

面向防疫的智能导诊机器人关键技术及应用

王腾¹, 潘晶², 董璐³, 孙长银¹

(1. 东南大学自动化学院, 江苏 南京 210018;

2. 上海钛米机器人股份有限公司, 上海 201203;

3. 同济大学电子与信息工程学院, 上海 201804)

摘要: 疫情防控期间, 智能机器人在医疗、巡检和配送等领域走上防疫一线, 成为抗疫战斗中一股极为重要的力量, 使得更多的人意识到智能机器人在解决公共卫生安全问题中的重要性。重点关注面向防疫的智能导诊机器人, 首先探讨智能导诊机器人的关键技术及其研究现状, 包括机器人消毒、人脸识别、语音交互、自主定位、路径规划等; 然后以钛米智能导诊机器人为应用案例, 阐述上述关键技术在实际应用中的落脚点; 最后对智能导诊机器人的发展趋势做出展望。

关键词: 机器人消毒; 人脸识别; 语音交互; 自主定位; 路径规划

中图分类号: TP242.6

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202119

Key technologies and applications of intelligent guiding robots for epidemic prevention

WANG Teng¹, PAN Jing², DONG Lu³, SUN Changyin¹

1. School of Automation, Southeast University, Nanjing 210018, China

2. Taimi Robotics Technology Co., Ltd., Shanghai 201203, China

3. School of Electromechanical and Information Engineering, Tongji University, Shanghai 201804, China

Abstract: During the period of epidemic prevention, intelligent robots play a significant role in the fields of medical treatment, inspection and distribution, making more people realize the importance of robots in solving public health and safety problems. Focusing on the intelligent guiding robots, the key technologies and research status of the intelligent guiding robots were discussed, including robot disinfection, face recognition, speech interaction, autonomous localization, path planning. Then, taking TMiRob intelligent guiding robot as an example to discuss how these key technologies were implemented in reality. Finally, the challenges and development trends of intelligent guiding robots were presented.

Key words: robot disinfection, face recognition, speech interaction, autonomous localization, path planning

1 引言

机器人的研发、制造和应用是衡量一个国家科技创新、高端制造水平以及综合国力的重要标志。现阶段, 美国、日本、德国等国家已将机器人作为一个战略高新技术并给予支持, 旨在大力发展颠覆性机器人技术和系统, 确保其在机器人领域的领导

地位, 如美国于 2011 年正式推出《国家机器人计划》, 日本于 2015 年公布了《机器人新战略》, 欧盟于 2016 年启动了《欧盟机器人研究计划》。与此同时, 我国于 2015 年部署了《中国制造 2025》, 明确指出需加快下一代机器人技术研发, 抢占机器人技术及产业发展的下一个制高点。机器人技术最早被应用于工业领域。随着人工智能、物联网、传

收稿日期: 2020-10-31; 修回日期: 2020-11-16

通信作者: 王腾, wangteng@seu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61803084, No.62041301)

Foundation Items: The National Natural Science Foundation of China (No.61803084, No.62041301)

感器以及 5G 通信技术的迅猛发展,当今机器人的用途越来越广泛,其开始从传统的工业领域向军事、医疗、服务等领域渗透,同时在城市安全、国防和公共服务等领域扮演着越来越重要的角色^[1-4]。

在新型冠状病毒肺炎疫情防控期间,减少人与人之间的接触成为防止疫情扩散的有效手段之一。在此背景下,利用机器人代替人工作业的需求量急剧增加,智能机器人在导诊、巡检、消毒和配送等领域走上防疫一线,发挥了重要作用。如智能导诊机器人在医院内部提供导诊分诊、就医流程引导、医疗知识宣传等工作;巡检机器人在公共场所开展人体测温、定点巡逻等任务;消毒机器人在医院、商场和户外等场所开展消毒工作;配送机器人将食物、商品和物资等送到指定地点。智能机器人的使用大大降低了人与人发生交叉感染的风险,让更多的人意识到智能机器人在解决公共卫生安全问题中的重要性。

本文重点关注面向防疫的智能导诊机器人的关键技术与应用。近年来,我国的服务机器人行业得到快速发展,其功能更完备、性能更稳定,因此有必要对智能导诊机器人的现状及趋势进行深入和详细的讨论和剖析。本文首先对面向防疫的智能导诊机器人的关键技术及研究现状进行梳理讨论;然后,结合应用案例讨论这些关键技术在实际应用中的落脚点;最后提出当前面向防疫的智能导诊机器人的关键技术的难点和面临的挑战,并给出其发展趋势。

2 关键技术及发展现状

面向防疫的智能导诊机器人主要研究如何借助人工智能、物联网、运动控制等技术,使机器人具备导诊、流行病史排查、语音互动、问路导航等功能,从而引导患者顺利就医。面向防疫的智能导诊机器人主要包含以下关键技术:机器人消毒、人脸识别、语音交互、自主定位、路径规划等。本节旨在对上述关键技术及其发展现状进行深入讨论、剖析及总结。

2.1 机器人消毒

在工作期间,与新型冠状病毒感染者的直接接触可能会导致智能导诊机器人成为病毒的传播者。为了降低交叉感染的风险,现阶段的防疫机器人主要从材料设计和机器人消毒两个方面降低病毒附着的可能性。从设计出发,英国涂层研发公司设计了一种可以破坏新型冠状病毒的抗微生物涂层,帮助清除机器人

表面的细菌。在机器人消毒方面,目前主要采用汽化过氧化氢消毒、紫外线消毒、等离子空气消毒等方式。例如,英国 Bioquell 公司推出了一系列消毒机器人产品,采用过氧化氢蒸汽对一定范围进行消毒;上海颂通科技有限公司推出的智能消毒机器人 KINO 采用过氧化氢高温闪蒸技术使被消毒表面形成微冷凝,从而达到消灭细菌的效果;丹麦“蓝色海洋”机器人公司设计了一款使用短波紫外线(UV-C)进行消毒的移动机器人,其能够自主移动到达室内的各个角落,杀灭 99.99% 的细菌;麻省理工学院研发的 UV-C 机器人的 4 个垂直光束会释放 UV-C,可在 30 min 内杀死 90% 的新型冠状病毒颗粒。需要注意的是,由于紫外线会对人体健康造成一定的危害,上述基于紫外线的消毒方法需在特定空间中进行,无法实现人机共存。因此,如何在人流密集的医院环境,实现快速、高效、安全的机器人消毒是面向防疫的智能导诊机器人的关键技术之一。

2.2 人脸识别

人脸识别是智能导诊机器人对患者进行身份认证的关键技术。人脸识别的本质是设计鲁棒的人脸特征描述,将待识别的人脸特征与大规模数据库进行比对,实现实时验证。现实生活中,光照、姿态、表情、年龄的变化,遮挡、伪装及图像成像质量等问题给人脸识别带来了很大的挑战,如图 1 所示。对人脸特征进行显性描述非常困难,而神经网络可通过学习的方式获得特征的隐性表达,适应性更强。因此,现阶段大量文献采用深度学习技术来学习鲁棒的多层次人脸特征。目前基于深度学习的人脸识别主要围绕神经网络结构设计及损失函数设计两方面展开研究^[5-14]。与传统算法相比,上述基于深度学习的人脸识别算法在公开数据集上的检测精度得到了很大提升。

但是值得一提的是,疫情期间,行人在公共场所需佩戴口罩来防止交叉感染。佩戴口罩导致人脸的大部分关键点受到遮挡,如图 1(e) 所示,使得现阶段基于深度学习的人脸识别算法精度大幅度下降。针对上述挑战,国内外多家科研机构进行了深入研究,在识别算法及数据集生成策略方面提出了一系列有效的解决方案。针对人脸识别算法,百度采用空间注意力机制进行特征学习,使算法着重关注未被遮挡且包含更多身份信息的眼部区域特征;北京旷视科技有限公司单独训练了一个人眼识别网络,并通过将人眼识别模型与

原有的人脸识别模型进行融合来减小佩戴口罩所产生的影响；京东在胶囊网络框架下^[15]，通过引入注意力机制，使人脸识别集中在眼睛和眉毛等未被遮挡部分；华为提出一种影像重构技术，可通过图像重构网络将佩戴装饰（如眼镜、口罩、帽子等）的人脸图像重构为未佩戴装饰的人脸图像，并利用常规人脸识别算法对重构后的图像进行识别；北京的卢深视科技有限公司采用全局特征与局部特征相结合的方法，充分利用未被遮挡部分的三维信息进行三维人脸识别；南京大学工程管理学院针对面部遮挡，提出了非凸范数约束下的鲁棒回归分析方法，在遮挡环境下的人脸识别问题上取得了显著成果^[16]；布拉德福德大学采用最新的卷积神经网络结构以及预先训练的VGG-Face^[5]模型来提取遮挡下的人脸特征，在此基础上使用余弦相似度和支持向量机两类分类器完成人脸识别^[17]。在数据集方面，百度提出基于人脸关键点的3D图像融合技术，使得生成的图片更加符合真实场景；北京旷视科技有限公司采用生成对抗网络来生成佩戴口罩的人脸图像；京东利用自研的106关键点模型提取人脸关键点，并根据关键点确定口罩区域，在此基础上为人脸添加随机颜色的虚拟口罩来生成人脸数据；北京的卢深视科技有限公司首先重构三维人脸模型，继而将口罩的三维模板模型与三维人脸模型做非刚性对齐，从而完成模拟佩戴口罩的操作。

2.3 语音交互

防疫期间，智能导诊机器人肩负着相关医疗咨

询服务的责任，通过人工智能解答，借助语音交互技术，快速应对网络问诊和信息上报的需求。智能导诊机器人中的语音交互主要涉及语音识别与语音合成，用于实现语音片段与文本之间的转换。相较于一般的自然语言处理任务，语音识别的特殊之处在于声学模型。因此，与语音识别相关的研究主要围绕声学模型展开。经典的声学模型GMM-HMM分别采用广义矩估计（generalized method of moment, GMM）和隐马尔可夫模型（hidden Markov model, HMM）对语音的观察概率及语音的时序进行建模，导致无法学习深层非线性特征变换。微软于2012年率先将前馈神经网络（feedforward neural network, FNN）引入声学模型建模^[18]，大大降低了语音识别的错误率，引领了DNN-HMM混合声学模型的风潮，并衍生出RNN-HMM、LSTM-HMM、CNN-HMM等混合声学模型^[19-23]。上述基于DNN-HMM的声学模型要求训练特征序列和标注特征序列必须是等长的，即训练数据中每一帧都要预先确定对应的标注。针对上述问题，谷歌Graves A等人^[24]将连接时序分类（connectionist temporal classification, CTC）技术引入语音识别，通过前向后向算法自动学习语音特征中的模型边界，解决训练特征序列与标注特征序列不等长的问题，实现了端到端语音识别。至此，业界开始在语音识别系统中大量使用CTC技术，如百度的Deep Speech^[25]和Deep Speech 2^[26]、谷歌的NSR^[27]等，均取得了比传统混合识别系统更好的性能。但是，CTC技术无法对输出之间的依赖关系进



图1 人脸识别中的困难样本示例

行建模,对上下文信息利用不充分。针对 CTC 技术的不足,谷歌 Graves A^[28]提出了 RNN-T 模型,巧妙地将语言模型与声学模型整合在一起进行联合优化。考虑到语音识别的本质是两个序列之间的转换,谷歌将机器翻译中广泛使用的基于注意力机制的 Seq2Seq 模型引入语音识别,提出了一种新颖的端到端的语音识别框架 LAS^[29-30]。尽管当前语音识别精度取得了较大的提升,但是重口音、方言、嘈杂环境、多人同时说话等场景下的语音识别仍表现不佳,这是当前语音交互面临的重要挑战。

针对语音合成,基于深度学习的语音合成逐渐成为主流方法,旨在利用深层神经网络的非线性建模能力,提升语音建模精度。谷歌于 2016 年提出一种基于完全概率和自回归的神经网络模型 WaveNet,用于产生原始音频波^[31]。由于每次只预测一个采样点,WaveNet 语音合成速度较慢。针对合成速度慢的问题,百度提出了 Deep Voice^[32],其采用模块化神经网络结构来模拟传统参数语音合成过程。虽然 Deep Voice 已经大大减少了人工特征数量,但仍需要使用音素标记、重音标记、基频等人工特征。为了降低语音合成对语言学知识的依赖,Tacotron^[33]采用融合注意力的编码器-解码器网络设计端到端的语音合成模型,直接将输入文本生成语音。Tacotron 发布不久后,百度的 Deep Voice 2 诞生了。Deep Voice 2 将 Deep Voice 中的声码器由 Griffin-Lim 替换成 WaveNet,同时引入说话人向量,以支持多说话人合成^[34]。与上一代不同,Deep Voice 3^[35]采用融合注意力的全卷积编码器-解码器框架,可完全并行计算,合成速度较快。谷歌的 Oord A V D 等人^[36]通过引入一种概率密度蒸馏方法,提出了一种并行 WaveNet 模型来提高语音合成的速度及精度。与此同时,谷歌和加州大学伯克利分校对 Tacotron 进行改进,提出了语义合成模型 Tacotron2^[37]。与上一代不同,Tacotron2 采用基于序列到序列的特征预测网络将字符嵌入梅尔标度

谱图,并将改进后的 WaveNet 作为声码器来合成时域波,以提高语音合成精度。现阶段的语音合成系统大多基于 Tacotron 或 Tacotron2 实现,虽然效果不错,但在合成音的表现力上,特别是语气和情感方面,仍存在不足。

2.4 自主定位

智能导诊机器人要完成指定任务,自主定位是其需要掌握的核心关键技术之一。目前,室内机器人主要依赖即时定位与地图构建(simultaneous localization and mapping, SLAM)技术实现自主定位。具有强鲁棒性和高精度的 SLAM 是移动机器人自主导航定位的首要需求。如图 2 所示,在纹理少、四季天气变化、光照剧烈变化、长走廊等情况下,采用单一传感器 SLAM 往往会导致 SLAM 失效或无用。为了提高移动机器人对不确定信息的抵抗能力,近年来,基于多传感器融合的 SLAM 技术成为研究热点。由于传感器自身的限制,感知信息存在不同程度的不确定性,直接使用感知信息很难得到准确的环境模型,通常需要对感知信息进行再处理,从而完成导航和定位。基于多传感器融合的大场景不确定信息下的机器人自主定位成为移动机器人定位导航领域的核心问题。

由于激光雷达能够获得高清的三维环境感知信息,目前激光雷达 SLAM 是比较成熟的定位导航方法。但是激光雷达具有价格昂贵、寿命短且无法获得语义信息等缺点,因此,激光雷达与其他传感器的融合成为目前 SLAM 领域的研究热点之一。Ye H 等人^[38]提出了一种基于紧密耦合的 Lidar-IMU 融合方法——LIO-mapping。Graeter J 等人^[39]提出了一种基于激光与相机的深度信息融合的里程计与光束平差法(bundle adjustment, BA)框架——LIMO。Zhang J 等人^[40]将普通的单目相机和其他设备(如 Lidar、RGB-D 相机)获得的深度信息相结合,提出了一种基于多传感融合的里程计。Zhang J 等人^[41]提出了一种视觉雷达里程计的通用框架 V-LOAM,

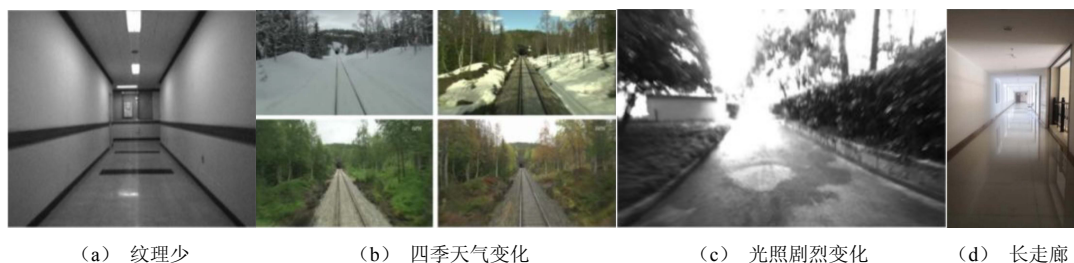


图2 单传感器 SLAM 棘手的场景示例

该框架探索了每个传感器的优点，并补偿了另一个传感器的缺点，通过视觉估计自身位姿来修正高频点云的畸变，同时基于扫描匹配的激光雷达测距仪来改进运动估计和点云配准，提高定位精度。参考文献[42]提出了一种立体视觉惯性激光雷达 SLAM，通过将紧耦合的立体视觉惯性里程计、Lidar 建图、Lidar 增强的闭环检测进行结合，实现实时生成闭环校正的 6 自由度 Lidar 姿态及 1 cm 体素密集地图。参考文献[43]提出了一种融合激光、IMU 与单目相机的里程计 LIC-Fusion，其可在任意两个传感器间进行在线的空间和时间的标定，并在 MSCKF 框架下对多传感器状态进行融合，实现位姿估计。为了实现鲁棒性的 SLAM，未来的 SLAM 必然是集成多个传感器的。不同传感器之间具有不同的数据类型、时间戳和坐标系表达式，如何实现快速、高精度的多传感器数据关联是未来多传感器融合 SLAM 的主要挑战。

2.5 路径规划

路径规划技术也是机器人研究领域的重要分支。最优路径规划是指依据某个或某些优化准则，在机器人工作空间中找到一条从初始状态到目标状态的最优路径。根据对环境信息的掌握程度不同，可以将机器人路径规划分为全局路径规划和局部路径规划。常见的路径规划算法主要包括人工势场法、A*算法、快速扩展随机树（rapidly-exploring random tree, RRT）、遗传算法（genetic algorithm, GA）、粒子群优化（particle swarm optimization, PSO）算法等^[44]。公共场合大量的行人会导致环境的高动态变化，这对机器人路径规划会产生极大的干扰。目前，绝大部分机器人路径规划算法尚不能满足高动态的需求。

随着人工智能的快速发展，基于强化学习的路径规划成为研究热点。基于强化学习的机器人路径规划演示如图 3 所示。强化学习不必基于系统平台的动力学模型，而是通过最大化累积奖励来学习最优策略，从而弱化了模型和环境的约束。在机器人导航任务中，奖励稀疏分布于应用场景中，通常只有当机器人达到目的地附近时才会得到一个正的奖励。奖励稀疏导致训练过程中样本利用率较低、算法收敛较慢，因此，无法直接应用各种典型的深度强化学习算法来解决机器人自主导航问题。针对上述挑战，研究者采用设计辅助任务的方式来监督深度卷积网络从环境中学习有用的特征作为状态，从而加速算法的收敛。Jaderberg M 等人^[45]提出了一种融合多种无监督辅助任务的强化学习模型 UNREAL。考虑到在环境中相邻图像间像素的剧烈变化通常对应重要的事件（如转弯），作者提出一种基于像素变化的虚拟奖励，使得机器人在运动过程中不仅接收来自外部环境的奖励，还接收基于像素变化的虚拟奖励。另外，为了确保能顺利到达目标位置，机器人应具有识别目标地点的能力。基于此，作者提出一种基于奖励预测的辅助任务来加快算法的收敛。在模型训练过程中，UNREAL 兼顾最大化环境及虚拟奖励以及最小化奖励估计误差，达到提高数据利用效率及训练效果的目标。受 UNREAL 的启发，Mirowski P 等人^[46]在 A3C 框架下提出了一种以深度估计为辅助任务的机器人导航算法。该算法设计融合 CNN 与 LSTM 的双分支神经网络结构，从而同时生成深度图与机器人运动策略。现阶段，基于深度强化学习的机器人路径规划主要在模拟环境下展开，仍处于探索试验阶段。与真实场景相比，模拟环境相对简单，导致上述导

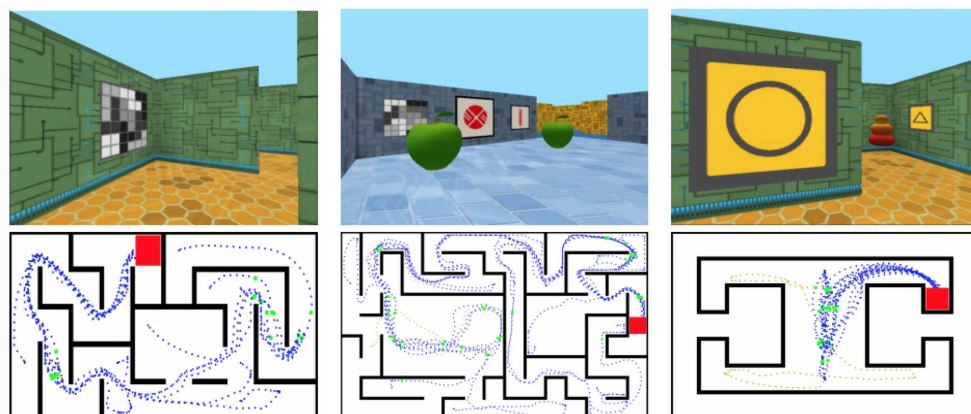


图3 基于强化学习的机器人路径规划演示

航算法泛化能力较差。如何提高基于深度强化学习的机器人导航算法的环境适应性是未来亟待解决的问题。

3 面向防疫的钛米智能导诊机器人

本节以钛米智能导诊机器人为应用案例,对上述关键技术在实际应用中的落脚点进行讨论。在新型冠状病毒肺炎疫情期间,作为一套综合运用物联网、人工智能、大数据等技术的系统(如图4所示),钛米智能导诊机器人通过机器消毒、人脸识别、语音交互、自主定位、路径规划等技术在医院门诊等地完成对患者的合理分流、智能导诊、医疗咨询及信息采集等任务。

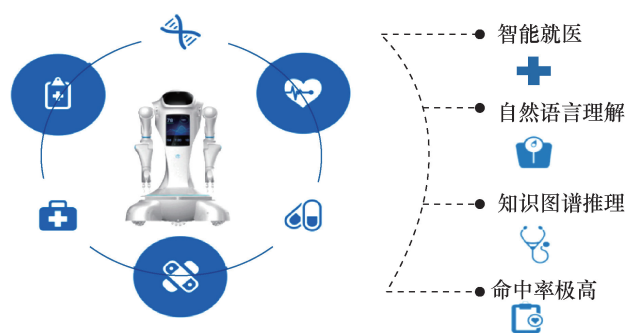


图4 钛米智能导诊机器人

(1) 机器消毒

为了降低交叉感染的风险,钛米智能导诊机器人首先从涂层入手,降低病毒颗粒附着的可能性;在此基础上,设置单独的消毒室,以固定时间频率采用紫外线消毒方法进行消毒。

(2) 人脸识别

钛米智能导诊机器人借助深度学习的优势来设计深度模型,以实现高精度人脸识别。考虑到疫情期间佩戴口罩导致人脸的大部分关键点受到遮挡的问题,相关人员从深度模型设计入手,采用眼部、眉毛等局部特征与全局人脸特征相融合的方式,并通过引入注意力机制帮助深度模型自动聚焦到信息丰富的眼部关键点,从而提高在口罩遮挡情况下的人脸识别精度。

(3) 语音交互

钛米智能导诊机器人以深度学习、超级计算和认知计算为基础,以自然语言交互为切入点,通过构建语音识别、自然语言理解及语义合成等模块实现完整的语音交互体系。针对患者存在口音、方言的问题,采用基于注意力机制的方言自适应优化算

法,达到使用少量方言数据即可快速优化识别效果的目标。另外,针对纯语音合成系统在长文本下通常显得机械、不似真人的问题,采用基于听感量化code编码语音合成技术来改善合成果。

(4) 自主定位

钛米智能导诊机器人搭载超声波、激光雷达、光学相机、IMU等多种传感器,通过多源异构数据融合对医院场景进行实时地图构建和自主定位。钛米智能导诊机器人主要采用SharpEdge精细化构图技术实现高精度、厘米级的地图构建;同时利用三维激光点云测距传感器及光学相机的观测信息与所创建的地图进行匹配,实时获得机器人的位置信息。

(5) 路径规划

钛米智能导诊机器人采用全局路径规划结合局部路径规划的思想,实现机器人在复杂医院、科研楼环境下的自主导航,完成患者引导、院内信息传达及医疗知识宣传等。

4 未来挑战与发展趋势

目前,关于防疫智能导诊机器人的研究引起机器人领域研究者的广泛关注,并成为非常前沿的热点问题之一。防疫作为智能移动机器人非常有潜力的落地应用之一,虽然已有部分研究成果投入使用,但是距离大规模的商业化应用还相距甚远。本节对当前面向防疫的智能导诊机器人面临的挑战及未来发展趋势进行讨论。

(1) 基于多模态数据融合的环境感知

感知能力是智能移动机器人理解周围环境、执行工作任务并完成智能决策的前提条件。实现灵活、稳定、可靠的场景理解,需要机器人通过视觉、听觉、触觉等多种异构传感器获得多模态数据,并进行融合,实现优势互补。然而,目前与智能移动机器人环境感知相关的研究主要针对视觉数据进行。因此,发展基于多模态异构数据融合的机器人环境感知技术是必然趋势。

(2) 高动态复杂环境下的自主导航

自主导航技术是机器人安全地完成各种指定工作任务的关键和核心技术。未来智能导诊机器人需要在各种结构化及非结构化的未知、复杂、高动态环境中工作,各种意外突发状况对机器人的自动导航智能化提出了较高的要求。目前的机器人导航方法大多基于传统方法,过分依赖先验环境知识,

导致机器人环境适应性较差。针对有环境变化的动态场景, 机器人导航避障技术无法做到实用。要解决这些问题, 必须使机器人具有自主学习的能力。随着人工智能技术的快速发展, 借鉴人工智能领域的方法(如强化学习、进化学习等)来增强机器人导航性能和对环境的适应性将是未来的研究重点。

(3) 面向情绪及意图识别的人机交互

智能导诊机器人在执行任务期间需要不停地与人进行沟通。机器人若要更好地服务于用户, 首先要充分理解用户的情绪及意图。在人机交互方面, 现有的机器人交互水平仅限于简单的语音识别、图像识别, 而对人的意图、行为、情绪等方面的认知比较匮乏, 亟待提高, 从而提供更快捷、人性化的服务。

5 结束语

本文从机器人消毒、大面积遮挡下的人脸识别、语音交互、自主定位与智能导航等方面对面向防疫的智能导诊机器人的关键技术及发展现状进行了详细讨论及剖析, 同时以钛米智能导诊机器人为例介绍了上述关键技术在实际系统中的落脚点。现阶段, 以深度学习为核心基础的新一代人工智能技术的快速发展为导诊机器人智能化发展提供了强有力的技术支撑, 在人脸识别、语音交互、自主定位与智能导航等领域取得了一系列有意义的成果, 并展现出优越的环境适应性及鲁棒性。但是, 现阶段的智能导诊机器人仍在重口音、方言、嘈杂环境等场景下的语音识别、大场景不确定信息下的自动定位、复杂动态环境下的机器人自主导航等方面存在不足。针对上述挑战, 未来将从基于多模态数据融合的环境感知、高动态复杂环境下基于人工智能技术的自主导航、面向情绪及意图识别的人机交互等方面展开对导诊机器人的研究。

参考文献:

- [1] 黄永安, 熊蔡华, 熊有伦. 智能机器人与应用的现状与发展趋势[J]. 国际学术动态, 2009: 38-39.
HUANG Y A, XIONG C H, XIONG Y L. Status and development trend of intelligent robots and applications[J]. International Academic Developments, 2009: 38-39.
- [2] 徐扬生. 智能机器人引领高新技术发展[J]. 企业科协, 2010(9): 28-31.
XU Y S. Intelligent robots lead high-tech development[J]. Enterprise Association, 2010(9): 28-31.
- [3] 谭民, 王硕. 机器人技术研究进展[J]. 自动化学报, 2013, 39(7): 963-972.
TAN M, WANG S. Research progress on robotics[J]. Acta Automatica Sinica, 2013, 39(7): 963-972.
- [4] HE J X, BAXTER S L, XU J, et al. The practical implementation of artificial intelligence technologies in medicine[J]. Nature Medicine, 2019: 30-36.
- [5] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: closing the gap to human-level performance in face verification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014.
- [6] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10000 classes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014.
- [7] PARKHI O M, VEDALDI A, ZISSERMAN A, et al. Deep face recognition[C]//Proceedings of the British Machine Vision Conference. [S.l.:s.n.], 2015.
- [8] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015.
- [9] DENG J, GUO J, ZAFEIRIOUS S. ArcFace: additive angular margin loss for deep face recognition[J]. arXiv preprint, 2018, arXiv:1801.07698.
- [10] CHEN D, CAO X, WANG L, et al. Bayesian face revisited: a joint formulation[C]//Proceedings of the 12th European Conference on Computer Vision-Volume Part III. Berlin: Springer, 2012: 566-579.
- [11] SANKARANARAYANAN S, ALAVI A, CHELLAPPA R. Triplet similarity embedding for face verification[J]. arXiv preprint, 2016, arXiv:1602.03418.
- [12] SANKARANARAYANAN S, ALAVI A, CASTILLO C D, et al. Triplet probabilistic embedding for face verification and clustering[C]//Proceedings of the 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems. Piscataway: IEEE Press, 2016.
- [13] KUMAR B G V, CARNEIRO G, REID I. Learning local image descriptors with deep Siamese and triplet convolutional networks by minimizing global loss functions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016.
- [14] TRIGUEROS D S, MENG L, HARTNETT M. Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss[J]. Image and Vision Computing, 2017(79).
- [15] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]//Proceedings of the 31st Conference and Workshop on Neural Information Processing Systems. [S.l.:s.n.], 2017.
- [16] ZHANG C, LI H, CHEN C, et al. Enhanced group sparse regularized nonconvex regression for face recognition[C]//Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE Press, 2020.
- [17] ELMAHMUDI A, UGAIL H. Deep face recognition using imperfect facial data[J]. Future Generation Computer Systems, 2019, 99: 213-225.
- [18] DAHLG E, YU D, DENG L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 20(1): 30-42.
- [19] VINIYALS O, RAVURI S V, POVEY D. Revisiting recurrent neural network for robust ASR[C]//Proceedings of the 37th International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2012.
- [20] MAAS A L, LE Q V, O'NEIL T M, et al. Recurrent neural networks for noise reduction in robust ASR[C]//Proceedings of the INTERSPEECH. Piscataway: IEEE Press, 2012.
- [21] GRAVESA, JAITLEY N, MOHAMED A R. Hybrid speech recognition with deep bidirectional LSTM[C]//Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. Pis-

- cataway: IEEE Press, 2013.
- [22] ABDEL-HAMID O, MOHAMED A, JIANG H, et al. Convolutional neural networks for speech recognition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2014, 22(10): 1533-1545.
- [23] SAINATH T N, VINYALS O, SAK H. Convolutional, long short-term memory, fully connected deep neural networks[C]//Proceedings of the 40th IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2015.
- [24] GRAVES A, FERNANDEZ S, GOMEZ F, et al. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks[C]//Proceedings of the 23rd International Conference on Machine Learning. [S.l.:s.n.], 2006: 25-29.
- [25] HANNUN A, CASE C, CASPER J, et al. Deep Speech: scaling up end-to-end speech recognition[J]. arXiv preprint, 2014, arXiv: 1412.5567.
- [26] AMODEI D, ANANTHANARAYANAN S, ANUBHAI R, et al. Deep Speech 2: end-to-end speech recognition in English and Mandarin[C]//Proceedings of the 33rd International Conference on Machine Learning. [S.l.:s.n.], 2016.
- [27] SOLTAU H, LIAO H, SAK H. Neural speech recognizer: acoustic-to-word LSTM model for large vocabulary speech recognition[C]//Proceedings of the INTERSPEECH. [S.l.:s.n.], 2017.
- [28] GRAVES A. Sequence transduction with recurrent neural networks[C]//International Conference on Machine Learning. [S.l.:s.n.], 2012.
- [29] CHAN W, JAITLY N, LE Q, et al. Listen, attend and spell: a neural network for large vocabulary conversation[C]//Proceedings of the 41st International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2016.
- [30] CHIUC C, SAINATH T N, WU Y, et al. State-of-the-art speech recognition with sequence-to-sequence models[C]//Proceedings of the 43rd International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2018.
- [31] OORD A V D, DIELEMAN S, ZEN H, et al. WaveNet: a generative model for raw audio[J]. arXiv preprint, 2016, arXiv: 1609.03499.
- [32] ARIKS O, CHRZANOWSKI M, COATES A, et al. Deep Voice: real-time neural text-to-speech[C]//Proceedings of the 31st International Conference on Machine Learning. [S.l.:s.n.], 2017.
- [33] WANG Y, SKERRY-RYAN R J, STANTON D, et al. Tacotron: towards end-to-end speech synthesis[C]//Proceedings of the 31st INTERSPEECH. [S.l.:s.n.], 2017.
- [34] ARIK S, DIAMOS G, GIBINASKY A, et al. Deep Voice 2: multi-speaker neural text-to-speech[C]//Proceedings of the 31st Conference and Workshop on Neural Information Processing Systems. [S.l.:s.n.], 2017.
- [35] PING W, PENG K, GIBINASKY A, et al. Deep Voice 3: scaling text-to-speech with convolutional sequence learning[C]//Proceedings of the 6th International Conference on Learning Representations. [S.l.:s.n.], 2018.
- [36] OORD A V D, LI Y, BABUSCHKIN I, et al. Parallel WaveNet: fast high-fidelity speech synthesis[C]//Proceedings of Machine Learning Research. [S.l.:s.n.], 2018.
- [37] SHEN J, PANG R, WEISSR J, et al. Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions[C]//Proceedings of the 43rd International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2018.
- [38] YE H, CHEN Y, LIU M. Tightly coupled 3D Lidar inertial odometry and mapping[C]//Proceedings of the International Conference on Robotics and Automation. [S.l.:s.n.], 2019.
- [39] GRAETER J, WILCZYNSKI A, LAUER M. LIMO: Lidar-Monocular visual odometry[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2018.
- [40] ZHANG J, KAESSE M, SINGH S. Real-time depth enhanced monocular odometry[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2014.
- [41] ZHANG J, SINGH S. Visual-Lidar odometry and mapping: low-drift, robust, and fast[C]//Proceedings of the International Conference on Robotics and Automation. [S.l.:s.n.], 2015.
- [42] SHAO W, VIJAYARANNAN S, LI C, et al. Stereo visual inertial Lidar simultaneous localization and mapping[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2019.
- [43] ZUO X, GENEVA P, LEE W, et al. LIC-Fusion: Lidar-inertial-camera odometry[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2019.
- [44] 霍凤财, 迟金, 黄梓健, 等. 移动机器人路径规划算法综述[J]. 吉林大学学报(信息科学版), 2018, 36(6): 639-647.
- HUO F C, CHI J, HUANG Z J, et al. Review of Path Planning for Mobile Robots[J]. Journal of Jilin University (Information Science Edition), 2018, 36(6): 639-647.
- [45] JADERBERG M, MNH V, MARIAN W, et al. Reinforcement learning with unsupervised auxiliary tasks[C]//Proceedings of the 5th International Conference on Learning Representations. [S.l.:s.n.], 2017.
- [46] MIROWSKI P, PASCANU R, VIOLA F, et al. Learning to navigate in complex environments[C]//Proceedings of the 5th International Conference on Learning Representations. [S.l.:s.n.], 2017.

[作者简介]



王腾(1988—), 女, 博士, 东南大学自动化学院讲师, 主要研究方向为机器人环境感知与视觉导航。



潘晶(1980—), 男, 上海钛米机器人股份有限公司董事长兼 CEO, 主要研究方向为智能医疗服务机器人设计与制造。



董璐(1990—), 女, 博士, 同济大学电子与信息工程学院副研究员, 主要研究方向为机器人智能控制。



孙长银(1975—), 男, 博士, 东南大学自动化学院教授、博士生导师, 主要研究方向为机器人环境感知与智能控制。