(假设所取维数为 q),则将 q 个特征向量组成转换矩阵 $\boldsymbol{W}$:

$$\boldsymbol{W} = [\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\cdots,\boldsymbol{\omega}_q]^{\mathrm{T}} \tag{7}$$

5)对所有的段级特征进行降维处理,得到新段级向量序列:

$$\boldsymbol{Y} = \boldsymbol{W}\cdot\boldsymbol{X} \tag{8}$$

新的段级向量序列 $\boldsymbol{Y}(i)$ 的维数为 q,达到了降维和去相关性的目的。

3 实验配置及结果分析

采用基于段级特征主成分分析的说话人识别系统框图如图 1 所示。在该图中,$\boldsymbol{F}$(Feature Vector)表示特征向量,$\boldsymbol{NF}$(New Feature Vector)表示经过 MDNLP 分段和 PCA 新方法变换后新的特征向量,$\boldsymbol{W}$ 表示转换矩阵,M(Model)表示训练得到的说话人模型。输入语音经过预处理和特征提取后得到帧级特征矢量集,然后用 MDNLP 对特征集进行分段,得到段级特征集。在训练阶段,由“PCA:转换矩阵”模块计算转换矩阵 $\boldsymbol{W}$,然后将 $\boldsymbol{W}$ 输入“PCA:特征变换”模块,得到新的特征向量序列,再用新的特征向量训练说话人模型,最后将训练好的说话人模型和该说话人的转换矩阵保存在数据库中。在说话人识别阶段,将段级特征序列输入“PCA:特征变换”模块,同时将转换矩阵依次输入该模块中,同测试特征向量相乘得到变

换后新的特征向量。然后,将新的测试特征输入“匹配得分”模块,与同时输入的说话人模型 M 进行匹配计算,得到对应于该说话人的模型得分。最后得到所有说话人的得分后,依据决策规则得到识别结果。

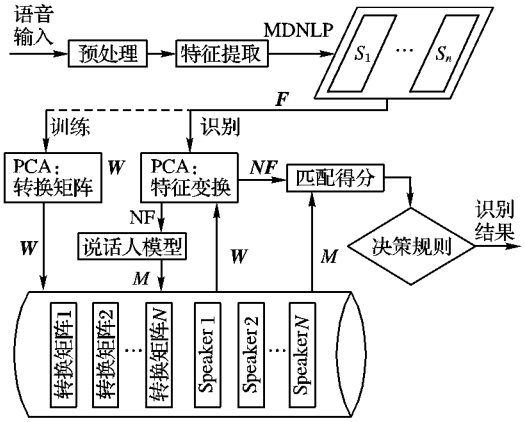


图 1　基于段级特征主成分分析的说话人识别系统框图

语音数据库使用一个含 28 人的语料库,每人共 20 条语音,每条语音包含两个中文词,时长 2 s。为了方便计算,其中 10 条语音用于训练,10 条用于识别,保证了训练与识别语音的不一致。采样率为 16 kHz,采用 8 bit 的量化。MDNLP 分段分为 4 段,其中 PCA 降维从 24 维降到 16 维。噪声库采用 NoiseX-92 专业噪声库中的平稳高斯白(White)噪声、粉(Pink)噪声和 Volvo 噪声。实验的 PC 配置为:Pentium Dual-Core CPU E5800 3.20 GHz,1.96 GB 内存。

纯净语音下系统性能对比实验结果如表 1 所示。由表 1 可见,以基于段级特征主成分分析的说话人识别算法进行训练和识别的说话人识别系统的识别率高于基线系统和分别只进行 MDNLP 分段和传统 PCA 降维的系统,同时训练速度和识别速度有了大幅度的提高。基于段级特征主成分分析的说话人识别算法的说话人识别系统训练时间仅为基线系统的 47.8%,测试时间也缩短为基线系统的 40.0%,同时识别率略有提高。

表 1　纯净语音下系统性能对比

测试项目	单位模型训练时间/s	单位样本测试时间/s	识别率/%
基线系统	2.971	1.004	88.89
MDNLP	1.683	0.503	89.10
传统 PCA	1.642	0.449	88.86
MDNLP + PCA 新方法	1.421	0.405	91.82

带噪环境下基于段级特征主成分分析的说话人识别算法的说话人识别系统性能分析实验结果如图 2 ~ 4 所示。由图 2 ~ 4 可以看出,在高斯白噪声和 Pink 噪声环境下,基于段级特征主成分分析的说话人识别算法的说话人识别系统与经传统 PCA 降维的识别系统识别率相近,均高于基线系统及经过 MDNLP 分段的识别系统;在 Volvo 噪声环境下,基于段级特征主成分分析的说话人识别算法的说话人识别系统识别率高于其他三个系统。

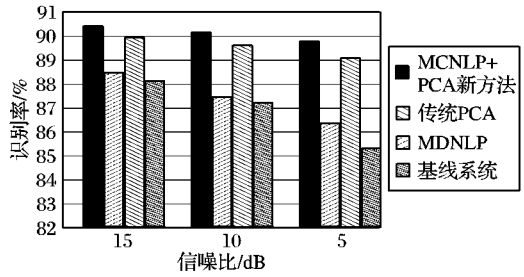


图 2　White 噪声下的识别结果

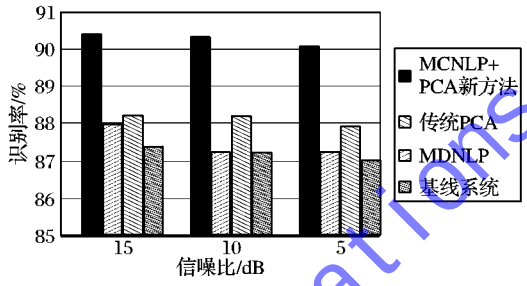


图 3　Pink 噪声下的识别结果

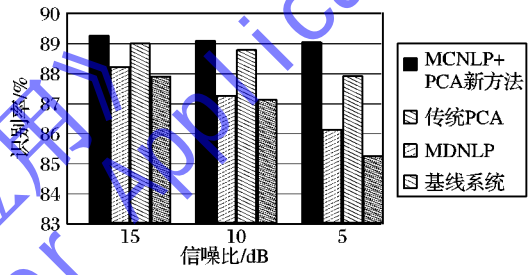


图 4　Volvo 噪声下的识别结果

4　结语

本文以现有的帧级语音特征为基础,提出了一种基于段级特征主成分分析的说话人识别算法。实验结果表明,该算法在提高系统鲁棒性的同时提高了系统运行速度,适用于实时说话人识别系统。

参考文献:

[1] 徐波. 语音识别技术与应用的发展趋势[J]. 中国计算机学会通讯, 2008, 4(2): 48 - 52.

[2] 吴朝辉, 杨莹春. 说话人识别模型与方法[M]. 北京: 清华大学出版社, 2009: 3 - 3.

[3] HUNG W W, WANG H C. On the use of weighted filter bank analysis for the derivation of robust MFCCs [J]. IEEE Signal Processing Letters, 2001, 8(3): 70 - 73.

[4] HOSSAN M A, MEMON S, GREGORY M A. A novel approach for MFCC feature extraction [C]// Proceedings of 2010 4th International Conference on Signal Processing and Communication Systems. Piscataway: IEEE Press, 2010: 1 - 5.

[5] KOPPARAPU S K, LAXMINARAYANA M. Choice of Mel filter bank in computing MFCC of a resampled speech [C]// Proceedings of 2010 10th International Conference on Information Sciences Signal Processing and their Applications. Piscataway: IEEE Press, 2010: 121 - 124.

[6] WANG H Z, XU Y C, LI M J. Study on the MFCC similarity-based voice activity detection algorithm [C]// 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce. Piscataway: IEEE Press, 2011: 4391 - 4394.

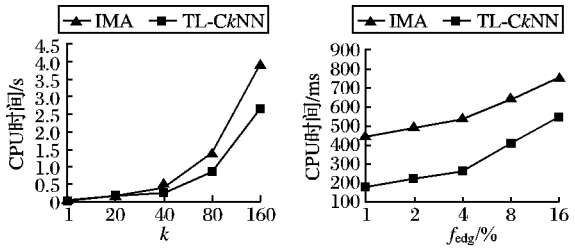
[7] GISH H, SCHMIDT M. Text-independent speaker identification [J]. IEEE Transactions on Signal Processing, 1994, 11(40): 18 - 32.

[8] YUAN Y J, ZHAO P H, ZHOU Q. Research of speaker recognition based on combination of LPCC and MFCC [C]// Proceedings of 2010 International Conference on Intelligent Computing and Intelligent Systems. Piscataway: IEEE Press, 2010: 765 - 767.

[9] ZBANCIOC M, COSTIN M. Using neural networks and LPCC to improve speech recognition [C]// Proceedings of 2003 International Symposium on Signals, Circuits and Systems. Piscataway: IEEE Press, 2003: 445 - 448.

(下转第 1968 页)

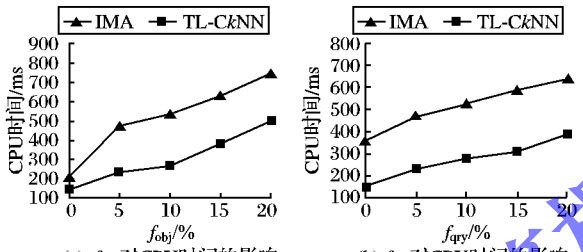
TL-C*k*NN 算法的 1.5 倍。实验结果表明TL-C*k*NN在 k 较大的情况下具有较 IMA 算法更好的性能。图 5(b)说明了边的灵敏度 f_{edg} 对查询处理时间的影响。当 f_{edg} 增大时,TL-C*k*NN 算法和 IMA 算法的查询处理时间都随之增加。这是因为权值发生变化的边越多,受到影响的查询也越多。



(a) k对CPU时间的影响 (b) f_{edg}对CPU时间的影响

图 5 k 和 f_{edg} 与 CPU 时间关系

在图 6(a)中,当 f_{obj} 增大,IMA 算法和 TL-C*k*NN 算法的查询处理时间也随之增加。原因是发生位置变化的移动对象越多,需要进行结果维护的查询也越多。图 6(b)比较了 f_{qry} 对 IMA 算法和 TL-C*k*NN 算法的查询处理时间的影响。实验结果表明 f_{qry} 增加会使 IMA 算法和 TL-C*k*NN 算法的性能下降。



(a) f_{obj} 对CPU时间的影响 (b) f_{qry}对CPU时间的影响

图 6 f_{obj} 和 f_{qry} 与 CPU 时间关系

在以上实验中,查询和移动对象在道路网中都采用的是均匀分布方式。图 7 比较了查询和移动对象在道路网中的不同分布方式对 CPU 时间的影响,其中:G 表示 Gaussian 分布,U 表示 Uniform 分布,Q 表示 Query(查询),O 表示 Object(移动对象)。结果表明,当查询为 Gaussian 分布,移动对象为 Uniform 分布时,TL-C*k*NN 算法具有最好的性能。在这种情况下,查询分布的比较集中,查询处理中可以充分利用其他查询的计算结果,避免了冗余搜索,从而节省了 CPU 时间。

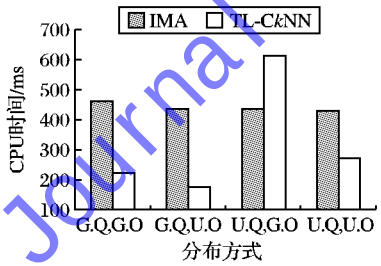


图 7 查询和移动对象的分布方式与 CPU 时间

4 结语

TL-C*k*NN算法的查询处理,由于充分考虑了系统中已存在的其他查询,并重用了其计算结果,从而在很大程度上避免了重复计算,缩短了查询的平均响应时间。

实际应用中,查询处理可能还要考虑查询和移动对象的运动方向,以及道路网的一些限制条件如单行道、某些路口不允许左右转弯等,因此如何基于更复杂的运动模型和道路网模型进行算法设计,使研究成果更贴近应用实际,将是下一步的重点研究内容。

参考文献:

[1] KOLAHDOUZAN M, SHAHABI C. Voronoi-based k nearest neighbor search for spatial network databases[C]// Proceedings of the 30th Very Large Data Base Conference. Toronto: VLDB Endowment, 2004: 840 - 851.

[2] 王淼,郝忠孝. 基于动态创建局部 Voronoi 图的连续近邻查询[J]. 计算机应用研究, 2008, 25(9): 2771 - 2774.

[3] HUANG X G, JENSEN C S, SALTENIS S. The islands approach to nearest neighbor querying in spatial networks[C]// Proceedings of the 9th International Symposium on Spatial and Temporal Databases. Berlin: Springer-Verlag, 2005: 73 - 90.

[4] XIONG X P, MOKBEL F M, AREF W G. SEA-CNN: scalable processing of continuous k-nearest neighbor queries in spatio-temporal databases[C]// ICDE 2005: Proceedings of the 21st International Conference on Data Engineering. Piscataway: IEEE, 2005: 643 - 654.

[5] YU X H, PU K Q, KOUDAS N. Monitoring k-nearest neighbor queries over moving objects[C]// ICDE 2005: Proceedings of 21st International Conference on Data Engineering. Piscataway: IEEE, 2005: 631 - 642.

[6] MOURATIDIS K, HADJIELEFTHERIOU M. Conceptual partitioning: an efficient method for continuous nearest neighbor monitoring[C]// Proceedings of ACM SIGMOD 2005. New York: ACM, 2005: 634 - 645.

[7] 卢秉亮,刘娜. 路网中移动对象快照 K 近邻查询处理[J]. 计算机应用, 2011, 31(11): 3078 - 3083.

[8] 梁茹冰,刘琼. 公路网移动终端的 KNN 查询技术[J]. 华南理工大学学报: 自然科学版, 2012, 40(1): 138 - 145.

[9] 赵亮,陈荦,景宁,等. 道路网中的移动对象连续 K 近邻查询[J]. 计算机学报, 2010, 33(8): 1396 - 1403.

[10] MOURATIDIS K, YIU M L, PAPADIAS D, *et al*. Continuous nearest neighbor monitoring in road networks[C] // Proceedings of the 32nd International Conference on Very Large Data Bases. Toronto: VLDB Endowment, 2006: 43 - 54.

[11] DEMIRYUREK U, BANAEI-KASHANI F, SHAHABI C. Efficient continuous nearest neighbor query in spatial networks using Euclidean restriction[C]// Proceedings of the 11th International Symposium on Advances in Spatial and Temporal Databases. Berlin: Springer-Verlag, 2009: 25 - 43.

[12] 廖巍,吴晓平,严承华,等. 多用户连续 k 近邻查询多线程处理技术研究[J]. 计算机应用, 2009, 29(7): 1861 - 1864.

[13] BRINKOFF T. A framework for generating network based moving objects[J]. Geoinformatica, 2002, 6(2): 153 - 180.

(上接第 1937 页)

[10] REYNOLDS D. An overview of automatic speaker recognition technology[C]// 2002 International Conference on Acoustics, Speech, and Signal Processing. Piscataway: IEEE Press, 2002: 4072 - 4075.

[11] LUO C H, WU X J, ZHENG F, *et al*. Segmentation-based method for text-dependent speaker recognition in embedded applications[C]// Proceedings of the Second Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Singapore: [s. n.], 2010: 466 - 469.

[12] 余利强,马道钧. 基于 PCA 技术的神经网络说话人识别研究[J]. 计算机工程与应用, 2010, 46(19): 211 - 213.

[13] XU H. Robust PCA via outlier pursuit[J]. IEEE Transactions on Information Theory, 2012, 58(5): 3047 - 3064.