软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(10):2509−2518 [doi: 10.13328/j.cnki.jos.005073]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

化身 1 号:一个快速的三维彩色全身人像捕捉系统*

程志全[1], 陈 寅[2], 宋 展[3], 吴 彤[4], 王 进[5]

[1](湖南化身科技有限公司,湖南 长沙 410205)

[2](中国人民解放军理工大学 国防工程学院,江苏 南京 210007)

[3](中国科学院 深圳先进技术研究院 智能设计与机器视觉研究室,广东 深圳 518055)

[4](国防科学技术大学 军事高科技培训学院,湖南 长沙 410073)

[5](湖南师范大学 计算机教学部,湖南 长沙 410006)

通讯作者: 程志全, E-mail: cheng.zhiquan@avatarscience.com

摘 要: 提出一个快速的三维彩色全身人像捕捉系统(名为化身 1 号):1.5s 采集数据,1min 内全自动地建立三维全身人像.化身 1 号使用了 8 台自制的结构光扫描仪,采集数据的精度达到 1mm.这 8 个扫描仪预先配置好空间位置:4 个抓拍头部,4 个抓拍身体.从而能够有效地覆盖采集成人对象的全身区域.在离线标定完 8 个扫描仪后,化身 1 号就可以捕获真实对象,在一个坐标系内处理各扫描仪采集的局部数据,历经全局注册和无缝隙的纹理拼接,鲁棒地建立高逼真度的纹理人像模型.该模型的几何精度达到 1mm,并具有照片真实感的色彩信息.实验结果表明,在高逼真度建模需求的前提下,化身 1 号的采集和处理速度均处于当前最优水平.

关键词: 三维彩色全身人像;捕捉系统;结构光;化身 1 号;几何精度

中图法分类号: TP391

中文引用格式: 程志全,陈寅,宋展,吴彤,王进.化身 1 号:一个快速的三维彩色全身人像捕捉系统.软件学报,2016,27(10): 2509−2518. http://www.jos.org.cn/1000-9825/5073.htm

英文引用格式: Cheng ZQ, Chen Y, Song Z, Wu T, Wang J. Avatar 1: A fast 3D color full-body capture system. Ruan Jian Xue Bao/Journal of Software, 2016,27(10):2509−2518 (in Chinese). http://www.jos.org.cn/1000-9825/5073.htm

Avatar 1: A Fast 3D Color Full-Body Capture System

CHENG Zhi-Quan[1], CHEN Yin[2], SONG Zhan[3], WU Tong[4], WANG Jin[5]

[1](Hu'nan Avatar Science Company, Changsha 410205, China)

[2](School of National Defense, PLA University of Science and Technology, Nanjing 210007, China)

[3](Laboratory of Artifical Design and Machine Vision, Shenzhen Institutes of Advanced Technology, The Chinese Academy of Sciences, Shenzhen 518055, China)

[4](College of Continuing Education, National University of Defense Technology, Changsha 410073, China)

[5](Computer Department, Hu'nan Normal University, Changsha 410006, China)

Abstract: This article presents a fast 3D color full-body capture system, named Avatar 1, which can scan a subject in 1.5 seconds and automatically provide the full-body model within 1 minute. It uses eight self-made structure-of-light scanners which are built based on structured lighting technique and have a measuring accuracy of 1mm. The eight scanners are configured to effectively cover the full-body surface of an adult, four scanners are arranged to capture the head, while four target the body. After calibrating the scanners in the offline step, it can capture the real subject and process all partial scanned data in the same coordinate. Through the global registration and

*基金项目: 国家自然科学基金(61272334, 61375041); 江苏省自然科学基金青年科学基金(BK20150723).

Foundation item: National Natural Science Foundation of China (61272334, 61375041); Jiangsu Natural Science Foundation of China Youth Science Fund (BK20150723).

收稿时间: 2016-01-13; 采用时间: 2016-03-29; jos 在线出版时间: 2016-08-08

CNKI 网络优先出版: 2016-08-09 15:38:08, http://www.cnki.net/kcms/detail/11.2560.TP.20160809.1538.013.html

© 中国科学院软件研究所 http://www.jos.org.cn

seamless texture stitching steps, it robustly reconstructs a high-quality full-body textured model, which has 1mm geometric accuracy and photo-realistic color information. The experiments demonstrate that Avatar 1 achieves state-of-the-art scanning and processing speed under high-quality result requirement.

Key words: 3D color full body; capture system; structured lighting; Avatar 1; geometric accuracy

自计算机图形学诞生以来,如何快速建立高逼真的三维彩色全身人像便成为科研人员关注的焦点.导致这一事实出现的原因有两个方面:一方面,三维彩色全身人像具有广阔的应用需求;另一方面,三维彩色全身人像的建模问题极其困难.

迄今为止,业界已经建立了两个代表性的三维彩色全身人像数据库:CEASAR[1]和 FAUST[2].受文献[1,2]的启发,进一步分析 CAESAR 和 FAUST 的数据实例,通过总结,本文发现,难以建立三维彩色全身人像的原因主要包括:

(1) 逼真度.建立的三维彩色全身人像模型具有较高的逼真度.逼真度包括几何和色彩两方面的信息.在几何方面,原始的采集数据应达到毫米级别;在色彩方面,其视觉效果应具有照片真实感.

(2) 鲁棒性.鲁棒地获取衣物和头发等材质,有效地采集多种材质的全身表面.此外,采集的数据能够尽可能地覆盖成人对象的全身区域,数据缺失少.

(3) 速度.速度进一步分为采集和处理两个方面.在采集上,速度达到秒级,用户无需长时间地保持静止状态;在处理过程中,建立三维全身人像的速度快,达到分钟级别,可尽快生成结果.

面向上述 3 个因素,本文提出一个快速的三维彩色全身人像捕捉系统(名为化身 1 号).化身 1 号以 1.5s 的速度采集成人对象的原始数据,1min 内全自动地建立以网格表示的三维彩色全身人像.该人像模型的几何精度达到 1mm,其纹理具有照片真实感的色彩信息.

化身 1 号由 8 台自制的结构光扫描仪所组成:4 个抓拍头部,4 个抓拍身体.综合考虑单个扫描仪的视场(field-of-view)因素,预先合理地配置好 8 台扫描仪的空间位置,使得全体扫描仪的视场合集能够有效地覆盖采集成人对象的全身区域.由于采用了高频二值条纹的结构光图像编码算法,自制的结构光扫描仪具有高鲁棒性,能够有效地采集黑色衣物和头发等材质表面.在各个扫描仪的局部视角下,快速建立 1mm 精度的三维原始点云,并获取了相应百万像素的彩色纹理(图像).

通过多个扫描仪的离线标定,各扫描仪采集的局部数据就能够在一个坐标系内匹配在一起.使用标定信息,化身 1 号捕捉成人对象(身高在 1.4m~2m 范围内),历经处理模型的全局注册和无缝隙的纹理拼接两个步骤,鲁棒地建立高逼真度的三维彩色(附有纹理)全身人像模型.

在实现化身 1 号的过程中,本文的技术贡献主要包括:

(1) 一种改进的高频二值条纹结构光图像编码算法.由于使用高频二值条纹的边缘结构信息计算三维点云,依此算法自建的扫描仪能够快速地采集对象,求解的几何精度达到 1mm.更重要的是,具有高鲁棒性:能够有效地采集黑色衣物和头发等材质.纠正了结构光扫描仪难以采集黑色材质的通常错误概念.

(2) 一种充分使用彩色信息的模型重建方法.在几何重建过程中,提出了一种使用彩色信息的多视角 SLAM 全局注册机制.SLAM(simultaneous localization and mapping)是即时定位与地图构建,在计算机视觉领域,深度相机典型应用 KinectFusion[3]可在定位相机的同时,实时地重建出采集对象的网格模型.本文进一步扩展了 KinectFusion 框架,在 SLAM 机制中,进一步使用彩色信息,以快速实现多视角局部数据的全局注册;在模型的色彩信息处理过程中,通过快速的、无缝隙的纹理拼接,达到照片真实感的视觉效果.

本文第 1 节综述相关的研究工作.第 2 节说明单个结构光扫描仪的实现方案,说明高频二值条纹的结构光图像编码算法.第 3 节阐述化身 1 号的 8 个扫描仪的空间配置、离线的扫描仪标定.第 4 节给出人像捕捉工作,重点说明全局注册和纹理拼接等数据处理步骤.第 5 节给出实验结果及其分析,以说明化身 1 号的逼真度、鲁棒性和速度.第 6 节总结本文工作,并讨论下一步的研究方向.

© 中国科学院软件研究所 http://www.jos.org.cn

1 相关工作

本节简要回顾三维彩色全身人像的相关研究,重点论述高精度的数据采集、高质量的几何重建和纹理拼接相关工作,尤其是那些具有当前最优或曾引领发展的代表性工作.

当前,采集人体数据的最优结果为毫米级别,相关设备包括:激光扫描仪、结构光扫描仪和高清相机阵列.CAESAR[1]全身人像数据库是采用激光扫描仪建立的,采集一个成人对象的速度一般为 17s.FAUST[2]则采用了高造价(百万)的高清相机阵列,能够以 0.02s 的速度采集对象.结构光扫描仪的采集速度快于激光扫描仪,它通过一组光栅,拍照式地快速获取对象表面的数据信息.从而,本文采用自制的结构光扫描仪搭建采集系统.但是,与市面上的常用结构光扫描仪不同,对于自制的结构光扫描仪,本文采用了高频二值条纹的结构光图像编码算法[4],缓解了结构光扫描仪难以采集黑色衣物和头发等材质的局限.更进一步地,本文只采用了 18 幅光栅图像,这与采用 30 幅光栅图像的原有算法[4]相比,采集速度更快.单个扫描仪的采集速度为 0.3s.

由于单个扫描仪的视场只能采集全身的部分区域,故需要从多个视角采集成人对象.从而,多视角的局部数据需要通过全局注册匹配在一起,重建出高质量的几何模型.对于此问题,可借助 SLAM 机制来求解.在开创性的 KinectFusion[3]工作中,在定位各个视角深度相机(扫描仪)的同时,将单视角的局部数据与静态场景实时地匹配在一起,实现全局场景的重建.为了改进 KinectFusion 的性能,研究人员提出了多种改进工作,如可扩展的体积分求解[5]、快速 CPU 求解[6].因而,KinectFusion 能够不在 GPU 加速的条件下,快速地处理大尺寸的静态场景.当前最优的 DynamicFusion[7]工作,能够实时地重建出动态变化的人体对象,将全局注册从刚性匹配提升到非刚性匹配.但是,与 KinectFusion 一样,DynamicFusion 仅使用了深度图像,没有使用彩色图像.本文采纳了快速的 SLAM 机制,并将其扩展,融入了色彩因素,建立的几何模型逼真度更高.

结构光扫描仪采集了单视角的百万像素的彩色图像.这些图像经过纹理映射生成人体表面的纹理细节,建立出照片真实感的模型.但是,多个视角的彩色纹理之间存在重叠区域,需要通过纹理拼接,无缝隙地生成一致的纹理贴图结果.此外,由于受光照等因素的影响,各视角的纹理亮度不一.上述问题,文献[8,9]提出了各自的无缝隙的纹理拼接算法:在网格模型的每个三角形上,选取出最合适视角的彩色图像.该问题本质上是一个离散能量最小化问题,组合优化求解的计算开销大.因此,文献[8,9]的求解时间都在 10min 以上.为此,本文使用了α扩展算法[10],快速地完成无缝隙的纹理拼接的计算.

2 自制的结构光扫描仪

使用高频二值条纹的结构光图像编码算法,本文自制了结构光扫描仪.在阐述完结构光扫描仪的组成结构后,说明改进的高频二值条纹结构光图像编码算法.

2.1 组成结构

自制的结构光扫描仪的基本组成部分(如图 1(a)所示)包括:CCD 相机(PointGrey 工业相机,帧速率为 90Hz,单帧分辨率为 2048×2048)、控制板(FPGA 视频图像处理开发板)和光栅投影仪(选型为明基投影仪,帧速率为 60Hz,分辨率为 1024×768).其中,CCD 相机与光栅投影设备之间的距离约为 50cm,CCD 相机和光栅投影设备的焦点在 2m 处.控制板的主要功能是:将预先存储的数字光栅通过视频接口输出至投影仪,输出每一帧画面的同时,向相机发送触发信号,由相机同步拍摄投影仪所投射的每一帧图像.

2.2 改进的高频二值条纹结构光图像编码算法

与文献[4]工作类似,自制的结构光扫描仪工作机理(如图 1(b)所示)如下:在控制板的控制下,光栅投影仪将编码后的高频(黑白)二值结构光栅条纹投影到被测对象表面上,光栅条纹受到对象表面高度的影响而发生变形,变形后的光栅条纹因此携带了对象表面的三维信息.CCD 相机获取相应的数字图像,经计算机执行相应算法,解码出光栅条纹变形所引起的变化量.从而,计算出被测对象的三维外形点云.

不同于文献[4]中的工作采用 30 幅光栅图像,改进后的高频二值条纹结构光图像编码算法仅采用了 18 幅光

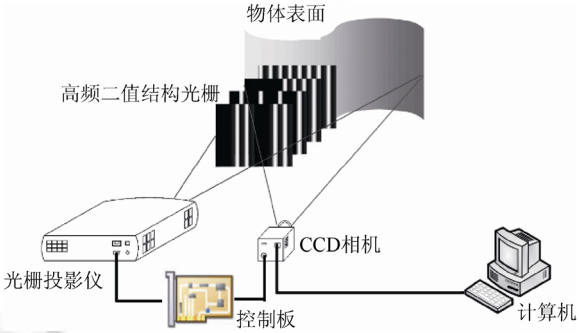
© 中国科学院软件研究所 http://www.jos.org.cn

栅图像.改进的原理在于:在原有的30幅光栅图像中,有12幅具有冗余性,对于三维点云的精度影响不大.所选用投影仪的帧速率为 60Hz,自制的结构光扫描仪投影了 18 幅光栅图像.因此,单个结构光扫描仪的采集时间为18/60=0.3s.

高频二值条纹结构光图像编码算法采用高频二值条纹的边缘特征,将编码特征的定位精度提高到了亚像素级别,因而其扫描结果具有高精度.

此外,在图像解码阶段,只需判别图像中投射黑白光栅条纹的边缘特征信息,而无需知道精确的图像灰度值.因此,对于光照、材质及颜色的影响具有较好的鲁棒性.这是自制的结构光扫描仪能够采集黑色衣物和头发等材质表面的根本原因.

(a) 硬件结构:按照从上至下的顺序,物理框架连接了 CCD 相机、控制板和投影仪 3 个组件



(b) 工作机理

Fig.1 The self-made structure-of-light scanner

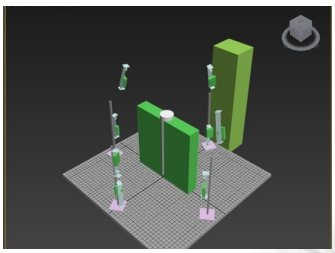
图 1 自制的结构光投影仪

3 化身 1 号的配置

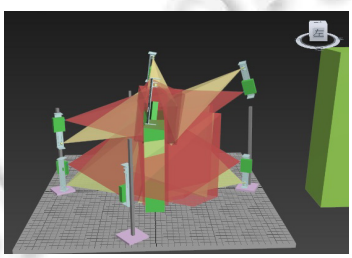
本文使用了 8 台扫描仪搭建了化身 1 号,在说明多台扫描仪的空间配置后,阐述多台扫描仪的离线标定.

3.1 多台扫描仪的空间配置

单个扫描仪的视场采集范围有限,故需要从多个视角采集成人对象,使得全体扫描仪的视场合集能够有效地覆盖采集成人对象的全身区域.如图 2 所示,本文采用 8 个结构光扫描仪,按照计算出的空间配置,采集的数据能够尽可能地覆盖成人对象的全身区域,具有较少的数据缺失.



(a) 8 台扫描仪的空间分布和简化的全身几何体



(b) 8 台扫描仪的视场并集,可有效覆盖全身几何体

Fig.2 The space configuration of 8 scanners

图 2 8 台扫描仪空间配置

在采集过程中,本文约定人体处于 A 姿态.成人的正常身高在 1.4m~2m 之间.从而,人体占用的空间范围可简化为躯体和头部两个子块组成的全身几何体(如图 2 坐标系中心的几何体所示).躯体部分简化为(绿色)长方

© 中国科学院软件研究所 http://www.jos.org.cn

体,长度为 1.4m,宽度为 0.6m,高度为 1.5m;头部简化为(白色)圆柱体,半径为 0.35m,高度为 1m.

单台扫描仪的视野范围为:水平视角 45°,垂直视角 58°.在 2m 的距离内,单台扫描仪可扫描的幅面为:宽度 1.6m,高度 2.2m.

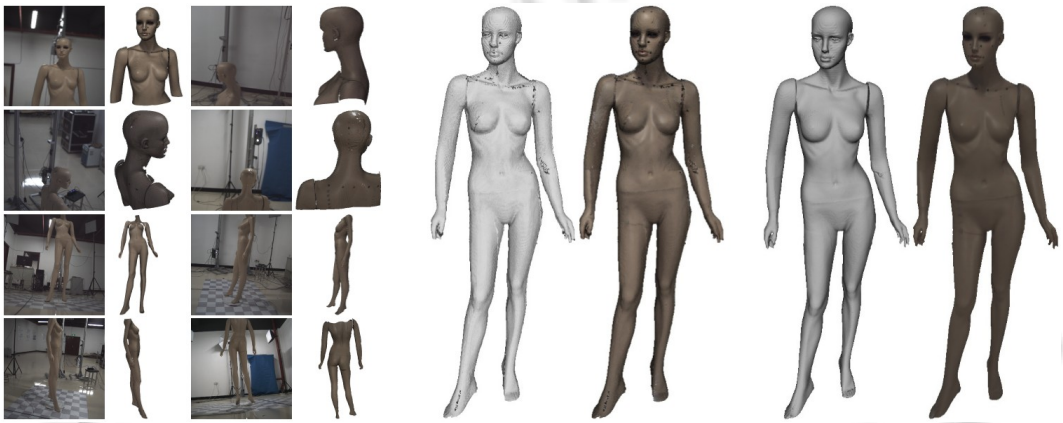
4 台扫描仪采集头部:1 台从正面采集人脸,3 台从侧面和后面的斜上位置采集头部,关注头发区域.4 台扫描仪从前后左右 4 个方向采集躯体.从而,8 台扫描仪的视场并集能够有效地覆盖 1.4m~2m 身高的成人空间范围.

3.2 离线标定

对于多台扫描仪,本文使用课题组既有工作[11]的离线标定方法,实现 8 台扫描仪的标定.计算扫描仪之间的标定矩阵,将各个扫描仪采集的三维点云数据转换到同一个参考坐标系中.

4 真实对象的捕获

下面,将以静态模特为实例,说明化身 1 号完成真实对象捕捉的过程:数据的快速采集、预处理、全局注册和纹理拼接.对于模特实例,图 3(a)显示了 8 幅原始彩色图像和 8 个优化后的三维彩色点云,图 3(b)显示了全局注册后的无色彩和有色彩的点云匹配结果,图 3(c)显示了重建后的几何模型(不带纹理)和最终的照片真实感全身人像结果(带纹理).



(a) 原始的彩色图像和预处理后的三维点云 (b) 不带色彩和带色彩的匹配点云 (c) 不带纹理和带纹理的重建结果

Fig.3 A mannequin

图 3 模特实例

4.1 数据的快速采集

在采集之前,8 台扫描仪的 CCD 相机均预先设定为彩色模式.采集开始后,按照以下方式工作.

(1) 以 1/90s 的速度,8 台扫描仪同时各自采集一张彩色图像;

(2) CCD 相机切换到黑白模式,该切换过程耗时 0.25s;

(3) 在避免扫描仪投影光栅图像的互干扰前提下,8 台扫描仪以两台为一组的顺序采集,抓拍 18 幅预设的高频二值光栅图像.由于真实人体对象是不透明的,能够遮挡上下相对的两台扫描仪,避免了其高频二值光栅图像的重叠.因此,上下相对的两台扫描仪同时工作.从而,黑白模式下采集的时间为 0.3s×8/2=1.2s.

(4) 在采集完毕后,CCD 相机再切换回彩色模式.

累加(1)(2)和(3)的时间,化身 1 号采集真实对象的总时间小于 1.5s.

4.2 数据的预处理

首先,对于每台扫描仪采集的高频二值光栅图像,使用第 2.2 节中描述的高频二值条纹结构光图像编码算法,求解出相应的三维点云.需要指出的是,由于结构光扫描仪使用了同一台 CCD 相机抓拍高频二值光栅图像

© 中国科学院软件研究所 http://www.jos.org.cn

和彩色图像,故很容易求解彩色图像与三维点云之间的映射关系.此外,对于单个扫描仪的三维点云,也可以表示为一幅深度图像.上述求解属于业内常用技术手段,不作详述.

其次,采用深度图像滤波方法[12],对各个三维点云进行增强.具体包括边界区域检测、外点去除和空间滤波3个步骤.本文采用Sobel算子来检测深度图像上与周围像素点存在较大深度差异的像素,并加以去除,这些像素点一般位于人体的边界上.

最后,采用课题组既有工作[11]的地面检测方法,自动删除地面数据.

4.3 全局注册

在几何重建过程中,本文扩展了 SLAM 机制,使用了彩色信息,将离线标定求取的标定信息作为初始值,快速实现多视角局部数据的全局注册.

对于各台扫描仪,获取了其视角下耦合在一起的深度-彩色图像.在齐次坐标系下,一个三维顶点可表示为 $\mathbf{p}=(X,Y,Z,1)^{\mathrm{T}}$. 使用逆向投影方程 π^{-1},使用像素 $\mathbf{x}=(x,y)^{\mathrm{T}}$ 和深度测量 $Z=Z(\mathbf{x})$,该顶点可表示为

$$\mathbf{p}=\pi^{-1}(\mathbf{x},Z)=\left(\frac{x-o_x}{f_x}Z,\frac{x-o_y}{f_y}Z,Z,1\right)^{\mathrm{T}} \tag{1}$$

其中, f_x 和 f_y 为相机的焦距, o_x 和 o_y 是相机的中心坐标.这些参数是相机的内参,是已知的.

那么,使用投影方程 π,该顶点对应的像素可表示为

$$\mathbf{x}=\pi(\mathbf{p})=\left(\frac{Xf_x}{Z}+o_x,\frac{Xf_y}{Z}+o_y\right)^{\mathrm{T}} \tag{2}$$

对于两个相邻的深度-彩色图像 I_1 和 I_2 上的共有顶点,利用投影方程 π 和相邻扫描仪间的标定矩阵 $\mathbf{A}$,可推导出 warping 函数 τ,基于 I_1 上的像素 $\mathbf{x}$ 和 $\mathbf{A}$,计算出 I_2 上的像素坐标 $\mathbf{x}'$.

$$\mathbf{x}'=\tau(\mathbf{x},\mathbf{A})=\pi_2\left(\mathbf{A}\pi_1^{-1}\left(\mathbf{x},Z_1(\mathbf{x})\right)\right)^{\mathrm{T}} \tag{3}$$

使用 warping 函数 τ,文献[6]中的工作为像素 $\mathbf{x}$ 定义了深度测量误差 r_Z.

$$r_Z=Z_2\left(\tau(\mathbf{x},\mathbf{A})\right)-Z_1(\mathbf{x}) \tag{4}$$

更进一步地,可表示为

$$r_Z=Z_2\left(\tau(\mathbf{x},\mathbf{A})\right)-\left[\mathbf{A}\pi_1^{-1}\left(\mathbf{x},Z_1(\mathbf{x})\right)\right]_Z \tag{5}$$

其中, $[\bullet]_Z$ 为一个顶点的深度值,即 $[\mathbf{p}]_Z=Z$.

通过如下的加权最小二乘方程,文献[6]中的工作求解优化的标定矩阵 $\mathbf{A}^*$ 为

$$\mathbf{A}^*=\arg\min_{\mathbf{A}}\sum_{i=1}^{N}w_i r_{Z,i}^2 \tag{6}$$

其中, w 为权重, N 为深度-彩色图像的数目,本文中 $N=8$. 文献[6]中的工作给出了上式的 CPU 实时求解方法.

除了考虑深度因素 r_Z ,本文又进一步考虑了彩色因素 r_C .将双元变量 $\mathbf{r}=(r_Z,r_C)$ 引入到加权最小二乘方程中.

$$\mathbf{A}^*=\arg\min_{\mathbf{A}}\sum_{i=1}^{N}w_i\mathbf{r}_i^2 \tag{7}$$

因此,与文献[6]中的工作类似,使用一阶泰勒展开,将上式线性化.进而,将离线标定求取的标定信息作为初始值,采用迭代的求解方法,求解出最终的标定信息,完成全局注册,快速建立以三角形网格表示的几何模型.

4.4 纹理拼接

建立的三角形网格模型,需要将多视角采集的彩色图像映射到它的表面上,无缝隙地拼接,生成一致的纹理映射结果.在上文中,已经计算了彩色图像与三角形网格之间的投影映射关系.基于此,本文将采用离散标签分配(discrete label assignment)的表示方式,通过能量最小化方程的求解,完成无缝隙的纹理拼接.

对于三角形网格模型 $\mathbf{M}$ 上的一个面片 f_j ($j=1,\ldots,F$, F 为面片的总数),分配了一个标签 l_j ,该标签指明了

© 中国科学院软件研究所 http://www.jos.org.cn

多视角彩色图像中的相关单一图像 I_i.本文共有 8 个视角下的彩色图像,故 $i=1,...,N, N=8$. 从而,多视角彩色图像的纹理拼接可形式化地表示为

$$\arg\min_{l_1,...,l_F}\sum_{j=1}^{F}E_{data}(l_j)+\lambda\sum_{\{f_j,f_k\}\in M,f_j\in B(f_k)}E_{smooth}(l_j,l_k) \tag{8}$$

其中,λ为数据项 E_{data} 和光滑项 E_{smooth} 之间的权重,经验设定为λ=10. $B(f_k)$ 是面片 f_k 的相邻面片集合.从而,$\{f_j,f_k\}\in M, f_j\in B(f_k)$ 的含义为 f_j 和 f_k 是相邻面片.

$$E_{data}=\sin^2\theta \tag{9}$$

在数据项 E_{data} 中,对于每个视角的彩色图像,θ为该视角投影仪的相机视线方向与三角形法向的夹角.从而,当夹角θ越小时,数据项的值越小,即该视角的彩色图像具有更优的纹理映射结果.

$$E_{smooth}(l_j,l_k)=\int_{e(j,k)}\left\|I_{l_j}\left(\phi_{l_j}(p)\right)-I_{l_k}\left(\phi_{l_k}(p)\right)\right\|^2 \tag{10}$$

光滑项 E_{smooth} 测量了相邻面片 f_j 和 f_k 的共享边 $e(j,k)$ 上的色彩误差积分.其中,ϕ_{l_j} 是将面片 f_j 映射到彩色图像 I_{l_j} 的函数.当相邻面片具有同一个分配标签时,光滑项的值为 0.

式(8)是一个典型的离散能量最小化方程,具有多种解法.文献[9]中的工作采用了加权树的消息传递解法(tree-reweighted message passing),收敛速度慢.为了加速式(8)的求解过程,本文采用了更快的α扩展算法[10],该算法可快速收敛,快速求解出最优解.

在纹理拼接后,本文采用图像的泊松融合算法[13],缓解各个视角彩色图像间的色彩强度差异,实现纹理图像间光照色彩的平滑过渡.并进一步使用该算法[13],对于因自遮挡导致的死角面片,通过色彩填充,新建相应的映射纹理.从而,实现了照片真实感的纹理映射效果.

5 实验结果

图 4 显示了 3 个具有代表性的成人实例结果.本节对其展开分析,说明化身 1 号的逼真度、鲁棒性和速度.

Fig.4 3D full-body results for typical adults. In each case, left: reconstructed geometric mesh, right: textured model. Personal details, such as faces, cloth wrinkles, and hairstyle, are well captured

图 4 典型成人对象的三维全身人像实例.在每个实例中,左侧为重建的几何模型,右侧为纹理模型.本系统能够很好地捕捉个性化的细节:如脸部、衣服褶皱和发型

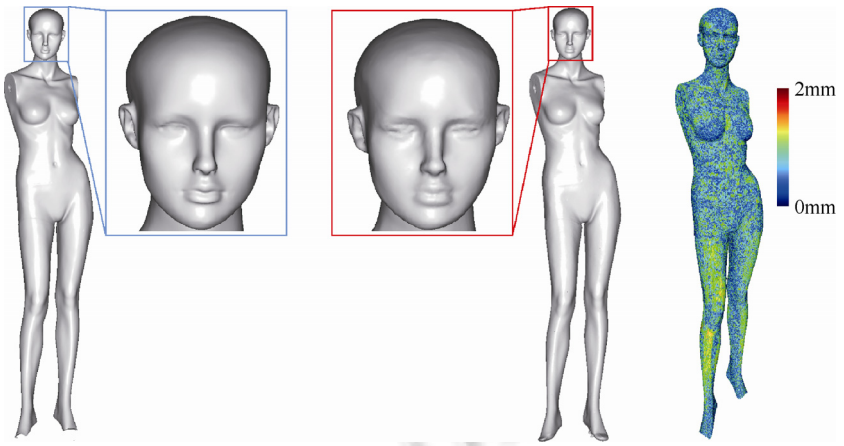
(1) 逼真度

三维彩色全身人像的逼真度包括几何和色彩两个方面.在几何方面,对于自制的结构光扫描仪,其采集精度为 1mm.依此建立的几何模型,也达到了 1mm 级别;在色彩方面,将相机抓拍的百万像素彩色图像作为纹理,从而,对于具有纹理贴图的人像模型,其视觉效果具有照片真实感.

为了进一步测量几何逼真度,本文在几何精度上做了对比实验(如图 5 所示).对于静态模特(没有手臂),预先使用了当前商用最高精度的结构光扫描仪建立了三维模板(如图 5(a)所示),此模板的几何精度达到了 0.1mm.本

© 中国科学院软件研究所 http://www.jos.org.cn

系统捕捉该静态模特,重建相应的模型(如图 5(b)所示).接下来,将重建模型与三维模板进行了以 L2 距离为标准的误差对比(用伪彩色图表示,如图 5(c)所示).实验结果表明,与模板相对比,本系统重建后的 L2 最大几何误差为 2mm,平均几何误差为 0.6mm.从而,说明了本系统的重建结果处于当前最优水平,即达到了 1mm 的几何精度.



(a) 使用高精度扫描仪建立的三维模板,几何精度为 0.1mm (b) 本文的重建模型 (c) 用伪彩色图表示的 L2 距离误差对比,平均几何误差为 0.6mm

Fig.5 The geometric accuracy comparison

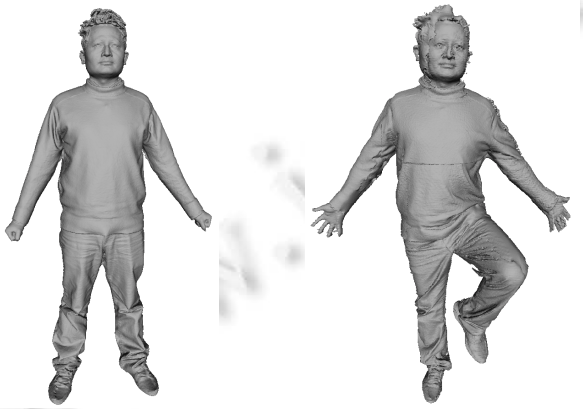
图 5 静态模特的几何精度对比

(2) 鲁棒性

由于采用了高频二值条纹结构光图像编码算法,自制的结构光扫描仪能够鲁棒地获取黑色衣物和头发等材质.从而,化身 1 号对光照、材质及颜色的影响具有较好的鲁棒性.

在化身 1 号中,8 台扫描仪的视场并集能够有效地覆盖 1.4m~2m 身高的成人空间范围,采集的成人对象全身数据缺失少.该有效覆盖一方面为高逼真度结果提供了保证,另一方面也提高了系统采集的鲁棒性.

由于化身 1 号的数据采集时间仅为 1.5s,成人对象能够尽可能地保持静止状态.从而,全局注册算法能够鲁棒地匹配各个视角的局部数据,进而建立高逼真度的结果.但是,如果在 1.5s 的采集过程中,成人对象发生动作变形,重建的结果会有失真.对于一个成人对象,图 6 显示了在其尽量保持静止与发生中等尺度变形下的结果.中等尺度变形时,该成人对象的躯体、头部和四肢均发生了明显的变形.实验结果表明,此时的重建结果出现了明显的拼缝.当变形尺度更大时,不难想象,失真度会更高.



(a) 尽可能静止状态下的结果 (b) 中等尺度变形下的结果

Fig.6 The effect of body deformation during the scanning

图 6 采集过程中人体变形因素的影响

© 中国科学院软件研究所 http://www.jos.org.cn

(3) 速度

化身 1 号运行在 Dell T7600 图形工作站上.表 1 描述了化身 1 号采集和处理各子步骤(预处理、全局注册和纹理拼接)的典型执行时间(单位:s).需要指出的是,各子步骤均未使用 GPU 加速.整体而言,对于采集的原始数据,处理的全部时间未超过 1min.迄今为止,这是有关报道的当前最快速度.

Table 1 The typical performing time of each step (s)

表 1 各步骤的典型执行时间 (s)

数据采集	预处理	全局注册	纹理拼接
1.5	26	12	17

6 结 论

本文提出了一个名为化身 1 号的三维彩色全身人像捕捉系统.本系统能够以 1.5s 的速度采集成人对象,1min 内建立高逼真的三维人像模型.该模型的几何精度达到 1mm,并借助纹理映射,达到照片真实感的视觉效果.本系统具有高鲁棒性,使用自制的结构光扫描仪,鲁棒地获取黑色衣物和头发等材质.

本系统还有几个待改进的地方.首先,由于自制的结构光扫描仪采用了白光,用户会因光亮过强而有短暂的不舒适感.其次,对于高反射的亮光皮鞋,自制的结构光扫描仪仍然难以鲁棒地建立高逼真度的三维点云.最后,在几何重建过程中,全局注册的前提是采集成人对象是保持静止的.当采集对象在采集期间发生非刚性变形时,会影响最终的几何重建结果.为此,需要采用非刚性的 SLAM 机制[7]以提高系统处理非刚性变形的能力.

References:

[1] Robinette K, Dannen H, Paquet E. The CAESAR project: A 3-D surface anthropometry survey. In: Webber B, ed. Proc. of the 2nd Int'l Conf. on 3D Digital Imaging and Modeling. IEEE Computer Society, 1999. 380–386. [doi: 10.1109/IM.1999.805368]

[2] Bogo F, Romero J, Loper M, Black MJ. FAUST: Dataset and evaluation for 3D mesh registration. In: O'Conner L, ed. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Columbus: IEEE Computer Society, 2014. 3794–3801. [doi: 10.1109/CVPR.2014.491]

[3] Izadi S, Kim D, Hilliges O, Molyneaux D, Newcombe R, Kohli P, Shotton J, Hodges S, Freeman D, Davison A, Fitzgibbon A. Kinectfusion: Real-Time 3D reconstruction and interaction using a moving depth camera. In: Agrawala M, ed. Proc. of the ACM Symp. on User Interface Software and Technology. New York: ACM, 2011. 559–568. [doi: 10.1145/2047196.2047270]

[4] Song Z, Chung R, Zhang X. An accurate and robust strip-edge based structured light means for shiny surface micro-measurement in 3D. IEEE Trans. on Industrial Electronics, 2013,60(3):1023–1032. [doi: 10.1109/TIE.2012.2188875]

[5] Chen J, Bautembach D, Izadi S. Scalable real-time volumetric surface reconstruction. ACM Trans. on Graphics (SIGGRAPH Issue), 2013,32(4):No.113. [doi: 10.1145/2461912.2461940]

[6] Steinbrucker F, Sturm J, Cremers D. Volumetric 3D mapping in real-time on a CPU. In: Hamel W, ed. Proc. of the Int'l Conf. on Robotics and Automation. IEEE Computer Society, 2014. 2021–2028. [doi: 10.1109/ICRA.2014.6907127]

[7] Newcombe RA, Fox D, Seitz SM. DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time. In: Grauman K, ed. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Columbus: IEEE Computer Society, 2015. 343–352. [doi: 10.1109/CVPR.2015.7298631]

[8] Lempitsky VS, Ivanov DV. Seamless mosaicing of image based texture maps. In: Kanade T, ed. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Columbus: IEEE Computer Society, 2007. 101–106. [doi: 10.1109/CVPR.2007.383078]

[9] Gal R, Wexler Y, Ofek E, Hoppe H, Cohen-Or D. Seamless montage for texturing models. Computer Graphics Forum, 2010,29(2): 479–486. [doi: 10.1111/j.1467-8659.2009.01617.x]

[10] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001,23(11):1222–1239. [doi: 10.1109/34.969114]

[11] Chen Y, Dang G, Cheng Z-Q, Xu K. Fast capture of personalized avatar using two kinects. Journal of Manufacturing Systems, 2014, 33(1):233–240. [doi: 10.1016/j.jmsy.2013.11.005]

© 中国科学院软件研究所 http://www.jos.org.cn

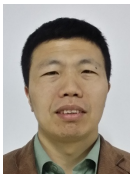
[12] Richardt C, Stoll C, Dodgson NA, Seidel H-P, Theobalt C. Coherent spatiotemporal filtering, upsampling and rendering of RGBZ videos. Computer Graphics Forum, 2012,31(2):247–256. [doi: 10.1111/j.1467-8659.2012.03003.x]

[13] Perez P, Gangnet M, Blake A. Poisson image editing. ACM Trans. on Graphics, 2003,22(3):313–318. [doi: 10.1145/882262.882269]

程志全(1977－),男,黑龙江五常人,博士,高级工程师,CCF 会员,主要研究领域为计算机图形学.

陈寅(1986－),男,博士,讲师,主要研究领域为计算机图形学.

宋展(1978－),男,博士,教授,博士生导师,主要研究领域为计算机视觉.



吴彤(1979－),男,博士,副教授,主要研究领域为计算机应用.

王进(1977－),男,博士,高级工程师,主要研究领域为计算机应用.

www.jos.org.cn

www.jos.org.cn