

声音感知在地面无人平台中的应用展望

刘文举¹, 杨宏斌², 胡鹏飞¹

(1. 中国科学院自动化研究所, 北京 100190; 2. 北方导航控制技术股份有限公司, 北京 100176)

摘要: 听觉是人类接受外界信息的一个重要来源.声音的感知和处理是人类智能化活动中不可缺少的一部分.因此,在机器人技术中,除图像以外,对声音的感知越来越受到重视.本文介绍了声音感知在地面无人平台技术中的应用.分别阐述了声音信息在环境感知,跟踪定位,及人机交互等方面的具体应用和技术特点.

关键词: 声音感知, 环境感知, 跟踪定位, 语音识别, 人机交互

0 引言

地面无人平台是一种用于军事目的的,集感知、控制和智能决策等理论与技术于一体,能够自主驾驶的智能平台.由于其能够大量减少战场上人员的伤亡,各国都纷纷在该领域投入了大量的人力和物力.无人平台研究的核心内容之一是智能行为决策,而智能行为决策的前提则是其行驶过程中对周边环境的自动感知.感知环境信息的手段可以有多种,如GPS、激光、雷达、红外线、视听觉信息等.其中视听觉信息的自动感知在无人平台的行驶中占有重要的地位.目前使用较多的是视觉感知信息,而较少利用听觉感知信息.事实上,听觉信息也是无人车系统不可或缺的重要决策依据信息,一方面听觉感知结果是对视觉感知结果的重要补充,通过二者的有效结合可以更准确地感知无人车辆的周边环境,这种辅助作用在诸如黑夜或隧道等可利用的视觉信息不理想的情况下,就表现得尤为突出;另一方面外部世界与无人车辆间的很多交互信息是基于声音的,如枪声炮声的识别定位,语音指令等,感知周围这些基于声音的交互信息,并做出正确的智能决策对无人车辆而言至关重要.

声音感知的目标是使机器能感知人耳听觉所能关注和理解的声音.声音的类型大体可分为语音和非语音,对不同的声音类型,其所采取的感知处理方法也不尽相同.有关语音感知方面的研究已较为丰富,如语音识别、说话人识别等.研究者们普遍认为非语音的声音也能传递有用的信息,通过对这些声音的分析和处理,能够有效获知相应的环境状态,为智能决策提供条件.

近年来,声音感知已经成为学术界的研究热点,各个国家和地区纷纷提出了与之相关的若干重大发展计划,例如欧洲的CHIL (Computers in the Human Interaction Loop) 和AMI (Augmented Multi-party Interaction) 计划,美国的VACE (Video Analysis and Content Extraction) 和CALO(Cognitive Assistant that Learns and Organizes)计划等[1].在这些计划和相关科研机构的推动下,CHIL于2006年和2007年先后举办了两次声音感知相关的评测CLEAR (Classification of Events, Activities and Relationships) [2] [3].在军事领域,美国西点军校和AFRL (Air Force Research Laboratory)实验室在这方面做了大量的工作.涉及军事环境下的声音获取,去噪,建模等各个方面[4] [5].

1 声音的环境感知及其建模

无人车或无人机作为一类面向信息化作战的智能移动机器人平台,必须具有一定的自主性.这种自主性体现在对战场环境的自动感知与理解、威胁规避、自主规划与决策、自主运动控制以及任务的自主执行等诸多方面.其中,战场环境的感知与理解作为其他环节的前提和依据,是其实现自主作战的基础.战场环境感知与理解的任务是使无人车具备对战场空间内各类相关信息的收集和认知能力,理解其所处的战场态势,指导其决策和行动.目前,图像是无人作战平台获取环境信息的主要数据来源.然而,周围环境声音也在一定程度上反应了相应的环境状态,传递了大量有用的信息.要想让无人战车在复杂的战场环境中应对自如,成为现代科技战场上真正的“耳目”,必须加入这部分听觉信息,才能使其“眼观六路,耳听八方”.声音信息用于环境感知包括两方面:一是声学事件检测 (Acoustic Event Detection),再者是特

定语义下的音频场景识别。

声学事件对应着能够表征一个完整事件的一段连续的声音,从人类听觉感知的角度,它是具有一定意义的最小声音单位,例如:笑声、脚步声、枪声、爆炸声、玻璃碎裂声、开关门声、狗吠声、鸟鸣声等。声学事件检测是通过采集设备拾取声音信号,经分析处理后获得其发生的开始时间、持续时间、场景类别等信息,并将其转化为相应的事件符号来表示,从而达到仿真耳听觉感知能力的目的。其关键研究内容包括:声学事件的定义、声学特征的选择、分类算法、检测算法等几方面。声学事件检测的研究工作总与其应用环境紧密联系。对于无人地面军事平台来说,重要的是枪声,炮声的检测,甚至根据炮声的不同来识别对方所用的武器装备。

音频场景是指语义上相关,且时间上相邻的若干声学事件组成的一个音频片段,它总是蕴含着高层抽象概念和特定的语义表达。音频场景分析是基于声学事件的检测,判断所处的场景。比如鸟鸣声比较频繁的地方更可能是丛林作战的环境。通过雷声和雨声可以辅助判断作战时的天气状况等。这些声音信息可以作为视觉信息的一个有益补充,更好的帮助决策。

在建模方面,研究者们尝试了各种声学事件建模方法,如基于支持向量机的方法[6];基于隐马尔可夫模型的方法[7];基于高斯混合模型的方法[8];基于贝叶斯网络的方法[9]等。但在军事背景下最大的技术难点在于无人车本身的引擎噪声的影响。低信噪比的情况会使声学事件检测的效果下降。

2 声音的跟踪定位

在战场上,对军事目标的跟踪定位,准确判断出发射源的方位和距离是地面无人平台的重要功能。如今的地面无人平台大多采用了雷达进行跟踪定位。在航海和航空中,雷达对周围感兴趣的目标有不受阻碍的视野。然而在陆地环境中,由于雷达天线装在地面车辆的顶上,附近的建筑物、树林或山峦将阻碍雷达视野,并且由车辆位置引起的地面斜度可能使雷达荧光屏显示充满连续的回波,这些回波叫地面杂波,它能将感兴趣的小目标掩蔽起来。声源定位可以在炮弹发射源的定位上补充雷达此方面的不足。它可以根据炮弹发射的声音,判断出发射源的方向和位置。现在利用声音来定位狙击手的系统已经被广泛应用,我国特警车辆上也已经装备了此类系统用于安保工作。在真实的战场环境下,由于炮弹发射的声音较大,在距离较远时,这种定位方法也比较有效。此外,利用声音来定位在硬件的要求上也相对简单,不需要大功率的信号发射接受装置。仅麦克风阵列就可以满足要求。

基于麦克风阵列的声源定位技术主要有两类方法:波束扫描(*beam scanning*)和基于时间延迟(*time delay estimation*, TDE)估计的定位方法。波束形成器实质上是一个空间滤波器,它通过操作各个麦克风的输出来增强来自特定方向的声音,同时抑制来自其他方向的噪声和干扰源信号。因此,用控制麦克风阵列波束形成器来扫描整个空间,寻找输出信号能量的最高值,从而估计出目标信号源的方向(假设目标信号源在空间局部有明显的能量优势)。使用两个麦克风阵列,并且将它们估计的信号源方向相交叉,就可以确定声源位置[10]。基于TDE的声源定位算法引入了两步架构的处理方法[11]。在第一步中,阵列中任意两个麦克风之间的相对到达时间差异(*time difference of arrival*, TDOA)被计算并形成当前时刻的TDOA集合。在第二步中,结合TDOA集合,利用已知的麦克风阵列的几何分布定位声源的位置。基于TDE的声源定位算法具有较高的计算效率,并可通过改变采样率和麦克风阵列的规格来调整定位精度,在应用中具有较大的灵活性。

3. 语音识别用于人机交互

地面无人平台所涉及的人机交互可包括两个层面,一个是来自指挥中心的远程交互,一个是近距离的与士兵间的人机交互。如下图中以色列宇航工业公司所研制的伴随步兵的轮式“机器骡”,在一定距离之内,士兵可以使用简单的命令,包括“停止”、“提取”和“紧随”等命令,操控无人平台[12]。



语音识别在手机上的成功应用，为地面无人平台的人机交互提供了另外一种方式。在复杂的战场环境下，采用包括语音识别在内的多元交互方式将提高指令传输的有效性。近距离的士兵，利用头盔上的无线耳机进行语音操控，完全不需要双手来辅助，不会耽误或打乱正在进行的战斗任务。因此在战场上利用语音识别技术进行人机交互具有天生的优势。对于远程的人机交互，利用语音识别技术能够方便地传输暗语，实现上直接对暗语建立声学模型即可。

语音识别技术起源于1952年贝尔实验室的Davis等人研制的特定说话人孤立数字识别系统。20世纪80年代隐马尔科夫模型和MFCC声学特征的提出又使其得到飞速发展。近年来，手机技术的快速发展加速了语音识别技术的应用。电影2012中语音识别在汽车里的应用也给人们留下了深刻的印象。语音识别被认为是下一代人机交互界面的一个趋势。在地面无人平台的人机交互中，语音识别的应用主要涉及的是简单指令的识别，与大词汇量的语音识别相比，算法难度相对较低。但由于战场环境的复杂多变，它也有自己面临的特殊问题。主要涉及前端对环境噪声的处理以及模型鲁棒性的考虑。

4 总结

地面无人平台作为高端智能化作战平台，其自主水平的提高在很大程度上受制于对战场环境信息的采集、融合、理解及利用能力。将声音感知引入到地面无人平台必将提高战场信息的利用效率，提升地面无人平台的自主能力和信息优势，为其实际作战应用奠定基础。

参考文献:

- [1] R. Stiefelhagen, K. Bernardin, R. Bowers, *et al*, *Multimodal Technologies for Perception of Humans*, Springer-Verlag Berlin Heidelberg, 2008.
- [2] CLEAR, 2006. Classification of Events, Activities and Relationships. Evaluation and Workshop. <<http://isl.ira.uka.de/clear06>>.
- [3] CLEAR, 2007. Classification of Events, Activities and Relationships. Evaluation and Workshop. <<http://www.clear-evaluation.org/>>.
- [4] Richard L. McKinley, "AUDITORY AND ACOUSTIC RESEARCH & DEVELOPMENT AT AIR FORCE RESEARCH LABORATORY (AFRL)," Sep.2010.
- [5] D.Cersovsky; E.Kleinschmidt; B.Mansager and B.Neta, "audio detection algorithms in combat simulation Detection Algorithms",

Mathl.Comput.Modelling Vol.23, No.1/2,pp.65-72,1996.

- [6] J.L. Rouas, J. Louradour, S. Ambellouis, "Audio events detection in public transport vehicle," in Proc. of IEEE Intelligent Transportation Systems Conference. 2006:733-738
- [7] G. Wichern, J. Xue, H. Thornburg, B. Mechtley, A. Spanias, " Segmentation, indexing, and retrieval for environmental and natural sounds, " IEEE Transactions on Audio, Speech and Language Processing. 2010,18(3): 688-707
- [8] A. Ito, A. Aiba, M. Ito, S. Makino, "Detection of Abnormal Sound Using Multi-stage GMM for Surveillance Microphone," in Proc. of IEEE International Conference on Information Assurance and Security. 2009:733-736
- [9] T. Giannakopoulos, A. Pikrakis, S. Theodoridis, "A Multi-Class Audio Classification Method With Respect To Violent Content In Movies Using Bayesian Networks," in Proc. of IEEE 9th Workshop on Multimedia Signal Processing. 2007:90-93
- [10] Woonseng G.; Eeleng T.; Kuo, S.M. , "Audio Projection," Signal Processing Magazine, IEEE , vol.28, no.1, pp.43-57, Jan. 2011
- [11] Liuhong; M. Shen , "Continuous sound source localization based on microphone array for mobile robots," Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on , vol., no., pp.4332-4339, 18-22 Oct. 2010
- [12] “机器骡”地面无人平台:<http://www.dsti.net/Information/News/53899>