

DOI: 10.19650/j.cnki.cjsi.J2107832

基于FPGA的三目半全局匹配算法设计与实现

刘阳,冯进良,黄伟,康世勋,马韵琪
(长春理工大学光电工程学院 长春 241000)

摘要:三目立体视觉系统能够克服双目立体视觉系统存在的遮挡等问题,进一步提高立体视觉系统测量精度。然而增加传感器数量会导致匹配算法计算量增大,影响系统实时性,从而限制三目立体视觉系统在各领域的实际应用。为此,本文提出一种三目半全局立体匹配算法及其硬件计算框架。首先,在对三目立体视觉系统基本模型深入分析的基础上,提出一种硬件友好的半全局匹配算法。随后,根据FPGA硬件并行化计算和流水线处理的特点,对片上系统整体框架及各计算模块结构进行设计。最后,基于Zynq-7000 SoC FPGA搭建一套完整硬件实验系统进行算法实现,分别使用数据集图像和真实场景图像对本文算法进行评估。实验结果表明,本文算法与传统双目半全局匹配算法相比,有效像素填充率提高17.31%,错误率降低13.06%,在真实场景下可实现60 fps实时立体匹配,能够满足各类应用场景的实际需求。

关键词:半全局匹配算法;三目立体视觉;硬件实时计算;现场可编程逻辑门阵列

中图分类号: TH741 TN47 **文献标识码:** A **国家标准学科分类代码:** 510.40

Triple semi-global matching algorithm and implementation on FPGA

Liu Yang, Feng Jinliang, Huang Wei, Kang Shixun, Ma Yunqi

(College of Opto-Electronic Engineering, Changchun University of Science and Technology, Changchun 241000, China)

Abstract: The trinocular stereo vision system could overcome the occlusion problem of the binocular stereo vision system. The measurement accuracy of the stereo vision system can be further improved. However, the large number of sensors may lead to the high computation load of the matching algorithm and affect the real-time performance of the system. Hence, the practical application of trinocular stereo vision systems is limited in various fields. In this article, a trinocular semi-global stereo matching algorithm and its hardware computational framework are proposed. Firstly, based on the in-depth analysis of the basic model of the trinocular stereo vision system, a hardware-friendly semi-global matching algorithm is proposed. Then, the overall framework of the on-chip system and the structure of each computational module are designed according to the parallelization and pipeline processing characteristics of field programmable gate array (FPGA) hardware computation. Finally, a complete hardware experiment system is established, which is based on Zynq-7000 SoC FPGA for algorithm implementation. The algorithm is evaluated by using dataset images and real scene images, respectively. Compared with the traditional binocular semi-global matching algorithm, experimental results show that the effective pixel filling rate of this algorithm is increased by 17.31%, the error rate is reduced by 13.06%, and the real-time stereo matching can be achieved at 60 fps in real scenes, which could meet the practical needs of various application scenarios.

Keywords: semi-global matching algorithm; trinocular stereo vision; hardware real-time computing; field-programmable logic gate array

0 引言

立体视觉系统因其相比单目视觉系统能够获取更多的三维信息而在各领域得到广泛应用,例如,三维重建^[1-2]、无人驾驶^[3-4]、机器视觉^[5]等。随着高分辨率图像

传感器应用成本不断降低,各领域对立体视觉系统的测量精度要求也越来越高,现有提高系统测量精度的主要途径为增加基线距离和传感器的种类与数量。然而,由于实际应用场景空间大小以及相机视场限制,基线距离能够增加的范围有限,因此,增加传感器的种类和数量成为提高立体视觉系统测量精度的主流选择^[6-8]。三目视

视觉系统能够克服在双目视觉系统中由于遮挡等因素导致的误匹配问题,因为双目系统受遮挡部分在三目系统的第 3 个相机中很可能是可见的,并且传感器数量增加使视觉系统能够获取更多目标信息从而增强系统测量精度和鲁棒性。三目立体匹配是三目视觉系统应用中的关键技术,通过三目立体匹配能够计算系统中 3 个相机获取图像的视差图,深度视差图记录了场景中物体与相机的实际距离,从而用于后续的三维重建与视觉测量。

但是,匹配过程计算量和结构复杂度增加也会影响视觉处理系统的实时性能。因此,提出一种适用于三目立体视觉系统的高精度立体匹配算法及其硬件实现,对三目立体视觉系统在各领域实际应用具有重要意义。

立体匹配主要可分为基于区域匹配和基于全局匹配。其中,全局匹配相较于区域匹配能够将整幅图片作为一个整体来考虑,生成的视差具有较好连续性,但全局匹配过程运算复杂且计算量大。针对这一问题,Hirschmüller^[9]提出了一种半全局立体匹配(semi-global matching, SGM)方法,将像素互信息作为初始代价,然后从 16 个方向逐渐逼近优化,这种方法在降低计算量的同时仍保持了全局连续性和匹配精度。

随着图像传感器数量和分辨率不断提高,匹配过程计算量也逐渐增加,通用计算机的中央处理器(central processing unit, CPU)顺序指令处理方式已不能完全满足系统的实时性需求,而图像处理器(graphics processing unit, GPU)在一些功耗敏感的应用场景下并不适用。现场可编程逻辑门阵列(field programmable gate array, FPGA)相比 CPU 和 GPU,具有低功耗、高能耗比、并行化处理和可重配置的优点,其并行化与流水线处理方式能够降低由于图像计算量增加给视觉处理系统实时性带来的影响^[10]。于是,有研究者借助 FPGA 并行处理结构提高半全局匹配处理实时性。例如,Honegger 等^[11]提出了一种 FPGA 与移动平台 CPU 组合异构计算平台进行立体匹配的方法,其中 FPGA 并行化处理匹配过程中复杂运算,低功耗移动平台 CPU 作为 FPGA 协处理器。Zhang 等^[12]针对 FPGA 匹配计算中存在吞吐量限制瓶颈的技术问题,通过优化存储架构、压缩片内存储开销和片外存储传输带宽,进一步提高了 FPGA 的硬件处理效率。Zhao 等^[13]利用高层次综合工具对基于 FPGA 的半全局立体匹配算法进行设计,最终实现了整套系统。

与此同时,研究者们试图在半全局匹配算法基础上进行各种改进,进一步提高算法准确度。Cambuim 等^[14]通过一种在图像预处理中增加梯度滤波和采样敏感的绝对差分方法,增强匹配计算过程鲁棒性并基于 FPGA 实现了整套系统。Facciolo 等^[15]提出了一种 MGM(more global matching)立体匹配算法,通过各个方向逼近路径中增加递归处理过程,克服了半全局匹配产生条纹伪影

的问题,随后,Rahnama 等^[16]对 MGM 算法进行了 FPGA 硬件实现。Schönberger 等^[17]提出了 SGMRest(semi-global matching rorest)算法,通过基于学习的方法对逐个像素进行分类,进一步提高了半全局匹配算法的精度。

上述半全局立体匹配算法大多数是基于双目视觉系统进行设计优化。随着视觉测量精度要求的提高以及传感器成本的降低,视觉测量系统也不仅仅局限于双目系统,基于三目立体视觉系统的匹配算法也需要进一步研究。Kallwies 等^[18]提出了一种基于半全局匹配的三目视觉立体匹配算法,通过视差图融合将三目视觉立体匹配分为水平匹配和垂直匹配两个过程,显著提高了立体视觉系统鲁棒性。随后,Kallwies 等^[19]又对前期工作进行了进一步优化,提出了一种 Triple-SGM(triple semi-global matching)算法,将三目视觉系统的半全局匹配视差融合过程提前到半全局优化之前,通过避免重复计算,进一步提升了算法性能。

相对于基于 FPGA 的双目立体视觉系统,目前关于三目立体匹配算法 FPGA 实现研究较少。Motten 等^[20]提出了一种基于层次分类器的 FPGA 三目立体匹配计算架构,通过图像降采样在不需外部拓展存储前提下实现了三目立体匹配计算。Akin 等^[21]提出了一种基于 FPGA 的三目立体匹配算法,通过自适应窗口实现三目图像区域匹配,最终在 FPGA 上实现了整套系统。Jia 等^[22]提出了一种自主导航三目视觉机器人,基于 FPGA+DSP 架构实现了三目立体视觉系统实时视差计算。

上述三目立体匹配算法与 FPGA 实现均是基于区域匹配算法实现的,基于半全局的立体匹配相比区域匹配具有更好的全局视差连续性,能够将整幅图像引入视差计算的过程中,从而获取更高精度的结果。如何提高三目立体视觉系统半全局匹配计算的实时性是三目系统在推广实际应用中亟待解决的关键问题。本文在对半全局匹配算法进行深入研究的基础上,提出一种硬件友好半全局三目匹配算法并进行 FPGA 实现。

1 三目半全局匹配算法

立体匹配是三目视觉系统处理过程中的关键步骤,匹配质量直接关系到视觉系统的测量精度。针对三目视觉系统立体匹配问题,本文首先对三目视觉系统基本模型进行分析,在双目半全局匹配算法的基础上,提出一种硬件友好的三目半全局立体匹配算法,实现三目视觉系统立体匹配。

1.1 三目视觉系统基本模型

本文所采用的视觉系统包括 3 个内参数相同的图像传感器,传感器的中心安装在同平面上并且对角线与外极线夹角为 45°,该系统获取的图像平行于 3 个相机所处

平面上,并通过文献[23]提出的极线矫正方法对三目立体视觉系统进行矫正,其基本模型如图1所示。

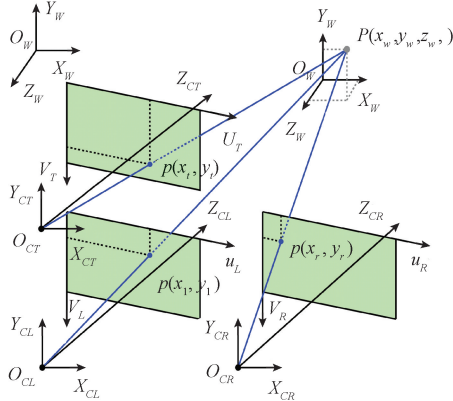


图1 三目视觉系统基本模型

Fig. 1 Model of the trinocular vision system

1.2 三目半全局匹配算法基本原理

图2为本文三目半全局匹配算法框架。算法主要包括代价初始化、代价融合、代价聚合和半全局优化4个部分。本文算法首先将上图像、左图像和右图像进行Census变换;随后,经代价初始化、代价融合、代价聚合,再进行半全局优化并通过赢者通吃(winner-takes-all, WTA)获取半全局优化过程最优视差;最后,进行后处理得到最终三目匹配视差结果。

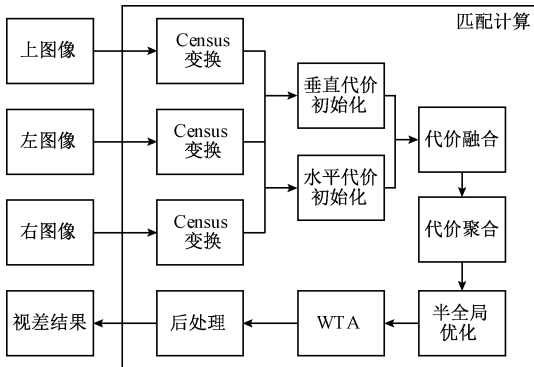


图2 三目半全局匹配算法框架

Fig. 2 Framework of the trinocular semi-global matching algorithm

Census变换能够将匹配过程中所使用像素的灰度绝对值转换为其邻域内的相对值,并且作为一种硬件友好变换,在增强算法鲁棒性的同时也能够降低数据量从而节省算法硬件开销^[24]。对于灰度值为 $I(p)$ 的像素 p 与其周围 7×7 邻域 $N(p)$ 内灰度为 $I(p')$ 的点 p' ,Census变换结果 $\xi(p,p')$ 为:

$$\xi(p,p') = \begin{cases} 1, & I(p') < I(p) \\ 0, & \text{其他} \end{cases} \quad (1)$$

于是可以得到邻域 $N(p)$ 内编码生成序列 $C(p)$:

$$C(p) = \bigotimes_{p' \in N(p)} \xi(p,p') \quad (2)$$

代价初始化涉及左图像与右图像、上图像与左图像两个匹配步骤,分别生成水平初始代价和垂直初始代价。其中,设定视差搜索范围为 $(0, d_{max})$,代价初始化过程中水平视差 d_h 与垂直视差 d_v 在此范围内进行搜索。在代价初始化与聚合过程中,初始像素点 p 向上下左右4个方向进行延伸建立水平支持臂 $HARM(p)$ 和垂直支持臂 $VARM(p)$,当超过最大视差搜索范围 d_{max} 或当前像素灰度值与像素 p 差异过大后停止。再将垂直支持臂上所有的点的水平支持臂组合起来作为像素 p 在代价初始化过程中的邻域 $W(p)$ 。

在生成水平初始代价过程中,通过计算在不同水平视差 d_h 下,像素 p 及其邻域 $W(p)$ 编码生成序列 $C_1(p)$ 与 $C_2(p + d_h)$ 之间的汉明距离,当水平方向左右图像像素及其邻域的相似度最高时,其汉明距离处于最小值^[25],将该最小值作为像素 p 的水平初始代价 $C_h(p, d_h)$:

$$C_h(p, d_h) = \min_{d_h \in (0, d_{max})} \left\{ \sum_{p' \in W(p)} H(C_1(p'), C_2(p' + d_h)) \right\} \quad (3)$$

同样地,可以计算得到垂直视差为 d_v 的初始代价 $C_v(p, d_v)$:

$$C_v(p, d_v) = \min_{d_v \in (0, d_{max})} \left\{ \sum_{p' \in W(p)} H(C_1(p'), C_2(p' + d_v)) \right\} \quad (4)$$

在获得水平方向和垂直方向视图的初始代价后,需要通过代价融合最终获得一幅完整的初始代价结果。代价融合过程应当尽可能早,这样能够在后续过程中发生重复计算,节约计算资源^[19]。本文在代价聚合之前进行代价融合,所采用的代价融合方法可以表示为:

$$C_i(p, d) = 0.5 \cdot C_h(p, d_h) + 0.5 \cdot C_v(p, f_b \cdot d_h) \quad (5)$$

其中, f_b 为针对将垂直方向视差变换到水平方向上进行统一融合处理中,水平基线距离和垂直基线距离可能存在不同而设置的参数:

$$f_b = \frac{b_v}{b_h} \quad (6)$$

式中: b_v 为垂直基线距离; b_h 为水平基线距离。本文涉及的三目视觉系统水平与垂直方向等基线,故 $b_v = b_h$, $f_b = 1$,于是有 $d = d_h = d_v$ 。

初始代价是在支持域内通过代价聚合得出的^[23]。对于每一个像素 p ,先进行水平聚合获得水平代价聚合结果 $C_{aggh}(p, d)$,再对水平聚合的结果进行垂直聚合,获得聚合结果 $C_{agg}(p, d)$,其计算过程如下:

$$C_{aggh}(p, d) = \sum_{q \in HARM(p)} C_i(q, d) \quad (7)$$

$$C_{agg}(p, d) = \sum_{q \in VARM(p)} C_{agg}(q, d) \quad (8)$$

式中: $HARM(p)$ 和 $VARM(p)$ 分别为像素 p 的水平支持臂和垂直支持臂。

半全局优化过程是沿着不同方向分别优化聚合代价^[9],对于方向 r , 其优化过程如下:

$$L_r(p, d) = C_{agg}(p, d) + \min \left\{ \begin{array}{l} L_r(p-r, d), \\ L_r(p-r, d \pm 1) + P_1, \\ \min_k L_r(p-r, k) + P_2 \end{array} \right\} - \min_k L_r(p-r, k) \quad (9)$$

式中: $L_r(p, d)$ 代表方向 r 上的路径代价; P_1, P_2 为函数不连续性的惩罚系数。

最终所有路径代价 $C_{final}(p, d)$ 为:

$$C_{final}(p, d) = \sum_r L_r(p, d) \quad (10)$$

随后通过 WTA 选取最优视差:

$$d_{res}(p) = \underset{0 \leq d \leq d_{max}}{\operatorname{argmin}} C_{final}(p, d) \quad (11)$$

在选取最优视差后,还需要根据实际情况进行后处理,具体包括间隙插值、孔洞填充和中值滤波等。

在本文算法设计过程中,充分考虑到算法的硬件可实现性,通过 Census 变换增强算法鲁棒性的同时,像素数据量从邻域内 n 个像素的 $n \times 8 \text{ bit}$ 缩减到 $n \text{ bit}$,节约了硬件开销。代价初始化函数选择计算汉明距离,仅涉及到硬件友好的异或和加法运算,没有复杂乘除运算。在半全局优化过程中,考虑到整幅图像的像素刷新方向问题,能够对优化方向进行缩减,保证了算法的实时性。

2 FPGA 算法实现

本文在三目半全局匹配算法设计的基础上,对片上系统进行整体设计规划,根据 FPGA 并行计算与流水线处理的特点进行算法硬件计算结构设计,最终完成整套算法的硬件实现。

2.1 整体架构设计

如图 3 所示,本文所采用的 Zynq-7000 片上系统包括可编程逻辑(programmable logic)和处理系统(processing system)两部分,可编程逻辑部分为 Artix-7 系列 FPGA,主要执行图像输入、图像输入信号转换和传输、匹配计算和图像输出显示。处理系统中两个 Cortex-A9 内核以 AMP 非对称多处理方式工作并在外部挂载了 1 GB 共享 RAM, CPU #0 作为主机运行操作系统,将深度信息生成点云数据, CPU #1 被定义为从机,作为可编程逻辑协处理器进行输入图像矫正。

可编程逻辑与处理系统之间通过 AXI 总线互联。上图像、左图像、右图像信号输入后分别被存入缓存, Video In to AXI-4 IP 核将输入的图像信号转换为 AXI-4 总线信

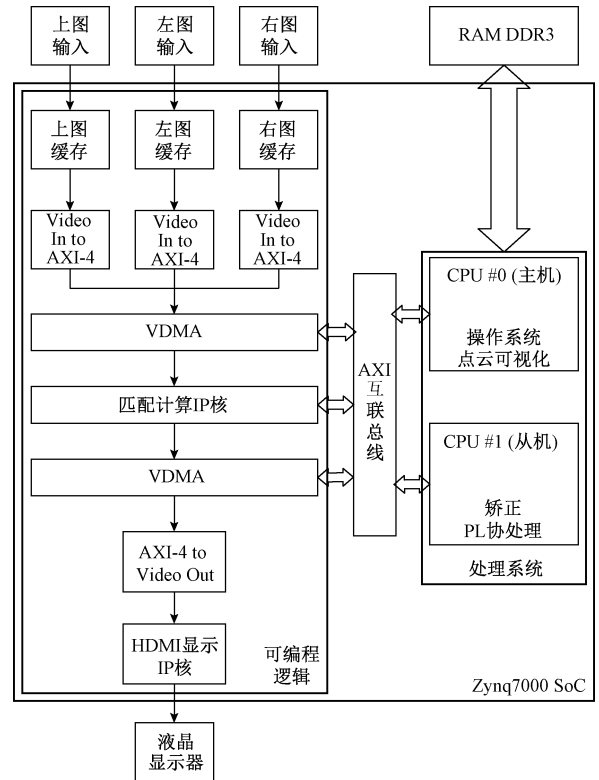


图 3 片上系统总体架构

Fig. 3 System-on-chip architecture

号并传入 VDMA,如果输入是未经矫正的图像,VDMA 将输入图像信号传输给协处理器 CPU #1 进行矫正,矫正后图像通过 VDMA 输入匹配计算 IP 核内进行匹配计算,计算结果由 AXI-4 to Video Out IP 核转换成 RGB 信号并通过 HDMI 接口输出在显示设备上。

2.2 代价初始化与聚合

输入的图像数据经过缓存后被读取,在代价初始化过程中通过计算分别获得左图像与上图像在垂直方向初始化代价 $C_v(p, d_v)$ 和左右图像水平方向初始化代价 $C_h(p, d_h)$,并通过代价融合获取融合后的代价 $C_r(p, d)$ 。在代价初始化与聚合过程中,本文将全部视差等级 N_d 分为 N_d/P_d 组,每次并行计算 P_d 级视差,为了进一步提升数据传输效率,像素数据在计算当前视差的代价后并不会被立即释放,经过寄存器延时后直接用于下一视差等级的代价计算,其计算架构如图 4 所示。

同样地,在图 5 中,代价聚合计算中 P_d 个模块并行执行 k 轮计算,先进行横向聚合再进行纵向聚合,在进行聚合的同时统计聚合像素数 $Count_p$,最终对计算结果进行归一化处理。

2.3 半全局优化设计

在文献[9]提出的半全局匹配算法中,优化过程包括对 16 个方向进行优化,但是部分优化方向与图像像素

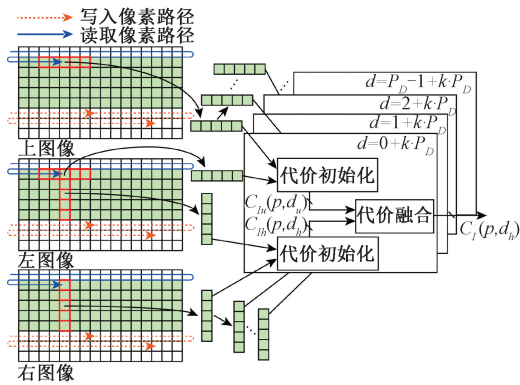


图4 匹配代价初始化与代价融合
Fig. 4 Cost initialization and cost fusion

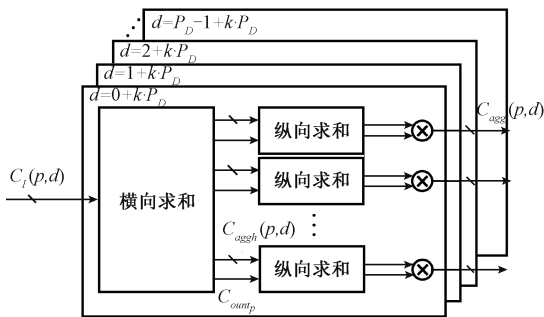


图5 匹配代价聚合
Fig. 5 Cost aggregation

刷新方向相反,如果进行16个方向的半全局优化需要图像一帧完全加载后才能进行,这样不仅需要更多硬件资源进行帧缓冲,而且会影响半全局匹配实时性。为了提高视觉系统实时性和优化资源消耗,本文将三目半全局匹配过程中的16个方向化简为如图6所示的4个方向 r_1, r_2, r_3, r_4 ,这种方向化简能在极少精度损失下提高系统匹配实时性^[26]。

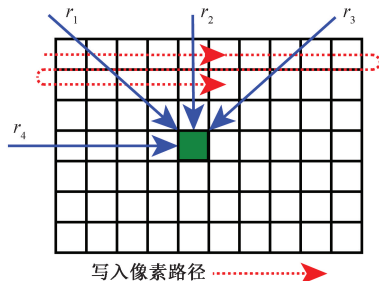


图6 半全局优化方向
Fig. 6 Direction of semi-global optimization

如图7,半全局优化阶段设计4个模块,从 r_1, r_2, r_3, r_4 4个方向分别优化,输入的 $L_r(p-r, d)$ 被延时单元寄存

并通过比较器得出 $L_r(p-r, d)$ 最小值,经比较得出的最小值与寄存值相减,并与 $C_{agg}(p, d)$ 相加后得出 $L_r(p, d)$ 结果值。

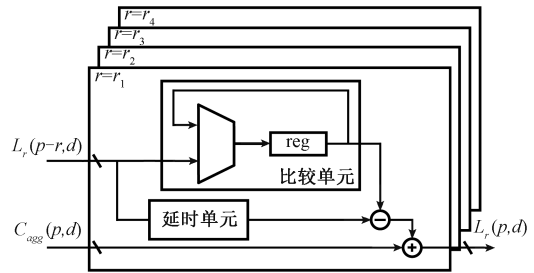


图7 半全局优化计算
Fig. 7 Semi-global optimization compute

3 实验结果与分析

为了测试整套算法的实际性能,本文搭建如图8所示硬件系统进行实验。所采用实验板载有Xilinx Zynq-7000系列xc7z020clg400-2芯片,其片上资源配置如表1所示。硬件系统外置存储卡和图像采集模块分别输入数据集图像和采集真实场景图像,其中图像采集模块基于3个OV5640传感器,安装内参数相同的镜头按垂直和水平方向等基线排布,最终计算结果通过HDMI接口输出显示到LCD显示器上。

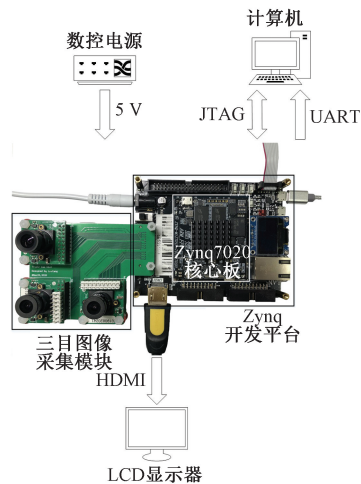


图8 实验系统硬件组成
Fig. 8 Hardware of the experimental system

本实验测试图像数据分别来源于数据集和真实场景,数据集图像为文献[19]所提供的计算机仿真模拟驾驶环境道路三目视觉实验集图像,真实场景图像为OV5640传感器所采集的图像,能够为算法准确性和鲁棒性测试提供充分依据。

表 1 片上系统主要资源

Table 1 System-on-chip resources

参数	数值
型号	xc7z020clg400-2
处理系统核心	双核 Cortex-A9
逻辑单元/K	85
查找表	53 200
BRAM	140
DSP Slices	220

3.1 数据集实验及结果分析

本实验从数据集图像中选取了 200 张驾驶场景下的三目视觉图像用于算法测试评估。图 9(a) 为测试使用数据集图像中截取的一帧图像,图 9(b)~(d) 分别为垂直匹配和水平匹配后生成的视差结果以及本文算法的视差结果。经比较可以看出,垂直和水平方向生成视差图存在大量无法匹配的边缘像素,而在本文算法所产生的融合视差结果图 9(d) 中,边缘全部是能够匹配的,并且图像中有效像素的数量也有明显提高。图 10 为图像序列中随机选取的 4 个不同场景下进行彩色化处理的实验结果。

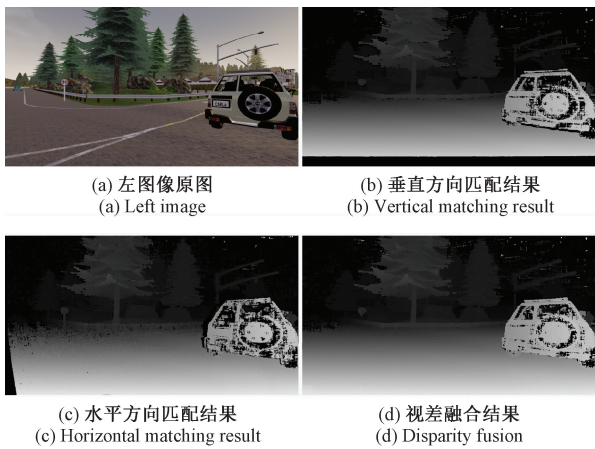


图 9 不同匹配方法的数据集实验

Fig. 9 Experiments on datasets with different methods

本文通过匹配计算结果与计算机建模生成真实视差相比较,将计算值与真实值误差大于 3 视差的结果作为错误视差,将错误视差占有所有像素的比率作为该视差图像的错误率。特别地,由于三目视觉系统能够克服双目视觉系统生成视差图像存在遮挡和无法匹配的边缘这一缺陷,本实验在对算法准确度进行评估的同时,还对生成视差图像有效像素填充率进行了比较和分析。在视差匹配过程中,一些视差结果与相邻值不连续的像素作为异常值被滤除,并且由于遮挡、光线、噪声等原因无法匹配,

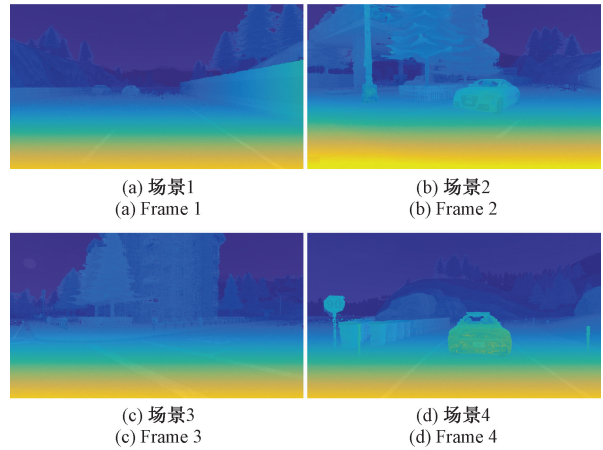


图 10 不同场景下的数据集实验

Fig. 10 Experiments with datasets in different scenarios

造成部分像素无有效视差值,将所有有效视差像素与图像总像素的比率作为有效像素填充率也是本文的一个重要指标。

如图 11 所示,由于本实验所使用的数据集图像水平分辨率大于垂直分辨率,在垂直匹配过程中生成的深度图像边缘会存在更多无法匹配的像素信息,所以水平匹配结果平均填充率高于垂直匹配结果 10.63%。通过本文融合方法获取的深度结果填充率平均为 90.40%,相比水平匹配和垂直匹配分别提高了 6.68% 和 17.31%;而在算法准确性方面,本文算法平均错误率为 85.99%,相比水平和垂直匹配的平均错误率分别降低了 10.46% 和 13.06%,数据集实验错误率结果如图 12 所示。

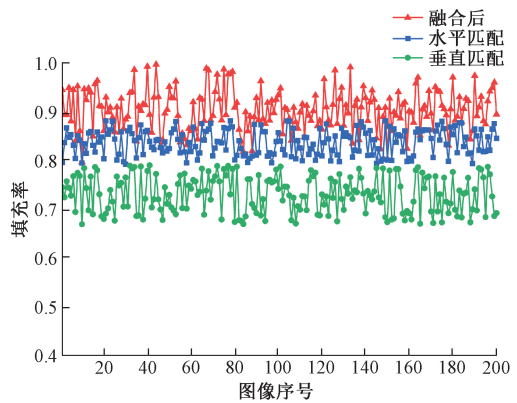


图 11 数据集实验填充率对比

Fig. 11 Comparison of experimental fill rates of data sets

3.2 真实场景实验及结果分析

数据集能够为算法准确性进行客观评估,但在一些实际场景中,可能存在噪声光照等因素影响,所以数据集图像并不能完整地评价算法在实际使用场景中的实际性能。本文通过图像传感器获取真实场景下不同分辨率图

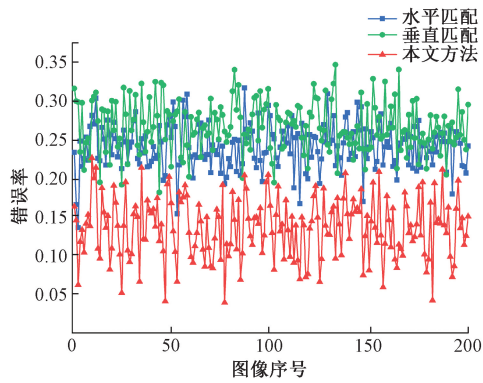


图12 数据集实验错误率对比

Fig. 12 Comparison of experimental error rates of data sets

像进行实时处理,对算法在真实场景下的计算速度、资源利用情况进行比较分析,如表2所示。

表2 不同分辨率的资源消耗

Table 2 Resource consumption at different resolutions

分辨率	帧率 /fps	视差等级	LUT 数量	BRAM 数量	DSP Slices 数量
640×480	60	64	30 270/ (56.89%)	54/ (38.57%)	37/ (16.81%)
		96	36 811/ (69.19%)	71/ (50.71%)	48/ (21.81%)
1 024×768	30	64	38 125/ (71.66%)	87.5/ (62.50%)	56/ (25.45%)
		96	44 115/ (82.92%)	126/ (90.00%)	59/ (26.81%)
1 920×1 080	15	64	41 652/ (78.29%)	98/ (70.00%)	63/ (28.63%)
		96	46 510/ (87.42%)	136/ (97.14%)	69/ (31.63%)

图13为本文实验系统对图像传感器实时采集分辨率为640×480真实场景图像以60 fps实时深度恢复结果,视差等级设定为96,各模块资源利用情况如表3所示。实验结果表明,在真实场景下,本文算法能够实现深度实时恢复并取得良好的效果。

本文与现有技术的三目立体匹配研究工作进行了对比,从研究方法的处理平台、分辨率、视差等级、帧率以及MDE/s (million disparity estimations per second, $W \times H \times D \times f$)进行了对比,其结果如表4所示。结果表明,本文算法及FPGA计算架构相比CPU平台和GPU平台具有更高的实时性能,与同类基于FPGA平台的研究工作相比也具有更高的性能指标,说明本文提出的算法



图13 真实场景实验

Fig. 13 Real-life scenario experiments

表3 各模块资源利用

Table 3 Resource utilization of each module

模块	LUT 消耗数量	BRAM 消耗数量	DSP Slices 消耗数量
图像采集	115/(2.16%)	0/(0%)	0/(0%)
图像传输	5 110/(9.60%)	21.5/(15.35%)	0/(0%)
匹配计算	30 186/(56.74%)	49.5/(35.35%)	48/(21.81%)
显示	165/(3.10%)	0/(0%)	0/(0%)
其他	1 235/(2.32%)	0/(0%)	0/(0%)

表4 实时性能比较

Table 4 Real-time comparison of the art-of state systems

方法	处理平台	分辨率	视差等级	帧率 /fps	MDE/s ² (10 ⁶)
EC-SV ^[18]	CPU	1 164×335	128	28.5	1 422
Triple-SGM ^[19]	GPU	926×276	128	40	1 308
MSVM-III ^[22]	FPGA+DSP	320×240	64	8	393
T-AWDE ^[21]	Virtex-7 FPGA	1 024×768	49	55	2 119
本文架构	Zynq FPGA	640×480	96	60	1 769
		1 920×1 080	96	15	2 985

及硬件计算架构加速效果明显,能够满足各种实时视觉测量的要求。

4 结 论

为了提高三目立体视觉系统匹配计算的实时性能,本文提出了一种适用于三目视觉系统的硬件友好半全局匹配算法,并基于Zynq-7000 SoC FPGA对整套算法进行

了硬件实现。本文算法借助 FPGA 并行计算和流水线处理,实现了三目视觉图像实时高精度立体匹配,实验结果表明本文算法的实时性能和计算精度能够满足各类应用场景的实际需求,拓展了三目立体视觉技术的应用范围,对三目立体视觉技术在各领域实际应用具有重要意义。在下一步工作中,考虑如何将前后帧的图像引入立体匹配过程中,借助帧间图像相关性进一步提高算法准确度和鲁棒性。

参考文献

- [1] 谢亮亮,屠大维,张旭,等. 深海原位激光扫描双目立体视觉成像系统[J]. 仪器仪表学报, 2020, 41(6): 106-114.
- XIE L L, TU D W, ZHANG X, et al. Deep sea in-situ binocular stereo vision imaging system with laser scanning [J]. Chinese Journal of Scientific Instrument, 2020, 41(6): 106-114.
- [2] HE Z, JIANG Z, ZHAO X, et al. Sparse template-based 6-D pose estimation of metal parts using a monocular camera [J]. IEEE Transactions on Industrial Electronics, 2020, 67(1): 390-401.
- [3] 张合新,徐慧,姚二亮,等. 动态场景下一种鲁棒的立体视觉里程计算法[J]. 仪器仪表学报, 2018, 39(9): 246-254.
- ZHANG H X, XU H, YAO ER L, et al. Robust stereo visual odometry algorithm in dynamic scenes [J]. Chinese Journal of Scientific Instrument, 2018, 39(9): 246-254.
- [4] DEHNAVI M, ESHGHI M. FPGA based real-time on-road stereo vision system [J]. Journal of Systems Architecture, 2017, 81: 32-43.
- [5] 涂俊超,张丽艳. 视觉辅助下的激光振镜模型参数分步标定[J]. 仪器仪表学报, 2019, 40(1): 236-247.
- XU J CH, ZHANG L Y. Two-step parameter calibration for galvanometric laser scanners using binocular stereo vision [J]. Chinese Journal of Scientific Instrument, 2019, 40(1): 236-247.
- [6] HINZMANN T, CADENA C, NIETO J, et al. Flexible trinocular; Non-rigid multi-camera-IMU dense reconstruction for UAV navigation and mapping [C]. IEEE International Conference on Intelligent Robots and Systems, 2019: 1137-1142.
- [7] SHIBA Y, ONO S, FURUKAWA R, et al. Learning-based active 3D measurement technique using light field created by video projectors [J]. IPSJ Transactions on Computer Vision and Applications, 2019, 11(1): 1-14.
- [8] MARTÍNEZ H G, GIORGI G, EISSFELLER B. Pose estimation and tracking of non-cooperative rocket bodies using Time-of-Flight cameras [J]. Acta Astronautica, 2017, 139(May): 165-175.
- [9] HIRSCHMÜLLER H. Stereo processing by semiglobal matching and mutual information [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 328-341.
- [10] LENTARIS G, MARAGOS K, STRATAKOS I, et al. High-performance embedded computing in space: Evaluation of platforms for vision-based navigation [J]. Journal of Aerospace Information Systems, 2018, 15(4): 178-192.
- [11] HONEGGER D, OLEYNIKOVA H, POLLEFEYS M. Real-time and low latency embedded computer vision hardware based on a combination of FPGA and mobile CPU [C]. IEEE International Conference on Intelligent Robots and Systems, 2014: 4930-4935.
- [12] ZHANG X, DAI H, SUN H, et al. Algorithm and VLSI architecture co-design on efficient semi-global stereo matching [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30 (11): 4390-4403.
- [13] ZHAO J, LIANG T, FENG L, et al. FP-stereo: Hardware-efficient stereo vision for embedded applications [C]. 30th International Conference on Field-Programmable Logic and Applications (FPL), 2020: 269-276.
- [14] CAMBUIM L F S, OLIVEIRA L A, BARROS E N S, et al. An FPGA-based real-time occlusion robust stereo vision system using semi-global matching [J]. Journal of Real-Time Image Processing, 2019, 17: 1447-1468.
- [15] FACCILOLO G, FRANCHIS C, MEINHARDT E. MGM: A significantly more global matching for stereo vision [C]. BMVC 2015, 2015.
- [16] RAHNAMA O, CAVALLERI T, GOLODETZ S, et al.

- R3SGM: Real-time raster-respecting semi-global matching for power-constrained systems [C]. 2018 International Conference on Field-Programmable Technology, 2018: 105-112.
- [17] SCHÖNBERGER J L, SINHA S N, POLLEFEYS M. Learning to fuse proposals from multiple scanline optimizations in semi-global matching [J]. Lecture Notes in Computer Science, 2018: 758-775.
- [18] KALLWIES J, WUENSCH H J. Effective combination of vertical and horizontal stereo vision [C]. 2018 IEEE Winter Conference on Applications of Computer Vision, 2018: 1992-2000.
- [19] KALLWIES J. Triple-SGM : Stereo processing using semi-global matching with cost fusion [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020: 192-200.
- [20] MOTTEN A, CLEASEN L, PAN Y. Trinocular disparity processor using a hierarchic classification structure [C]. IEEE/IFIP VLSI-SoC, 2015: 247-250.
- [21] AKIN A, CAPOCCIA R. Trinocular adaptive window size disparity estimation algorithm and its real-time hardware [C]. 2015 VLSI-DAT, 2015: 4-7.
- [22] JIA Y, LI M, AN L, et al. Autonomous navigation of a miniature mobile robot using real-time trinocular stereo machine [J]. Rissp, 2003: 417-421.
- [23] 李秀智,张广军. 三目视觉图像的极线校正方法[J]. 光电工程, 2007(10): 50-54.
LI X ZH, ZHANG G J. Epipolar rectification in trinocular vision images [J]. Opto-Electronic Engineering, 2007(10): 50-54.
- [24] ZABIH R, WOODFILL J. Non-parametric local transforms for computing visual correspondence [J]. Lecture Notes in Computer Science, 1994: 151-158.
- [25] MEI X, SUN X, ZHOU M, et al. On building an accurate stereo matching system on graphics hardware [C]. 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), IEEE, 2011: 467-474.
- [26] WANG W, YAN J, XU N, et al. Real-time high-quality stereo vision system in FPGA[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(10): 1696-1708.

作者简介



刘阳, 2018年于安徽师范大学获得学士学位, 现为长春理工大学硕士研究生, 主要研究方向为图像处理与计算机视觉。

E-mail: liuyang@mails.cust.edu.cn

Liu Yang received his B. Sc. degree from Anhui Normal University in 2018. He is currently a master student at Changchun University of Science and Technology. His main research interests include image processing and computer vision.



冯进良 (通信作者), 1987年于长春光学精密机械学院获得学士学位, 1999年于长春理工大学获得硕士学位, 2004年于中国科学院长春光学精密机械与物理研究所获得博士学位, 现为长春理工大学教授, 主要研究方向为仪器总体设计与仿真。

E-mail: jinliang.feng@cust.edu.cn

Feng Jinliang (Corresponding author) received his B. Sc. degree from Changchun Institute of Optical Precision Machinery in 1987, received his M. Sc. degree from Changchun University of Science and Technology in 1999, and received his Ph. D. degree from Changchun Institute of Optics and Precision Mechanics and Physics, Chinese Academy of Sciences in 2004. He is currently a professor at Changchun University of Science and Technology. His main research interest is instrument design and simulation.