

DOI: 10.19650/j.cnki.cjsi.J1905510

混响环境下移动机器人语音控制方法及系统实现*

李艳生, 刘园, 张毅, 杨美美

(重庆邮电大学先进制造工程学院 重庆 400065)

摘要:为了满足移动机器人交互控制多样性需求,提升移动机器人的语音控制性能,设计基于语音的移动机器人控制系统。通过对控制信号传递过程和机器人使用环境中语音信号噪音源分析,制定出移动机器人语音控制系统组成方案,给出前端语音识别部分实现的主要流程,重点对语音增强的解混响算法进行设计。充分利用语音潜在的谱特征,将非负矩阵分解和深度神经网络结合提出一种解混响算法,先通过矩阵分解得到语音信号特征,再生成特征矢量来训练激活函数,降低深度神经网络模型的训练复杂度,经过对比分析表明该算法对解决语音混响问题具有优势。编写控制软件并嵌入语音识别算法,搭建工业移动机器人语音控制平台来验证语音控制系统有效性,在混响环境下不同人对机器人多个动作进行语音控制实验,结果表明该系统能够实现移动机器人语音控制,所提出的语音识别方法可使机器人在0.3、0.6和0.9 s的混响条件下动作正确平均执行率分别能达到96%、95%和93%以上。

关键词: 移动机器人; 控制系统; 深度神经网络; 语音识别

中图分类号: TH113 **文献标识码:** A **国家标准学科分类代码:** 460.4099

Speech control method and system realization of mobile robot in reverberation environment

Li Yansheng, Liu Yuan, Zhang Yi, Yang Meimei

(School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: In order to meet the diversity requirements of mobile robot interaction control and improve the speech control performance of mobile robot, a speech-based mobile robot control system is designed. Through analyzing the control signal transmission process and speech signal noise source of the robot in using environment, the composition scheme of the mobile robot speech control system is made up. The main flow for the implementation of front end speech recognition section is given. And the de-reverberation algorithm for speech enhancement is emphatically designed. By fully utilizing the potential spectral features of speech, a de-reverberation algorithm is proposed based on combined non-negative matrix factorization and deep neural network. Firstly, the speech signal features are obtained through matrix decomposition, and then the feature vector is generated to train the activation function, which reduces the training complexity of the deep neural network model. The comparison analysis shows that the proposed algorithm possesses superiority in solving the speech reverberation problem. The control software was written, which was embedded in the speech recognition algorithm. A speech control platform of industrial mobile robot was built to verify the effectiveness of the speech control system. In the reverberation environment, speech control experiments were conducted, in which different people performed multiple actions on the robot. The results show that the system can realize the speech control of mobile robots. The proposed speech recognition method can achieve the average correct execution rates of actions of 96%, 95% and 93%, respectively for the mobile robot under the reverberation conditions of 0.3, 0.6 and 0.9 s.

Keywords: mobile robot; control system; deep neural network; speech recognition

0 引言

近年来,随着移动机器人技术的迅速发展,移动机器人产品得到广泛应用,在生活领域中居室内的家庭移动机器人人们的生活带来极大便利,如扫地机器人和娱乐机器人等,而用来运送货物的工业移动机器人可作为智能生产工具,提高了企业的生产效率和制造水平,其主要分为有轨式和无轨式两种。传统的移动机器人控制需要预先编写程序并载入控制系统,通过面板屏幕或机械按钮来交互控制,要求手眼紧密协作,使操控复杂繁琐,控制人员容易疲劳^[1]。而语音控制与传统的控制方式不同,不需要控制人员接触机器人机械本体,通过语音就可以在相对较大的空间范围内下达控制命令,解放人们的双手,尤其是对于老年和残障人士,语音提供了一种较好的控制方式^[2]。随着人工智能技术的兴起,移动机器人传统的按照预编程和机械面板按钮控制方式,已经不能满足现代人们生活生产的需要,而移动机器人语音控制渐渐成为移动机器人控制中热门的研究方向,这种控制方式已经被不同的人群广泛接受^[3-4]。

移动机器人语音控制系统的设计首先要识别语音信号,提高语音的易懂度和听觉质量,增强语音识别的鲁棒性,而传统的语音信号处理方法,在低信噪比非稳定环境下存在噪声冗余的缺点^[5-6]。因此,为了提高语音增强技术的适用范围,学者们不断研究新的语音增强算法^[7]。其中非负矩阵分解算法(non-negative matrix factorization, NMF)利用语音信号自身的非负性质,能够描述语音信号潜在特征的高质量字典原子,是语音降噪的常用方法^[8-10]。近年来,基于深度神经网络的语音处理方法受到学者关注,深度网络架构能够很好地映射输入语音和增强语音之间的非线性关系,但是此类方法需要大量训练样本且需要较高的模型匹配度,否则去混响效果会受到很大的影响^[11-12]。随着语音识别方法的不断创新,移动机器人的语音控制技术也在发展,张汝波等^[13]利用聚类方法和离散隐马尔可夫模型对语音控制命令进行识别,在履带移动机器人上完成了有效性验证,王晓华等^[14]将小波包变换方法引入到语音识别的特征参数提取中,在机器人操作系统(robot operating system, ROS)下实现了移动机器人语音控制。由于深度学习需要大量训练样本,实现语音识别方法与移动机器人控制技术完美融合还需要深入研究。

为了满足机器人交互控制多样性需求,提升移动机器人的语音控制性能,本文设计了一种移动机器人控制用的语音系统,首先分析语音控制需要解决的主要问题,制定整个控制系统的设计方案,重点针对语音混响问题,结合非负矩阵分解和深度神经网络特点来提出语音处理算法,给出其主要实现步骤,并通过对不同方法和指标的对

比分析来证明算法的先进性。最后编制语音控制软件,搭建工业移动机器人语音控制系统,实验证明所提出方法和设计系统的有效性。

1 语音控制系统方案

移动机器人语音控制系统在控制人员和机器人底层电机控制器之间起到连接和转换的作用,在控制过程中系统首先准确识别出人的语音指令,然后将指令转换成机器人可直接执行的运动电信号(电压或脉冲等),通底层控制器根据运动电信号驱动底盘电机运转,最终实现机器人位置和方向按照语音指令改变。其中前端的语音识别子系统是影响整个移动机器人语音控制系统性能的关键环节,因为人发出的语音在空气中传播,机器人的麦克风传感器接收到语音信号会受到噪声污染,降低了移动机器人语音指令的正确识别率,影响机器人动作的有效执行。

移动机器人控制人员的语音信号在传播过程中会受到自身反射信号叠加产生混响,是噪声的主要来源。混响信号实际模型如图1所示,信号源主要有机器人语音控制信号 S ,噪声信号 n_1 和干扰信号 n_2 ,传播过程中遇到物体表面会发生反射与原始信号混叠。在相对封闭的环境下可以只考虑语音控制信号的混响噪声。如家庭机器人的使用环境在室内,空间较小,语音信号传播会遇到家具和墙壁反射生成混响。又如工业移动机器人的工作区域较大,工厂内机器种类多,噪声混杂,这种情况下很难实现远距离语音信号控制,但可以将移动机器人语音控制系统的语音识别部分和机器人本体分离,将麦克风传感器放置在控制人员的工作室内,传感器输出信号通过无线网络输送给机器人本体,这样可以避免在长距离空气传播中语音信号衰减和工厂内机器强噪声干扰。但是在相对封闭的控制室内,与家庭机器人的室内应用环境相似,语音信号会受到自身信号的反射叠加污染产生混响。

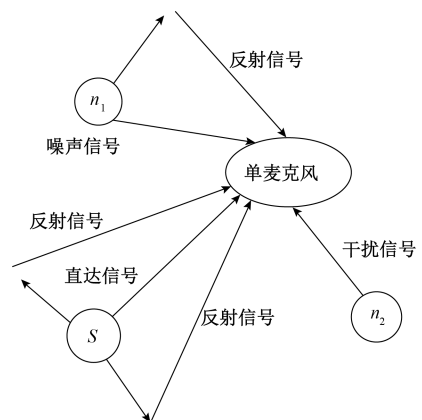


图1 混响信号模型

Fig.1 Reverberation signal model

本文设计的移动机器人语音控制系统主要分为前端语音处理识别部分和后端机器人移动控制部分,移动机器人语音控制系统整体方案如图 2 所示,前端语音信号从输入到输出,包括信号采集、信号处理、语音增强和语音识别 4 个过程。信号采集通过麦克风传感器实现,采集的信号经过一阶高通数字滤波器预加重,提升语音信号的高频分辨率,同时对低频信号部分进行过滤,再经交叠分段来分帧并加窗平滑语音帧的边缘,然后通过端点检测来确定一段语音信号的中语音的起始点和结束点的位置。经过预处理的信号解混响实现语音增强,再提取语音特征和与声学模型比对,输出识别结果。后端移动机器人移动控制部分将前端语音识别结果转换成的运动控制器输入信号,驱动底盘电机实现机器人运动。系统方案中前端语音解混响增强过程是本系统设计的关键环节,将在下文详细阐述。

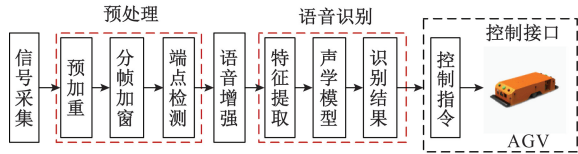


图 2 移动机器人语音控制系统方案

Fig.2 Scheme diagram of the mobile robot speech control system

2 语音解混响算法设计

传统的 NMF 语音去噪算法虽然能很好地提取语音信号的潜在特征,但由于该算法是线性的,无法很好地描述混响信号到纯净语音之间复杂的非线性关系^[15]。而基于深度神经网络(deep neural network, DNN)的语音解混响算法语音解混响质量高,但深度神经网络没有利用语音信号潜在的频谱结构,并且需要足够大的样本,增加了网络模型训练的复杂度^[16]。因此,结合深度神经网络与 NMF 优点,提出一种基于 DNN-NMF 的语音解混响算法,嵌入到本文设计的移动机器人语音控制系统中。

2.1 语音信号的 NMF 处理

NMF 算法能够将混响信号投影到较低维度,降低深度神经网络训练的复杂性,并且 NMF 算法能够获得语音信号的潜在特征,使 DNN 的训练能获得更好的初始值。先用 NMF 算法对语音信号进行预处理,将语音信号经过短时傅里叶变换得到幅度谱矩阵,然后将其分解为字典矩阵 \mathbf{W} 和系数矩阵 \mathbf{X} 乘积,同时保证分解前后的误差最小,矩阵分解过程的目标函数如下所示:

$$(\mathbf{W}, \mathbf{X}) = \underset{\mathbf{W}, \mathbf{X}}{\operatorname{argmin}} D(\mathbf{V} \parallel \mathbf{W}\mathbf{X}), \text{ s.t. } \mathbf{W} \geq 0, \mathbf{X} \geq 0 \quad (1)$$

其中, $D(\mathbf{V} \parallel \mathbf{W}\mathbf{X})$ 表示矩阵 \mathbf{V} 与字典矩阵和系数矩

阵乘积 $\mathbf{W}\mathbf{X}$ 之间距离的函数。

最小化目标代价函数时首先初始化 \mathbf{W} 和 \mathbf{X} , 初始值通过随机生成非负矩阵生成, 然后使用乘性更新规则对 \mathbf{W} 和 \mathbf{X} 进行更新直至收敛, 更新规则为:

$$\begin{cases} \mathbf{X} \leftarrow \mathbf{X} \otimes \frac{\mathbf{W}^T \mathbf{V}}{\mathbf{W}^T \mathbf{E}} \\ \mathbf{W} \leftarrow \mathbf{W} \otimes \frac{\mathbf{V}}{\mathbf{E}\mathbf{X}^T} \end{cases} \quad (2)$$

式中: \otimes 表示矩阵或向量之间的点乘; \mathbf{E} 表示元素值均为 1 的矩阵; $(\cdot)^T$ 表示矩阵的转置。

通过 NMF 乘性更新规则得到纯净语音信号的字典矩阵 \mathbf{W}_s , 并将其保存下来作为增强阶段的先验信息; 然后通过 NMF 对混响信号进行乘性更新迭代得到混响信号的字典矩阵 $\hat{\mathbf{W}}(r, t)$ 和激活系数矩阵 $\hat{\mathbf{X}}(r, t)$; 最后将激活系数矩阵 $\hat{\mathbf{X}}(r, t)$ 用作深度神经网络的输入特征, 由于语音信号存在时间动态特性, 在获得每个信号帧 t 的激活系数矩阵后, 建立一个以时帧 t 为中心, 宽度为 $(2d+1)$ 的系数矩阵向量, 将该向量作为激活系数对 DNN 进行训练。

2.2 激活系数的 DNN 训练

DNN 架构的训练可以将通过 NMF 得到的激活系数矩阵进行再优化, 学习获得混响信号激活系数和纯净语音激活系数之间的非线性映射函数。该 DNN 架构由 3 个 Sigmoid 隐藏层、1 个 Softmax 输出层和 1 个语音信号重构层组成。输出层用来得到训练获得的激活系数矩阵 $\hat{\mathbf{X}}(r, t)$, 语音信号重构层用来获得增强得到的语音信号 $\hat{\mathbf{S}}(k, t)$, 表示为:

$$\hat{\mathbf{S}}(k, t) = \mathbf{G}(k, t) \otimes \mathbf{Y}_h(k, t) = \frac{\sum_r \mathbf{W}_s(k, r) \hat{\mathbf{X}}(r, t)}{\sum_{r, \tau} \mathbf{H}(k, \tau) \mathbf{W}_s(k, r) \hat{\mathbf{X}}(r, t - \tau)} \otimes \mathbf{Y}_h(k, t) \quad (3)$$

DNN 网络训练的目标函数用增强语音 $\hat{\mathbf{S}}(k, t)$ 与纯净语音 $\mathbf{S}(k, t)$ 之间的最小均方差表示。

$$\mathbf{J} = \sum_{k, t} [\mathbf{S}(k, t) - \hat{\mathbf{S}}(k, t)]^2 \quad (4)$$

权值修正的方法是根据误差来反向传播的, 误差是统计的平方和项累积, 估计网络权重的成本函数 \mathbf{J} 的梯度为:

$$\frac{\partial \mathbf{J}}{\partial \mathbf{W}} = \frac{\partial \mathbf{J}}{\partial \hat{\mathbf{S}}} \frac{\partial \hat{\mathbf{S}}}{\partial \mathbf{X}} \frac{\partial \mathbf{X}}{\partial \mathbf{W}} \quad (5)$$

网络参数经式(5)的迭代更新, 可训练得到一个深度神经网络来实现语音解混响。

2.3 算法框架

综上所述, 所提出的 DNN-NMF 算法框架如图 3 所示。先通过 NMF 乘性更新算法得到纯净语音信号的字典

典矩阵,建立语音信号的字典矩阵 W_s ;通过 NMF 对混响信号进行乘性更新得到混响信号的字典矩阵 $\tilde{W}(r,t)$ 和激活系数矩阵 $\tilde{X}(r,t)$;然后通过深度递归网络获得混响信号 NMF 激活系数与干净语音等效激活系数之间的非线性映射函数,以此获得优化后的混响信号激活系数矩阵 $\hat{X}(r,t)$;最后使用纯净语音字典 W_s 和优化后的激活系数矩阵 $\hat{X}(r,t)$ 重构语音信号。

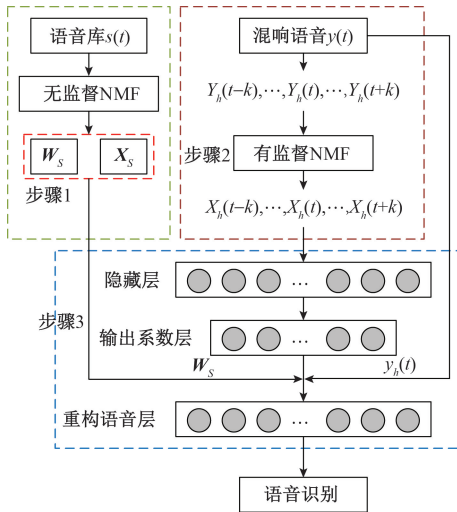


图3 解混响算法

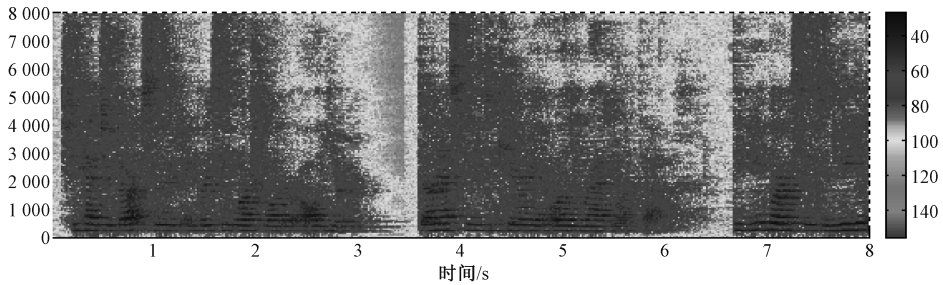
Fig.3 Dereverberation algorithm

算法中字典矩阵 W_s 和系数矩阵 X_s 的初始值使用随机生成非负矩阵的方法获得,迭代更新次数设为 200,非

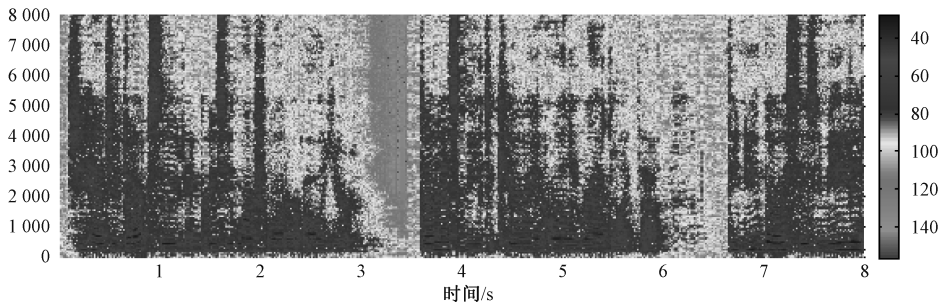
负字典原子个数 $R=1\ 000$,因此 NMF 字典矩阵的维数设置为 $257 \times 1\ 000$ (频率 \times 基数)。步骤 2 中混响信号字典矩阵维数设置与步骤 1 相同。步骤 3 中 DNN 训练模型里 d 设置为 5,即输入特征选取以当前帧 t 为中心的 11 个帧的上下文窗口进行提取,隐藏层设置为 3 层,每层 1 024 个节点,DNN 训练模型的初始参数由无监督的受限玻尔兹曼机(restricted Boltzmann machines, RBM)网络逐层训练得到,DNN 使用每帧维度为 1 000 的 NMF 激活系数向量进行训练,即激活系数矩阵的维度为 $1\ 000 \times 11$ 。为了证明所提出算法抑制混响的有效性,混响语音信号的语谱和基于 DNN-NMF 的解混响语如图 4 所示,从图 4 中可以看出 DNN-NMF 算法能够很好地处理混响信号的拖尾现象,增强语音信号的语谱特征。

2.4 算法对比分析

为了验证所提出的 DNN-NMF 算法解混响优势,选用 DNN、NMF、联合长短期记忆网络(long short-term memory network, LSTM)和 NMF 的算法^[17](LSTM-NMF)进行对比。其中 DNN 是输入为混响信号谱,包括 1 个输入层,3 个隐藏层,1 个输出层的深度训练模型;NMF 是 $R=1\ 000$,迭代次数为 200 的稀疏表示模型;LSTM-NMF 是首先通过 LSTM 网络模型重构语音信号,之后再通过 NMF 对重构语音信号进行后处理的深度训练模型。DNN 和 LSTM-NMF 均是在语音信号上直接进行训练的算法。进行仿真实验来表明各个算法性能,采用相同类型和相同数量的训练集和测试集,对每种算法进行处理。



(a) Reverberation speech spectrum



(b) De-reverberation speech spectrum

图4 语音谱对比

Fig.4 Comparison of the speech spectrums

评价语音处理算法的性能指标可从语音识别和语音增强两个角度进行评价,算法仿真选用从语音增强的角度进行评价主要考虑以下几点原因。

1) 机器人具有仿生拟人的特性,其语音控制系统要适应不同工作场景,语音中的噪声来源不仅限于混响,而纯音信号具有语义最本质的特征,为了保证移动机器人的广泛适用和拟人性,设计的算法以提取语音纯净程度为重要指标。

2) 纯音信号的识别相对容易,其准确识别是机器人听懂语音命令的最基本能力,假设纯音信号可准确识别的条件下,利用谱特征与深度神经网络结合,将带混响的语音信号先映射为纯净语音信号来进行识别,映射输出的语音质量可在一定程度上反映出识别准确率。

3) 若仅从语音识别角度来评价算法性能,使用语音识别的错词率(word error rate, WER)不能反映出纯音信号映射输出的中间结果,无法体现机器人仿生拟人性和应用潜力,所以先通过算法仿真从语音增强角度进行评价,混响环境的语音识别率通过后面移动机器人语音控制系统进行实验测试。

为了突出机器人的仿生拟人性和证明算法的去噪映射能力,采用评价主观音质的语音增强性能指标^[18]。图 5(a)所示为选用语音质量的感知评估(perceptual evaluation of speech quality, PESQ)评价效果,图 5(b)所示为选用频率加权分段语音信噪比(frequency-weighted segmented speech signal-to-noise ratio, FWSSSNR)评价效果,图 5(c)所示为选用倒谱距离(cepstrum distance, CD)评价效果。其中 PESQ 和 FWSSSNR 指标是数值越高,表示语音增强的性能越好,CD 指标是数值越低,表示语音增强的性能越好。

图 5(a)中 PESQ 指标结果显示在混响时间 $T_{60} \leq 0.6$ s 时 DNN 和 NMF 算法对 PESQ 值并无影响, LSTM-NMF 和 DNN-NMF 算法对 PESQ 值的影响也不明显,这是因为短期混响不会导致音质的显著下降,这与早期混响有助于提高音质的结论是一致的,但是随着混响时间的增长,当 $T_{60} = 0.9$ s 时, DNN、NMF、LSTM-NMF、DNN-NMF 均提高了 PESQ 值,提高了解混响的性能,但相对其他算法,本文提出的 DNN-NMF 算法能更好地提高 PESQ 的值。图 5(b)中 FWSSSNR 指标结果显示 DNN 算法在 $T_{60} = 0.3$ s 和 $T_{60} = 0.6$ s 时的评估度量值不好,这是由 FWSSSNR 评估度量的性能引起的,因为 DNN 的其他评价指标值的结果较好,而 LSTM-NMF 和 DNN-NMF 能很好地提高 FWSSSNR 值。从图 5(c) CD 指标中可以看出 DNN-NMF 的值优于 LSTM-NMF。对比分析可以得出本文提出的 DNN-NMF 算法的解混响性能最优。

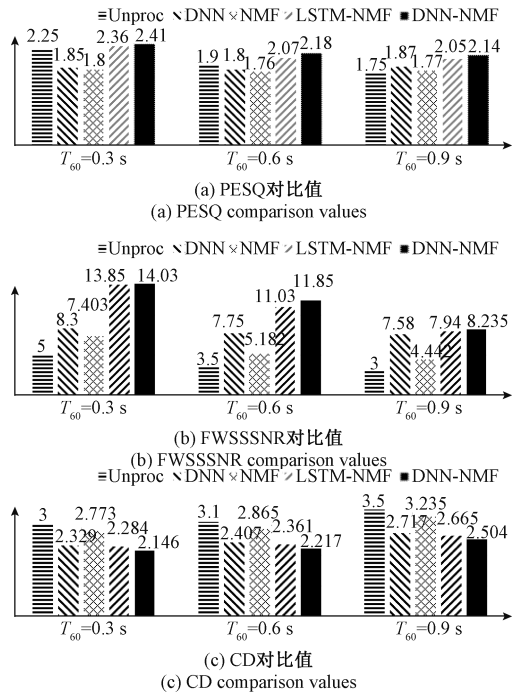


图 5 不同算法评价指标值对比

Fig.5 Comparison of the evaluation index values among different algorithms

3 系统实现与实验

为了证明本文方法和仿真分析的有效性,搭建平台进行实验来验证十分必要^[19],基于提语音控制系统方案,搭建移动机器人控制平台。控制对象为工业移动机器人,移动机构包括 2 个标准轮和 1 个万向轮,底层控制器通过驱动标准轮实现是机器人运动。机器人下层放置主控板、激光传感器,上层安装电池。语音控制系统的前端语音识别部分的实现是通过在上位机采用 C++ 编写软件采集语音和信号处理,嵌入了本文提出的解混响算法,软件先采集语音信号进行模型训练,保存最优参数后再进行语音命令识别,最后将识别的结果通过无线局域网发布给机器人下位机。后端机器人移动控制部分中,下位机的语音模块接收到语音识别结果,将控制命令传给机器人控制板仲裁,语音信号的优先级低于碰撞传感器、激光避障传感器和超声波传感器信号,最后按照控制命令驱动电机运动。

在封闭的室内环境中采集语音信号,混响时间分别约为 0.3、0.6 和 0.9 s,应用语音控制系统对机器人进行前进、后退、左转和右转运动控制实验,机器人每个动作执行的实验数据均由 20 个男生和 20 个女生语音控制完成,不同算法下的机器人每个动作正确执行率如表 1 所示。对机器人语音控制的实验数据进行量化分析,对比

可知在 0.3 s 混响条件下 4 种语音算法都能使机器人的动作正确平均执行率达到 90% 以上,而在大混响环境下,机器人动作正确平均执行率都有所下降。在 0.6 s 混响条件下 NMF 算法的机器人动作正确平均执行率已经将低到 90% 以下,在 0.9 s 混响条件下 DNN 算法的机器人动作正确平均执行率已经接近于 90%,而 LSTM-NMF 算法和 DNN-NMF 算法仍然能保持较好的效果,特别的 DNN-NMF 解混响算法在 0.3、0.6 和 0.9 s 的混响条件下,机器人的正确动作平均执行率分别能达到 96%、95% 和 93% 以上,高于 LSTM-NMF 算法,因此本文提出的结合语音谱特征和深度神经网络的 DNN-NMF 解混响算法在机器人控制应用上效果较好。

表 1 机器人动作正确执行率

Table 1 Correct execution rates of the robot actions

混响时间/s	语音指令	DNN/%	NMF/%	LSTM-NMF/%	DNN-NMF/%
0.3	前进	94.78	91.15	95.09	97.13
	后退	92.22	89.07	96.34	96.90
	左转	93.13	92.74	94.54	96.54
	右转	95.79	90.49	97.02	98.18
	停止	94.55	88.25	95.44	96.22
0.6	前进	93.55	88.76	94.05	96.31
	后退	91.14	88.64	93.72	93.04
	左转	93.02	90.62	95.31	96.30
	右转	93.58	89.33	94.01	95.41
	停止	91.89	88.01	94.84	95.19
0.9 s	前进	92.15	86.36	92.87	94.22
	后退	90.44	87.86	90.92	91.21
	左转	91.32	88.92	93.66	94.11
	右转	90.38	88.45	92.62	93.41
	停止	90.09	87.81	93.21	93.77

4 结 论

本文制定出移动机器人语音控制系统的整体方案。通过对控制信号传递过程和机器人应用环境分析,将系统分解成前端语音识别和后端移动机器人运动控制两部分,并指出语音处理部分的解混响算法设计是语音控制系统实现的关键环节。本文提出一种基于非负矩阵分解和深度神经网络的语音解混响算法。该算法能够利用语音信号的潜在谱特征降低深度神经的的训练复杂度,给出了算法实现的主要步骤,经对比分析证明了本文算法解混响效果最优。本文实现了移动机器

人语音控制系统。以工业移动机器人作为控制对象,设计语音控制软件,嵌入解混响算法,通过多组实验测试表明,本文语音处理算法控制效果较好,通过语音控制机器人的动作执行率高。

参考文献

- [1] 黄晓辰,张小俊,孙凌宇,等.核电站多功能水下打捞机器人系统研究[J].仪器仪表学报,2017,38(3):552-559.
HUANG X CH, ZHANG X J, SUN L Y, et al. Research on the reactor multi function robot for object salvaging underwater[J]. Chinese Journal of Scientific Instrument, 2017,38(3):552-559.
- [2] 刘策,刘小峰.助老服务机器人多模式友好交互系统设计与实现[J].电子测量与仪器学报,2019,33(1):177-182.
LIU C, LIU X F. Design and implementation of multi-mode friendly interactive system for elderly assistance robot [J]. Journal of Electronic Measurement and Instrumentation,2019,33(1):177-182.
- [3] 房爱青,李长勇,江宏,等.基于语音的移动机器人车体控制系统设计[J].现代制造技术与装备,2016(7):31-32.
FANG AI Q, LI CH Y, JIANG H, et al. Design of mobile robot body control system based on speech [J]. Modern Manufacturing Technology and Equipment, 2016(7):31-32.
- [4] SAON G, KUO H K J, RENNIE S, et al. The IBM 2015 english conversational telephone speech recognition system [J]. Eurasip Journal on Advances in Signal Processing, 2015, 2008(1): 1-15.
- [5] ZHANG L, XU X, CHEN H, et al. Supervised single-channel speech dereverberation and denoising using a two-stage model based sparse representation [J]. Speech Communication, 2018, 97: 1-8.
- [6] 张雄伟,李轶南,郑昌艳,等.语音去混响技术的研究进展与展望[J].数据采集与处理,2017,32(6):1069-1081.
ZHANG X W, LI Y N, ZHENG CH Y, et al. Speech dereverberation: Review of state-of-the-arts and prospects [J]. Journal of Data Acquisition and Processing, 2017, 32(6): 1069-1081.
- [7] 刘文举,聂帅,梁山,等.基于深度学习语音分离技术的研究现状与进展[J].自动化学报,2016,42(6):819-832.
LIU W J, NIE SH, LIANG SH, et al. Deep learning based speech separation technology and its developments[J]. Acta Automatica Sinica, 2016, 42(6): 819-832.
- [8] MOHAMMADIHA N, DOCCLO S. Speech dereverberation using non-negative convolutive transfer function and

- spectro-temporal modeling [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(2): 276-289.
- [9] 李艳生,刘园,张毅.基于感知掩蔽的重构非负矩阵分解单通道语音增强算法[J].计算机应用,2019,39(3):894-898.
LI Y SH, LIU Y, ZHANG Y. Reconstructed NMF single channel speech enhancement algorithm based on perceptual masking [J]. Journal of Computer Applications, 2019, 39(3): 894-898.
- [10] MOHANAN N, VELMURUGAN R, RAO P. Speech dereverberation using NMF with regularized room impulse response [C]. IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2017: 4955-4959.
- [11] ZHAO Y, WANG D L, XU B, et al. Late reverberation suppression using recurrent neural networks with long short-term memory [C]. IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2018: 5434-5438.
- [12] 韩伟,张雄伟,闵刚,等.基于感知掩蔽深度神经网络的单通道语音增强方法[J].自动化学报,2017,43(2):248-258.
HAN W, ZHANG X W, MIN G, et al. A single-channel speech enhancement approach based on perceptual masking deep neural network [J]. Acta Automatica Sinica, 2017, 43(2): 248-258.
- [13] 张汝波,刘冠群,吴俊伟,等.移动机器人语音控制技术研究与实现[J].华中科技大学学报(自然科学版),2013,41(S1):348-351
ZHANG R B, LIU G Q, WU J W, et al. Research and implementation on speech control technology for mobile robot [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2013, 41(S1): 348-351.
- [14] 王晓华,要鹏超,广夏桐,等.基于 ROS 的移动机器人语音控制设计与实现[J].传感器与微系统,2019,38(5):100-103.
WANG X H, YAO P C, GUANG X T, et al. Design and implementation of speech control for mobile robot based on ROS [J]. Transducer and Microsystem Technologies, 2019, 38(5): 100-103.
- [15] CHUNG H, PLOURDE E, CHAMPAGNE B. Single-channel enhancement of convolutive noisy speech based on a discriminative NMF algorithm [C]. IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2017: 2302-2306.
- [16] HOU J C, WANG S S, LAI Y H, et al. Audio-visual speech enhancement based on multimodal deep convolutional neural network [J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, 2(2): 117-128.
- [17] 刘斌,陶建华.联合长短时记忆递归神经网络和非负矩阵分解的语音混响消除方法[J].信号处理,2017,33(3):268-272.
LIU B, TAO J H. A research to speech dereverberation method based on BLSTM recurrent neural networks and non-negative matrix factorization [J]. Journal of Signal Processing, 2017, 33(3): 268-272.
- [18] 周伟力,贺前华,李艳雄,等.复杂环境下基于干净语音的音质评价方法[J].华中科技大学学报(自然科学版),2016,44(7):121-126.
ZHOU W L, HE Q H, LI Y X, et al. Quasi-clean speech-based speech quality evaluation method under complex environments [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2016, 44(7): 121-126.
- [19] 李艳生,杨美美,芦娜,等.同心连接器的力学性能测试技术和系统研究[J].仪器仪表学报,2017,38(2):328-335.
LI Y SH, YANG M M, LU N, et al. Research on testing technology and system of mechanical properties for a concentric connector [J]. Chinese Journal of Scientific Instrument, 2017, 38(2): 328-335.

作者简介



李艳生 (通信作者), 分别在 2009 年和 2016 年于北京邮电大学获得学士学位和博士学位, 现为重庆邮电大学副教授, 主要研究方向为机器人技术和智能制造技术。

E-mail: liyansheng@cqupt.edu.cn

LI Yansheng (Corresponding author) received his B. Sc. degree in 2009 and Ph. D. degree in 2016 both from Beijing University of Posts and Telecommunications, respectively. Now, he is an associate professor in Chongqing University of Posts and Telecommunications. His main research interests include robotics and intelligent manufacturing technology.



刘园, 在 2015 年于河南工程学院获得学士学位, 现为重庆邮电大学硕士研究生, 主要研究方向语音处理和机器人控制技术。

E-mail: 1767237325@qq.com

Liu Yuan received her B. Sc. degree in 2015 from Henan University of Engineering. Now, he is a M. Sc. candidate in Chongqing University of Posts and Telecommunications. Her main research interests include speech processing and robot control technology.