

基于 FPGA 的双目立体视觉系统

周文晖 杜 歆 叶秀清 顾伟康

(浙江大学信息与电子工程学系, 杭州 310027)

摘 要 立体视觉的目的之一就是为获得周围场景的 3 维信息,其关键在于匹配算法。然而即便是使用目前先进的通用处理器,其计算致密视差图所需的时间仍无法满足高速自主导航的需求。为了解决这个问题,提出了一种基于现场可编程门阵列(FPGA)的双目立体视觉系统的设计方案,同时介绍了系统的硬件结构,并在讨论区域匹配的快速算法的基础上,提出了基于 FPGA 的像素序列和并行窗口算法框架,用以实现零均值像素灰度差平方和(ZSSD)的匹配算法。该算法是先将视频信号经解码芯片生成场景立体图像对,并由 FPGA 来完成立体图像对的几何校正和 ZSSD 匹配算法,然后将获得的致密视差图通过 PCI 总线发送至上位机。实践表明,该算法效果好、速度快,不仅具有较强的鲁棒性,并且硬件系统性能稳定、可靠。此外,该方案还适用于像素灰度差的绝对值和(SAD)和像素灰度差的平方和(SSD)等多种传统区域匹配算法的快速实现和实时处理。

关键词 双目立体视觉 快速匹配算法 零均值像素灰度差平方和 现场可编程门阵列

中图分类号: TP301.6 **文献标识码**: A **文章编号**: 1006-8961(2005)09-1166-05

Binocular Stereo Vision System Based on FPGA

ZHOU Wen-hui, DU Xin, YE Xiu-qing, GU Wei-kang

(Department of Information and Electronic Engineering, Zhejiang University, Hangzhou 310027)

Abstract The goal of stereo vision is to recover the three-dimensional information of a scene, and the core of stereo vision is to find the corresponding pixels. However, the correspondence search is too time-consuming for high speed robot vehicles autonomous navigation, even if state-of-art general-purpose processors are used to accelerate it. Aiming at this problem, this paper presents a design scheme for binocular stereo vision system based on field programmable gate arrays(FPGA), its hardware structure is introduced, and a fast zero mean sum of squared differences(ZSSD) stereo matching algorithm is discussed. Moreover, a pixel-serial and window-parallel architecture based on FPGA processor is proposed to achieve ZSSD matching algorithm, which is suitable for parallel processing. The stereopsis is captured by two same decoder chips. After rectification and ZSSD matching implemented by FPGA, dense disparity map is computed and sent to general-purpose computer by PCI bus. The proposed scheme is very robust and exhibits great performance, such as high speed. And the hardware system has high stability and reliability. In addition, this scheme is also applicable to fast and real-time processing of other conventional area-based stereo matching algorithms, such as sum of absolute differences(SAD), sum of squared differences(SSD), etc.

Keywords binocular stereo vision, fast stereo matching, zero mean sum of squared differences(ZSSD), field programmable gate arrays(FPGA)

1 引 言

双目立体视觉系统的重要任务之一就是两个相同摄像机获取的立体图像对中恢复场景 3 维信息,

并利用场景的 3 维信息识别目标、道路、划分障碍区。

立体匹配算法是立体视觉中的关键算法。基于区域的匹配算法由于可直接获得致密的视差图,因而更广泛地应用于快速或实时的立体视觉系统中。区域匹配算法的性能在很大程度上要依赖于相似性测

基金项目: 国家自然科学基金项目(60105003)

收稿日期: 2004-03-31; **改回日期**: 2005-02-21

第一作者简介: 周文晖(1977 ~),男,1999 年获杭州大学学士学位,现为浙江大学通信与信息系统专业博士生。从事图像处理、机器视觉,及信号处理等方面的研究。E-mail: sunshine@hznc.com。

度因子的选取,而最常用的测度因子有像素灰度差的绝对值和 (sum of absolute differences, SAD), 及像素灰度差的平方和 (sum of squared differences, SSD)。Wu 的研究表明,采用零均值归一化交叉相关 (zero mean normalized cross correlation, ZNCC) 和零均值像素灰度差的平方和 (zero mean sum of squared differences, ZSSD) 均可获得更好的结果^[1]。考虑到硬件实现的难易度,本文采用 ZSSD 的匹配方法。

由于匹配算法复杂度很高,即需要处理大量数据,采用优化技巧加速后,算法复杂度虽可减少至 $O(MND)$ ^[2],但由于采用通用计算机或数字处理芯片的串行处理方式仍无法满足高速、实时处理的性能要求,因而需要开发专用的硬件并行处理系统。

2 双目立体视觉算法描述

2.1 定标和图像几何校正

区域匹配算法往往都建立在摄像机模型为线性,且左、右图外极线与图像扫描线平行的基础上,以满足外极线约束的要求。然而,由于在实际匹配时这些假设都是不成立的,因此在匹配运算之前,需要先对摄像机系统进行定标,即先计算两个摄像机的内、外参数,然后才能根据定标得出的原始图像待校正像素 4 个相邻像素点的校正系数和原始图像待校正像素的位置来完成立体图像对的几何校正,以使用于校正镜头的非线性畸变和外极线配准。显然定标的精度将直接影响匹配结果。张正友提出的 2 维模板定标方法^[3],虽简化了定标过程,但依然较为复杂。所幸的是,在被动立体视觉中,由于摄像机系统位置固定不变,因而定标过程可离线完成。这样实时操作中仅需读取定标中得到的校正系数即可完成立体图像对的几何校正(如式(1)所示)。

$$\hat{I}(u, v) = \sum_{i=1}^4 C(u_i, v_i) \cdot I(u_i, v_i) \quad (1)$$

其中, $\hat{I}(u, v)$ 为校正后得到的匹配窗口 W 内的像素值, $I(u_i, v_i)$ ($i=1, \dots, 4$) 为原始图像中相应像素的 4 个相邻像素点的像素值, $C(u_i, v_i)$ ($i=1, \dots, 4$) 是相应的校正系数。 $I(u_i, v_i)$ 和 $C(u_i, v_i)$ 均由定标确定。

2.2 区域匹配算法

区域匹配算法是先以基准图的待匹配点为中心创建一个窗口,然后在对准图中,以相应外极线上的像素点为中心创建同样大小的滑动窗口,并以两者间的相似性作为匹配代价,由此可得到一条匹配代

价曲线,而其极值处即为最佳匹配点。一般 SAD (或 ZSAD) 和 SSD (或 ZSSD) 取最小值的点,而 NCC (或 ZNCC) 则取最大值的点。Wu 的研究表明,采用 ZNCC 和 ZSSD 虽均可以获得较好的结果^[1],但由于 ZNCC 计算每点都需要乘法、除法和开根号运算,不易于硬件实现,因而本文采用 ZSSD 算法。

设右摄像机拍摄的图像为基准图,左摄像机拍摄的图像为对准图, $f_{m,n}$ 为基准图中 (m, n) 点的亮度值, \bar{f} 为基准图上匹配窗口 W 内的像素均值,在对准图 g 上也做相应的定义,则两个窗口的 ZSSD 值可写为

$$ZSSD_{m,n,d} = \sum_{(u,v) \in W} ((f(u,v) - \bar{f}) - (g(u+d,v) - \bar{g}))^2 \quad (2)$$

其中, d 为左摄像机拍摄的图像上窗口在外极线上移动的距离,即视差。 (u, v) 是以待计算像素 (m, n) 为中心的匹配窗口 W 内的像素坐标。将式(2)扩展可得到

$$ZSSD_{m,n,d} = \sum_{(u,v) \in W} f^2(u, v) + \sum_{(u,v) \in W} g^2(u+d, v) - w \cdot (\bar{f})^2 - w \cdot (\bar{g})^2 - w \cdot \bar{f} \cdot \bar{g} + M_{m,n}(d) \quad (3)$$

其中, w 为窗口大小,

$$M_{m,n}(d) = \sum_{(u,v) \in W} f(u, v) \cdot g(u+d, v) \quad (4)$$

由式(3)和式(4)可知,计算每点的 ZSSD 值需获得窗口内像素的均值、平方和及两窗口像素的乘积和。

2.3 Box 滤波

McDonnell 提出的 Box 滤波方法^[4]是一种递归的 2 维均值滤波的快速算法。该算法对每个输出像素只需 4 次运算,且与窗口大小无关。Stefano 将 Box 滤波应用于立体匹配中,由于大大减少了计算量,因而提高了运算速度^[5]。由于计算窗口的均值或像素平方和时可充分利用前(或上)一个像素窗口的结果,因此只需完成两行(或列)的像素运算即可获得结果(如图 1 所示)。

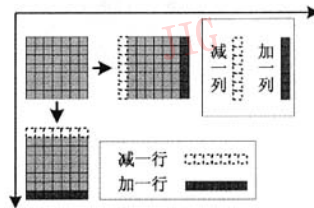


图 1 基于 Box 滤波的立体匹配加速方法

Fig. 1 The acceleration method based on Box filter

3 算法框架及其实现

基于现场可编程门阵列 (field programmable gate arrays, FPGA) 的立体视觉算法设计的关键在于,在片内 RAM 带宽和容量制约的前提下,充分利用 FPGA 的并行性计算体系结构及合理的流水线设计来实现并行算法。此外,匹配窗口大小的选取也受片内 RAM 带宽和容量制约。在 Box 滤波算法中,不仅需要暂存上一个像素的计算结果,还需暂存其匹配窗口中每行的计算中间值,而且,窗口的大小也直接影响着存放中间结果所需的位数。为减轻片内

RAM 的制约,可以将大部分暂时不用的中间结果放到外部 RAM 中,在参与运算前读入即可。

根据式(3),在 ZSSD 算法中,匹配窗口的像素均值、平方和及两窗口间像素乘积和可并行计算。本文以两窗口间像素乘积和 $M_{m,n}(d)$ 为例介绍该算法,其算法框图如图 2 所示,设窗口半径为 r ,在计算第 (m,n) 点的视差为 d 的 $M_{m,n}(d)$ 的值时,可利用第 $(m,n-1)$ 点视差为 d 的结果,减去第 p 行的乘积和 $sum(p,d)$,再加上第 q 行的乘积和 $sum(q,d)$ 来获得。在实际运算中,图像像素是顺序送入 FPGA,为了便于操作,图像读写地址均采用高位为列,低位为行的地址形成方式。

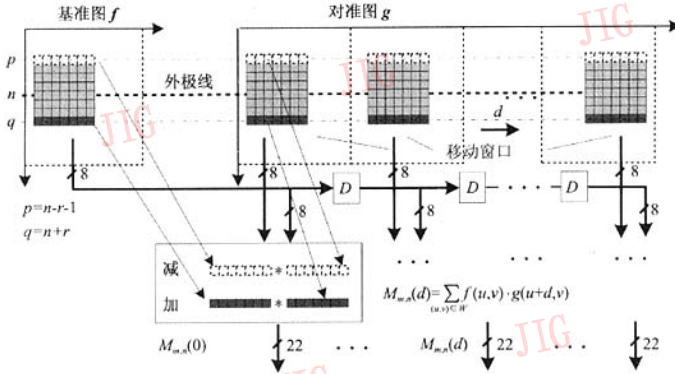


图 2 左、右匹配窗口像素乘积和算法框图

Fig. 2 Algorithm framework of the sum of multiplied corresponding pixels in left and right match windows

具体的实现过程如图 3 所示,当左、右摄像机拍摄的图像的第 q 行像素顺序送入 FPGA 时,即可在像素运算级中完成 8 位乘法运算,并可根据视差 d

值进行窗口截取及求和,同时从外部 RAM 中读取 $sum(p,d)$ 值,并完成 $M_{m,n-1}(d)$ 减 $sum(p,d)$ 的运算,待 $sum(q,d)$ 运算完成后与之相加,即获得 $M_{m,n}$

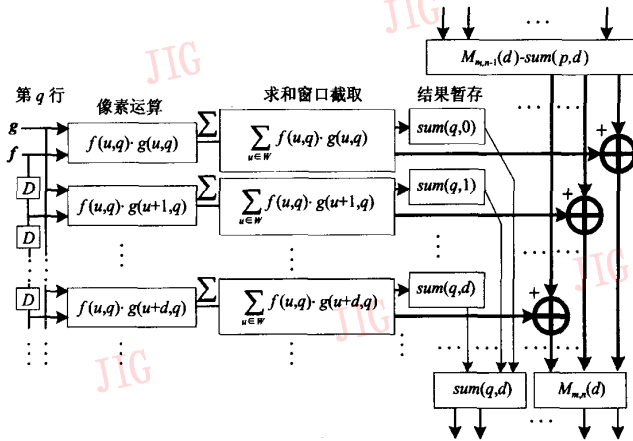


图 3 左、右匹配窗口像素乘积和算法实现

Fig. 3 Algorithm implementation of the sum of multiplied corresponding pixels in left and right match windows

(d),并写入暂存,然后参与ZSSD的后续计算及 $M_{m,n+1}(d)$ 的计算,同时将 $sum(q,d)$ 写入外部RAM,以便在计算第 $q+r+1$ 行时读取。在计算窗口内的像素均值及平方和时,也采用类似的Box滤波方法来实现实时运算。

4 硬件结构设计方案

双目立体视觉硬件系统框图见图4,具体按功能可划分为视频解码模块、视频处理模块、数据传输模块3个模块,其中,视频解码模块用于实现对双摄像机的同步采集,形成立体图像对;视频处理模块则利用FPGA的并行性计算体系结构及其合理的流水线设计来实现立体视觉的并行算法;数据传输模块则借助PCI总线线性突发传输的特点,以实现数据的高速传输。

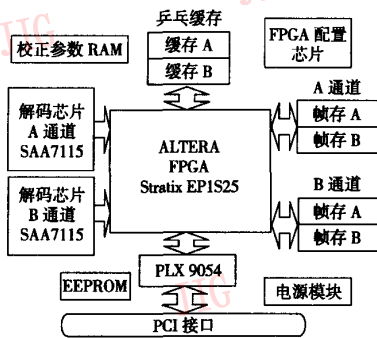


图4 双目立体视觉硬件系统框图

Fig.4 The framework of binocular stereo vision hardware system

视频解码模块采用两块Philips公司的视频解码芯片SAA7115^[6],该芯片不仅提供了9位低噪声的ADC转换器,包括抗混叠滤波器、高性能的增强型梳状滤波器等,同时还提供了高性能的缩放单元,用于实现行、场的任意截取和缩放,以及亮度、对比度、饱和度的控制电路。该芯片还支持多种视频标准。

作为整个系统核心的视频处理模块,不仅是整个系统的逻辑控制单元,更是立体视觉并行算法的实现单元,因而FPGA的选择决定了整个系统性能。算法需求则是FPGA选择的另一准则。如式(3)所示,ZSSD算法需要大量的乘法和加减运算。Altera公司Stratix™系列^[7]EP1S25F780器件片内RAM总容量达到1944576 bits,具有10个DSP块,80个 9×9 的嵌入式硬件乘法器,能满足立体匹配算法实

现的要求。

此外,为了充分利用FPGA的并行计算结构和流水线特点,存储模块采用乒乓缓存结构。乒乓操作的最大特点就是保证经过缓冲的数据流能无时间停顿地送入处理模块,其非常适合对数据流进行流水线处理,以及实现数据的无缝缓冲和实时处理。

PLX公司的PCI 9054芯片^[8],是一块兼容PCI v2.2规范的32位33MHz总线接口控制器,其不仅具有先进的数据流水线架构,包含两个DMA引擎,还具有两个独立的DMA通道。其可编程局域总线运行时钟可达50MHz。突发传输速度可达132MB/s。实际运行过程中,图像传输速率约46MB/s,可满足实时传输要求。

5 系统性能分析

两路PAL制视频信号经SAA7115芯片解码和缩放,仅取奇数帧信号,每40ms生成一组大小为 $320 \times 240, 8\text{bit}$ 的立体图像对。系统时钟采用50MHz,匹配窗口大小为 9×9 ,视差最大搜索范围为32。

由于校正后图像的4个边缘有不同程度的扭曲,因而为了避免边缘无像素值的点引起误配,这些行和列不参与匹配。此外,图像下面的1/3部分是自主导航车的车头,因此亦可无需匹配该部分像素点。

由于式(1)中的 $I(u, v_i)$ 无规律可循,因而只有在在一帧图像采集完成后,才能开始校正程序,并且每校正一个像素后,随即送入匹配模块。整个算法采用AHDL硬件描述语言编写。实践表明,完成一个像素的校正约需11个时钟周期,包括读取4个校正系数和4个原图像素值。匹配过程中,每个像素值约需4个时钟周期才能完成求和(包括计算平方和及左、右图像的像素乘积和)。另外,匹配窗口的相似性的比较可并行完成。整个算法所需时间约16.9ms,PCI传输时间约6.4ms,能满足实时处理需求。而相同的算法在双P4至强服务器上,并利用OpenMP并行机制加速,其所需平均时间约34.6ms。

此外,由于SAD和SSD效果受左、右图像像素亮度差异影响较大,ZSSD匹配算法可通过减去窗口均值来克服这个问题,以增加匹配算法的鲁棒性。图5为实验中的一帧实时输出结果,视差结果几乎没有误配点,而且能很好地反映出路上的行人和车辆。长期的实验结果表明,该方案算法效果好,速度快,不仅具有较强的鲁棒性,而且整个硬件系统稳定、可靠。



(a)校正后左图像

(b)校正后右图像

(c)视差图

图 5 双目立体视觉系统输出结果

Fig. 5 The result of binocular stereo vision system

参考文献 (References)

- 1 Wu Q X, McNeill S J, Pairman D. Fast algorithms for correlation-relaxation technique to determine cloud motion fields [A]. In: Proceedings of Digital Image Computing: Techniques and Applications [C], Brisbane, Australia, 1995: 330 ~ 335.
- 2 Sun Changming. A fast stereo matching method [A]. In: Proceedings of Digital Image Computing: Techniques and Applications [C], Auckland, New Zealand, 1997: 95 ~ 100.
- 3 Zhang Z Y. Flexible camera calibration by viewing a plane from unknown orientations [A]. In: Proceedings of IEEE International Conference on Computer Vision [C], Corfu, Greece, 1999, 1: 666 ~ 673.
- 4 McDonnell M J. Box-filtering techniques [J]. Computer Graphics and Image Processing, 1981, 17: 65 ~ 70.
- 5 Stefano L Di, Marchionni M, Mattoccia S, et al. A fast area-based stereo matching algorithm [A]. In: Proceedings of the 15th IAPR/CIPPRS International Conference on Vision Interface [C], Calgary, Canada, 2002: 146 ~ 153.
- 6 Philips Corporation. SAA7115 CVIP2 Datasheet [M]. Amsterdam, Holland: Philips Corporation, 10 October, 2001.
- 7 Altera Corporation. Stratix device handbook [M]. San Jose CA, USA: Altera Corporation, July 2003.
- 8 PLX Corporation. PCI 9054 data book [M]. Sunnyvale, CA, USA: PLX Corporation, version 2.1, January, 2000.