

基于 Fisher 比的梅尔倒谱系数混合特征提取方法

鲜晓东, 樊宇星*

(重庆大学 自动化学院, 重庆 400044)

(* 通信作者电子邮箱 fanyuxing88@126.com)

摘要: 针对语音识别中梅尔倒谱系数(MFCC)对中高频信号的识别精度不高, 并且没有考虑各维特征参数对识别结果影响的问题, 提出基于 MFCC、逆梅尔倒谱系数(IMFCC)和中频梅尔倒谱系数(MidMFCC), 并结合 Fisher 准则的特征提取方法。首先对语音信号提取 MFCC、IMFCC 和 MidMFCC 三种特征参数, 分别计算三种特征参数中各维分量的 Fisher 比, 通过 Fisher 比对三种特征参数进行选择, 组成一种混合特征参数, 提高语音中高频信息的识别精度。实验结果表明, 在相同环境下, 新的特征与 MFCC 参数相比, 识别率有一定程度的提高。

关键词: 识别精度; 梅尔倒谱系数; 逆梅尔倒谱系数; 中频梅尔倒谱系数; Fisher 准则

中图分类号: TP391 **文献标志码:** A

Parameter extraction method for Mel frequency cepstral coefficients based on Fisher criterion

XIAN Xiaodong, FAN Yuxing*

(College of Automation, Chongqing University, Chongqing 400044, China)

Abstract: Concerning the low identification precision of Mel Frequency Cepstral Coefficients (MFCC) parameters in high frequency signals and the problem that the influence of each dimension feature parameters has not been considered to identify, the method of extracting features based on MFCC, IMFCC (Inverted MFCC) and MidMFCC (Mid-frequency MFCC) combined with Fisher criterion was adopted. Extracting MFCC, IMFCC and MidMFCC parameters from speech signals and calculating the Fisher ratio of components of three parameters, useful parameters were chosen by using Fisher standard and a mixture feature was constructed to improve mid-frequency and high frequency recognition accuracy. The experimental results show that the new feature has better recognition results compared with MFCC in the same environment.

Key words: recognition accuracy; Mel Frequency Cepstral Coefficient (MFCC); Inverted MFCC (IMFCC); Mid-frequency MFCC (MidMFCC); Fisher criterion

0 引言

在语音识别中, 最终的训练和识别环节是针对语音的特征参数进行的, 因此特征提取在识别过程中作用非常关键^[1]。

目前, 常用的特征参数主要有线性预测系数 (Linear Prediction Cepstral Coefficient, LPCC)、感性预测系数 (Perceptual Linear Predictive, PLP) 及梅尔倒谱系数 (Mel Frequency Cepstrum Coefficient, MFCC), 其中 LPCC 与 PLP 对环境的依赖性较高, 抗噪能力差^[2], 所以目前应用较多的是 MFCC。MFCC 系数描述的是语音在频率上的能量分布, 使用一串在低频区域交叉重叠排列的三角滤波器组, 获得语音的频谱信息, 但是由于低频区域使用的滤波器数量较多, 分布密集, 同时中高频区域使用的滤波器数量较少, 分布稀疏, 造成 MFCC 随着频率的提高, 计算精度下降, 对最终的识别结果会造成影响, 特别是当识别基元选择为音节或词组时, 在两个读音相近的音节或词之间误识率较高。目前对于语音的特征参数的提取, 提出了许多改进的算法, 如针对短时傅里叶变换无法兼顾时间和频率分辨率, 采用分数阶傅里叶变换或小波变换等调频信号的处理方法代替短时傅里叶变换对语音信号进

行处理^[3], 或者针对 MFCC 参数对于语音信号描述的不完整性, 在其基础上进行改进。文献[4]采用分数阶傅里叶变换代替短时傅里叶变换, 三角 Mel 滤波器组替换为听觉滤波器组进行特征参数的提取, 但是在这种算法中变换阶数的选取对于分数阶傅里叶变换的性能有至关重要的影响。Sandipan 等^[5]针对高频信息, 提出了逆梅尔倒谱系数 (Inverted MFCC, IMFCC), 结合 MFCC 与 IMFCC 作为新的特征参数。韩一等^[6]提出了针对中频能量的滤波器组, 得到一种新的参数, 结合 MFCC 和 IMFCC 组成混合参数进行情感识别。袁正午等^[7]采用 IMFCC、中频梅尔倒谱系数 (Mid-frequency MFCC, MidMFCC) 和 MFCC 相结合的改进算法, 使用混合滤波器组, 进行语音识别。

评价特征参数中不同分量对识别的贡献一般有两种方法, 通过各个分量的 Fisher 比得到各分量的区分能量, 或者通过增减分量的方法判断每个特征分量的贡献^[8]。甄斌等^[9]采用增减特征分量的方法研究了 MFCC 各维倒谱分量对说话人识别和语音识别的贡献, 但是通过增减特征分量的方法计算量较大, 同时受环境及语音识别的影响较大, 因此利用 Fisher 准则来分析特征向量, 确定特征分量的可分离性是一种常用的特征选择方法。张芸等^[10]利用 Fisher 准则对 LPCC

收稿日期: 2013-08-15; 修回日期: 2013-10-08。

作者简介: 鲜晓东 (1966-), 女, 重庆人, 副教授, 硕士, 主要研究方向: 无线传感器网络、移动机器人控制、信号处理; 樊宇星 (1988-), 男, 山西原平人, 硕士研究生, 主要研究方向: 语音信号处理。

和 MFCC 特征参数进行结合构造一种新的特征参数应用于说话人的识别。胡雪峰等^[11]利用 Fisher 准则结合 MFCC 与 IMFCC 构造出了新的混合特征参数,但这种参数没有考虑到中频能量分布情况。

本文首先计算 MFCC 参数、IMFCC 参数和 MidMFCC 参数,然后利用 Fisher 准则,选择其中可分离程度较大的特征分量,组成一种新的特征参数。实验结果表明,这种参数提高了语音的识别率。

1 语音特征参数的提取

1.1 Mel 倒谱系数的提取

梅尔倒谱系数(MFCC)模拟人耳的听觉特性,即对于低频语音信号的感知分辨能力相比高频信号更强,将语音线性频谱转化为基于 Mel 频标的非线性性频谱,然后转换到倒谱域^[1]。在 1000 Hz 以下,Mel 刻度与 Hz 频率刻度近似遵循线性关系,而在 1000 Hz 以上的声音频率范围 Mel 刻度与 Hz 频率坐标遵循近似对数关系,两种坐标的变换关系如式(1):

$$F_{mel} = 1127 * \ln(1 + f_{Hz}/700) \tag{1}$$

MFCC 是采用滤波器组的方法计算出来的,在 Mel 域均匀分布一组三角带通滤波器,称为 Mel 滤波器组,变换到 Hz 频域,滤波器的中心频率和带宽随着频率的变化而变化,在 1000 Hz 以下滤波器分布密集,而在 1000 Hz 以上,滤波器分布比较稀疏,频域滤波器分布如图 1 所示。

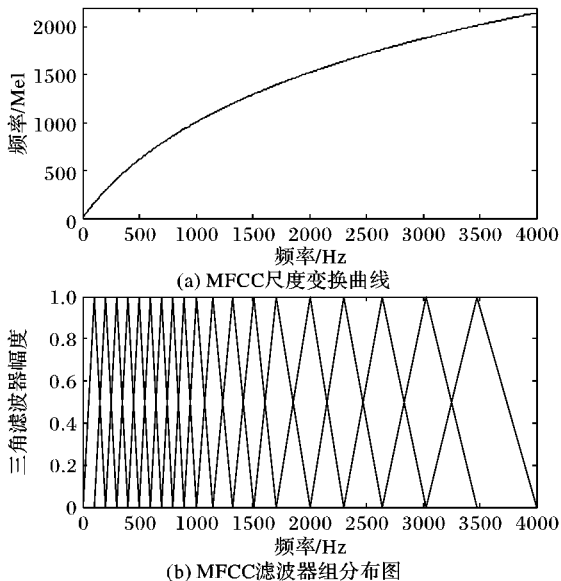


图 1 MFCC 尺度变换关系和滤波器组分布图

图 1(a)表示 MFCC 中尺度变换关系,(b)是三角滤波器组分布图,人耳对于不同频率的声音,感知能力是不一样的:频率在 1000 Hz 以下的声音,人耳对其感知能力基本满足线性关系;频率高于 1000 Hz 时,人耳对声音的感知能力与声音的频率近似满足对数关系。而在 Mel 频率域,人耳对于不同 Mel 频率的声音近似满足线性关系。

MFCC 参数提取流程:

- 1) 语音信号经过加窗处理后变为短时信号,利用快速傅里叶变换(Fast Fourier Transform, FFT)将这些时域信号转换为频域信号,并计算短时能量谱。
- 2) 将短时能量谱用一组三角滤波器组进行滤波,得到一

组滤波系数,其中每个滤波器的中心频率在 Mel 刻度上均匀分布,带宽为相邻中心频率的差值。

3) 对得到的滤波系数根据式(2)进行余弦变换,去除各维信号之间的相关性,将信号映射到低维空间以得到特征参数。

$$c(m) = \sum_{n=0}^{M-1} S(n) \cos\left(\frac{\pi m(n+0.5)}{M}\right); 1 \leq m \leq M \tag{2}$$

其中: $c(m)$ 表示 MFCC 第 m 维特征参数, $S(n)$ 表示滤波器系数, M 表示滤波器的阶数。

1.2 IMFCC 和 MidMFCC 的提取

在 MFCC 参数中,滤波器组主要分布于低频部分,注重于语音的低频频谱,对于中高频的频谱信息计算精度较差,造成最终提取的参数不能最大限度地表征语音的特征。为了提高中高频的计算精度,先后提出了逆梅尔倒谱参数(IMFCC)^[5]和中频梅尔倒谱参数(MidMFCC)^[6]。这两种参数与经典的 MFCC 的不同主要在 Mel 刻度与 Hz 刻度的转换上,IMel 刻度和 MidMel 刻度与 Hz 刻度的转换关系如式(3)、(4)所示。

$$F_{Imel} = 2146.1 - 1127 * \ln\left(1 + \frac{4000 - f_{Hz}}{700}\right) \tag{3}$$

MidMel 刻度与 Hz 刻度之间的关系为:

$$F_{MidMel} = \begin{cases} 1073.05 - 527 * \ln\left(1 + \frac{2000 - f_{Hz}}{300}\right), & 0 < f_{Hz} \leq 2000 \\ 1073.05 + 527 * \ln\left(1 + \frac{f_{Hz} - 2000}{300}\right), & 2000 < f_{Hz} \leq 4000 \end{cases} \tag{4}$$

在参数提取过程中,刻度的转换关系主要体现在滤波器组的分布情况上,因此在 IMFCC 参数中,滤波器主要分布于高频部分,表征了高频部分的频谱特征;而 MidMFCC 参数中,滤波器在中频部分分布比较密集,在低高频部分分布比较稀疏,滤波器组结构如图 2、3 所示。IMFCC 和 MidMFCC 的提取过程与 MFCC 的提取过程一致。

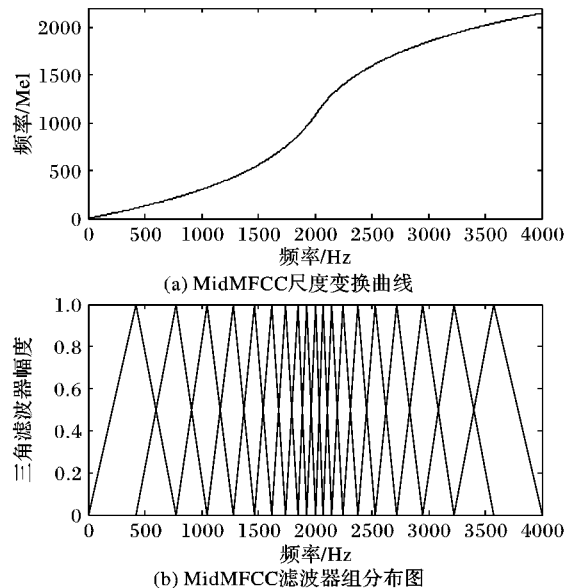


图 2 MidMFCC 尺度变换关系和滤波器组分布图

图 2 和图 3 分别表示了针对中频和高频的 Hz 频率与 Mel 频率的变换曲线及滤波器组的分布。在图 2(a)中:在声音频率低于 1500 Hz 和高于 2500 Hz 部分,Mel 频率与 Hz 频

率呈对数关系,表示人耳对声音的感知能力与声音的 Hz 频率近似满足对数关系;而在 1 500 Hz 与 2 500 Hz 之间,Mel 频率与 Hz 频率呈对数关系,表示人耳对声音的感知能力与声音的 Hz 频率呈线性关系。在图 3 (a) 中:声音频率低于 3 000 Hz 时,Mel 频率与 Hz 频率呈对数关系,表示人耳对声音的感知能力与声音的 Hz 频率近似满足对数关系;而在 3 000 Hz 以上,Mel 频率与 Hz 频率呈对数关系,表示人耳对声音的感知能力与声音的 Hz 频率呈线性关系。

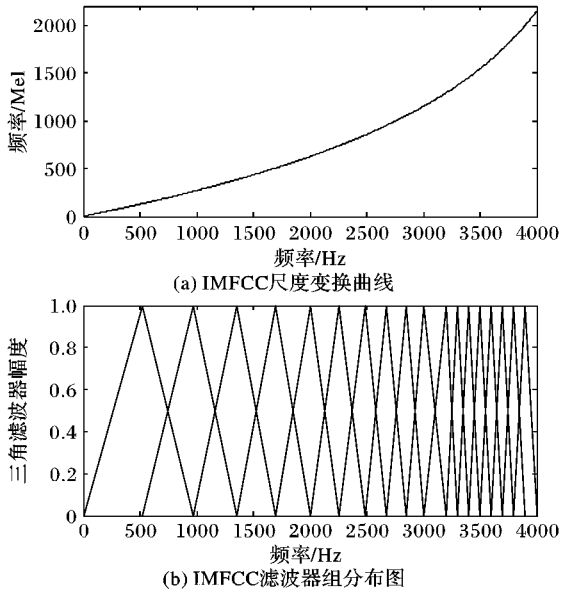


图3 IMFCC 尺度变换关系和滤波器组分布图

2 Fisher 比准则

以上三种特征参数分别表征语音信号三个频段的特点,可以结合这三种特征对语音进行描述,但是,直接将它们叠加组合会产生一些问题。首先,增加了特征参数的维数,如果每种参数提取 12 维,那么最终得到的是 36 维参数,增加了计算量,同时各维特征参数对识别的贡献不同,有些参数可能包含的信息量较少,有些包含的可能是冗余信息,如果将它们同等对待,最终会影响识别性能。所以必须对各维参数对识别结果的影响程度进行评价,得到对识别影响最大的参数,然后组合三组特征作为新的特征参数。通过计算特征分量的 Fisher 比得到各分量的区分能力^[12]。

Fisher 判别准则:

$$r_{\text{Fisher}} = \frac{\sigma_{\text{between}}}{\sigma_{\text{within}}} \quad (5)$$

其中: r_{Fisher} 是特征分量的 Fisher 比; σ_{between} 表示特征分量的类间方差,即不同语音特征分量均值的方差; σ_{within} 表示特征分量的类内方差,即同一语音特征分量的方差的均值。

$$\sigma_{\text{within}} = \sum_{i=1}^N \frac{1}{n_i} \sum_{c \in \omega_i} (c_k^i - m_k^i)^2 \quad (6)$$

$$\sigma_{\text{between}} = \sum_{k=1}^N (m_k^i - m_k)^2 \quad (7)$$

其中: k 表示特征参数的维数, $k = 1, 2, \dots, 12$; m_k 表示语音特征第 k 个分量在所有类上的均值; m_k^i 表示语音特征第 k 个分量第 i 类的均值; ω_i 表示第 i 类的语音特征序列; M, n_i 分别表示语音特征序列的类别数及各类的样本数; c_k^i 表示第 i 类语音特征序列的第 k 个分量。

特征分量的类间方差反映了不同语音样本之间差异的程度,而类内方差反映了同一语音样本之间的密集程度,对于特征分量,从类内方差和类间方差两个方面表征了它的可分离程度,图 4 给出了三种特征参数中维数与 Fisher 比之间的关系,从中可以看出不同维数分量对识别的贡献程度是不同的。

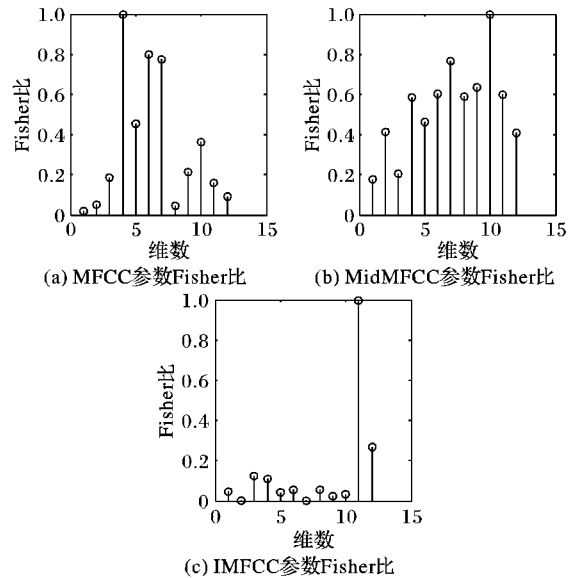


图4 MFCC、IMFCC 和 MidMFCC 各维分量的 Fisher 比

混合特征参数的提取过程如下:

- 1) 对语音数据进行预处理,包括预加重、分帧加窗和端点检测;
- 2) 分别对预处理后的语音数据在三种滤波器组的情况下提取特征参数 MFCC、IMFCC 和 MidMFCC;
- 3) 分别从 MFCC、IMFCC 和 MidMFCC 三种参数中选择 Fisher 比即区分度最大的 6 个分量组合成新的 18 维的混合特征参数。

3 实验分析

从说话的方式可以将语音识别分为孤立词识别、连接词语音识别和连续语音识别:孤立词语音中主要是将词语作为整体进行识别;连接词语音通常涉及的是数字串的认可;连续语音识别的对象是自然连续的语音,词与词之间没有停顿,像日常口语一样,因此会出现大量的连音和变音现象,识别时比较复杂。因此,一种新的特征参数不仅在孤立词下能够提高识别率,同时在连续语音的情况下,对识别性能应该有所改善。为了验证本文提出的混合特征参数的有效性,主要针对孤立词和较复杂的连续语音进行试验,分别设计了两个实验进行仿真验证,实验一在孤立词的情况下进行,实验二是在语音段长度较长的情况下对本文提出的特征参数有效性进行验证。

3.1 实验一

实验中所用到的语音数据采集自实验室的同学,创建了一些控制词汇(包括启动、左转、右转、停止)的语音库,录制 4 个人(3 男 1 女)的语音,每个词每人录制 7 次,其中采样频率为 22 050 Hz,共有 112 个语音数据,将其中的 80 个作为测试语音,每个词有 20 个测试样本,其余的作为训练样本。

实验中首先需要确定所选的滤波器的阶数,通常滤波器组的阶数在 18 到 24 之间,所以本文在考虑计算速度的同时

采用 19 阶滤波器组,通带范围 0 ~ 4000 Hz,这是由于正常情况下普通人发出的声波在 4000 Hz 以下,帧长 256 个采样点,帧移 128 个采样点,分别得到 12 阶 MFCC 系数、MidMFCC 系数及 IMFCC 系数,然后分别在三种参数中选择 Fisher 比最大的 6 个组成 18 维的混合特征参数。在识别实验中,选取 MFCC 特征参数和文献[7]的特征参数作为本文参数的比较算法。

训练和识别环节采用混合高斯模型(Gaussian Mixture Model, GMM),GMM 的混合度为 8,利用高斯模型对训练样本的混合特征参数进行训练,得到 4 个语音声学模型,然后对测试语音进行识别,与经典 MFCC 特征参数和文献[7]的特征参数相比,识别率如表 1 所示。

表 1 实验一基于不同特征参数的识别结果

测试模板	基于 MFCC 特征参数			文献[7]的特征参数			本文提出的特征参数		
	总数	正确数	误识数	总数	正确数	误识数	总数	正确数	误识数
启动	20	19	1	20	20	0	20	20	0
左转	20	16	4	20	15	5	20	17	3
右转	20	17	3	20	17	3	20	18	2
停止	20	18	2	20	20	0	20	20	0

表 2 实验一三种特征参数识别结果比较

特征参数	识别率/%
MFCC	87.50
文献[7]的特征参数	90.00
本文的特征参数	93.75

由结果可以得到:

1)根据文献[7]的算法和本文算法得出的特征参数与 MFCC 相比,在相同环境下,识别率上都有不同程度的提高,这主要是这两种方法都增强了中频与高频能量的计算精度,这说明虽然语音的主要特征信息集中于低频部分,但是中高频部分对于识别结果有影响。

2)从表 1 看到,基于 MFCC 参数的识别在四个词中都存在误识,其中“左转”和“右转”误识次数较多,这主要是由于本文是以词为识别基元,在这两个词中存在相同的音节,对最终的识别结果有影响。基于文献[7]的算法提取的特征参数虽然在识别率上提高了 2.5%,但是在“左转”和“右转”这两个词之间识别效果并没有明显的改进,并且误识一般发生在这两个词之间,即“左转”识别成“右转”,同时将“右转”识别成“左转”,这说明在滤波器组分布上简单地对三种滤波器进行组合,对识别结果有所改善,但是若音节的读音比较相近,那么与 MFCC 相比,改善效果并不明显,甚至可能加大误识数。本文算法提取的特征参数在音节读音相近如左转和右转之间,能够减小它们之间的误识数,说明本文通过 Fisher 比选择特征参数能够提高识别率,同时改进相近字节之间的识别效果。

3.2 实验二

在实验二中采用字数较多的词组来验证提出的算法的效果,这些词组为“重庆大学自动化学院”“重庆大学电气工程学院”“重庆大学机械工程学院”和“重庆大学光电学院”。相比实验一的数据,这些词组读音的相似度很高,同时字数比较多。每段语音的录制情况与实验一中相同每人录制 7 次,将其中的 80 个作为测试语音,每个词有 20 个测试样本,其余的

作为训练样本,语音识别过程其他参数的设定与实验一相同,结果如表 3 所示。

表 3 实验二三种特征参数识别结果比较

特征参数	误识总数	识别率/%
MFCC	19	76.25
文献[7]的特征参数	13	83.75
本文的特征参数	9	88.75

1)实验中语音数据的特点是语音较长,语速与正常交谈的语速相同,同时语音模板之间读音的相似度较高,仅仅只有两个字不同,因此,相比实验一,实验二中的数据更加符合日常的交谈特征,同时本文只采用声学模型进行验证,没有结合语言模型,造成三种算法的识别率都有所下降。

2)由表 3 可以得到本文的特征提取算法相比传统的 MFCC 算法和文献[7]的算法最终的识别率高,并且在没有语言模型的情况下,也可以达到 88.75%,说明不管是在孤立词还是在较长的语音段的情况下,本文算法提取的特征参数更能表征语音的特征,这主要是因为本文算法结合每一维特征参数在不同语音模板之间的方差来构造特征参数。

4 结语

本文针对经典的 MFCC 参数对于语音中高频信息描述不完全的问题,利用 Fisher 比分布选择 MFCC、IMFCC 和 MidMFCC 参数中可分程度最大的特征分量构造了新的混合参数。最后采用 GMM 进行语音的识别,结果表明,这种特征参数相比经典 MFCC 参数,在相同的环境条件下有较好的识别效果。

参考文献:

- [1] HAN J, ZHANG L, ZHENG T. Speech signal processing[M]. Beijing: Tsinghua University Press, 2004: 101 - 124. (韩纪庆, 张磊, 郑铁染. 语音信号处理[M]. 北京: 清华大学出版社, 2004: 101 - 124.)
- [2] WEI Y, ZHANG X. A PLP speech feature extraction method in noisy environment[J]. Journal of Taiyuan University of Technology, 2009, 40(3): 221 - 224. (魏艳, 张雪英. 噪声条件下的语音特征 PLP 参数的提取[J]. 太原理工大学学报, 2009, 40(3): 221 - 224.)
- [3] YIN H, XIE X, KUANG J. Acoustic features based on auditory model and adaptive fractional Fourier transform for speech recognition[J]. Acta Acustica, 2012, 37(1): 97 - 103. (尹辉, 谢湘, 匡镜明. 基于听觉模型与自适应分数阶 Fourier 变换的声学特征在语音识别中的应用[J]. 声学学报, 2012, 37(1): 97 - 103.)
- [4] ZHANG W Q, HE L, HOU T, et al. Fractional Fourier transform based auditory feature for language identification[C]// Proceedings of the 2008 IEEE Asia Pacific Conference on Circuits and Systems. Piscataway: IEEE, 2008: 209 - 212.
- [5] SANDIPAN C, ANINDYA R, SOURAV M. Capturing complementary information via reversed filter bank and parallel implementation with MFCC for improved text-independent speaker identification [C]// Proceedings of the 2007 International Conference on Computing: Theory and Application. Piscataway: IEEE, 2007: 463 - 467.

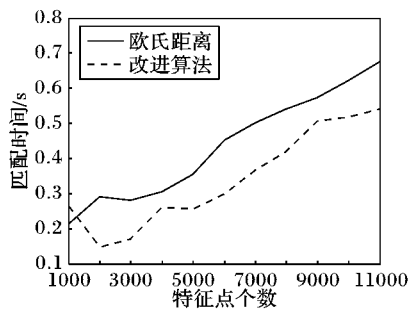


图7 两种算法关联时间对比

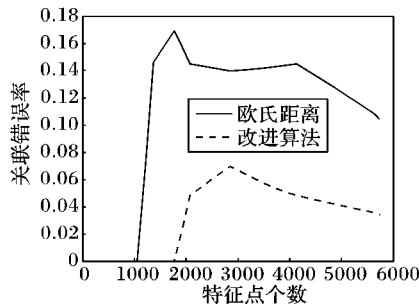


图8 两种算法的关联误差对比

4 结语

本文通过仿真实验分析了视觉 SLAM 系统中数据关联的一种改进方法,该方法将路标位置信息与欧氏距离相结合的方式对地图库路标与当前采集到的图像匹配特征点进行数据关联,该方法在关联时间和关联准确性上都有一定的改善。但是随着机器人工作范围的扩大,机器人的自身位姿和路标位置的计算都将产生一些误差,这种误差可能会对路标位置的判断产生一些影响,造成一些点关联点不在圆形范围之内,错失一些匹配点。因此,机器人自身位姿与路标位置的精确计算是值得进一步研究的问题。

参考文献:

- [1] LEONARD J J, DURRANT-WHYTE H F. Mobile robot localization by tracking geometric beacons[J]. IEEE Transactions on Robotics and Automation, 1991, 7(3): 376 - 382.
- [2] TRIEBEL R, BURGARD W. Improving simultaneous mapping and localization in 3D using global constraints[EB/OL]. [2013-06-20].

- [3] PENG L, SHI S, HONG X. A SLAM algorithm based on monocular vision and odometer[J]. Computer Simulation, 2008, 25(10): 172 - 175. (王彭林, 石守东, 洪小伟. 基于单目视觉和里程计的 SLAM 算法研究[J]. 计算机仿真, 2008, 25(10): 172 - 175.)
- [4] MIKOLAJCZYK K, SCHMID C. Performance evaluation of local descriptors[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615 - 1630.
- [5] KE Y, SUKTHANKAR R. PCA-SIFT a more distinctive representation for local image descriptors[C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2004, 2: 506 - 513.
- [6] BICEGO M, LAGORIO A, GROSSO E, et al. On the use of SIFT features for face authentication[C]// Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop. Piscataway: IEEE, 2006: 35 - 35.
- [7] DAVISON A J, MURRAY D W. Simultaneous localization and mapping using active vision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 865 - 880.
- [8] LOWE D G. Object recognition from local scale-invariant features [C]// Proceedings of the 7th IEEE International Conference on Computer Vision. Piscataway: IEEE, 1999, 2: 1150 - 1157.
- [9] DURRANT-WHYTE H, RYE D, NEBOT E. Localization of autonomous guided vehicles[M]. London: Springer, 1996: 613 - 625.
- [10] LINDBERGER T. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention[J]. International Journal of Computer Vision, 1993, 11(3): 283 - 318.
- [11] LINDBERGER T. Feature detection with automatic scale selection [J]. International Journal of Computer Vision, 1998, 30(2): 79 - 116.
- [12] F Y, SONG B, WU J. An improved scale invariant feature transform algorithm[J]. Journal of Harbin Engineering University, 2010, 31(5): 632 - 636. (付永庆, 宋宝森, 吴建芳. 边缘分类 SIFT 算法[J]. 哈尔滨工程大学学报, 2010, 31(5): 632 - 636.)

(上接第 561 页)

- [6] HAN Y, WANG C, YANG Y. Speech emotion recognition based on MFCC [J]. Journal of Chongqing University of Posts and Telecommunications: Natural Science, 2008, 20(5): 597 - 602. (韩一, 王国胤, 杨勇. 基于 MFCC 的语音情感识别[J]. 重庆邮电大学学报: 自然科学版, 2008, 20(5): 597 - 602.)
- [7] YUAN Z, XIAO W. Improved speech recognition algorithm based on MFCC feature[J]. Computer Engineering and Applications, 2009, 45(33): 108 - 110. (袁正午, 肖王辉. 改进的混合 MFCC 语音识别算法研究[J]. 计算机工程与应用, 2009, 45(33): 108 - 110.)
- [8] KANEDERA N, ARAI T, HERMANSKY H. On the importance of various modulation frequencies for speech recognition [C]// Proceedings of Eurospeech. Rodos, Greece: [s. n.], 1997, 9: 1079 - 1082.
- [9] ZHEN B, WU X, LIU Z, et al. On the importance of components of the MFCC in speech and speaker recognition[J]. Journal of Peking University: Natural Sciences, 2001, 37(5): 371 - 378. (甄

- 斌, 吴玺宏, 刘志敏, 等. 语音识别和说话人识别中各倒谱分量的相对重要性[J]. 北京大学学报: 自然科学版, 2001, 37(5): 371 - 378.)
- [10] ZHANG Y, LI X, ZHENG Y, et al. Speaker recognition based on Fisher discriminant criterion[J]. Journal of Lanzhou University: Natural Sciences, 2007, 43(2): 92 - 97. (张芸, 李昕, 郑宇, 等. 一种基于 Fisher 准则的说话人识别方法研究[J]. 兰州大学学报: 自然科学版, 2007, 43(2): 92 - 97.)
- [11] HU X, ZHANG X. Speaker recognition method based on Mel frequency cepstrum coefficient and inverted Mel frequency cepstrum coefficient[J]. Journal of Computer Applications, 2012, 32(9): 2542 - 2544. (胡雪峰, 张璇. 基于梅尔频率倒谱系数与翻转梅尔频率倒谱系数的说话人识别方法[J]. 计算机应用, 2012, 32(9): 2542 - 2544.)
- [12] HU Y, LOIZOU P C. Subjective evaluation and comparison of speech enhancement algorithms [J]. Speech Communication, 2007, 49(7/8): 288 - 601.