

基于视觉、听觉系统的声源水平方位定位

陈 涛¹, 张明路¹, 付灵丽²

(1 河北工业大学 机械工程学院, 天津 300130; 2 河北工业大学 计算机科学与软件学院, 天津 300130)

摘 要: 针对空间声源水平方位定位的精度问题, 提出了利用基于软域值的小波分析方法对声源信号进行去噪处理. 在此基础上, 利用机器人的听觉系统对目标声源的水平方位进行粗定位; 通过双目立体视觉系统对粗方位进行一定的矫正、补偿, 实现方位的精定位. 实验证明, 提出的由粗到精的定位策略和方法具有较高的定位精度.

关键词: 软阈值去噪; 水平方位角; 双目立体视觉; 方位补偿

中图分类号: TP242.6

文献标识码: A

文章编号: 1000-7180(2010)04-0009-04

Azimuth Localization of Sound Source Based on Video-vision System

CHEN Tao¹, ZHANG Ming-lu¹, FU Ling-li²

(1 Institute of Mechanical Engineering, Hebei University of Technology, Tianjin 300130, China;

2 Institute of Computer Science and Software, Hebei University of Technology, Tianjin 300130, China)

Abstract: For the azimuth accuracy of sound source localization, this paper adopted wavelet analysis to denoise signal with soft-threshold. Based on this, estimated the rough azimuth degree using the auditory system of robot. In order to gain the fine direction of azimuth, paper provided the error compensation model of azimuth with binocular stereo vision system. Through experiments, the provided coarse-to-fine strategy and algorithm can generates more accurate results for sound source localization.

Key words: soft-threshold denoise; azimuth; binocular stereo vision system; direction compensation

1 引言

目标定位技术是智能机器人目标识别研究的前提和重点. 空间声源的方位定位主要从声源的水平方位、垂直方位和距离等方面进行评估. 这些定位指标中, 水平方位的定位相对准确, 也最为重要. 目前, 对于水平方位的定位主要基于三种策略: 声波到达时间差(TDOA)、可控波束响应功率和特征子空间估计. 出于算法易用性和实时性的考虑, 这里采用基于 TDOA 的算法策略.

视觉定位系统对声源目标的方位定位更加的直观和准确, 不过漫无目的地进行全空间的目标搜索, 系统需要处理过多的冗余数据, 不利于机器人的实时作业. 而声波具有特有的衍射效应和不易受光线的影响等优点, 可以利用机器人的听觉定位系统先

于视觉系统对声源目标进行粗定位. 有了粗定位的初始方位, 就可以指导机器人进行快速的目标范围确定及方位补偿.

2 信号预处理

机器人的听觉系统由头部左右耳处对称的耳廓装置和嵌入其中的麦克风传感器组成. 麦克风采集到的原始信号容易受环境中噪声、回音等因素的影响, 不利于声源目标方位的精确判断, 因此在进行具体的定位之前, 需要对声音信号进行适当的去噪处理^[1]. 目前采用最多的去噪方法是利用小波分析对噪声信号进行去噪处理, 主要有以下几个步骤:

(1) 多层小波分解. 选用适当的小波对信号进行多尺度分解, 得到各尺度上的近似(低频段 L)分量和细节(高频段 H)分量(见图 1). 这里可以采用

收稿日期: 2009-03-16; 修回日期: 2009-05-16

基金项目: 国家“八六三”计划项目(2006AA04Z221)

db 小波进行三层分解.

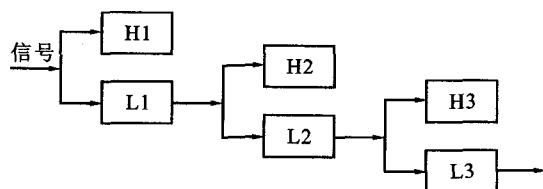


图1 信号的小波分解

(2) 阈值操作. 对尺度1到尺度 N (N 为小波分解的层数) 上的细节高频分量和大尺度上的近似分量选取不同的软阈值进行消噪处理^[2].

(3) 多层小波重构. 根据尺度 N 上的近似分量和从尺度1到 N 上经过处理的细节分量计算出信号的小波重构.

阈值化处理的关键在于选择合适的阈值门 δ . 如果阈值门太小, 处理后的信号仍有噪声存在; 阈值太大, 重要的信号特征将被滤掉, 引起偏差. 大多数阈值的选取过程是针对一组小波系数, 即根据本组小波系数的统计特性, 计算出一个阈值 δ . Donoho等提出了一种典型的阈值选取方式, 从理论上说明阈值与噪声的方差成正比:

$$\delta_j = \sqrt{2 \log N_j} \sigma_j \quad (1)$$

$$\sigma = \text{median}(|d_j|) / 0.6745 \quad (2)$$

d_j 为 j 层上的小波系数, N_j 为 j 层的小波系数的数量. 该阈值的选取方式能够获得较高的信噪比, 起到很好的消噪作用. 对信号进行去噪处理后, 还需进行分帧处理, 这样可以为提取特征参数做准备. 分帧的处理过程主要采用交叠分段的方法, 使帧与帧之间能够平滑过渡, 保持信号的连续性.

3 声源水平方位粗定位

声源水平方位的定位主要利用机器人听觉系统的“双耳效应”原理. 当声源偏向左耳或右耳, 即偏离两耳正前方的中轴线时, 声源到达左、右耳的距离存在差异, 这将导致到达两耳的声音信号在时间、声级和相位上存在差异. 这种差异为声源的方位评估提供了定位线索.

对于信号频率低 ($< 1.5\text{kHz}$) 的声源, 两耳麦克风接收的信号不会产生模糊的相位, 此时两耳时间差 (ITD) 线索占主导地位. 当信号频率较大时, 波长较短, 两耳之间存在过多的周期信号, 单一的 ITD 线索定位比较模糊. 此时, 需要借助于双耳效应的另一个定位线索: 两耳声压差 (ILD)^[3,4].

ITD 和 ILD 的定位模型为

$$\begin{cases} \text{ITD}(\theta, f) \approx \beta_f r (\sin \theta + \theta) / c \\ \text{ILD}(\theta, F) \approx \alpha_f \text{Fun}_M(\theta) \end{cases} \quad (3)$$

式中, α_f 和 β_f 为模型的尺度因子, $\text{Fun}_M(\theta)$ 为某频率的 ILD- 方位角评估函数. r 为头部半径, c 为声速.

信号频率 $< 1.5\text{kHz}$ 时, 两耳麦克风接收的信号之间不会产生模糊相位, 可以单独通过 ITD 的模型进行 θ 角的评估. 此时 ITD 可以通过两耳信号的互相关求取. 在信号领域, 互相关是用来计算 A 通道和 B 通道信号的相关系数, 从而了解两个信号间的相似情况. 互相关的表达式为

$$R_{lx}(\tau) = \sum_{n=0}^{N-1} x_l[n] x_r[n - \tau] \quad (4)$$

互相关函数是时延的函数, 与具体时间点无关. 利用互相关可以计算出一个信号相对于另一个信号移动的采样点数, 并由此求出信号的时延. 有了两耳的时延后, 通过反解方程可以得出 ITD 模型定位的声源水平方位.

当信号频率大于 2kHz , 就需要借助与 ILD 模型. 决定 ILD 模型好坏的关键是如何选择供参考的 $\text{Fun}_M(\theta)$ 函数. Viste 采用的是正弦函数模型^[5], 即 $\text{Fun}_M\theta = \sin(\theta)$. 机器人听觉系统含有左右对称耳廓装置, 利用耳廓的结构特性可以提高定位的精度. 耳廓结构具有独有性, 即不同的耳廓对信号的影响不同. 对于同一耳廓, 不同的频率, ILD- 方位角评估函数呈相似的曲线; 同一频率时, 不同的声源高度, 评估函数也呈相似的曲线.

对于某一耳廓结构, 提取的 ILD- 方位角评估函数模型如图 2 中粗线所示. 细线表示不同声源高度下的该耳廓结构的评估函数曲线, 虚线表示正弦函数模型曲线. 实验用的耳廓结构的 $\text{Fun}_M(\theta)$ 模型函数:

$$\text{Fun}(\theta, f) = \frac{1}{\alpha_M} (a\theta^4 + b\theta^3 + c\theta^2 + d\theta + e) \quad (5)$$

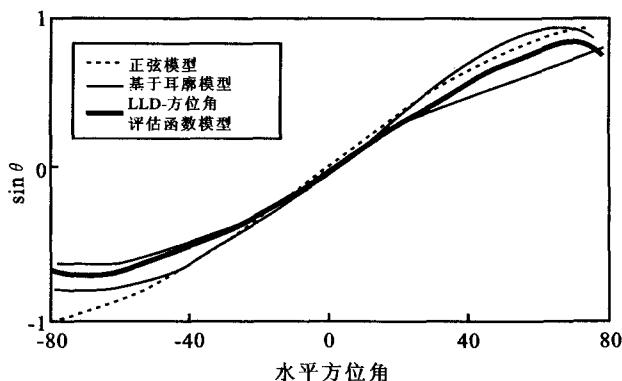


图2 评估函数模型曲线

对于机器人系统来说,耳廓结构是固定的.因此,只要建立某一频率时的模型函数,再利用不同频率时信号之间的关系,就可以对不同频率的声源方位进行判断.

对于高频信号而言,需要将两种模型结合起来进行综合定位.定位时,这里还需引入 ILD 和 ITD 模型的另外一种表达式.利用两耳的平均功率求出声压差,利用相位差求出两耳时间差.

$$\text{ILD}(t, f) = 10 \lg \left[\frac{\sum k s_r^2(k)}{\sum k s_l^2(k)} \right] \quad (6)$$

$$\text{ITD}_p(t, f) = \frac{1}{2\pi f} \left(\angle \frac{X_R(t, f)}{X_L(t, f)} + 2\pi k \right) \quad (7)$$

式中, $\sum s^2(k)$ 表示信号的平均功率; $X(t, f)$ 为信号 x 的快速傅里叶变换; k 为两耳之间的相位周期.由于 ILD 具有很强的分散性,不会由于频率的变化而变得极其模糊;而 ITD 计算的方位角具有更好的精确性.因此将 ILD 线索评估的方位作为粗方位,利用 ITD 模型来优化 ILD 评估的结果,即

$$\begin{aligned} \theta(t, f) &= \text{solve}(\text{ITD}_{T,k}), k \\ &= \arg \min | \text{ITD}_L(\theta_L, f) - \text{ITD}_P | \end{aligned} \quad (8)$$

定位的具体步骤如下:

- (1) 利用式(6)的 ILD 模型评估出方位角;
- (2) 将评估的方位角代入式(3)的 ITD 模型,求出 ITD;
- (3) 在式(7)中,利用计算的 ITD 评估出最佳的相位周期数 k ,即取最小整数;
- (4) 将评估的 k 值,代入式(7)进行 ITD 计算;
- (5) 反解式(3)的 ITD 模型,求出最终的声源水平方位.

4 声源水平方位角补偿

单一的听觉系统定位的声源水平方位角还不够精确,通过机器人的双目立体视觉系统可以对听觉系统定位的粗方位进行补偿,进一步地提高了定位的精度.

图3为双目立体视觉的成像原理图^[6-7],两摄像头光轴成平行位置. P 点为声源目标, $P_1(u_1, v_1)$ 和 $P_2(u_2, v_2)$ 为声源目标在摄像头成像平面的坐标, O_1 和 O_2 为左右眼摄像头的焦点, f 为焦距,两摄像头之间的基线距为 b . 由成像原理可知, P 点在左摄像机坐标系中的坐标为:

$$x_l = \frac{b - u_1}{|u_1 - u_2|}, y_l = \frac{b - v_1}{|u_1 - u_2|},$$

$$z_l = \frac{b \cdot f}{|u_1 - u_2|} \quad (9)$$

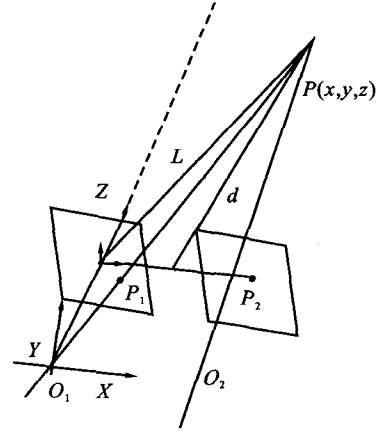


图3 双目立体视觉成像原理

图4为水平方位误差补偿示意图, C_1 、 C_2 代表左右摄像头, O 为头部中心, P 为声源目标, (x, y, z) 为该声源目标在左摄像头 C_1 坐标系中的坐标, C 为 C_1 摄像头坐标系的原点. 平面 I 称为成像平面, O 点所在平面 II 称为中心平面. DE 距离为 d_1 , OE 长度为 d_2 , 两摄像头中心之间的距离为 $2d_2$, 头部中心到摄像头连线的距离为 l . 当机器人旋转的角度为声源目标的真实水平方位 θ 时, 旋转后, 声源目标 P 的水平面的映射 A 将位于 C_1C_2 的中垂线. 实际应用中听觉评估的方位为 θ , 机器人旋转的方位为 θ' , θ' 是 θ 的取整形式.

$$\theta' = \text{fix}(\theta) \quad (10)$$

以 $\theta' < \theta$ 为例, 若旋转后 P 点位于 O' 左侧, 说明旋转的角度小于真实的方位角. 根据图4的几何关系, 建立误差补偿模型.

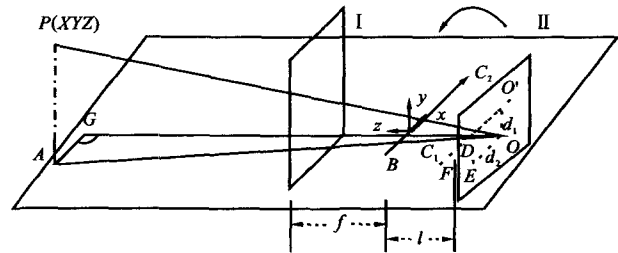


图4 水平方位误差补偿模型

P 在以 C 点为原点的坐标系中的坐标为 (x, y, z) , 在以 O' 为原点的坐标系中的坐标为

$$x' = x - d_2, y' = y, z' = z + l \quad (11)$$

将 P 点转化到以 O 点为原点的坐标系中的坐标:

$$x_o = x - d_2, y_o = y + d_1, z_o = z + l \quad (12)$$

此时的水平方位为

$$\text{AG} = x - d_2, \text{GO} = z + l \quad (13)$$

$$\Delta\theta = \arctan \frac{AG}{GO} = \arctan \frac{x - d_2}{z + l} \quad (14)$$

补偿后声源水平方位为 $\theta' + \Delta\theta$ ($\theta' < 0, \Delta\theta < 0$). 当 P 点位于 O' 右侧, 说明旋转的角度大于真实的水平方位, 此时 $\Delta\theta > 0$, 最终补偿后的水平方位为 $\theta' + \Delta\theta$ ($\theta' < 0, \Delta\theta > 0$).

当 $\theta > 0$ 时, 机器人向右旋转 θ' , 此时 $\theta' > 0$. 如 P 点同样位于 O' 左侧, 说明旋转的角度大于真实的方位角, 利用坐标公式计算后, 得到的补偿后水平方位为 $\theta' + \Delta\theta$ ($\theta' > 0, \Delta\theta < 0$). 如 P 点位于 O' 右侧, 说明旋转的角度小于真实的方位角, 则需要加上误差补偿角, 同样为 $\theta' + \Delta\theta$ ($\theta' > 0, \Delta\theta > 0$). 综合以上情况, 补偿后的方位角应为 $\theta' + \Delta\theta$.

5 实验结果

文中的主要实现算法通过 Matlab7.0 实现. 进行水平方位判断前, 先通过机器人头部的转动自由度(逆时针 90°)判断声源的前后方向. 旋转前后, 左耳麦克风声压减小, 右耳麦克风声压增加. 可知声源目标位于机器人的前方.

以正前方的声源目标为例, 此时坐标为 (2.5 cm, -2cm, 95cm), 声源目标水平方位角为 0° . 利用听觉定位系统评估的声源水平粗方位为 1.9° . 机器人头部逆时针旋转 1° 后, 声源目标在左摄像头坐标中的坐标为 (1.76cm, -2.01cm, 95.01cm), 根据补偿模型, 求得相应的补偿方位为

$$\Delta\theta = \arctan \frac{1.76 - 2.5}{95.01 + 5} = -0.997^\circ.$$

最终评估的声源目标水平方位为

$$\theta' + \Delta\theta = 1^\circ + (-0.997^\circ) = 0.003^\circ.$$

最终的误差为 0.003° . 保持声源高度不变, 当水平方位从 0° 移动到 80° 时, 对应的方位评估误差如图 5 所示.

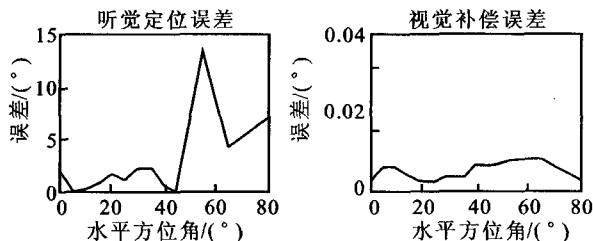


图 5 评估误差

可见利用听觉、视觉定位系统综合定位的方法具有较高的定位精度, 方位误差可以控制在 0.02° 以内.

6 结束语

文中根据机器人的自身特点, 利用多种定位线索, 并结合听觉系统和双目立体视觉系统, 提出了由粗到精的声源目标水平方位角评估策略和算法. 并通过实验证明定位的结果具有一定的可靠性, 能够满足机器人的作业要求. 此外定位的结果可以作为与其他传感系统(视觉系统、嗅觉系统)信息融合的基础.

参考文献:

- [1] 于梅, 殷兵, 何国栋, 等. 一种基于小波-Contourlet 变换的图像去噪算法[J]. 微电子学与计算机, 2008, 25(12): 100-102.
- [2] 夏平, 刘馨琼, 向学军, 等. 基于自适应提升方案的回波信号消噪算法研究[J]. 微电子学与计算机, 2008, 25(7): 201-204.
- [3] Xie Bosun, Zhong Xiaoli, Rao dan. Head-related transfer function database and its analyses[J]. Science in China Series G: Physics, Mechanics & Astronomy, 2007, 50(3): 267-280.
- [4] Tom Francart, Jan Brokx, Jan Wouters. Sensitivity to interaural level difference and loudness growth with bilateral bimodal stimulation[J]. Audiol Neurotol, 2008(13): 309-319.
- [5] Harald Viste. Binaural localization and separation techniques[D]. Swiss: Federal Institute of Technology, 2004.
- [6] Ogawa Y, Yamaoka D, Yamada H. Binocular stereo vision processing based on pulse coupled neural networks[C]// SICE 2004 Annual Conference. Japan, 2004, 1(1): 311-316.
- [7] Davison A. Real-time simultaneous localization and mapping with a single camera[C]// Proceedings Ninth IEEE International Conference on Computer Vision (ICCV). France, 2003: 1403-1410.

作者简介:

陈涛 男, (1981-), 博士研究生. 研究方向为机器人目标定位和信息融合.

张明路 男, (1964-), 教授. 研究方向为移动机器人、多智能体和目标识别.

付灵丽 女, (1966-), 副教授. 研究方向为虚拟现实和模式识别.