

像平面偏移位置可变的 3D 视频采集方法研究

王 伟¹, 任 屹¹, 谢 翔¹, 李 国 林²

(1 清华大学 微电子所, 北京 100084; 2 清华大学 电子工程系, 北京 100084)

摘 要: 在 3D 视频采集的过程中, 往往需要摄像机系统的物镜光轴中心和像平面中心有一定的偏移量 h . 论证了偏移量 h 对 3D 显示的重要影响, 提出了利用 CMOS 图像传感器自身功能特性实现对 h 灵活控制的方案, 并实验验证了 3D 视频采集中对偏移量 h 的控制. 实验平台使用自行设计的基于 FPGA 的双目立体视频采集存储系统, 该系统为 3D 视频高效采集提供良好的平台支持.

关键词: 镜头光轴偏移量; 视差; 深度; 基线; 双目视觉

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-7180(2014)03-0048-04

Research on Production Rules for 3D Acquisition with a Variable Shift of Image Plane

WANG Wei¹, REN Yi-yi¹, XIE Xiang¹, LI Guo-lin²

(1 Institute of Microelectronics, Tsinghua University, Beijing 100084, China;

2 Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: Before we capture 3D video, we may need the cameras with a parallel sensor-shifted setup. In this paper, at first, we discussed the importance of the small shift h of the sensor targets and got formulas to calculate h by the mathematical method. Secondly, we presented a program to flexibly control h by using CMOS image sensor in practice. At last, we proved that the program was right by using a hardware platform designed by ourselves. It is a two-channel video capture and storage system based on FPGA. This platform can make the way of 3D acquisition better and more effective.

Key words: shift of the sensor targets; disparity; depth; baseline; binocular stereo vision

1 引言

3D 视频已经进入人们的日常生活, 在立体视频的采集过程中, 往往会由于摄像机系统本身的缺陷导致采集数据存在缺陷. 比如, 采集时摄像机基线距离(记为 t_c)设置不当, 导致在观看立体视频时, 图像匹配点视差太大而不宜观看. 又如, 采集时镜头光轴中心相对像平面中心的偏移量选取不当, 导致在显示端难以获得舒适的 3D 视频效果. 很多立体视觉系统并没有考虑 3D 采集时需要设置最佳摄像机系统参数, 只是对多路摄像头采集到的图像的共同区域进行分析^[1-2]. 针对这些问题, 本文对立体视频采集方法中有关偏移量 h 的影响进行探讨^[3], 并根据

CMOS 图像传感器自身功能特性, 提出一种像平面偏移位置灵活可变的 3D 视频采集方案以及系统实现, 更有效地采集 3D 视频数据.

2 中心偏移量 h

3D 采集通常以双目采集为主要方式. 双目采集的摄像机有三种摆设方式, 如图 1 所示. 图 1(a)所示的是模仿人眼的摆设方法. 此方法有两个问题造成后续处理困难: ①立体窗口^[4]不重合; ②立体窗口内的物体在左右两幅图像上具有垂直方向的视差. 图 1(b)的两台摄像机光轴平行(t_c 为两摄像头基线距离), 解决了垂直方向视差的问题, 但也存在明显的问题, 即双路摄像机拍摄的图像仅有一部分是重

合的,重合部分可以用来分析场景深度,非重合部分则对3D合成没有意义.同时注意到,这样拍摄的左右图像中心点显示到屏幕上时,零视差点对应的是无穷远处的物点,所有非零视差的点都显示到屏幕的前端.这和人眼观看3D视频的舒适地带原则是违背的^[5].图1(c)的设置模式中,镜头光轴中心相对像平面中有一偏移量 h .这种方法的好处在于:左右视频图像对应的立体窗口是理想重合的.由于像平面中心位置平移,新的视差可以看成是原来视差^[6]平移的结果,于是,采集端的零视差面可以映射到显示端的零视差面,使得显示端和采集端可以更好地匹配.本设计以图1(c)的设置作为研究基础.

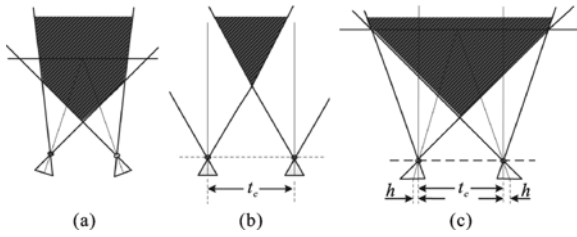


图1 3D采集的摄像机摆设方式比较

2.1 中心偏移量 h 的影响与调整

图2展示了采集端摄像机系统和3D显示端的一些重要参数.图2(a)所示为采集端摄像机系统的参数: t_c 为双摄像头之间的基线距离, h 为镜头光轴中心相对像平面的偏移量, f 为摄像机焦距, Z_c 为零视差面所在的深度位置,即位置 Z_c 处的物点将对应到3D显示屏的表面位置,位置大于 Z_c 的物点将对应到显示屏后,小于 Z_c 的物点将对应到显示屏前.若定义视差为 d ,任意深度距离为 Z ,那么可以计算得到

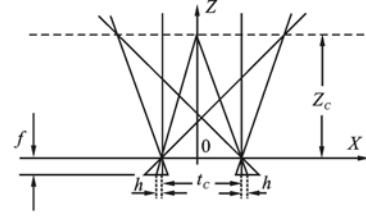
$$d = 2h - t_c \cdot f / Z = t_c \cdot f \cdot (1/Z_c - 1/Z) \quad (1)$$

可见采集端摄像机系统的主要参数为 f , t_c , h 或 Z_c .本设计假设 f 不可调,但 t_c , h (或 Z_c)可调.

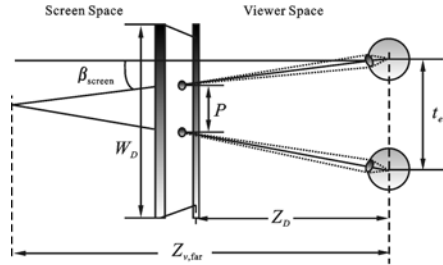
对于显示端来讲,如图2(b)、(c)所示, Z_v 为任意一对匹配点重建得到的深度,图中给出的 $Z_{v, \text{far}}$ 和 $Z_{v, \text{near}}$ 是 Z_v 的两个极端,它们由显示端视差 P ,人的双眼间距 t_e ,人眼距离屏幕的距离 Z_d 共同决定:

$$Z_v = \frac{z_D \cdot t_e}{t_e - P} \quad (2)$$

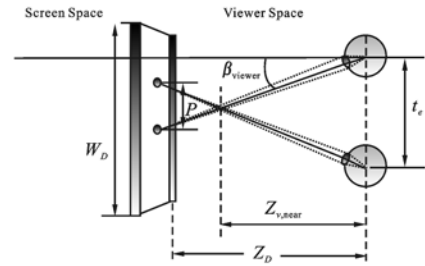
显示端图像对于采集端图像就是一个放大的过程,若已知摄像机图像传感器的宽度为 W_s ,即可知道放大比 $S_M = W_D / W_s$,进而 $P = S_M \cdot d$.代入式(1)可得重建深度 Z_v 与实际深度 Z 之间的关系:



(a) 3D采集端的相关参数



(b) 3D显示端最远重建距离及相关参数



(c) 3D显示端最近重建距离及相关参数

图2 3D视频采集端和显示端的重要参数

$$Z_v = \frac{z_D \cdot t_e \cdot Z}{S_M \cdot f \cdot t_c - Z \cdot (2 \cdot S_M \cdot h - t_e)} \quad (3)$$

可见, Z_v 和 Z 之间不一定是线性关系.这一特性可以直接利用以获得较好的视觉效果.

在3D显示端对 Z_v 有实际的约束.人眼在观看3D视频时,双眼聚焦于屏幕,而双眼汇聚则可能出现在屏后、屏上或屏前.由于和观看真实世界的习惯不同,3D视频处理不当就会导致双眼经常出现疲劳、疼痛等不适.研究表明,观看3D视频时,只有围绕屏幕周围一段区域内才是效果舒适的.这一段区域可以根据人眼转向角的变化程度来进行确定^[7].如图2(b)、(c)所示,屏幕将空间分为Screen Space和Viewer Space两个区域,人眼转向角描述为观看3D物点时,以眼睛正前方为参考线,观察到屏幕上对应物点眼睛转过的角度,记为 β .图2(b)和(c)中 β_{screen} 和 β_{viewer} 分别代表了Screen Space和Viewer Space两个区域中观察到 $Z_{v, \text{far}}$ 和 $Z_{v, \text{near}}$ 深度物点时的人眼转向角.根据经验,为了观看3D不产生复视效果,Screen Space和Viewer Space人眼转向角的总变化不应超过各自的允许值.这两个允许值分别为^[7]

$$\Delta\beta_{\text{viewer}} = \arctan(t_e/(2 \cdot Z_{v,\text{near}})) - \arctan(t_e/(2 \cdot Z_D))$$

$$\approx \frac{t_e}{2} \cdot \left(\frac{1}{Z_{v,\text{near}}} - \frac{1}{Z_D} \right) = 0.0077 \text{ rad} \quad (4a)$$

$$\Delta\beta_{\text{screen}} = \arctan(t_e/(2 \cdot Z_D)) - \arctan(t_e/(2 \cdot Z_{v,\text{far}}))$$

$$\approx \frac{t_e}{2} \cdot \left(\frac{1}{Z_D} - \frac{1}{Z_{v,\text{far}}} \right) = 0.0069 \text{ rad} \quad (4b)$$

换句话说,在显示端为了观看到舒适的 3D,重建的深度范围(从 $Z_{v,\text{near}}$ 到 $Z_{v,\text{far}}$)应该满足人眼转向角总变化的约束。

在采集 3D 视频时,一般会遇到如下两种需求:①对实际的深度场景进行等比例重建;②不等比例重建时,将已知的实际场景重建到 3D 显示端的舒适区域(从 $Z_{v,\text{near}}$ 到 $Z_{v,\text{far}}$)之内。对于第一种情况,在式(4)中,令 $2 \cdot S_M \cdot h - t_e = 0$ 即可,这对偏移量 h 提出了如下要求:

$$h = t_e / (2 S_M) \quad (5)$$

对于第二种情况,如果实际场景的深度范围为 Z_{near} 到 Z_{far} ,而根据人眼转向角的约束要求重建深度为 $Z_{v,\text{near}}$ 和 $Z_{v,\text{far}}$,则根据式(3), Z_{near} 和 Z_{far} 分别对应于 $Z_{v,\text{near}}$ 和 $Z_{v,\text{far}}$,如是可获得对摄像头基线距离 t_e 和中心偏移量 h 的要求:

$$t_e = \frac{Z_D \cdot (1/Z_{v,\text{near}} - 1/Z_{v,\text{far}})}{S_M \cdot f \cdot (1/Z_{\text{near}} - 1/Z_{\text{far}})} \cdot t_e \quad (6)$$

$$h = \frac{t_e}{2 S_M} \left[\frac{Z_D}{Z_{\text{far}} - Z_{\text{near}}} \left(\frac{Z_{\text{near}}}{Z_{v,\text{near}}} - \frac{Z_{\text{far}}}{Z_{v,\text{far}}} \right) + 1 \right] \quad (7)$$

针对以上两种简单的采集需求,我们通过显示端的约束得到了对 h 的调整要求:或者是固定的 h 如式(6),或者是根据不同场景深度 Z_{near} 、 Z_{far} 实时调整的 t_e 和 h ,如式(6)~(7)。

3D 采集时,为了解决视感冲突的问题,还会遇到另一种需求。视感冲突的现象描述为:对于一个出现在立体窗口边缘的物体,可能利用双眼通过视差感受到物体在屏幕前方,但是由于单眼告诉我们的感觉是物体在屏幕后方,因为物体的一半被屏幕框挡住了。这种现象会带来坏的 3D 体验。改善此问题的一种方法是先调整采集端参数 Z_c (零视差深度)的位置,使边缘物体重建到屏幕的后方。确定了 Z_c ,然后根据式(1)可以得到 t_e 和 h 的关系(假定 f 保持不变),若再有一组感兴趣的深度和实际深度的对应关系即可求出 t_e 和 h 的具体值。

2.2 对偏移量 h 的调整控制

上面讨论了三种 3D 视频采集要求情况下 h 的

不同选择方法,有时 h 是固定的,有时 h 是需要实时调整的。如果拍摄系统的中心偏移量 h 可随时调整,系统将是一个灵活系统。下面考察如何利用 CMOS 图像传感器的功能特性实现对 h 的灵活控制。

图像传感器自身的配置可以实现中心偏移量 h 的控制,如图 3 所示。

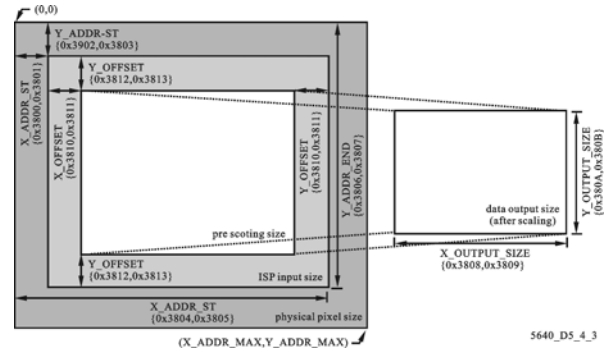


图 3 OV5640 的图像输出窗口配置^[8]

这是 OmniVision 公司的 OV5640 芯片的图像窗口配置图。窗口起始点 X_ADDR_ST 、 Y_ADDR_ST 和终止点 X_ADDR_END 、 Y_ADDR_END 以及 ISP 窗口的 X_OFFSET 、 Y_OFFSET 等的数值可通过相关寄存器直接配置。图像实际输出窗口为右侧白色区域。这里用到了传感器的图像缩放功能。简单来说,经过图像传感器配置输出的 pre-scaling size 窗口,需要经过图像缩放功能得到定义的 Data output size 窗口。这一功能同样是图像传感器内部经过配置相关寄存器来完成的。OV5640 支持的最大分辨率为 2592×1944 ,每个像素点的大小为 $1.4 \mu\text{m} \times 1.4 \mu\text{m}$,如果根据理论先计算得到 h ,那么就可以计算输出像平面具体平移的像素数 m 。若缩放的倍数为 n ,则

$$m = n \cdot \frac{h(\mu\text{m})}{1.4(\mu\text{m})} \quad (8)$$

因此,要想实现 h 的控制,可以通过配置改变上图中 X_ADDR_ST 、 X_ADDR_END 的位置,而其他寄存器取适当的值并保持不变。

此方法有两个优点:①操作非常简单,实现具体的 h ,通过简单的寄存器写数即可;② h 的可控性非常灵活,其值以 $1.4 \mu\text{m}$ 为步长连续可调。

3 实验验证与结果分析

为了验证上述方案的可行性,设计了一套双目立体视频数据采集存储系统,通过寄存器配置方法改变中心偏移量 h ,对得到的图像匹配对进行了数据分析。

3.1 验证系统设计

本系统中,CMOS 图像传感器的型号为 OV5640, FPGA 选用 Xilinx 公司的 Virtex-4 XC4VLX100 FPGA,存储器选用 Samsung 公司的 K9F8G08U0M. 其结构如图 4 所示.

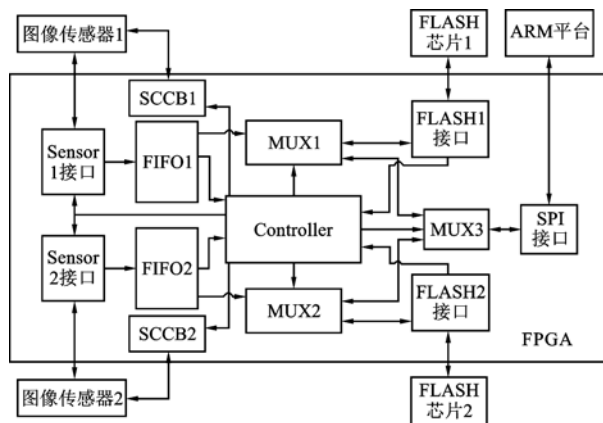


图4 验证系统的整体结构图

该结构中,各个模块的功能如下:FPGA 为控制核心,控制器 Controller 为总控制,Controller 通过 SCCB 启动图像传感器工作, Sensor 接口模块判断何时记录数据, FIFO 对图像数据进行缓存,并将有效图像数据通过 MUX1(MUX2)送到 FLASH 接口模块. MUX1(MUX2)的功能是选择 FLASH 接口需要的初始化参数(FLASH 擦除或读写一些相关参数). FLASH 接口模块的作用是接收 MUX1(MUX2)的相关参数,内部实现 FLASH 编程操作的命令时序,对 FLASH 芯片进行相应的数据操作. MUX3 的作用是将 FLASH 芯片的读操作参数赋给对应的 MUX1 或者 MUX2,准备读取该 FLASH 芯片的数据. SPI 接口的作用是通过 SPI 协议将图像数据读取到 ARM 平台上处理.此外,Controller 和每个模块之间都有相应的逻辑联系.

3.2 关于中心偏移量 h 的测试结果分析

本系统最高可支持 720P, 25 f/s 的双路图像数据输出. 实验中主要实现了 640×480 分辨率的像平面偏移. 图像传感器的像素阵列 2592×1944 , 为了观察较明显的像平面偏移, 使用了较小的 640×480 分辨率. 具体的实现方法为: ①双 CMOS 图像传感器尽量保持平行, 在图像传感器的配置中首先选取像平面中心为传感器物理中心, 采集并获取双目图像. ②保持场景稳定, 摄像头位置不变. 同时改变双 CMOS 图像传感器的配置, 左路的像平面向左平移, 右路的像平面向右偏移. 采集并获取双目图

像. 其结果如图 5 所示.

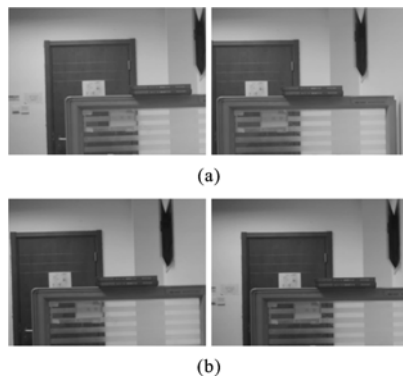


图5 双目采集时实现 h 的实验结果

实验中, 图像传感器之间的基线距离为 65 mm, 保持不变. 图 5(b) 是图 5(a) 实际偏移 256 个像素阵列的结果. 显而易见, (b) 中的图像匹配对可以合成 3D 的信息明显多于 (a). 配置中使用的 Pre-scaling size 大小为 1536×1152 , data output size 大小为 640×480 . 故上图结果是实际图像传感器阵列缩小为原来 5/12 倍的结果. 经过实际计算测试, 图 5 的窗口水平偏移量约为 107 个像素, 换算为实际的图像传感器像素位置差约为 256.8 个像素点. 这与实际结果相符. 经过测试, 本系统利用 640×480 分辨率测试的最大水平像素多达 426 个像素点, 实际的传感器像素偏移个数为 1 024. 因此, h 的灵活可控性得以保证, 而且可变化的范围也非常理想.

在实际应用中, 假定显示端的屏幕大小为 14 寸, 宽高分别为 47 cm, 29 cm. 对于 640×480 分辨率的最佳观看距离^[9] 约为 $7 \times 29 \text{ cm} = 2.03 \text{ m}$. 可计算得最佳的 $Z_{v, \text{near}}$ 和 $Z_{v, \text{far}}$ 分别为 1.35 m 和 3.5 m. 若要把实际场景深度为 0.7 m 到 10 m 的范围进行重建到这个范围内, 需要的 t_c 和 h 为 51 mm, 0.03 mm. 那么图像传感器水平偏移的像素数约为 21. 若等比例重建场景时, h 为 0.061 mm, 约为 44 个像素点. 均在可控的范围之内.

4 结束语

本文通过分析指出, 双摄像头系统的光轴中心相对像平面中心的偏移量 h 如果灵活可调, 3D 系统则有良好的视觉效果. 本文提供了对中心偏移量 h 的控制方式, 实验表明, 通过对中心偏移 h 的控制, 双摄像头拍摄图像可以得到有效调控, 从而对后期 3D 处理效果提供了支持, 3D 摄像机系统将具有更高效的 3D 视频采集性能.

(下转第 55 页)

了进一步提高.

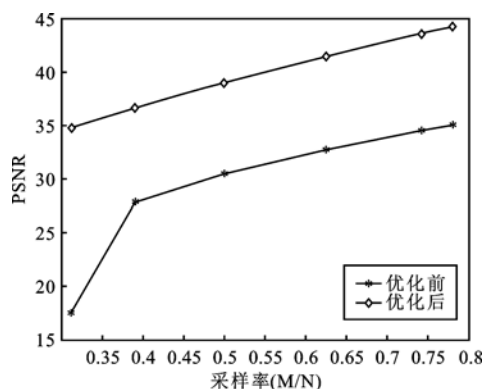


图3 在不同采样率下的两种压缩感知方案的 PSNR 值的比较

5 结束语

提出了一种基于压缩感知的网络编码算法,实现了网络性能的优化与信息传输.并对传统的压缩感知方法进行优化,在保证低网络编解码复杂度和所占内存空间减小的前提下,提高了彩色图像信息重构效果.下一步的研究方向是寻找更优化的压缩感知方法与网络编码相结合,使得网络编解码复杂度更低,重构效果更优,并且尝试使用语音、视频等其它类型的信息进行传输,以满足多种类型信号传输的要求.

参考文献:

- [1] Ahlswede R, Cai N, Li S R, et al. Network information flow[J]. IEEE Transactions on Information Theory, 2000, 46(4):1204-1216.
- [2] 徐光宪,付晓.基于稀疏矩阵的低复杂度安全网络编码算法[J].计算机工程,2012,38(9):55-57.
- [3] Donoho D. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4):1289-1306.
- [4] 宋晓霞,石光明.低冗余的压缩感知观测[J].西安电子科技大学学报:自然科学版,2012,39(4):144-148.
- [5] 张波,刘郁林,张建新.基于最佳联合稀疏表示的分布式压缩感知算法[J].西南科技大学学报,2012,27(2):65-68.
- [6] Baraniuk R. A lecture on compressive sensing [J]. IEEE Signal Processing Magazine, 2007, 24(4):118-121.
- [7] Tropp J, Gilbert A. Signal recovery from random measurements via orthogonal matching pursuit [J]. IEEE Transactions on Information Theory, 2008, 53(12):4655-4666.
- [8] 祁乐.基于压缩感知理论的图像采样和重构[D].大连:大连理工大学,2011:1-53.

作者简介:

钱辰 女,(1989—),硕士研究生.研究方向为图像处理.
田小平 男,(1963—),教授.研究方向为数字图像处理与图像通信.

(上接第51页)

参考文献:

- [1] Andorko I, Corcoran P, Bigioi P. Hardware implementation of a real-time 3D video acquisition system. in Optimization of Electrical and Electronic Equipment (OPTIM) 2010 [C]//12th International Conference on. Romania, 2010.
- [2] 王心焕.基于FPGA的实时立体视觉系统研究[D].浙江:浙江大学,2011.
- [3] Zilly F, Kluger J, Kauff P. Production rules for stereo acquisition[J]. IEEE, 2011, 99(4):590-606.
- [4] 孙延禄.立体电影摄影的基本规律及立体电影摄影机(一)[J].影视技术, 2003(6):41-46.
- [5] Smolic A. Three-dimensional video postproduction and processing[J]. IEEE, 2011, 99(4):607-625.
- [6] 章毓晋.图像工程(下册):图像理解[M].北京:清华大学出版社,2007:107-108.
- [7] Yeh Y, Silverstein L D. Limits of fusion and depth judgment in stereoscopic color displays[J]. Human Factors: The Journal of the Human Factors and Ergonomics Society, 1990, 32(1):45-60.
- [8] OmniVision Corporation. OV5640 datasheet[M]. OmniVision Corporation, Last Modified, 2011.
- [9] 章毓晋.图像工程(上册):图像处理[M].北京:清华大学出版社,2005:34-35.

作者简介:

王伟 男,(1987—),硕士研究生.研究方向3D视频采集与处理.
任屹奕 男,(1987—),博士研究生.研究方向为集成电路设计、图像处理.
谢翔 男,(1971—),副研究员.研究方向为集成电路系统设计、图像处理、无线通信、生物医学电子学等.
李国林 男,(1970—),副教授.研究方向为互连分析以及RFIC设计.