

1 秒采集的三维头像摄影系统

李方召¹⁾, 金士尧¹⁾, 林 帅¹⁾, 陈 寅¹⁾, 李 宝¹⁾, 程志全^{2)*}

¹⁾(国防科学技术大学计算机学院 长沙 410073)

²⁾(湖南化身科技有限公司 长沙 410013)

(cheng.zhiquan@gmail.com)

摘 要: 为了实现高精度、快速度、低成本的三维头像建模, 提出一个人体头部的三维摄影系统. 该系统能够 1 s 采集用户的三维信息, 3 min 内建立照片真实感的三维模型. 通过配置多目 RGB-Depth 相机摄影系统, 保证 1 s 内完成对人体头部的全方位数据采集, 进而对所得到的点云数据进行数据增强、配准、表面重建和着色, 最终生成高逼真的三维头像模型.

关键词: 三维头像; 1 s 采集; 高逼真度

中图法分类号: TP391

“3D Head Maker” Captured in a Second

Li Fangzhao¹⁾, Jin Shiyao¹⁾, Lin Shuai¹⁾, Chen Yin¹⁾, Li Bao¹⁾, and Cheng Zhiquan^{2)*}

¹⁾(School of Computer, National University of Defense Technology, Changsha 410073)

²⁾(Hunan Avatar Science Company, Changsha 410013)

Abstract: In order to achieve high accuracy, fast, low-cost 3D head modeling, in this paper, we proposed a 3D head maker system, which could capture the subject in a second, reconstruct the 3D model with photo-realistic quality in 3 minutes. The system is composed of several RGB-depth cameras, configured together to cover the subject in the full field, to capture the 3D raw data in a second. The data is further processed with the consolidation, matching, surface reconstruction and coloring steps, the high-quality 3D head model is built at the end.

Key words: 3D head; a second capture; high quality

头像是具有较大头部的摄影作品, 即肩部以上的人像. 常见的实例为 1 寸照、2 寸照、护照照片和大头贴等. 由于众多应用领域的需求驱动, 自 20 世纪 70 年代至今, 三维头像的建模成为计算机图形学领域的研究热点^[1]. 由于人体头部复杂的组成结构与物理属性, 三维头像建模一直是最富有挑战性的课题之一.

随着三维扫描技术与计算能力的飞速发展,

三维扫描的头像建模方法受到业界的广泛关注. 该方法建立的三维模型精度高, 操作相对简单, 应用范围广. 但当前三维扫描系统的一个典型问题是扫描时间过长, 对于人这种易于变化的非刚性对象而言, 轻微的动作会导致扫描失真, 因此, 如何提高头像采集的速度是非常重要的问题.

本文基于新型 RGB-Depth 相机, 通过多目相机的合理空间配置, 保证相机拍摄范围完全、合理

收稿日期: 2014-04-30; 修回日期: 2014-10-10. 基金项目: 国家自然科学基金(61103084, 61272334, 61202334). 李方召(1989—), 男, 硕士研究生, 主要研究方向为计算机图形学; 金士尧(1937—), 男, 教授, 博士生导师, 主要研究方向为计算机仿真; 林 帅(1987—), 男, 博士研究生, 主要研究方向为计算机图形学; 陈 寅(1986—), 男, 博士研究生, 主要研究方向为计算机图形学; 李 宝(1982—), 男, 博士, 讲师, 主要研究方向为计算机图形学; 程志全(1977—), 男, 博士, 研究员, 论文通讯作者, 主要研究方向为可视计算、计算机图形学、虚拟现实与可视化.

覆盖人体头部空间, 搭建 1 s 内完成头像采集的三维摄影系统, 从而对数据进行处理, 最终生成照片真实感的三维头部模型。

1 相关工作

三维头像的建模是计算机图形学领域的经典问题。先驱工作可追溯到 20 世纪 70 年代的 Parke 的参数化手工建模方法^[1]。近年来, 随着激光扫描仪、结构光和深度相机等新型设备的出现, 自动、快速地获取人体三维数据成为可能, 并迅速成为获取头像数据的重要硬件平台, 科研人员正在广泛开展三维头像获取生成的技术研究。

使用激光扫描仪, 构建全身采集系统的代表性工作为 Cyberware 的 full head scanner 系统^[2]。其工作过程如下: 激光扫描仪从上到下扫描, 通过激光测量点来求得物体的表面特征。优点是精度特别高(0.01 mm), 但速度慢(10 s)、造价高(市面价格人民币 50 万元)。

InSpec 公司基于结构光三维扫描仪的人体建模, 根据空间覆盖原理, 采用合理的扫描仪配置, 从多个角度进行抓拍, 有效覆盖采集对象的表面。由于结构光三维扫描仪的单帧采集具有一定的视角范围, 这种抓拍式的工作方式, 使得多个结构光扫描仪组成的采集系统速度适中(3 s)、造价中等偏上(人民币 20 万元)、扫描的精度级别为 0.1 mm。

以 Captured Dimension 公司为代表, 多个高清 RGB 相机组成了近景照相机系统, 该系统利用近景摄影测量原理, 采用多台 RGB 数码相机组成相机阵列, 从不同角度进行拍摄, 可以在一瞬间捕捉到人头的三维模型。该系统的特色是多台相机(至少 20 台), 全方位瞬间成像, 三维成像的精度高。但是, 因为必须使用多台相机同步工作, 故系统的造价高; 此外, 系统结构复杂, 搭建困难。

低成本的 RGB-Depth 相机的头像采集系统为 KinectFusion^[3], 其绕着扫描对象移动该设备, 根据需要尽可能多地扫描对象, 捕获完整的三维信息, 重建出高质量的三维模型。与其他建模系统相比, 该系统价格低廉、使用便捷。但是, 这种便捷性是以时间来换取的, 也就是说, 数据获取过程中要求人体对象保持长时间的静止。例如, 采集完一个普通的头像, 一般需要 3 min。显而易见, 这种要求对采集的人体对象, 是十分苛刻和不合理的。

借助于多台 Kinect, 科研人员已尝试采集三

维人体信息。例如, Tong 等^[4]借助 3 台 Kinect 和转台搭建了人体的采集系统, 人站立于转台上, 经过 30 s 的采集时间后建立相应的三维模型。本文工作的目标在于三维头像的采集与建立, 与文献[4]关注的全身是很好的补充。

通过以上论述不难发现, 采集设备直接决定了静态人体采集生成系统的建模精度、采集时间和造价。但是, 在保证建模精度的前提下, 一款能够广泛应用的三维头像系统必然是尽可能降低采集时间和造价。以此事实为基础, 本文使用低成本的 RGB-Depth 相机构建多目相机的摄影系统, 1 s 内完成采集, 重建出高质量的三维头像模型。

2 多目相机的空间配置

2.1 RGB-Depth 相机的精度分析

本文使用华硕的 Xtion Pro Live 相机, 与 Kinect 一样, 它们都是新型的 RGB-Depth 相机。关于 RGB-Depth 相机的深度值计算, 其精度分析详见文献[5]。文献[5]中, 几何误差 δ 可大致表示为

$$\delta = 2.85 \times 10^{-6} d^2 \quad (1)$$

其中 d 为相机中心点到采集顶点间的距离。因此在采集过程中, 相机与采集对象间的距离要足够近, 以满足采集精度。

2.2 多目 RGB-Depth 相机配置

建立多视角多 RGB-Depth 相机的系统环境来覆盖被拍摄对象空间是一种典型的集合覆盖问题, 集合覆盖问题是一种典型的 NP-hard 问题。具体到本文中的集合覆盖问题中, 考虑到一般人头的特征, 基本属于球型, 其空间范围为半径为 0.15 m 的球体 C 。

针对上述假设, 可将多视角多 RGB-Depth 相机系统环境的数学模型描述为: 在能够接受的上限误差 δ 小于给定阈值 μ 的前提下, 有效地配置 N 个 RGB-Depth 相机, 使其视角范围 F 的并集能够全部有效地覆盖对象活动空间范围空间的球体 $S(C)$, 在以上条件下得到最小值 N 和每个 RGB-Depth 相机的配置位置及角度。其形式化数学定义为

$$\begin{cases} \arg \min_N \left(U_{n=1}^N F(n) \supseteq S(C) \right) \\ \text{s.t. } \delta < \mu \end{cases} \quad (2)$$

经过严密的理论计算和反复的实践验证, 本文选择使用 5 台 RGB-Depth 相机的配置, 在保证精度 $\mu = 1$ mm 的前提下对头部全覆盖。具体配置

方案如图 1~2 所示.

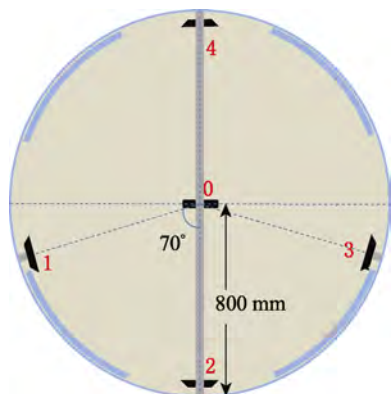


图 1 摄影系统配置俯视图

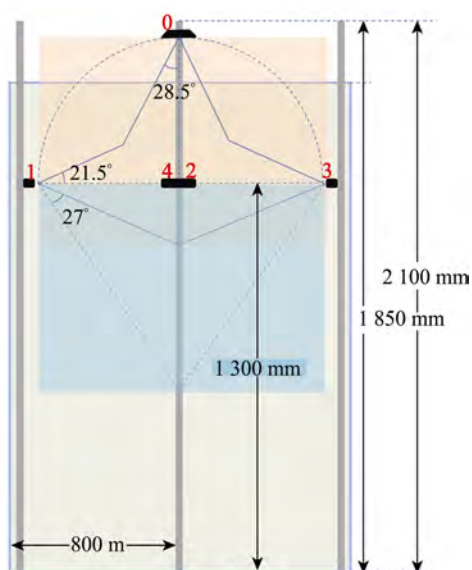


图 2 摄影系统配置正视图

图 1 所示为摄影系统的俯视图, 显示水平 5 个 RGB-Depth 相机的具体位置及其拍摄范围. 如上文所述, 人体头部区域可视为一个直径 30 cm 的球体空间. 根据理论计算, 3 个 RGB-Depth 相机就能够全覆盖人头部的四周, 但是实验过程中我们发现, 由于 3 个相机的重叠区域较少, 处理后的数据精度十分不理想; 首先, 在人头部摄影过程中, 面部细节十分重要; 其次, 人耳朵部位数据较隐蔽, 重建效果不理想. 因此, 本文在头部正面和耳朵两侧各设置了 1 个 RGB-Depth 相机, 面部两侧的 RGB-Depth 相机为了衔接面部细节和耳朵细节数据. 此外, 为了有效地处理用户的发型, 在头顶部位又添加一个相机, 以保证数据的完整. 最后经过实际检验, 得到上述相机的配置(系统的正视图如图 2 所示).

2.3 多目 RGB-Depth 相机的标定

多目相机的标定是指对于每个相机生成的点云数据计算出统一坐标系下的位置, 从而完成各局部点云数据的配准, 在计算机世界中再现用户的全部三维数据.

本文使用一个静态模板(一个模特头像的三维模型), 通过“整体模板-局部采集数据”的迭代最近点算法(iterative closest point, ICP)^[6]完成多目相机标定. 该整体模板为使用扫描仪建立的高精度三维模型. ICP 算法如下: 给定 2 个来自不同坐标系的三维数据点集, 找出这 2 个点集的空间变换, 以便它们能进行空间配准. 假定用 P 和 Q 表示空间中的 2 个点集, P 中各点经过三维空间变换后与点集 Q 中点一一对应, 对应配准变换为使

$$f(\mathbf{R}, \mathbf{t}) = \arg \min \sum_{i=1}^N \|Q_i - (\mathbf{R}P_i + \mathbf{t})\|^2 \quad (3)$$

的目标函数最小. \mathbf{R} 为三维旋转矩阵, \mathbf{t} 为平移向量.

ICP 算法的实质是基于最小二乘法的最优标定算法, 它重复进行“确定对应关系点集—计算最优刚体变换”的过程, 即找到目标点集与参考点之间的旋转 \mathbf{R} 和平移 \mathbf{t} 变换, 使得两配准数据中间满足某种程度度量准则下的最优配准.

对于静态模板, 图 3 显示了多目相机采集到的数据的配准结果和最终的重建模型. 可以看出, 多目相机很好地标定在一起, 为本文系统的实际采集过程提供了很好的配准参数. 需要指出的是, 该模板的配准过程是系统调试过程中预处理的, 即一次标定的结果可在系统使用过程中多次利用.

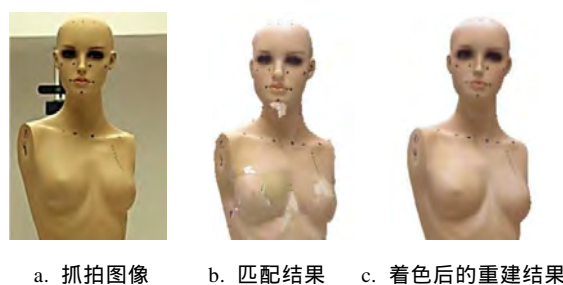


图 3 静态模板实例

3 三维头像的建模过程

三维人像摄影与建模系统的工作流程如下: 三维人像的摄影, 三维点云的增强处理, 三维点云的配准, 三维网格模型的表面重建和着色.

1) 三维人像的摄影. 对于每个相机, 程序控制

它们分时同步捕捉用户的 RGB 和 Depth 图像, 进而根据 Depth 图像所提供的深度值生成附带颜色的三维点云. 如图 4 所示, 从左至右分别是摄影系统中各个相机所拍摄的原始图像.

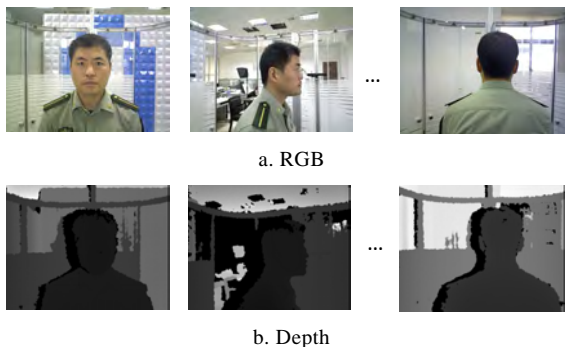


图 4 三维人像摄影原始图像

2) 三维点云的增强处理. 通过双边滤波等方法降低三维点云的噪声, 增强三维数据的质量. 通过去除外点操作, 删除掉不是目标用户的三维数据(如背景外点).

为了保证面部细节精确, 本文在数据采集过程中将会对头部的 RGB-Depth 相机进行设置, 对头部进行多帧数据采集, 再通过对此多帧深度值进行平均得到最终的深度数据, 这样就能够保留充分的细节数据.

此外, 由于 RGB-Depth 相机设备精度的限制, 本文得到的最初点云数据常常会有较多较大的空洞, 本文采用联合双边滤波算法^[7]来优化这些深度数据. 通过联合双边滤波后的深度点云数据仍然包含一些非人体数据, 例如地面和墙壁等, 这些冗余的信息需要单独处理掉. 为了去除这些冗余的背景深度点云信息, 本文采用对深度值进行限制的方法, 即深度值大于一定阈值的点云数据均要被切除. 由于本文的 RGB-Depth 相机设置为各个相机到被拍摄人体头部距离相等, 因此通过深度值来删除多余的环境等外点效果很好. 通过实验得出, 这个深度值的阈值设置为 80mm 效果最好.

如图 5 所示, 我们在实验中发现, 采集到的三维数据在边缘位置容易出现较多的错误点, 颜色值错误或者深度值错误, 对下一步的模型重建和着色都有很大的影响. 本文采用 Sobel 算子^[7]检测出数据的边缘, 并设置阈值, 对于深度值超过阈值的点予以删除, 这样就能够保证保留下来的点均是精确度较高的点.

3) 三维点云的配准分成 2 个步骤:

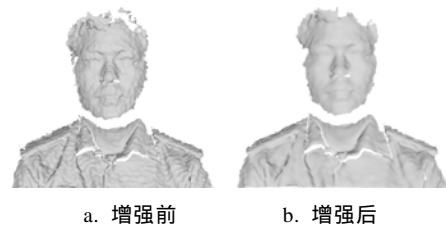


图 5 点云的增强处理实例

Step1. 利用 5 个相机的配准参数, 进行初步配准;

Step2. 得到每个相机拍摄的点云数据后, 进一步使用 ICP 算法进行局部精确配准.

图 6 所示为三维点云配准前后的实例, 从人像的五官等标志部位可以看出, 配准效果很好.

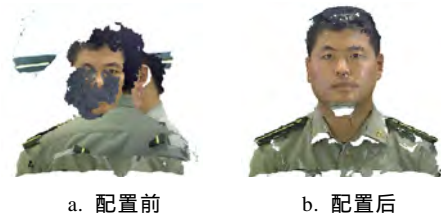


图 6 点云配置实例

4) 三维网格模型的表面重建. 通过表面重建算法重建出三维网格模型. 泊松表面重建算法^[8]对数据噪声不敏感、算法鲁棒性好、重建结果的质量高, 因此本文采用该算法对虚拟人像的三维点云数据进行表面重建. 实验结果如图 7a 所示.



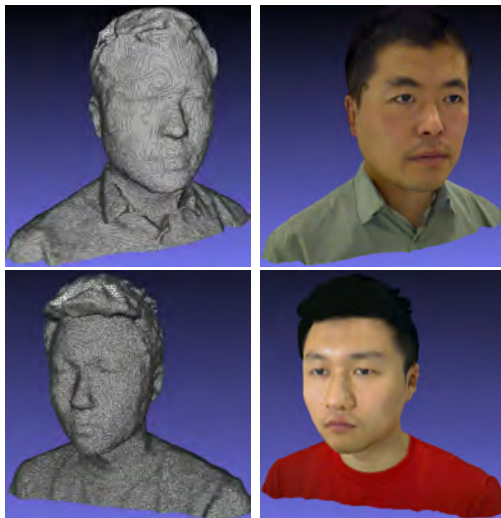
图 7 重建结果

5) 三维模型的着色. 利用三维点云的颜色信息完成三维网格模型的着色. 如何从含有深度和颜色信息的三维扫描重建模型把颜色反投影到模型上, 就是点云的着色问题. 由于不同深度相机的拍摄角度和曝光强度不同, 如果直接将颜色设置给最近的顶点上, 就会导致在这些相机数据的拼接处颜色不连续, 出现区域性的明暗光斑, 点云着色结果很不理想. 本文使用 Chuang 等^[9]的算法, 通过从最近的点云数据中求出颜色梯度, 得到一个矢量场来描述纹理在网格上的局部变化, 求解泊松方程使得颜色梯度值在这个矢量场中得到

最好的匹配, 从而产生一个无缝的均匀纹理表面. 这种算法对于环境光照的均匀性要求不高, 只要不存在严重的光照盲区, 就能够保证较好的着色效果. 实验结果如图 7b 所示, 可以看出, 没有出现因各个相机曝光程度的不同而产生的光斑, 得到的是一个颜色过渡光滑的模型.

4 实验结果及分析

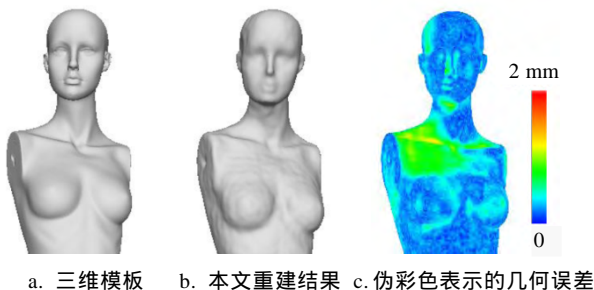
在 Visual C++2010 平台上, 本文使用 Open GL 语言对上述过程进行了编程实现. 2 个头像摄影系统的建模实例如图 8 所示.



a. 重建的几何模型 b. 着色的彩色模型

图 8 头像摄影系统的建模实例

使用预先扫描的三维模板, 本文测量了三维头像摄像系统的几何误差, 如图 9 所示. 图 9a 中, 三维模板使用激光扫描仪获取, 几何精度为 0.01 mm; 图 9b 显示了本文建立的三维几何模型; 本文测量



a. 三维模板 b. 本文重建结果 c. 伪彩色表示的几何误差

图 9 三维模板的几何误差

了模板与模型上最近顶点间的 L_2 几何距离, 并用伪颜色绘制, 如图 9c 所示, 可以看出, 最大的几何误差不超过 2 mm.

在硬件配置为 P4 2.8GHz CPU, 2GB 内存和 512MB NVIDIA 显卡的 PC 机上, 对本文系统进行了充分的测试. 在多目相机一次配置完毕后, 即可进行三维头像的采集和建模工作. 平均而言, 采集时间为 0.9 s, 建模时间 < 3 min.

5 结 语

本文实现了三维头像摄影系统, 其主要特征为 1 s 内完成三维照相. 该系统利用多台低成本的 RGB-Depth 相机, 通过多相机的空间配置完成用户的 RGB-Depth 数据获取; 进而通过数据优化处理、匹配、表面重建和着色等步骤, 生成照片真实感的三维头像.

参考文献(References):

- [1] Parke F I. A parametric model for human faces [OL]. [2014-04-30]. <http://content.lib.utah.edu/cdm/singleitem/collection/ospace/id/1613/rec/55>
- [2] Cyberware. Head & face color 3D scanner [OL]. [2014-04-30]. <http://cyberware.com/products/scanners/px.html>
- [3] Izadi S, Kim D, Hilliges O, *et al.* Real-time 3D reconstruction and interaction using a moving depth camera[C]// Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. New York: ACM Press, 2011: 559-568
- [4] Tong J, Zhou J, Liu L G, *et al.* Scanning 3D full human bodies using kinects [J]. IEEE Transactions on Visualization and Computer Graphics, 2012, 18(4): 643-650
- [5] Khoshelham K, Elberink S O. Accuracy and resolution of kinect depth data for indoor mapping applications [J]. Sensors, 2012, 12 (2): 1437-1454
- [6] Besl P J, McKay N D. A method for registration of 3-D shapes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(2): 239-256
- [7] Richardt C, Stoll C, Dodgson N A, *et al.* Coherent spatiotemporal filtering, upsampling and rendering of RGBZ videos [J]. Computer Graphics Forum, 2012, 31 (2): 247-256
- [8] Kazhdan M, Bolitho M, Hoppe H. Poisson surface reconstruction [OL]. [2014-04-30]. <http://research.microsoft.com/en-us/um/people/hoppe/proj/poissonrecon>
- [9] Chuang M, Luo L J, Brown B J, *et al.* Estimating the Laplace-Beltrami operator by restricting 3D functions [J]. Computer Graphics Forum, 2009, 28(5): 1475-1484