

# 一种基于SIFT特征的机器人环境感知双目立体视觉系统

宋海涛<sup>†</sup>, 何文浩, 原 魁

(中国科学院 自动化研究所, 北京 100190)

**摘 要:** 针对双目立体视觉系统在机器人环境感知领域中存在的立体匹配以及测量精度问题, 设计一种基于 SIFT 特征的双目立体视觉测量系统. 利用 SIFT 特征良好的旋转、尺度、光照不变性等特性, 有效地解决双目立体视觉系统的匹配问题, 同时将 SIFT 算法在由 FPGA 和 DSP 组成的嵌入式系统上实现, 显著地提高测量系统的实时性. 提出一种基于二次多项式的误差补偿方法, 对系统的测量结果进行补偿, 弥补双目立体视觉系统测量误差随测量距离增加而增加的不足, 从而提高系统测量精度. 通过实际的测量实验、移动机器人环境感知实验以及与传统双目立体视觉产品的对比实验结果表明, 系统能够很好地解决双目立体视觉的立体匹配和精度问题, 并且能较好地应用于移动机器人环境感知任务中.

**关键词:** 机器人; 环境感知; SIFT; 双目立体视觉; FPGA; DSP

**中图分类号:** TP249

**文献标志码:** A

## A stereo vision system based on SIFT feature for robot environment perception

SONG Hai-tao<sup>†</sup>, HE Wen-hao, YUAN Kui

(Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** To solve the stereo matching and measurement accuracy problems which exist in stereo vision systems in the field of robot environment perception, this paper designs a stereo vision system based on scale-invariant feature transform (SIFT) features. The matching problem is solved by using the rotation, scale, illumination invariance of the SIFT features, and the SIFT algorithm is implemented in the embedded system based on FPGA and DSP, thus the real-time problem of the system is effectively solved. Then in order to offset the problem that the measurement error increases with the distance, an error compensate method based on quadratic polynomial is proposed, which can compensate the measurement results of the stereovision system and improve the measurement accuracy of the system. The measurement experiment, robot environment perception experiment and contrast experiment between existing stereovision products are carried out, and the results show that, the stereovision system can commendably solve the matching and accuracy problems, and also can be better applied to the environment perception task of mobile robots.

**Keywords:** robot; environment perception; SIFT; stereo vision; FPGA; DSP

## 0 引 言

双目立体视觉测量系统模仿人类利用双眼同时观察一个目标,能够在获得二维图像的同时获得场景的三维信息.系统具有成本低、结构简单、使用方便等优点,因此在机器人环境感知领域有着非常广泛的应用前景<sup>[1-5]</sup>.然而,在双目立体视觉系统中,如何实时、准确地确定空间一点在两个摄像机像平面上的像点之间的匹配关系这一关键问题一直未能得到很好的解决,因此,双目立体视觉测量方法也一直未能得到很好的推广应用.

尺度不变特性变换 (Scale-invariant feature transform, SIFT) 算法是 Lowe<sup>[6-7]</sup> 于 1999 年首先提出,并于 2004 年总结完善后正式提出的一种基于尺度空间的特征点检测算法. SIFT 特征点对图像的旋转和尺度变化具有不变性,对三维视角变化和光照变化具有很强的适应性,局部特征在场景杂乱时仍保持不变性,特征之间相互区分的能力强,这些特点有利于特征点之间的匹配,非常适合用于解决双目立体视觉中的立体匹配问题.

虽然 SIFT 算法是目前研究最多、应用最广的特

收稿日期: 2017-12-27; 修回日期: 2018-04-22.

基金项目: 国家 863 计划项目 (2015AA042307); 国家自然科学基金项目 (61421004).

责任编辑: 牛玉刚.

<sup>†</sup>通讯作者. E-mail: haitao.song@ia.ac.cn.

征匹配方法之一,但是应用SIFT算法进行特征匹配存在一个很大的问题,即高维度SIFT特征描述符的复杂计算过程造成算法实时性差的问题.为了克服传统SIFT算法存在的不足,目前研究的主要改进方法包括:降低特征描述向量维度简化算法复杂性<sup>[8-10]</sup>和应用硬件设计加速SIFT算法<sup>[11]</sup>.

本文针对上述SIFT算法实时性和双目立体视觉测量系统在机器人环境感知应用中存在的问题,设计一种基于SIFT特征的双目立体视觉测量系统.该系统只需得到建立匹配关系的SIFT特征点处的三维信息,而不是整幅图像的三维信息,这样的设计既能避免整幅图像三维信息带来的数据冗余给移动机器人环境感知系统造成数据处理压力,又能为移动机器人环境感知系统提供足够多关键位置点的三维信息,有望在移动机器人环境感知系统中得到推广应用.

### 1 测量算法与系统

#### 1.1 测量原理及模型

双目立体视觉测量方法是由两幅图像获取物体三维几何信息的方法.在计算机视觉系统中,双目立体视觉测量系统一般由两台摄像机从不同角度同时获取周围景物的两幅数字图像,或由一台摄像机在不同时刻从不同角度获取周围景物的两幅数字图像,并基于视差原理恢复出物体三维几何信息,重建周围景物的三维形状与位置.双目立体视觉测量的基本模型如图1所示.

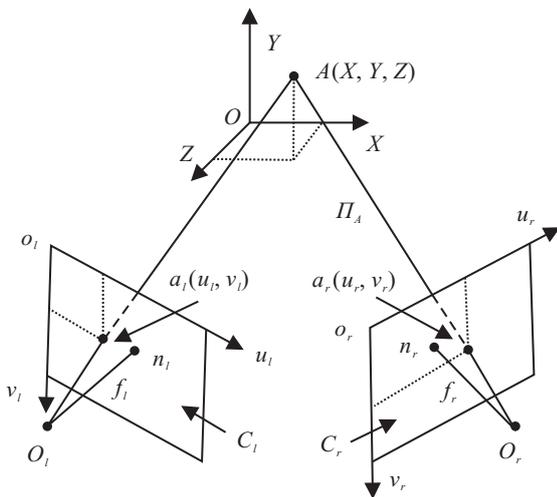


图1 双目立体视觉测量模型

由图1可以得到双目立体视觉测量的基本计算模型为

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \frac{1}{Z_l} \begin{bmatrix} f_{xl} & 0 & u_{0l} & 0 \\ 0 & f_{yl} & v_{0l} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_l \\ Y_l \\ Z_l \\ 1 \end{bmatrix}, \quad (1)$$

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \frac{1}{Z_r} \begin{bmatrix} f_{xr} & 0 & u_{0r} & 0 \\ 0 & f_{yr} & v_{0r} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_r \\ Y_r \\ Z_r \\ 1 \end{bmatrix}. \quad (2)$$

其畸变模型为

$$\begin{aligned} \tilde{u}_l &= \\ u_l &+ (u_l - u_{0l})[k_{1l}(x_l^2 + y_l^2) + k_{2l}(x_l^2 + y_l^2)^2] + \\ &2p_{1l}y_l + p_{2l}[(x_l^2 + y_l^2) + 2x_l^2], \\ \tilde{v}_l &= \\ v_l &+ (v_l - v_{0l})[k_{1l}(x_l^2 + y_l^2) + k_{2l}(x_l^2 + y_l^2)^2] + \\ &p_{1l}[(x_l^2 + y_l^2) + 2y_l^2] + 2p_{2l}x_l; \end{aligned} \quad (3)$$

$$\begin{aligned} \tilde{u}_r &= \\ u_r &+ (u_r - u_{0r})[k_{1r}(x_r^2 + y_r^2) + k_{2r}(x_r^2 + y_r^2)^2] + \\ &2p_{1r}y_r + p_{2r}[(x_r^2 + y_r^2) + 2x_r^2], \\ \tilde{v}_r &= \\ v_r &+ (v_r - v_{0r})[k_{1r}(x_r^2 + y_r^2) + k_{2r}(x_r^2 + y_r^2)^2] + \\ &p_{1r}[(x_r^2 + y_r^2) + 2y_r^2] + 2p_{2r}x_r. \end{aligned} \quad (4)$$

两台摄像机之间的空间位置关系可以用结构参数表示为

$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_l \\ Y_l \\ Z_l \\ 1 \end{bmatrix}. \quad (5)$$

以上各式中,  $f_{xl}$ 、 $f_{yl}$ 、 $u_{0l}$ 、 $v_{0l}$  和  $f_{xr}$ 、 $f_{yr}$ 、 $u_{0r}$ 、 $v_{0r}$  分别为左右摄像机的内参数;  $k_{1l}$ 、 $k_{2l}$ 、 $p_{1l}$ 、 $p_{2l}$  和  $k_{1r}$ 、 $k_{2r}$ 、 $p_{1r}$ 、 $p_{2r}$  分别为左右摄像机的畸变系数;  $R$  为两摄像机坐标系之间的旋转矩阵;  $T$  为两摄像机坐标系之间的平移向量. 在参数已知的情况下, 可根据立体图像对中建立匹配的特征点对图像坐标唯一地确定相应空间点的三维世界坐标. 由式(3)和(4)可以看出, 摄像机的畸变模型计算过程极其复杂, 不利于嵌入式系统实现, 因此, 本文方法通过标定得到的畸变系数离线计算得到整幅图像的畸变校正查找表, 通过查找表对图像进行校正. 经查找表校正后, 空间点三维坐标计算复杂度大大降低. 将式(1)和(2)展开代入(5), 使其成为关于  $Z_l$  的方程组, 并由其计算得到  $Z_l$  的值, 从而得到空间点的三维坐标.

#### 1.2 测量系统

图2左侧所示为本文使用的双目立体视觉测量实验系统. 它由两台相同的CCD摄像机组成, 摄像机安装在一个间距可调的齿轮齿条移动台上. 图2右侧

为实验室自主开发的嵌入式图像处理卡实物图,硬件主要包括FPGA、DSP、SRAM、SDRAM、FLASH、两路视频输入、一路视频输出、串口、百兆以及千兆以太网接口等,其几何尺寸为 $158 \times 123 \text{ mm}^2$ ,本文测量算法在该系统上实现。

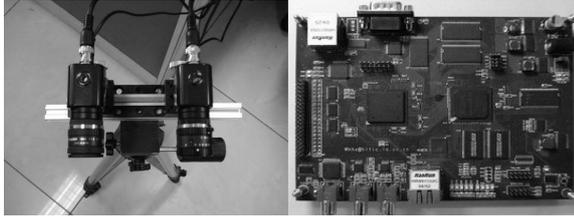


图2 测量实验系统

### 1.3 测量系统标定

由上述原理可知,已知两台摄像机的内参数和结构参数,就可以根据匹配点的图像坐标唯一地确定相应空间点的三维世界坐标。同样地,本文方法中的误差补偿模型也具有参数,摄像机参数和补偿模型参数都可以通过标定得到。因此,本文系统的标定过程包括两个部分,即摄像机参数的标定和补偿模型参数的标定。

需要标定的摄像机参数包括上文介绍的两台摄像机的内参数、畸变参数以及两台摄像机之间的结构参数。摄像机参数的标定使用Jean-Yves Bouquet提供的Matlab工具箱得到。本文所提出的误差补偿模型为二次多项式模型,因此采用曲线拟合的方法估计补偿模型的参数。

## 2 基于二次多项式的误差补偿方法

通常情况下,双目立体视觉测量系统计算得到的三维坐标系是与左摄像机坐标系重合的。然而,在实际情况下,摄像机光心的位置很难精确定位,通常引入一个与摄像机平面平行的参考平面,并将原点沿Z坐标正方向平移至参考平面。在实际测量时,深度信息是指被测点到参考平面的垂直距离。

对于摄像机而言,在测量过程中很难保证摄像机的光轴线垂直于被测目标平面,当发生倾斜时,这种倾斜无法用肉眼进行调整,因此,导致测量误差随摄像机距目标距离的增大而不断增大。此外,双目立体视觉测量系统的基线距B对视觉测量系统精度的影响也较为复杂,其影响是非线性的。一般而言,对于工作距离较大的系统,要求系统的基线距必须也较大,当系统结构已经确定时,系统工作距离越大,测量精度越低。因此,为了减少摄像机畸变、被测平面与摄像机光轴不垂直以及基线距有限等因素对测量结果的影响,并满足移动机器人环境感知任务对于较大测

量范围的需求,本文尝试引入某种误差补偿模型对双目立体视觉系统的测量结果进行补偿,以提高系统的测量精度。

图3为实际采集得到的一组双目立体视觉测量系统误差与实际测量距离的曲线拟合结果。由图3可见,误差拟合曲线与二次多项式曲线极其接近,因此本文提出一种基于二次多项式的误差补偿模型,以提高系统的测量精度。该方法首先使用上文介绍的公式计算得到空间点的三维坐标,然后利用二次多项式补偿模型对Z坐标进行补偿得到最终的三维坐标。

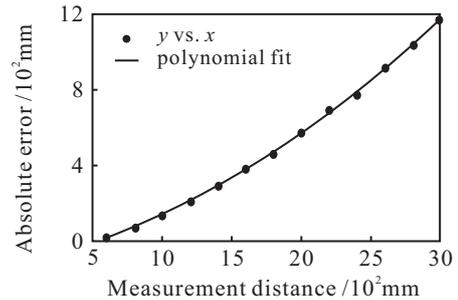


图3 测量距离与绝对误差曲线拟合结果

引入的二次多项式误差补偿模型为

$$e = p_1 Z_c^2 + p_2 Z_c + p_3. \quad (6)$$

其中: $p_1$ 、 $p_2$ 、 $p_3$ 为补偿参数; $Z_c$ 为空间点到参考平面的垂直距离; $e$ 为实际距离与测量结果之间的绝对误差,有

$$e = Z_c - Z, \quad (7)$$

$Z$ 表示由双目立体视觉测量方法直接计算得到的结果。经过补偿得到的测量结果表示为

$$Z_c = \frac{-(p_2 - 1) - \sqrt{(p_2 - 1)^2 - 4p_1(p_3 + Z)}}{2p_1}. \quad (8)$$

## 3 测量算法在FPGA和DSP中的实现

本文所使用双目立体视觉测量算法主要包括图像校正、SIFT特征提取、特征描述向量计算以及特征点匹配几个主要模块,下面将分别进行详细介绍。

### 3.1 图像校正

本文方法的图像校正过程通过查找表实现。畸变校正查找表通过标定得到的畸变参数离线计算得到,每台摄像机对应一张查找表。

畸变校正查找表是一张二维表,其大小与图像大小相同,查找表中的数据为16位无符号数,数据结构如图4所示,高8位表示原始图像中像素点在校正过程中 $u$ 坐标的变化量,低8位表示 $v$ 坐标的变化量,每个8位数据中的最高位表示相应像素点的移动方向,0表示正向移动,1表示反向移动。使用查找表对图像进行校正的方法可以在FPGA上使用流水线结

构来实现实时校正.

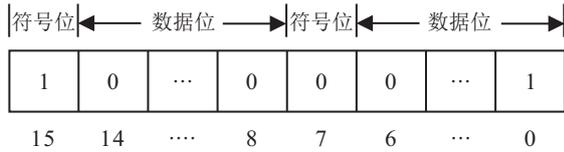


图4 畸变校正查找表数据结构

### 3.2 SIFT特征点提取

SIFT特征点的提取过程主要包括构建尺度空间、检测DoG尺度空间极值点和精确定位关键点并去除不稳定的特征点3个步骤<sup>[6]</sup>,计算量较大,难以实现实时处理.为了解决SIFT算法的实时计算问题,科研人员主要从3个方面进行探索,即对算法本身进行改进<sup>[9]</sup>、利用GPU进行计算<sup>[11]</sup>以及在FPGA中用硬件电路实现SIFT算法<sup>[12-18]</sup>.将SIFT算法在硬件电路中实现是解决算法实时性的重要途径.通过充分挖掘计算的并行性、设计合理的流水线结构,采用FPGA检测SIFT特征点的速度能够比通用计算机平台软件方法快成百上千倍.但是,文献[12-18]都存在一个比较突出的问题,即计算精度不足.这些文献所提出的方案普遍没有采用足够多的定点数位,因此很难达到与软件相当的计算精度,从而限制了其实用性.

本文所使用的FPGA为低成本产品,其内部硬件资源有限,为了能够在FPGA上实现SIFT特征点的检测,一方面要通过算法改进降低算法的计算量,另一方面要设计合理的电路结构,提高硬件资源的使用效率.本文根据嵌入式硬件系统的工作特点,对SIFT特征点检测算法进行改进,具体的改进体现在以下几方面:

- 1) 为了减少FPGA资源的消耗量,不对图像进行增大操作,而是直接用标准差较小的高斯函数对原始图像进行高斯滤波得到较高的尺度层;
- 2) 采用文献[6]介绍的次优方案进行尺度层选择,即以 $2^{1/2}$ 的尺度层间距构造高斯金字塔,并用 $\sigma = \sqrt{2}$ 的高斯函数得到每个Octave中的第1个尺度层;
- 3) 舍去过低和过高的尺度层,仅选择第1个Octave的第2个尺度层以及第2个Octave中的两个尺度层,共3个尺度层进行特征点的检测;
- 4) 使用面密度插值法和双重极值点约束检测特征点的位置;
- 5) 取消特征点的“精确定位”过程.

通过以上对原算法的有效改进,将SIFT特征点检测在完全同步的流水线结构上实现,高斯滤波、

DoG图像的求取、极值点的检测分别处在流水线的不同级别上,实现了计算在时间上的并行性;3个尺度层上的特征点可以同时进行检测,构成3条流水线,实现了计算在空间上的并行性,从而使算法的性能得到显著提高.

### 3.3 特征描述向量的计算

SIFT特征描述向量的计算过程包括关键点主方向的计算以及对局部图像区域的坐标旋转<sup>[6]</sup>,这些操作计算过程复杂、计算量大.在FPGA中实现,不但需要占用较多的硬件资源,而且需要借助DSP的辅助处理,不可避免地会影响整个算法的执行时间.另一方面,由于SIFT特征描述向量的计算主要是为了保证特征点的旋转不变性,而本文系统得到的两幅图像之间并不存在较大的旋转、视角、尺度以及光照变化,不需要如此复杂的高维度描述向量.

本文根据区域匹配原理对特征描述向量的计算过程进行简化,并降低描述向量的维度,具体计算过程如下:

- 1) 以特征点为中心选取一个 $9 \times 9$ 的邻域;
- 2) 对邻域内的像素点进行 $3 \times 3$ 窗口的中值滤波;
- 3) 直接使用滤波后邻域内的像素值构成一个81维的特征描述向量;
- 4) 将特征描述向量进行归一化处理,进一步消除光照变化的影响.

通过以上简化,不仅能够大大提高特征描述向量的计算速度,减少存储向量所需的存储空间,而且可以进一步减少后续特征点匹配的计算量,使得整个算法的执行时间更短.

### 3.4 特征点匹配

立体图像对中特征点之间匹配正确与否以及匹配精度将直接决定整个测量算法的成败和最终的测量精度.由于本文方法简化了特征描述向量的计算过程,降低了描述向量的维度,使得描述向量中包含的信息量比原始算法有所减少.为了保证匹配的正确率,本文使用双向匹配的搜索策略对左右图像中的特征点进行匹配(即左右一致性检验).具体匹配步骤如下:

- 1) 在左图像中选择一个待匹配点;
- 2) 在右图像中搜索与左图像选择的点构成一对候选匹配点,本文使用差值平方和(Sum of squared difference, SSD)作为匹配代价函数;
- 3) 在左图像中搜索与2)中得到的右图像候选匹配点匹配代价函数最小的点.

若搜索得到的点与最初选择的待匹配点相同,则

得到一对匹配点,否则本次匹配失败.

### 4 实验分析

为了验证基于SIFT特征的双目立体视觉测量方法的有效性,本节首先对实验系统进行标定;然后进行测量精度的验证实验,并将其安装在移动机器人平台完成自动入库实验;最后将本文方法与其他SIFT改进及实现方法进行比较,并将未来可能基于本文方法设计的立体相机与现有双目立体相机产品进行对比.所有标定以及测量实验都在实验室自然环境下进行,实验中的真实距离是指测量系统前端到被测表面的垂直距离,这一距离通过刻度尺事先测量得到,其测量精度为1 mm.

#### 4.1 标定结果

##### 4.1.1 摄像机参数标定结果

本文摄像机标定过程使用棋盘格平面作为标定参照物.左右两台摄像机分别同时采集10幅不同视角、不同距离上的标定参照物图像.标定得到的摄像机参数如下所示:

$$\begin{aligned} f_{xl} &= 885.533\ 89, f_{yl} = 967.034\ 82, \\ u_{0l} &= 176.571\ 15, v_{0l} = 162.909\ 34, \\ k_{1l} &= -0.060\ 10, k_{2l} = -2.539\ 66, \\ p_{1l} &= 0.003\ 66, p_{2l} = 0.001\ 86. \end{aligned} \tag{9}$$

$$\begin{aligned} f_{xr} &= 884.273\ 82, f_{yr} = 966.599\ 21, \\ u_{0r} &= 162.935\ 44, v_{0r} = 170.813\ 27, \\ k_{1r} &= -0.160\ 79, k_{2r} = -1.064\ 27, \\ p_{1r} &= 0.005\ 16, p_{2r} = -0.006\ 05, \end{aligned} \tag{10}$$

$$R = \begin{bmatrix} 0.999\ 67 & 0.022\ 67 & -0.012\ 10 \\ -0.022\ 84 & 0.999\ 64 & -0.014\ 22 \\ 0.011\ 77 & 0.014\ 49 & 0.999\ 83 \end{bmatrix}, \tag{11}$$

$$T = [-111.030\ 39\ 1.923\ 23\ 0.999\ 83]^T. \tag{12}$$

##### 4.1.2 补偿参数标定结果

在已知摄像机参数的情况下,沿着深度方向(Z坐标方向),在600~3000 mm范围内,每隔200 mm对已知垂直于光轴的平面进行一次测量,将测量结果与实际测量得到的被测平面到平行于摄像机光平面的参考平面距离进行比较,实验数据如表1所示.

由表1可见,所得到的绝对误差相当大,这是因为表1数据中的实际距离与测量距离指的不是同一个距离.测量距离指的是摄像机光心到被测物体表面的距离,而实际距离指的是位于测量系统前端的、与摄像机光平面平行的参考平面到被测表面的距离.虽然可以标定得到摄像机光心到参考平面的距

表1 未补偿的实验数据

实际距离/mm	测量距离/mm	绝对误差/mm	相对误差/%
600	583.616 146	16.383 54	2.73
800	727.869 19	72.130 81	9.02
1000	861.174 1	138.825 9	13.88
1200	987.577 1	212.422 9	17.70
1400	1 105.452 1	294.547 9	21.04
1600	1 226.234 6	373.765 4	23.36
1800	1 340.673 0	459.327 0	25.52
2000	1 429.403 4	570.596 6	28.53
2200	1 514.342 4	685.657 6	31.17
2400	1 630.635 3	769.364 7	32.06
2600	1 690.813 2	909.186 8	34.97
2800	1 765.474 8	1034.526 2	36.95
3000	1 829.348 0	1 170.652 0	39.02

离,但是这一标定过程本身会携带测量误差,因此,本文将整体考虑,采用误差补偿的方式予以消除.当实际距离(即参考平面到被测平面之间的垂直距离)为600 mm时,测量结果的相对误差为2.73%;当实际距离为3000 mm时,相对误差达到39.02%.整体呈现出随着测量距离的增加测量误差逐渐增大的趋势,这与前面的理论分析结果一致.

由表1实际距离和绝对误差进行曲线拟合,标定得到补偿参数为

$$\begin{aligned} p_1 &= 8.698e - 5, p_2 = 0.166\ 8, \\ p_3 &= -114.8. \end{aligned} \tag{13}$$

#### 4.2 测量实验

使用前文标定得到的参数进行实际测量实验.实验在600~3000 mm的范围内进行,每隔100 mm使用刻度尺测量一次参考平面到已知平面的距离,其精度为1mm,并将被测平面上所有特征点的深度数据(Z坐标值)取平均值作为该次测量的测量结果.补偿后测量结果的相对误差最大值为2.028%,平均误差为0.589%,图5为本次实验的相对误差曲线.

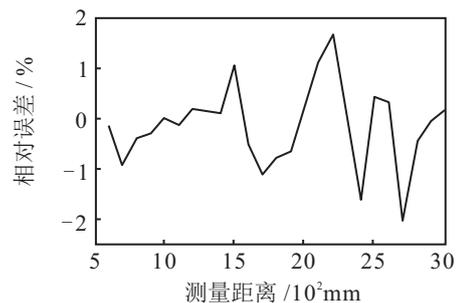


图5 测量误差曲线

### 4.3 移动机器人自动入库实验

本文设计的基于SIFT特征的双目立体视觉测量系统能够获得场景中建立匹配的SIFT特征点上的三维信息,所得到的特征点可能分布在整个场景,并且场景越复杂特征点越多,而通常场景中越复杂的部分对环境理解越有帮助.因此,该系统既能够得到环境内足够多特征点的三维信息,又能够避免整幅图像三维信息庞大的数据量而导致的处理压力.本节将基于SIFT特征的双目立体视觉测量系统安装在实验室的自主移动机器人平台上,使其在室内环境下完成简单的自动入库任务.

在该任务中,移动机器人需要根据整个场景的三维信息点分布情况估算出车库的方位,规划入库路径,在入库的过程中,不断对车库的方位进行更新并修正入库路径,进入库内在规定的停放距离处停止,完成最终的入库任务.图6为移动机器人完成自动入库任务的视频截图.

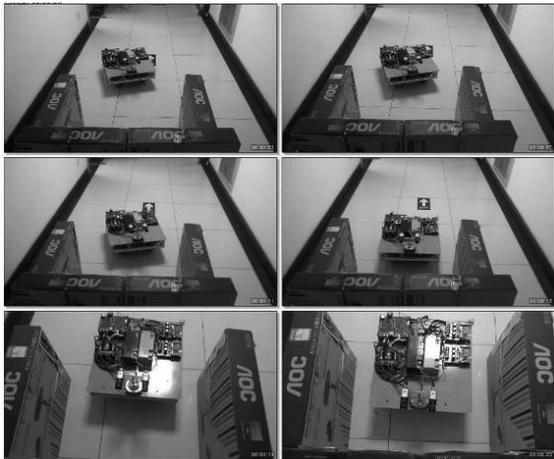


图6 移动机器人自动入库任务视频截图

从实验结果中可以看出,本文所提出的基于SIFT特征的双目立体视觉测量方法能够很好地为移动机器人提供必要的环境信息,使移动机器人能够快速地对环境进行理解并规划出相应的策略.

### 4.4 与其他SIFT改进算法比较

本节对SIFT算法、SIFT的改进算法SURF、GPU上实现的SIFT算法以及本文在FPGA上实现的SIFT算法的运算速度进行比较,其中SIFT和SURF算法在Intel Core i5-4210U型号CPU下完成,GPU实现的数据来自文献[11],所使用的GPU型号为NVIDIA Geforce 8 800M,具体比较结果如表2所示.其中,本文方法在FPGA上将所有的图像处理任务实现在同一时钟的流水线上,使得处理结果与输入视频流同步,只存在几行图像像素时钟的延迟,因此,表中的运算时间为DSP的计算时间.

表2 几种不同的SIFT改进及实现方法运行速度对比表

方法名称	运算速度/s	方法名称	运算速度/s
SIFT(CPU)	2.25	SIFT(GPU)	0.03
SURF(CPU)	0.89	本文方法	0.03

由表2数据可以看出,虽然在现今的电脑硬件条件下,SIFT算法的速度得到大幅提高,但是其处理速度仍需要几秒钟,随着GPU性能的提升,GPU实现的SIFT算法其执行速度必定可以超越本文方法,但是GPU的功耗较高,无法满足移动机器人环境感知系统对于低功耗的要求,因此综合考量,本文方法对于移动机器人环境感知系统而言,将会是一种比较有优势的解决方案.

### 4.5 与现有双目立体视觉产品比较

立体相机可以得到物体的深度信息,能够弥补单个视觉传感器缺少距离信息的缺陷,因此在机器人环境感知领域得到较广泛的应用.市面上已经出现了一些双目立体视觉传感器产品,本文的主要目的是研制开发新型双目立体视觉传感器产品.图7分别给出了基于本文方法的双目立体相机设计草图(上)、PointGrey公司的Bumblebee2(左中)、Stereolabs公司的ZED(右中)以及Nerian公司的Karmin2立体相机(下).基于本文方法的双目立体视觉传感器与上述其他双目立体视觉传感器的主要特点比较如表3所示.

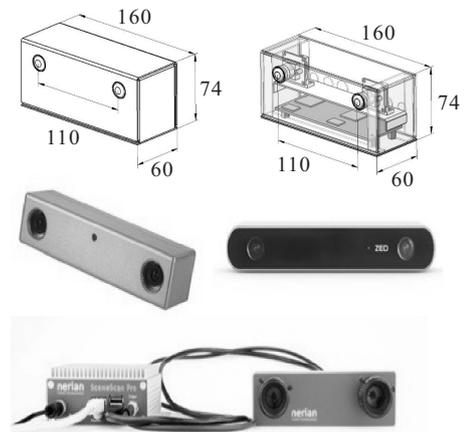


图7 本文系统设计草图及双目立体视觉传感器产品

由表3可见,现有的双目立体视觉测量方法所获得的都是整幅图像的三维数据,这对于机器人环境感知任务而言是一个极其庞大的数据量,要想在这些数据中获取有用的环境信息,感知系统需要处理大量的数据,这必然会占用较多的系统时间,从而影响感知系统的处理速度,难以满足服务机器人实时环境感知任务的要求.此外,现有的立体视觉传感器还普遍存在需要上位机(造成体积和功耗过大)、处理速度慢、测量精度低等问题,难以用于普通的服务机器人.

表3 双目立体视觉传感器主要特点对比表

名称	Bumblebee2	ZED	Karmin2	本文方法
分辨率	658×488 max	2 208×1 242 max	800×800 max	取决于物体的图像特征
基线距离/mm	120	120	100	110
测量范围	较小	较大	较小	较大
测量精度	较低	低	较低	较高
测量速度	较低	较低	较高	较高
提供图像特征	不可以	不可以	不可以	可以
体积	较大	较大	较小	较小
功耗	较大	较大	较小	小
价格	较高	较高	较高	较低
主要优点	可以实时获取全视场的深度信息,并能够对镜头畸变和相机位置偏差进行自动校正	可以实时获取全视场的深度信息,并得到高分辨率的景深图像	使用配套的FPGA套件可以实时得到深度图像	体积小、成本低、功耗低、能够实时获取场景中关键点处的深度信息并同时提供二维图像特征,具有较高的测量精度
主要缺点	深度信息由计算机得到,系统功耗和体积较大;深度信息过密,数据量过大,处理难度较大	深度信息由计算机的GPU运算得到,系统功耗和体积较大;深度信息过密,数据量过大,处理难度较大	深度信息过密,数据量过大,处理难度较大	深度信息的提取依赖于观测对象的图像特征,难以得到比较稠密的深度信息

4.6 结果分析

以上实验的主要目的是为了对本文设计的基于SIFT特征的双目立体视觉测量系统进行可行性以及测量精度的验证.通过距离测量实验结果可以看出,使用本文方法能够在较大的距离范围内得到稳定且较高的测量精度,不会出现其他双目立体视觉测量系统中存在的测量精度随测量距离增大而降低的问题.

与现有的立体视觉传感器相比,本文系统既能很好地解决双目立体视觉中立体匹配的难题,又能通过对测量结果的补偿进一步提高系统测量精度.虽然所提出方法只能得到环境中部分特征点处的深度信息,但对于一般的环境感知任务而言,感知系统只需要获得场景中某些关键位置点的三维信息,如角点、边缘点以及一些图像变化较大的位置等,而SIFT特征点恰恰出现在这些关键位置,因此利用SIFT特征点处的深度信息可以较好地满足服务机器人的环境感知任务.

本文方法中的图像校正、SIFT特征提取等需要进行大量计算的图像处理任务在FPGA中实现,而DSP则完成特征点的匹配以及三维信息的计算,使

得测量系统的处理速度得到大幅提升.因此,基于本文系统的环境感知传感器具有体积小、使用方便、实时性好、精度较高、功耗较低等特点,可以很好地满足移动机器人对于快速地进行环境理解的需求,具有非常好的应用前景.

5 结论

本文设计了一种基于SIFT特征的双目立体视觉测量系统,其测量算法在由FPGA和DSP组成的嵌入式图像处理系统上实现.该系统能够获得两幅图像之间建立匹配的SIFT特征点处的三维信息,移动机器人可以利用这些三维信息对所处环境进行快速理解,实现相应的控制策略.

所提出方法对SIFT特征点的提取过程进行了修改,简化了特征描述向量的计算复杂度和维度,并将其在FPGA和DSP上进行实现,使得测量算法的执行速度得到大幅提升,可以较好地满足移动机器人环境感知系统的需要.同时,为了弥补简化特征描述向量导致匹配准确率下降的问题,引入了双向匹配搜索策略.此外,针对双目立体视觉测量误差随着测量距离增加而增加的问题,提出一种基于二次多项式的误差补偿方法,有效地提高了系统的测量精度.最后通过

实验验证了系统用于移动机器人环境感知系统的可行性和有效性。

#### 参考文献(References)

- [1] Wu Y, Rui N, He B. 3D reconstruction model of underwater environment in stereo vision system[J]. *Oceans*, 2013, 8(4): 1-4.
- [2] Fan X, Liu X, Wang X, et al. An automatic robot unstacking system based on binocular stereo vision[C]. *Proc of the Int Conf on Security, Pattern Analysis, and Cybernetics*. Wuhan: IEEE, 2014, 6: 86-90.
- [3] Qu J, Zhang F, Fu Y, et al. Approach movement of redundant manipulator using stereo vision[C]. *Proc of the IEEE Int Conf on Robotics and Biomimetics*. Dali: IEEE, 2014: 2489-2494.
- [4] Puheim M, Bundzel M, Madarasz L. Forward control of robotic arm using the information from stereo-vision tracking system[C]. *Proc of the IEEE Int Symposium on Computational Intelligence and Informatics*. Budapest: IEEE, 2013: 57-62.
- [5] Nakazato K, Touma Y, Hagiwara H, et al. FPGA-based stereo vision system using census transform for autonomous mobile robot[C]. *Proc of the Int Conf on Informatics, Electronics & Vision*. Fukuoka: IEEE, 2015: 1-4.
- [6] Lowe D G. Object Recognition from local scale-invariant features[C]. *Proc of the IEEE Int Conf on Computer Vision*. Fukuoka: IEEE, 1999, 2: 1050-1057.
- [7] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *Int J of Computer Vision*, 2004, 60(2): 91-110.
- [8] Ke Y, Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors[C]. *Proc of the IEEE Conf on Computer Vision and Pattern Recognition*. Washington DC: IEEE, 2004, 2: 66-75.
- [9] Bay H, Tuytelaars T, Gool L V. SURF: Speeded up robust features[J]. *Computer Vision & Image Understanding*, 2006, 110(3): 404-417.
- [10] Heikkil M, Pietik Inem M, Schmid C. Description of interest regions with local binary patterns[J]. *Pattern Recognition*, 2009, 42(3): 425-436.
- [11] Chariot A, Keriven R. GPU-boosted online image matching[C]. *Proc of the Int Conf on Pattern Recognition*. Tampa FL: IEEE, 2008: 1-4.
- [12] Chati H D, Muhlbauer F, Braun T, et al. Hardware/Software co-design of a key point detector on FPGA[C]. *Proc of the IEEE Symposium on Field-Programmable Custom Computing Machines*. Napa CA: IEEE, 2007: 355-356.
- [13] Bonato V, Marques E, Constantinides G A. A parallel hardware architecture for scale and rotation invariant feature detection[J]. *IEEE Trans on Circuits and Systems for Video Technology*, 2008, 18(12): 1703-1712.
- [14] Qiu J, Huang T, Ikenaga T. A 7-round parallel hardware-saving accelerator for gaussian and dog pyramid construction part of SIFT[C]. *Proc of the Asian Conf on Computer Vision*. Xi'an: Springer-Verlag, 2009: 75-84.
- [15] Qiu J, Huang T, Ikenaga T. A FPGA-based dual-pixel processing pipelined hardware accelerator for feature point detection part in SIFT[C]. *Proc of the Int Joint Conf on INC, IMS and IDC*. Seoul: IEEE, 2009: 1668-1674.
- [16] Chang L. A hardware architecture for SIFT candidate keypoints detection[C]. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Guadalajara: Springer-Verlag, 2009: 95-102.
- [17] Yao L, Feng H, Zhu Y, et al. An architecture of optimised SIFT feature detection for an FPGA implementation of an image matcher[C]. *Proc of the Int Conf on Field-Programmable Technology*. Sydney: IEEE, 2009: 30-37.
- [18] Pettersson N, Petersson L. Online stereo calibration using FPGAs[C]. *Proc of the IEEE Intelligent Vehicles Symposium*. Las Vegas: IEEE, 2005: 55-60.

#### 作者简介

宋海涛(1987—),男,助理研究员,博士,从事机器人视觉、嵌入式视觉系统的研究, E-mail: haitao.song@ia.ac.cn;

何文浩(1982—),男,副研究员,博士,从事嵌入式系统设计与开发的研究, E-mail: wenhao.he@ia.ac.cn;

原魁(1957—),男,研究员,博士生导师,从事智能机器人、智能传感与智能信息处理技术、人机接口技术、虚拟现实技术等研究, E-mail: kui.yuan@ia.ac.cn.

(责任编辑: 郑晓蕾)