

基于FPGA的立体视觉匹配的高性能实现

丁菁汀^{*①} 杜歆^① 周文晖^② 刘济林^①

^①(浙江大学信息与通信工程研究所 杭州 310027)

^②(杭州电子科技大学计算机学院 杭州 310018)

摘要: 立体视觉系统在3维场景信息感知中起着重要的作用。其中立体匹配算法的运算复杂度较高,实时处理需要硬件实现匹配运算。但在现有的不多实现中,性能要求和硬件资源的矛盾突出。随着分辨率的增加,对处理速度和视差搜索范围都有更高的要求。对此,该文提出了一种立体匹配硬件实现结构,通过并行化算法子模块和合理安排流水结构来提高性能。匹配算法引入了自适应相关窗口的匹配策略,提升了深度不连续区域的视差质量。该方法结合左右一致性校验准则,可有效去除大部分错误匹配结果。整个匹配流程在单片现场可编程门阵列(FPGA)上实现,并在有限硬件资源条件下将视差搜索范围扩大到128像素。系统时钟60 MHz时,对于512×512分辨率的立体图像,系统可以实现60帧/秒以上的处理速度。

关键词: 图像处理; 立体视觉; 实时性; FPGA; 自适应相关窗口

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2011)03-0597-07

DOI: 10.3724/SP.J.1146.2010.00943

High Performance Implementation of Stereo Vision Matching Based on FPGA

Ding Jing-ting^① Du Xin^① Zhou Wen-hui^② Liu Ji-lin^①

^①(Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China)

^②(College of Computer, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: Stereo vision system plays important role in three-dimensional information perception. Due to the high computational complexity, real-time processing of stereo vision needs to use dedicated hardware. However, performance requirements conflict with hardware resources in existing implementations. With the resolution increased, system requires larger disparity range and higher processing speed. In this paper, a stereo vision implementation is proposed using fine-grain pipelined structure and sub-module parallelism to improve performance. The implemented matching algorithm used adaptive correlation window strategy to raise disparity quality at object borders and integrated left-right consistency check to reduce possible errors in general. The entire stereo matching process is realized using a single chip of Field Programmable Gate Array (FPGA) and extended disparity search range to 128 pixels under limited resources. The matching process is capable of generating disparities at more than 60 frames per second on 512×512 images when clocked at 60 MHz.

Key words: Image processing; Stereo vision; Real-time; FPGA; Adaptive correlation window

1 引言

立体视觉技术是一种重要的深度信息获取方法,在无人车自主导航、3维测量和物体建模中都有广泛的应用。其中匹配算法是立体视觉的关键技术。匹配算法需要对大量的数据进行重复操作,因此具有很高的计算复杂度。在PC平台的快速实现依赖于单指令多数据流(SIMD)优化^[1]或采用GPU(Graphic Processing Unit)并行处理^[2,3]的方法。

SIMD优化的方法将占用大量的CPU处理时间,使之无法完成系统的后续任务;而GPU所带来的高功耗问题也限制了其应用场合。因而,更具实用性的实时实现方法是通过硬件电路构建匹配模块。

其中开创性的工作是构建在PARTS上的立体视觉系统^[4],整个硬件平台由4×4网格互联的FPGA组所构成。后续,文献[5]和文献[6]提出了在FPGA上基于SAD匹配准则的算法实现。与像素差绝对值和(SAD)及像素差平方和(SSD)等度量因子相比,Census变换对图像中的噪声和亮度不一致性有很高的鲁棒性,能获得更优的视差结果。最近,文献[7]采用基于Census变换的匹配算法实现了640×480分

2010-08-31收到,2010-11-10改回

国家自然科学基金(60534070,90820306)资助课题

*通信作者:丁菁汀 martindd113@gmail.com

分辨率图像的实时处理,但其实现结构需要耗费大小与视差范围成正比的片内存储空间,限制了视差搜索范围的扩展。

本文在单片 FPGA 上设计并实现了一种基于 Census 变换的改进立体匹配算法。与之前提出的设计方案相比,当前算法实现的主要优点是:首先,在具体硬件实现算法中结合了自适应匹配窗口策略和左右一致性约束等相应改进。其次,在硬件实现中通过并行化算法子模块和合理安排流水结构,实现了对高分辨率图像更大视差搜索范围的实时处理。

2 基于相关的立体匹配算法

基于相关的立体匹配方法是一种快速、致密的局部匹配方法。相关匹配的基本假设是匹配窗口内的像素具有相同的深度值。一个典型的算法流程可分为以下几步:(1)计算匹配代价;(2)匹配代价在相关窗口内进行累积;(3)视差搜索和校验。

2.1 非参数化方法

非参数化方法是由文献[8]提出的,其思想是对图像进行非参数化变换(Census 或 Rank),提取包含局部纹理信息的描述符。

Census 变换如式 1。首先以某一像素为中心构建一个变换窗口 N , 函数 ξ 对中心像素的灰度值 $I(m,n)$ 与窗口内其余像素的灰度值 $I(x,y)$ 进行比较。

$$\xi [I(m,n), I(x,y)] = \begin{cases} 0, & I(m,n) < I(x,y) \\ 1, & I(m,n) \geq I(x,y) \end{cases} \quad (1)$$

窗口内所有像素的比较结果串连(\otimes)成一个表征中心像素 (m,n) 与周围像素灰度值的比较关系的比特向量 $\text{code}(m,n)$ 。

$$\text{code}(m,n) = \otimes_{[x,y] \in N} \xi [I(m,n), I(x,y)] \quad (2)$$

最后,以不同视差 d 条件下两个比特向量之间的汉明距离作为左右视图像素间的匹配代价。

2.2 自适应相关窗口

相关窗口以待匹配像素为中心,对窗口内像素的匹配代价值进行累积。通过找寻累积代价值最小的相关窗口来搜索相应的视差值 d 。因而,相关窗口尺寸直接影响最终的匹配结果。通常选用较大尺寸的相关窗口可以减少在弱纹理区域的错误匹配。另一方面,相关窗口在横跨视差不连续区域时,由于违背了深度一致性假设会引入较多的匹配误差。为了更准确地提取出 3 维场景中物体的边缘信息,

本文采用了文献[9]提出的自适应变换相关窗口的匹配方法。

图 1 示意了自适应相关窗口结构与匹配过程。算法中的相关窗口共由 9 个子窗口组成,作为代价值累积的候选区域。图中圆形与梯形区域分别对应不同深度值的景物在图像中的投影。如图所示,当横跨深度不连续区域时,相关过程会存在歧义性。此时相关窗口将自适应进行子窗口选取。最终的累积代价 C 为中心子窗口代价值 C_{mid} 与其余 4 个最小的子窗口代价值 C_{sub_i} 的累加和。

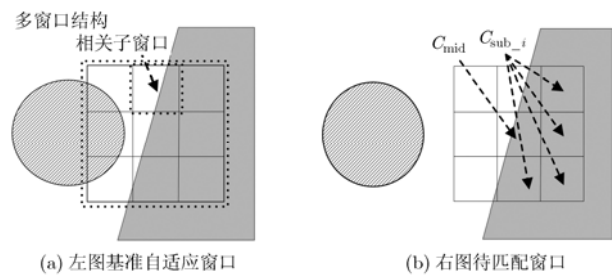


图 1 自适应相关窗口匹配示意

2.3 左右一致性校验

在相关匹配中,当图像中的物体存在遮挡时,左右视图像素之间将不存在对应关系。左右一致性校验对于因遮挡产生的错误匹配结果具有最高的检测率和最低的误检率,能有效地滤除无效视差。校验过程中要求来自左右视图的一对像素匹配对,只有在以左图和右图两次匹配过程中存在一致的匹配关系才认为是有效的匹配结果。

3 算法硬件实现

图 2 为算法硬件实现的整体流程框图。整个流程可大致分为非参数化变换、自适应相关匹配两个主要的阶段。在实现中采用了并行展开的流水结构,通过缓存空间特定的读写操作,构建多个并行 Census 变换和自适应匹配模块来提升处理速率。同时系统降低了单个视差模块的硬件资源消耗,扩展有限资源下的视差搜索范围。

3.1 缓存空间数据读写

Census 变换与匹配代价值累计本质上都是基于窗口的运算。假设操作窗口的高度为 h , 宽度为 w 。为了实现流水的节拍操作,需要对 h 行的输入数据进行缓存。缓存完成后,存储的数据按列方向输出到后续模块的寄存器组。当 w 列数据输出后初始数据窗口构建完毕,此时通过寄存器间数值比较或累

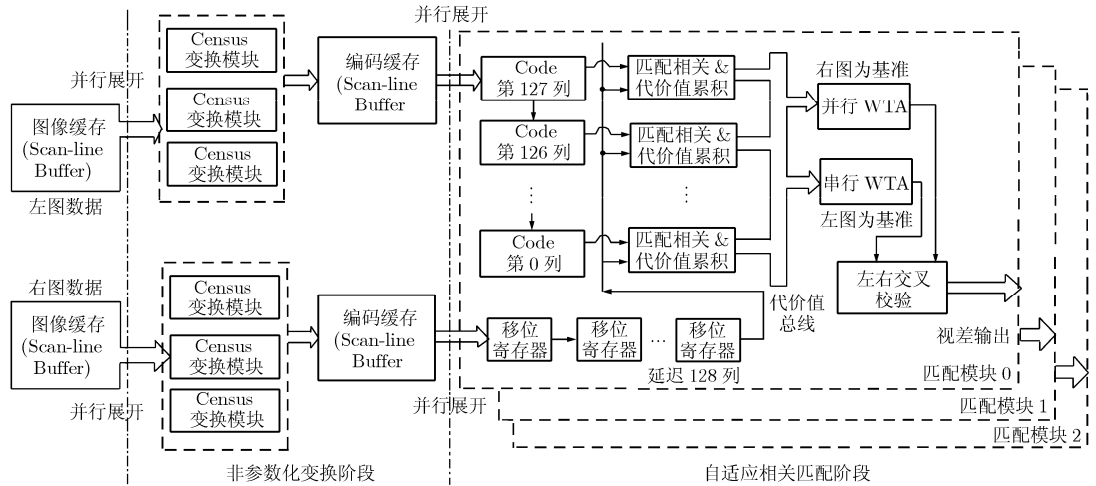


图 2 匹配算法流程图

加就可获得相应的 Census 编码或匹配代价值。

数据窗口在水平移动过程中使用移位寄存器可以避免窗口数据的重复读取，但是在下移过程中每个数据对应窗口中的不同行会被重复读取 h 次。耗费过多内存读取周期将限制系统的处理速率。为此，文本采用了典型的并行展开思想，见图 3。在窗口水平滑动过程中，每列读取 $h + 2$ 个数据来同时构建 3 个并行的数据窗口，同步执行 3 行的窗口运算。以 Census 变换为例，在 $h + 2$ 个时钟的列周期内系统将完成 3 个窗口的数据更新，同时使能变换信号产生编码数据。3 个窗口的变换使能信号(位于 Idle 周期)之间依次存在一个时钟周期的延迟。当移动至每行末尾时 3 个变换窗口同时下移 3 行，进行下一周期窗口计算。扩展后处理频率将达到 $3 / (h + 2)$ 像素/单位时钟，当 h 较大时处理速度提升接近 3 倍。

系统中缓存模块由 FPGA 片内双口 RAM 块构成。整个缓存模块分为两个部分，其中 $h + 2$ 个行缓存为数据输出空间，其余 3 个行缓存用于数据更新。

图 4 为缓存模块的读写时序。当缓存初始化完成后，窗口水平移动输出数据，同时输入端进行后续图像数据的更新，确保窗口下移过程的无缝进行。

3.2 匹配模块描述

自适应相关匹配过程在算法中处于核心地位。在相关过程中，需要单独计算不同视差条件下匹配窗口累计代价值。文献[7]在 Hamming 模块后对不同视差下的匹配代价分别行缓存，以进行窗口累计。此结构的优点是可以同时得到不同视差下的匹配代价值，具有很高的计算效率。但缺点在于所需的缓存模块数量与视差搜索范围相同。综合结果显示即使采用当前较大片内存储空间的 FPGA 芯片也仅能满足 64 像素的搜索范围。为了在有限资源消耗下实现更大视差搜索范围，本文实现结构在 Hamming 模块之前构建 Census 编码缓存模块进行全局行缓存。编码缓存输出时通过移位寄存器实现左右视图不同视差编码之间的对应匹配。图 5 描述了不同视差下的匹配代价值累计电路结构，图中 D 模块为采用移

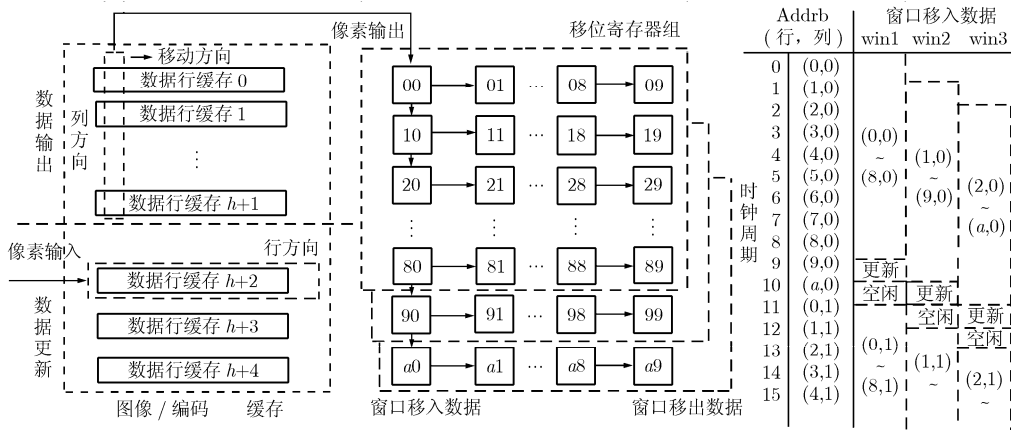


图 3 并行展开窗口操作示意，窗口尺寸为 9×9 像素

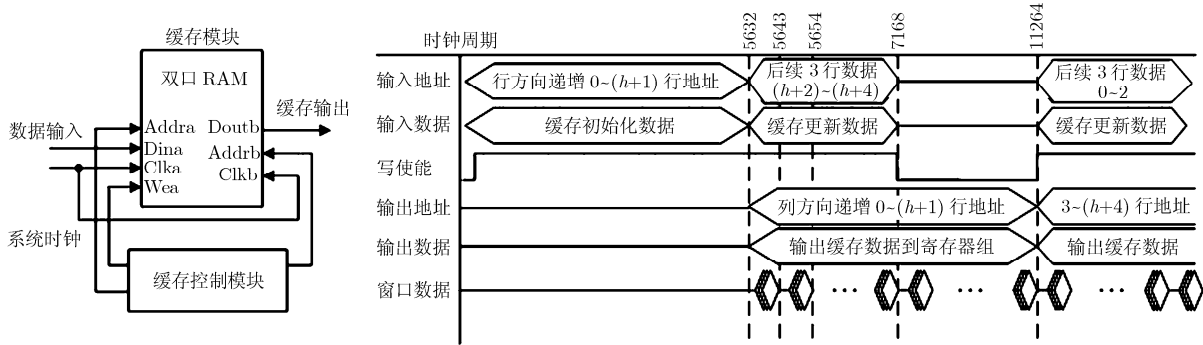


图 4 缓存模块的读写时序图 (时钟周期计数对应于图像宽度 512 像素、窗口高度 $h = 9$)

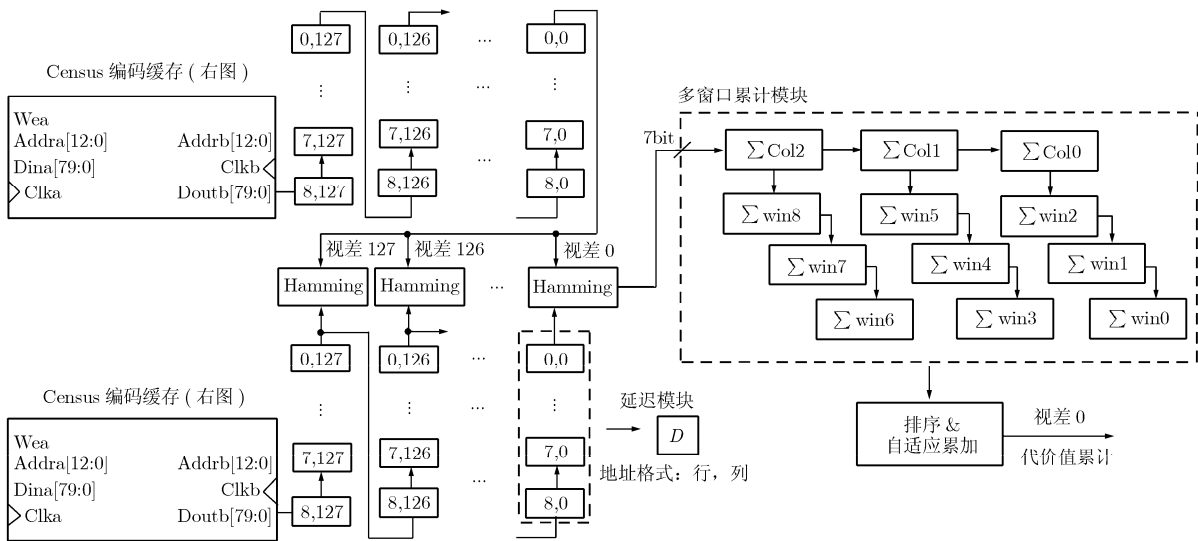


图 5 不同视差下的像素匹配结构

位寄存器实现的列延迟单元。在当前结构下，不同视差条件下仅需对自适应窗口内的代价值进行缓存，将所需的行缓存空间复杂度从 $O(n)$ 降为 $O(1)$ 。

3.2.1 自适应窗口代价值累加 多候选窗口代价值累加过程可参见图 5 右侧模块。首选，在列方向上构建了 3 个代价值缓存模块 Col_n 来完成列代价值的累积。列累积完成后，缓存模块 Col_n 的代价值输入到不同子窗口 win_n 组成的累积网络。在每一个列周期 win_i 从所连接的 win_j 模块移入最新的列累积值，同时将最早输入的列累加值移出，实现窗口滑动过程。

在子窗口代价值累积完成之后，候选子窗口需要进行一个代价值排序的排序过程，在硬件电路实现中采用了 Batcher bitonic 排序网络结构。8 个比较选择单元模块在一个时钟周期内进行先后两次相互比较选择，两个时钟周期后将 4 个最小的代价值输出至后续累加模块。

3.2.2 左右一致性校验 左右一致性校验需要同步获取左图和右图为基础的匹配结果。如果简单地

原有单方向的匹配电路进行复制，将耗费大量的硬件资源。本文借鉴了文献[10]中匹配结果复用的思想，设计了特定的流程结构使得两个方向的匹配过程在一次相关中实现。图 6 示意了相应的算法思想，为了简单起见，图中只显示了每行前 128 个像素的

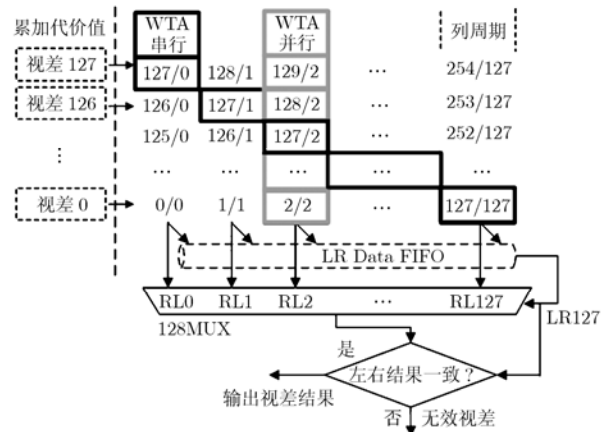


图 6 像素匹配结构与一致性检测示意 (表达式 u/v 表示左图一行中第 u 个像素和右图同一行中第 v 个像素之间匹配)

匹配过程。

图中术语 WTA 为‘Winner Take All’的缩写, 在相关匹配中意为只选取最优匹配的视差值作为最终结果。在单位列周期内, 自适应累加模块输出同一右图像素 v 为基准(R-L)的匹配代价, 通过并行 WTA 就可以得到右图像素 v 的视差值。从图 6 中可以发现在对角线方向为同一左图像素 u 为基准(L-R)的匹配代价值。由于相应匹配代价值位于不同的列周期, 因而可通过一系列串行的比较和缓存进行匹配代价值的依次比较, 即串行 WTA。串行 WTA 的流程示意图 7, 此过程类似于冒泡排序, 在进行 127 次选择和比较后得到最小代价和相应视差值。最后, 前 128 周期内 R-L 匹配结果按先后顺序输入多路选择器, 串行 WTA 输出的 L-R 匹配结果作为相应选择信号。只有当选择器输出结果在数值上与选择信号相同时判定匹配结果为有效视差¹⁾。

4 实验结果与性能分析

目前算法设计实现为处理 512×512 像素分辨率图像, 采用 9×9 像素大小的 Census 变换窗口。相关窗口由 9 个 3×3 像素大小的子窗口所组成。系统设定的时钟频率为 60 MHz, 最大视差搜索范围为 128 像素。由于采用了并行扩展的流水结构, 单位像素的处理时间可以缩减到接近于 3 个时钟周期。最终运算处理能力达到 16 Mpps, 在当前分辨率条件下处理速度约为 64 帧/秒。算法实现平台为 Xilinx-XC5VLX155T 芯片, 表 1 列出了 ISE 10.1 综合产生的硬件资源占用情况, 分别以查找表、触发器、片内存储单元和硬核乘法器消耗的绝对数量和相对百分比进行说明。

表 2 为立体匹配硬件实现的性能对比, 帧率按 60 MHz 时钟归一化。从表中可以看出 Jim 等人^[7]给出了当前最快速率的设计实现。但如前所述, 其匹配结构在 64 像素视差下已消耗 192 块片内 36 k BlockRam, 限制了向更大视差搜索范围的扩展。本文的设计结构平衡了高分辨下算法实时性和大视差搜索范围下资源消耗之间的矛盾。此外, 当前设计首次将自适应相关窗口引入硬件实现中, 后续评估

结果显示此项改进将提升算法的准确性。

整体模块代码在 ModelSim 6.2 b 平台上进行了仿真。采用 Middlebury 数据集的‘Art’立体图像对进行算法评估。图 8 显示了基准视差结果和仿真得到的视差图像, 视差图像忽略了每行的前 128 个像素的视差值, 以去除边界效应。由于算法不涉及乘/除的运算, 没有定点精度的影响, 因而仿真结果与在通用处理器上的运算结果完全一致。

在算法视差质量的评估中选用了以下 3 个性能指标: 有效视差像素的查找率(Found), 视差结果的准确率(Accuracy)和偏差均方根值(RMS)。在立体视觉应用中有效视差的准确率以及 RMS 值直接关系到系统的可靠性。图 9 显示了仿真后得到视差结果的评估对比。从评估结果中可以看出, 左右一致性校验去除了大量的错误匹配, 使得 RMS 值大幅降低。而在一致性校验基础上引入自适应相关窗口算法, 则使得最终视差结果在 3 个指标上都有相应提高。

图 10 为移植实际硬件平台后在真实场景中的测试结果。在当前基线长度下, 128 像素搜索范围较好覆盖了近距离景物的视差值。最终获取的视差图像准确的反映了场景中的结构信息和物体的边缘轮廓。

5 结束语

本文描述了一种改进的立体视觉匹配算法在 FPGA 上的设计和实现。算法针对传统相关匹配方法的不足, 引入自适应相关窗口和左右一致性校验等相应机制。在硬件实现通过并行化算法子模块和合理安排流水结构, 实现对 512×512 像素分辨率图像超过每秒 60 帧的处理速度。同时在有限资源消耗下将视差搜索范围扩展到了 128 像素, 提升了深度方向的分辨率和探测距离。通过合理安排匹配流程, 在一次相关中实现左右一致性校验。实验结果显示最终视差结果滤除了大部分遮挡区域的错误匹配, 同时提升了视差不连续区域的边缘质量。

本文实现的匹配算法模块作为立体视觉系统的核心部分, 将应用于陆地自主车的导航中以实时提供场景的深度信息。下一步的研究将围绕实际应用中的具体场景, 对算法进行评测和相应调整。

¹⁾实际上, 考虑到亚像素精度的缺失, 在校验过程中允许存在左右匹配结果间一个像素的偏差。

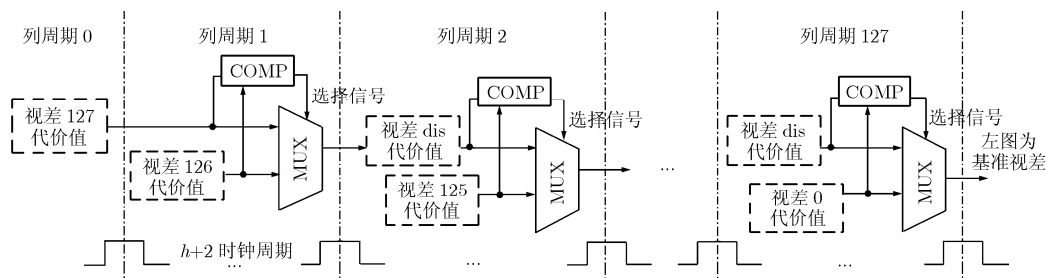


图 7 串行 ‘WTA’ 匹配结构示意图

表 1 XC5VLX155T 芯片的硬件资源消耗

Slice LUTs	Slice flip-flops	On-chip Ram Blocks		Dsp48E slices
46082 / 97280 (47.37 %)	26549 / 97280 (27.29 %)	18 kbits	36 kbits	0 / 128 (0 %)
		0 / 0 (0 %)	44 / 212 (20.75%)	

表 2 立体视觉硬件实现性能对比

实现方案	图像尺寸	匹配准则	视差范围	窗口尺寸	帧率 (60MHz)
Ambrosch ^[5] <i>et al.</i>	320×240	SAD	100	3×3	392 fps
Chen ^[6] <i>et al.</i>	640×480	SAD	32	N/A	30 fps
MSVM-III ^[11]	640×480	SSD	64	7×7	30 fps
Jin ^[7] <i>et al.</i>	640×480	Census	64	Census (11×11) Corr (15×15)	146 fps
本文方案	512×512	Census	128	Census (9× 9) Corr (自适应)	64 fps

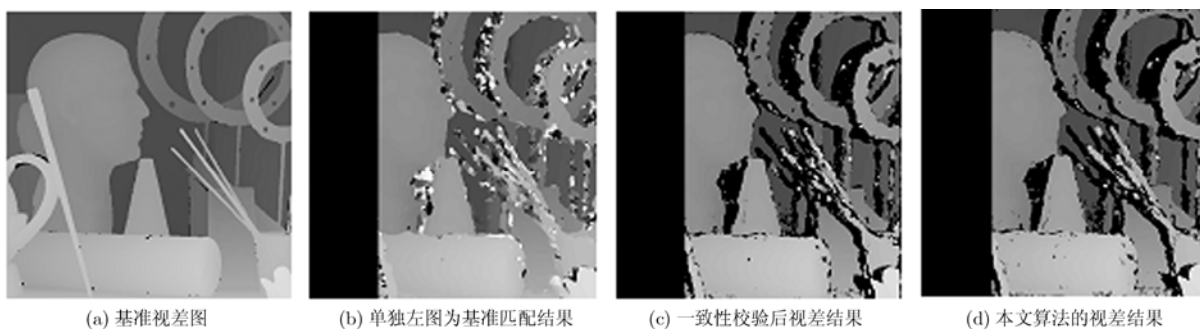


图 8 视差结果对比 (图像中暗点为一致性校验滤除的无效视差)

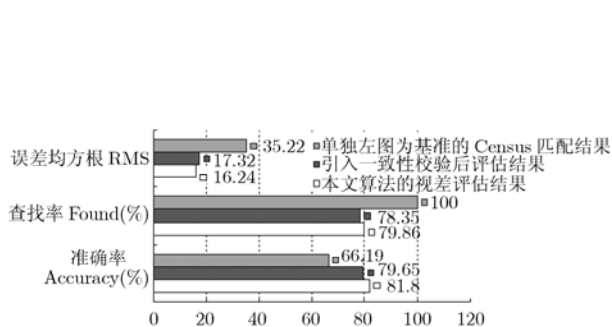


图 9 算法的评估结果

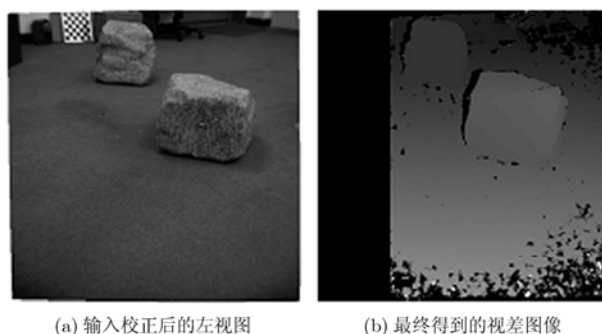


图 10 真实场景的视差结果

参 考 文 献

- [1] Forstman S, Kanou Y, and Ohua J, *et al.* Real-time stereo by using dynamic programming [C]. Proceedings of the Conference on Computer Vision and Pattern Recognition, Washington, DC, Jul.02, 2004, Vol.3: 29.
- [2] Gong M and Yang Y. Near real-time reliable stereo matching using programmable graphic hardware [C]. Proceedings of the Conference on Computer Vision and Pattern Recognition, San Diego, Jun.20, 2005, Vol.1: 924-931.
- [3] Wei Yu, Tsuhan C, and Hoe J C. Real time stereo vision using exponential step cost aggregation on GPU [C]. Proceedings of the International Conference on Image Processing, Cairo, Nov. 2009: 4281-4284.
- [4] Woodfill J and Von-Herzen B. Real-time stereo vision on the PARTS reconfigurable computer[C]. Proceedings of 5th IEEE Symposium on FPGAs for Custom Computing Machines, Napa California, Apr. 16, 1997: 201-210.
- [5] Ambrosch K, Humenberger M, and Kubinger W, *et al.* Hardware implementation of an SAD based stereo vision algorithm [C]. Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Minneapolis, Jun. 17, 2007: 1-6.
- [6] Chen L and Jia Yunde. A Parallel reconfigurable architecture for real-time stereo vision [C]. Proceedings of the International Conference on Embedded Software and Systems, Hangzhou, May 25, 2009: 32-39.
- [7] Jin S, Cho J, and Phan X, *et al.* FPGA design and implementation of a Real-Time stereo vision system[J]. *IEEE Transaction on Circuits and Systems for Video Technology*, 2010, 20(1): 15-26.
- [8] Zabin R and Woodfill J. Non-parametric local transforms for computing visual correspondence [C]. Proceedings of 3rd European Conf. Computer Vision, Stockholm, May 1994: 150-158.
- [9] Hirschmüller H, Innocent P R, and Garibaldi J. Real-time correlation-based stereo vision with reduced border errors [J]. *International Journal of Computer Vision*, 2004, 47(1-3): 229-246.
- [10] Mühlmann K, Maiser D, and Hesser J, *et al.* Calculating dense disparity maps from color stereo images, an efficient implementation [J]. *International Journal of Computer Vision*, 2004, 47(1): 79-88.
- [11] Jia Yun-de, Zhang Xiao-xun, and Li Ming-xiang, *et al.* A Miniature Stereo Vision Machine(MSCM- III) for dense disparity mapping [C]. Proceedings of 17th International Conference on Pattern Recognition, Cambridge, August, 2004: 728-731.
- 丁菁汀: 男, 1985年生, 博士生, 研究方向为FPGA硬件开发、计算机视觉。
- 杜 歆: 男, 1975年生, 博士, 副教授, 研究方向为计算机视觉、三维重建、模式识别。
- 周文晖: 男, 1977年生, 博士, 副教授, 研究方向为智能机器人导航、智能计算。
- 刘济林: 男, 1947年生, 教授, 博士生导师, 主要从事图像处理、模式识别、智能交通等方面的研究工作。