

一种实用的多目立体视觉系统设计方法

A Practical Stereo Vision System with Multiple Cameras

王忠立 裴明恩 刘万春 贾云得

(北京理工大学计算机科学与工程系 北京100081)

Abstract This paper presents an implementation of a practical multi-camera stereo vision system. The system consists of a stereo vision head and a firewire bus module. The stereo head contains multiple CMOS cameras which captures images synchronously. The firewire bus transfers the images to a computer at video rate for post vision tasks.

Keywords Multi-camera stereo vision, CMOS camera, IEEE 1394

1 引言

立体视觉是获取物体三维形状信息的重要方法之一,已经在工业测量、物体建模、视觉导航和物体识别等领域得到了广泛的应用。目前的立体视觉系统通常由多个摄像机、图像采集卡、外同步信号控制以及多路转换器等组成,系统复杂、成本高,操作不方便,限制了立体视觉技术在实际中的应用。本文基于 CMOS 图像传感器和 IEEE 1394 总线接口技术,

探讨了一种实时、低成本、简单可靠的多目立体视觉系统的设计与实现方法。我们认为,随着 CMOS 图像传感器成像质量的不断提高,本文提出的视觉系统设计方法会得到越来越广泛的应用。

2 系统组成

如图1所示多目立体视觉系统主要由两部分组成:多目立体视觉成像头模块与 IEEE 1394 传输模块。

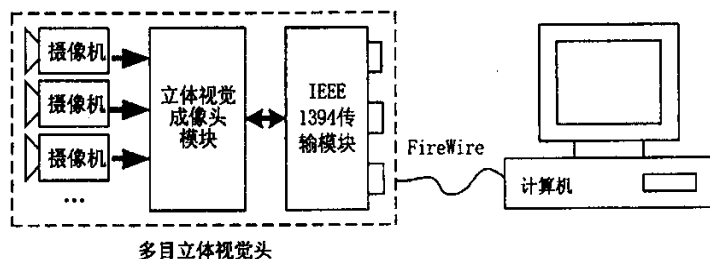


图1 立体视觉系统结构框图

多目立体视觉成像头采用多个 CMOS 图像传感器,同步获取图像(各个摄像机之间信号完全同步),并将图像存储在该模块的缓存中。CMOS 图像传感器与 CCD 相比,具有功耗低、集成度高、速度快、访问灵活、低成本的特点^[1],非常适合于小型化、低成本、低功耗、高集成度的成像应用场合。另外,直接数字图像输出易和后续的数字信号处理电路接口。系统设计时,可以根据不同的视觉任务要求增加或者减少摄像机的个数。

IEEE 1394 传输模块读取多目立体视觉成像头模块中缓存的图像数据,然后通过 IEEE 1394 接口将图像数据实时传送到计算机或其他后端处理设备。由计算机或其他后端处理设备根据视觉任务的不同对多目图像数据进行不同处理。IEEE 1394 高速串行总线支持同步传输、异步传输两种事务传输

模式和点到点(Peer to Peer)的传输,总线上的设备能自主执行事务,而不需要 CPU 的干预,传输速度可以达到 400Mb/s^[2],在不对图像数据进行压缩的条件下,能够实现图像数据的实时传输,还可以传送数据及设备控制指令,非常适合用于多目图像数据的实时传输任务。

3 多目立体视觉成像头设计

3.1 多目立体成像头组成概述

图2为具有多个 CMOS 图像传感器的立体视觉成像头模块组成框图。在本设计中,我们采用了多片 Omnivision 公司的 CMOS 图像传感器 OV7120,和 FPGA、图像数据存储单元(SRAM)、单片机控制器等,构成多目立体视觉的图像采集部分。OV7120 可以以两种分辨率模式输出,VGA 模式下为 640 ×

480, QVGA 模式下为 320×240 , 前者为 30 帧/秒, 后者可达 60 帧/秒。在上述分辨率输出模式下, 还可以通过设置其控制寄存器改变输出窗口大小。OV7120 有逐行扫描和隔行扫描两种模式, 参数设置、状态读

取由 SCCB 总线控制。系统采用了 ATMEL 公司的 AT89C51 单片机(内部集成了 4KB 的闪存), 用两个 IO 口模拟 SCCB 总线时序对 OV7120 内部寄存器进行读写控制。

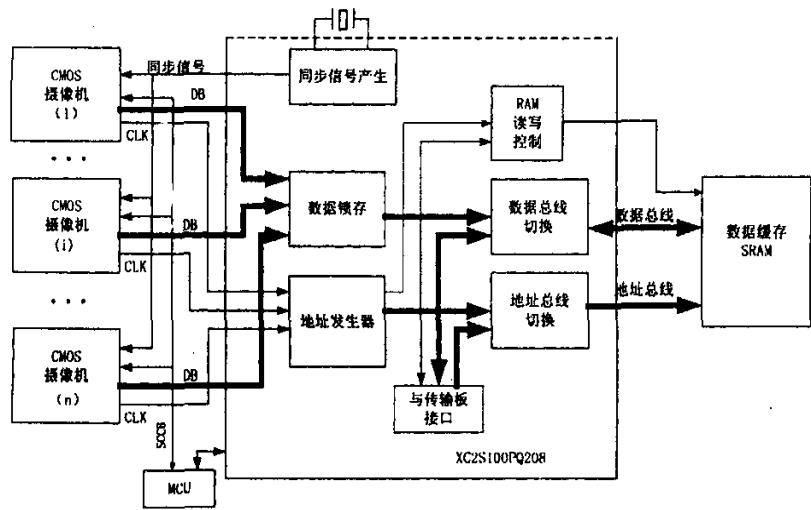


图2 多目立体视觉头组成框图

FPGA 采用的是 Xilinx 公司的 Spartan II 系列的 XC2S100, XC2S100 是一款 10 万门级的 FPGA, 有 2700 个逻辑单元和 600 个可配置逻辑块, 用户可用 I/O 引脚与封装有关, 最高可达 196, 内部有 10 个 4096bit 的 BlockRAM, FPGA 完成摄像机同步信号产生、摄像机输出信号接口、图像数据锁存地址的产生、SRAM 的读写控制逻辑、地址切换、和传输模块

的接口等功能。

3.2 同步采集原理

对于多目立体视觉的图像采集, 确保多个摄像机之间的图像同步非常重要, 尤其是对运动、变形物体的图像采集, 否则会对后续的深度计算和三维恢复有较大影响。图像同步采集是设计中需要重点解决的问题之一。

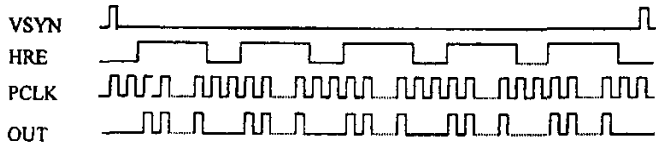


图3 OV7120的工作时序

如图3所示, OV7120 有 3 个关键的同步信号: 帧同步 VSYNC、行同步 HREF、像素时钟 PCLK。为确保多个摄像机的同步, 需要保证各个摄像机的三个信号 (VSYNC、HREF、PCLK) 同步。OV7120 有两种工作模式: 主模式和从模式。工作在从模式下的 OV7120 可以接收工作在主模式下的 OV7120 输出的同步信号, 以协调两者之间的同步, 但是我们实验发现, 工作在主模式下的摄像机与工作在从模式下的摄像机同步信号差别较大, 而工作在从模式下的多个摄像机能很好地保持同步。因此将所有的 OV7120 摄像机都设置为从模式, 采用 FPGA 作为摄像机外同步信号的发生器, 将 FPGA 产生的外同步信号 (帧同步 VSYNC 和行同步 CHSYNC) 输入到多个工作在从模式下的 OV7120 摄像机, 这样就

解决了多个摄像机的同步问题。通过实际分析测试, 多个摄像机的信号完全一致, 达到了图像同步的目的。

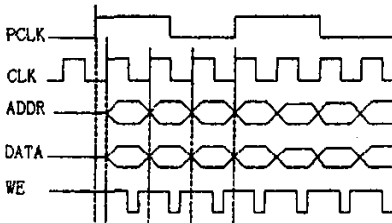
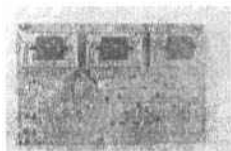


图4 图像数据锁存时序示意图

图像数据锁存、锁存地址发生、读写控制逻辑之间的时序如图4所示。三个摄像机的输出时钟信号是一致的, 可以用其中一个摄像机的输出时钟为参考,



a) 三目立体视觉头



b) 五目立体视觉头

图5 多目立体视觉头实物照片

按图中所示的时序工作。图像缓存采用两片 SRAM, 保证图像采集和传输同时进行。一片 SRAM 正在进行当前图像数据存储时, 另一片 SRAM 所存储的前一帧图像数据可供 IEEE 1394 传输模块读取并传输至 PC 机中。图5为三目立体视

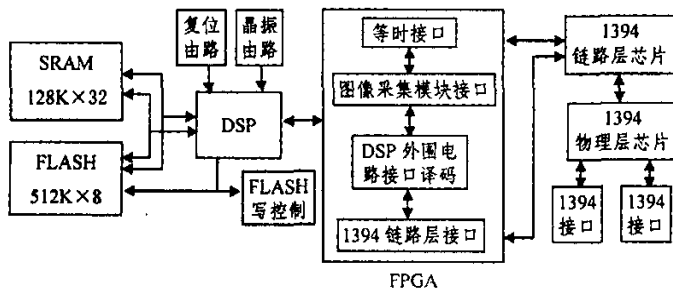


图6 IEEE 1394传输模块结构框图

在数字摄像机中使用的 DCAM 规范 (1394-Based Digital Camera Specification) 是一种针对单个摄像机的 CSR 体系^[4], 包括设备配置信息 (ConfigROM) 的规定。该规范对基于 1394 的数字摄像机的状态读取、控制、数据传输格式等都做了严格规定。但不适合于多个摄像机的应用场合, 我们对它作了相应的修正, 使其适合于多目摄像机的应用场合。

传输模块中, DSP 主要完成以下任务: (1) 异步事务。如 ConfigROM 信息的读取, 对应于摄像机状态及控制用的 CSR 体系的实现; (2) 等时事务控制。当其它节点发出等时传输开始请求时, 启动本节点的等时传输开始, 及相应的控制任务; (3) 串行总线管理。包括循环控制器 (CM)、等时资源管理器 (IRM)、总线管理器 (BM) 的相关功能。



a) IEEE 1394 传输模块实物图片



b) 多目立体视觉系统实物图片

图7 IEEE 1394传输模块和多目立体视觉系统实物图片

基于上述功能, 1394 传输模块可以像单个数字摄像机一样和计算机相连, 通过计算机内的应用程序读取视觉头的状态、控制摄像机的参数、图像的实际采集等, 又可以和其它具有 1394 功能的节点组成小型视觉系统, 应用于嵌入式系统、微小型视觉系统中。

觉头和五目立体视觉头^[3]的实物照片。

4 IEEE 1394传输模块

IEEE 1394 传输模块的硬件部分主要由 DSP 工作电路、IEEE 1394 的物理层 (PHY) 与链路层 (LLC) 接口、DSP 与链路层接口、等时传输控制, 以及与图像采集模块之间的接口等组成。其中, DSP 外围电路接口译码、与 1394 链路层接口、等时传输控制, 以及与图像采集模块之间的接口用一片 Xilinx 的 FPGA XC2S100 来实现。DSP、IEEE 1394 的物理层芯片、链路层芯片均采用 Ti 公司的 TMS320C32、TSB12LV32、TSB41LV03。图6为 IEEE 1394 传输模块的结构框图。

为了完成计算机控制多目视觉头的图像采集等任务, 需要在计算机一端编写视觉头的底层驱动程序和应用层接口程序。在 Windows 2000 操作系统下, 基于微软提供的 OHCI, 采用 DriverStudio 开发了 WDM 模式的 1394 底层驱动, 实现了多目立体视觉硬件系统的 PnP 功能, 以及 1394 的异步、等时传输等通用功能。应用层程序接口负责与用户的交互操作, 将用户发出的读取状态、控制命令传递到驱动程序层, 并发送给立体视觉头, 并将返回结果提供给用户。图像数据的读取主要通过等时传输方式实现, 用户可根据视觉任务对采集的图像作相应的处理和计算。在 200Mb 的带宽下, 实现了 5 幅 320×240 黑白图像以 30 帧/S 的视频速度实时采集。图7为 IEEE 1394 传输模块和三目立体视觉系统的实物图片。

结束语 基于 CMOS 图像传感器与 IEEE 1394 接口构成的多目立体视觉头解决了多个摄像机的图像同步问题, 并且通过 IEEE 1394 串行总线实现了图像数据的实时传输, 为立体视觉系统的应用提供了一个高效率、低成本、小型化的多目图像获取平台。系统具有可扩展性好、通用性好的优点。可广

(下转第 67 页)

根据倒谱参数,判定语音流的有无采用最差相似度准则。它的基本思想是人耳对语音的感知依靠当前和过去声音的比较,根据这个性质,可以求出相邻帧的差异,据此判断语音流的出现。D-J-RASTA-PLP 倒谱有效地提取了语音的参数,可以很方便地根据这些参数计算出相邻帧的差异度。即: $p(C_i) = \frac{1}{5} \sum_{p=1}^5 (C_p - C_p^{-1})^2$, 其中 C_p 为第 i 帧的第 p 阶倒谱。

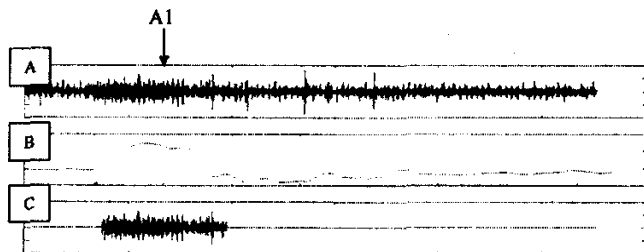


图6 D-J-RASTA-PLP 信噪比较差的情况下语音流的提取

图6表明 A 是 1s 的原始语料,其中包含复杂的背景噪声,仅在 A1 处有微弱的语音流出现。传统的方法很难检测出。但在图 B 中我们可以很清楚地看到有语音流出现是倒谱参数有了明显的变化。为语音流的提取提供了良好的特征。图 C 是根据倒谱参数提取的语音流。

通过以上两个例子,我们可以看到,本文提出的方法可以很好地抑制噪声,与人的听觉特性相一致,当环境中出现语音时,倒谱参数就有相应的变化,正如人耳听到有人说话会不由自主地倾听一样。

讨论 目前在语音信号处理中主要关注的问题之一是如何使在干净环境下表现良好的系统,在恶劣环境下也有突出的表现。我们相信造成系统性能恶化的原因之一是环境中不同频率噪声的影响。为了消除这种影响出现了很多优秀的去噪方法,像小波去噪、谱减法、Kalam 滤波等等。这些方法在一定程度上表现非常优秀。但它们往往忽略了听觉的特性,无论如何听觉在语音的交流中起到了不可替代的作用^[6]。本文对听觉特性给予了足够的重视,并取得了良好的检测效果。但正如所有的事物一样,优秀只是相对的,对每一个系统来说都有它的不足所在,本系统最大的不足就是信噪比驱动下 J 值的确定。只有在 J 值比较合适的情况下,系统才能得到比较满意的效果。在实验中,用自适应的方法对信噪比进

根据研究工作的需要,我们需要对多语种、多信噪比的语音进行检测。图5是一些试验结果。

图5表明 A 是 1s 的原始语料,从中可以看出有很多的脉冲噪声。A1 出为英文 'O', B1 为英文 'K'。图 B 是 D-J-RASTA-PLP 倒谱参数,在相应的脉冲噪声处倒谱参数起伏很小。而真正的语音处,到谱参数却变化的相当有规律,很好地体现了语音的谱包络。图 C 是对语音流的提取。

行了实时估计,取得了不错的效果,但这也付出时空上的代价。期待有一种良好的解决环境信噪比的方法,以使系统性能更优。总之,无论从抗噪性来看还是从应用的角度来看,D-J-RASTA-PLP 有很好的应用前景。

参考文献

- 1 Hermansky H. Perceptual Linear Predictive (PLP) analysis of speech. *Journal of Acoustical Society of America*, 1990, 87(4): 1738~1752
- 2 Hermansky H, Morgan N. RASTA processing of speech. *IEEE Trans. Speech and Audio processing*, 1994, 2(4): 578~589
- 3 Vuurens, Hermansky H. Data-driven design of RASTA-like filter. *Proc. Euro-speech*, Rhodes, Greece, 1997
- 4 Kanedera N, Arai T, Hermansky H, Pavel M. On the importance of various modulation frequencies for speech recognition. In: *Proc. of EURO-SPEECH'97*, Rodos, Greece, 1997
- 5 Hirsh H. Estimation of noise spectrum and its application to SNR estimation and speech enhancement. [Technique report TR-93-012]. *Issi uc Berkeley*, 1993
- 6 Hermansky H. Should recognizer have ears?. *Speech communication*, 1998, 25(1-3): 3~27
- 7 Greenberg S. The ears have it: the auditory basis of speech perception. *Proc. IcphS-95*, 1995, 3: 34~44
- 8 Rabiner L R, Sambur y M R. An Algorithm for Determining the Endpoints of isolated utterance. *The Bell System Technical Journal*, 1975, 54(2)

(上接第57页)

泛应用于桌面视觉、物体建模、视觉导航等领域。

参考文献

- 1 赵文伯, 刘俊刚. CMOS 图像传感器发展现状. *半导体光电*, 20, (1): 11~18

- 2 Don Anderson 著, 姜议龙等译. *FireWire 系统体系*. 中国电力出版社, 2001
- 3 杨云飞. 多目立体视觉图像获取和三维恢复技术. [北京理工大学硕士学位论文]. 2001
- 4 1394-based Digital Camera Specification, Version 1. 20. July 23, 1998, 1394 Trade Association