

基于 Kinect 深度信息的实时 三维重建和滤波算法研究

陈晓明, 蒋乐天, 应忍冬

(上海交通大学 电子工程系, 上海 200240)

摘要: 分析了基于 Kinect 输出的深度数据进行场景实时三维重建的算法。针对实现过程中出现的深度图像噪声过大的问题, 根据其信号结构的特点给出了改进的双边滤波算法。新算法利用已知的深度图像噪声范围, 将权值函数修改为二值函数, 并结合 RGB 图像弥补了缺失的深度信息。实验表明, 新算法无论是在降噪性能还是计算效率上, 都大大优于已有的双边滤波, 其中计算速度是原始算法的 6 倍。

关键词: 实时三维重建; Kinect; 三维点云; 噪声分析; 深度图像; 双边滤波; 联合双边滤波

中图分类号: TP391; TP301.6

文献标志码: A

文章编号: 1001-3695(2013)04-1216-03

doi:10.3969/j.issn.1001-3695.2013.04.069

Research of 3D reconstruction and filtering algorithm based on depth information of Kinect

CHEN Xiao-ming, JIANG Le-tian, YING Ren-dong

(Dept. of Electronic Engineering, Shanghai Jiaotong University, Shanghai 200240, China)

Abstract: This paper analyzed and improved 3D reconstruction algorithm using the depth information from Kinect. To reduce noise, it proposed an improved bilateral filtering algorithm based on the signal structure. This new algorithm used a two-valued function to compute the weights of the filter, because the range of depth image data was already known. It also combined the RGB values and depth information of surrounding pixels to complement some missing depth information. The results show that the proposed algorithm has much better performance and efficiency, namely 6 times as fast as the original algorithm.

Key words: real-time 3D reconstruction; Kinect; 3D point cloud; noise analysis; depth image; bilateral filter; joint bilateral filter

三维重建技术是计算机视觉、人工智能、虚拟现实等前沿领域的热点和难点,也是人类在基础研究和应用研究中面临的重大挑战之一,被广泛应用于文物数字化、生物医学成像、动漫制作、工业测量、沉浸式虚拟交互等领域。现有的三维重建技术,按照获取深度信息的方式,可分为被动式技术和主动式技术。被动式技术利用自然光反射,一般通过摄像头拍摄图片,然后通过一系列的算法计算得到物体的三维坐标信息,如 structure from motion^[1] 和 multi-view stereo^[2]。Structure from motion 技术利用不同时间的图像建立对应关系,因此只适用于刚性物体;Multi-view stereo 技术适用于刚体,但是计算量非常大,现阶段很难做到实时。主动式技术包含一个光源,直接测量物体的深度信息,因而很容易做到实时效果,如采用结构光技术的 Kinect^[3] 和采用 time of flight 技术的 CamCube^[4]。而相对于 CamCube,采用结构光技术的 Kinect 价格更便宜,更容易推广。

本文着重解决动态场景建模,选用 Kinect 技术作为三维重建的方法。Kinect 具有反应速度快、价格便宜的优点,但是得到的深度图像精度低,并包含大量噪声。因此在三维重建之前需要对得到的深度图像进行滤波,降低噪声。相对于 RGB 图像滤波,深度图像滤波的研究工作还较少。文献[5]中采用双

边滤波的方法降低噪声,采用多图像融合的方法修补未得到的深度信息,取得了很好的效果。但是该方法只适合静态场景,对于动态场景并不适合。对于缺失的深度信息,本文结合 RGB 图像将其补全,并针对深度图像对双边滤波器进行改进,处理一般的噪声,提高了性能和计算速度。

1 三维重建算法

图1是三维重建流程。Kinect 设备的 RGB 摄像头和红外摄像头分别得到 RGB 图像和深度图像。为了消除因为两个摄像头位置不同而产生的图像中心不一致,首先将深度图像经过一定的坐标变换与 RGB 图像对齐,然后计算出空间点的 X、Y 坐标,最后将三维点云数据(X, Y, Z, R, G, B)通过 PCL 开源库显示。

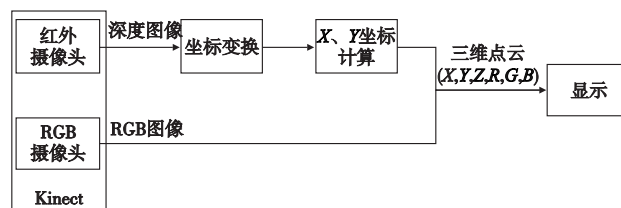


图1 三维重建流程

收稿日期: 2012-08-21; 修回日期: 2012-09-28

作者简介: 陈晓明(1987-), 男, 浙江临海人, 硕士, 主要研究方向为嵌入式、图像处理(xmchen2009@gmail.com); 蒋乐天(1975-), 男, 副教授, 博士, 主要研究方向为嵌入式系统、软件可靠性和可用性; 应忍冬(1975-), 男, 副教授, 博士, 主要研究方向为导航信号处理、嵌入式系统、SoC、数字信号处理。

下面针对图中信号处理流程中各个环节进行具体分析。

1.1 Kinect 摄像头信号标定

在使用 Kinect 设备之前,需要对其进行标定,包括 RGB 摄像头和红外摄像头的标定,以及确定深度与 RGB 图像间的坐标变换关系。摄像头标定是一项比较成熟的技术,这里不详细叙述。文献[6]中,标定前的深度图像误差已经很小,标定后没有显著提高,表明 Kinect 设备内部的芯片已经对 RGB 摄像头和红外摄像头进行了标定,因此不需要额外的矫正,得到的图像即为经过矫正后的结果。

RGB 图像和深度图像分别是以 RGB 摄像头和红外摄像头为中心拍摄到的图片。因此,如果需要得到包含颜色信息的三维点云,就需要将深度图像通过一定的变换,对应到 RGB 图像中。文献[6]中详细介绍了标定原理及具体方法,这里不加详述。实际上,OpenNI 软件内部为深度图像提供了简单的函数 `setViewPoint()`,能够方便地进行图像转换。因此本文采用 OpenNI 内部函数来实现深度图像对齐。

1.2 X、Y 坐标计算

图2是真实世界坐标系与深度图像投影坐标系的示意图。在真实世界坐标系中,是以 Kinect 设备为坐标原点的右手坐标系,而在深度图像的投影坐标系中,是以深度图像原点作为坐标原点的。因此,需要根据测量得到的深度信息 Z 坐标,计算出真实世界坐标系中的 $X、Y$ 坐标。

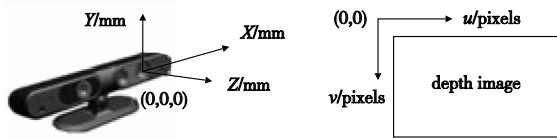


图2 两种坐标系示意图

经过标定后的摄像头,可以等同为理想的针孔成像模型。在本文中,设定 RGB 图像和深度图像的分辨率均为 640×480 ,根据简单的相似三角形原理,得到 $X、Y$ 坐标分别为

$$X = (u - 320) \times Z \times \frac{1}{f} \quad (1)$$

$$Y = (v - 240) \times Z \times \frac{1}{f} \quad (2)$$

式(1)(2)中, $X、Y$ 分别为真实世界坐标系中的横坐标与纵坐标; $u、v$ 为深度图像中的横坐标与纵坐标; Z 为测量得到的深度信息; f 为图像的焦距,其值为

$$f = f_{\text{SXGA}} \times \frac{X \text{ ResOutput}}{X \text{ ResSXGA}} = 1050 \times \frac{640}{1280} = 525 \quad (3)$$

为了提高计算速度,使用乘法代替除法,将 $1/f$ 预先计算出来,其值为 0.0019047619。

1.3 三维点云显示

得到三维点云数据(X, Y, Z, R, G, B)后,利用 PCL 开源库的 viewer 类显示,如图3所示。

由于深度图像存在噪声,呈现的场景中会出现黑色斑点。从图3中可以看到,呈现的场景中出现大量黑色斑点,严重影响了场景观看及进一步处理。图3左下方的黑色显示器部分,出现了不连续的块状,对照图3及相应的深度图4,发现这是由于这部分未得到深度信息所致;而图像右下角的椅子后面,由于红外线被椅子挡住,不能测量出深度信息,因此会出现一片阴影区域,这是由于测量原理所导致的,无法消除。因此,在三维点云进一步的处理之前,需要对其进行降噪处理,填补未得到的深度信息。

从深度图4可以看到,深度图像具有很强的边缘性,前景与背景有很大的深度差,需要在降噪的时候保留这些边缘信息。如果处理不当,将边缘平滑,在最后的点云图显示时,会在

前景与背景之间出现过渡结构,不仅影响显示,而且改变了场景内容。另一方面,由于算法需要做到实时性,要求滤波算法效率要高,但是又不能影响算法精度。本文针对这些问题,一一提出了自己的解决方案。

2 噪声分析

Kinect 是以红外线发出人眼看不见的 class1 镭射光,经过镜头前的光栅将镭射光均匀分布投射在测量空间中,再经过红外线摄像头记录下空间中的各个散斑,得到原始资料后再利用专用芯片计算出深度图像。该技术的关键是镭射光散斑,当镭射光照射到物体表面的时候,会形成随机的放射斑点,称之为散斑。如果不能在物体表面形成反射散斑,也就不能得到该物体的深度信息,这是第一种深度图像噪声。发生这种情况一般有两种原因:a)特殊材料,如透明玻璃、镜子、吸收红外线材料;b)特殊结构,不能反射散斑。

图5是从 Kinect 红外摄像头中得到的包含散斑的原始图像,图中白色斑点即为镭射光散斑;图6为计算得到的深度图像。图5、6中,透明玻璃窗处因为将镭射光投射过去而没有形成镭射光散斑,故无法得到深度数据。

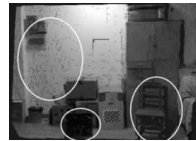


图3 场景三维点云显示图

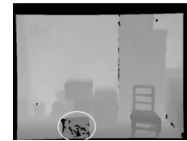


图4 场景深度图



图5 红外摄像头得到的原始图像

第二种深度图像噪声是常规的图像噪声,是由 Kinect 软硬件本身所造成的噪声。图7是利用 Kinect 拍摄白色墙壁得到的深度图,图8是将深度图像左上角放大得到的三维可视化图像。从图8中可以看出,即使测量平面也会得到噪声。一般来说,Kinect 深度图像的分辨率为 $3 \sim 4 \text{ mm}$,即大于 $3 \sim 4 \text{ mm}$ 的深度数值差才会被视为两者具有不同深度。



图6 对应的深度图像



图7 平面墙深度图

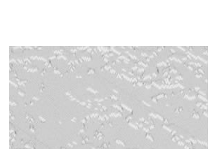


图8 平面墙深度图三维可视化

3 去噪

一般图像滤波可以用式(4)表示为

$$I'(x, y) = \frac{1}{w_p} \sum_{i, j \in \Omega} w(i, j) * I(i, j) \quad (4)$$

其中: I' 为滤除噪声后的图像; Ω 为像素(x, y)的领域范围,一般情况下是以(x, y)为中心的矩形区域; $w(i, j)$ 为滤波器在(i, j)点处的权值; w_p 是唯一化参数; I 是噪声图像,其中

$$w_p = \sum_{i, j \in \Omega} w(i, j) \quad (5)$$

对于双边滤波^[7-9],权值系数由空间域权值 w_s 和图像灰度域权值 w_r 组成,即

$$w = w_s \times w_r \quad (6)$$

其中:

$$w_s = \exp\left(-\frac{(i-x)^2 + (j-y)^2}{2\sigma_s^2}\right) \quad (7)$$

$$w_r = \exp\left(-\frac{(I(i, j) - I(x, y))^2}{2\sigma_r^2}\right) \quad (8)$$

式(7)(8)中, σ_s, σ_r 为基于高斯函数的标准差,决定了双边滤波器的性能,限定了像素点的相对空间位置、灰度变化范围。

根据式(6)~(8),双边滤波本质上是一种改进的高斯滤波,在滤波的同时利用灰度变化信息来保留边缘信息。其权值根据领域灰度差作调整,即在 Ω 领域范围内,如果点 (i,j) 处的灰度与 (x,y) 处越接近,权值 w 越接近高斯滤波权值。在 σ_s 固定的情况下, σ_r 过大,不同灰度差对应的权值 w 均较大,则失去了利用灰度变化保留边缘信息的效果,双边滤波器退化为高斯滤波器; σ_r 过小,权值 w 对不同灰度差过于敏感,则失去了滤波的效果。因此,对于输入图像,需要对输入图像的噪声范围有一个大概认识,设置合理的参数。针对 Kinect 得到的深度图像,将超过 4 mm 的深度差视为不同深度的物体表面,其权值小于 0.5,则 σ_r 经计算为 34。

使用双边滤波方法,假设图像分辨率为 640×480 , Ω 为 5×5 正方形,图像帧率为 30 帧,则要计算 w_r 权值,1 s 大约需要计算 2.3 亿次指数运算,大大地限制了滤波速度。而且,对于 Kinect 得到的深度图像,因为噪声范围已知,因此不需要使用这么复杂的 w_r 来计算权值。在本文中,对式(8)进行了修改,使用简单的二值函数来替换,如式(9)所示。

$$w_r = \begin{cases} 1 & |I(x,y) - I(i,j)| \leq 4 \text{ mm} \\ 0 & |I(x,y) - I(i,j)| > 4 \text{ mm} \end{cases} \quad (9)$$

式(9)使用一个减法、一个取绝对值和一个比较运算替换,大大减少了运算量,并能够理想地做到深度差值大于 4 mm 即判断为不同表面,将其权值设为 0。

如果 $I(x,y) = 0$,那么使用修改过的式(9)计算得到的 $I'(x,y) = 0$,即不能利用周边已知的深度图像值进行补全。文献[10]利用高分辨率图像对低分辨率图像进行升采样,以弥补低分辨率图像中的信息缺失。考虑到 RGB 图像中的信息是完整的,本文将 RGB 图像和深度图像结合,将缺失的深度图像补全。假设深度信息与灰度信息具有相同的变化趋势,对滤波器权值修改如式(10)所示。

$$w = w_s \times w_g \quad (10)$$

$$\text{其中: } w_g = \exp\left(-\frac{(G(i,j) - G(x,y))^2}{2\sigma_g^2}\right) \quad (11)$$

式(11)中, G 为深度图像对应的色度图像灰度值; σ_g 为基于高斯函数的标准差,限定了色度图像的变化范围。

至此,得到整个滤波过程如下:对于每个深度值,若为 0,则利用式(7)(10)(11)补全深度信息;若不为 0,则利用式(6)(7)(9)进行降噪。该算法在降噪的同时很好地保留了深度信息,并弥补了未得到的深度信息。

4 实验结果及实时性分析

本文使用 C++ 编程并实现了双边滤波和本文的滤波算法,并比较两者的性能。为了方便比较两者性能,两者均设定领域 Ω 为 5×5 的正方形, σ_s 均设为 3。双边滤波算法中取 σ_r 为 34,这样正好能够使得 4 mm 以上的深度差权值小于 0.5。本文滤波算法中取 σ_g 为 1。

针对原始三维点云数据,图 9(a)(b)是使用双边滤波后得到的深度图像及对应点云图,图 9(c)(d)是使用本文滤波算法后得到的深度图像及对应点云图。与图 3 比较,图 9(b)中的黑色斑点大大减少,说明双边滤波对于第二种常规的深度图噪声具有明显的降噪效果。但是通过比较图 4 与图 9(a),场景中的第一种噪声——缺失的深度信息,即图中的黑色块状基本上没有变化,说明双边滤波对第一种噪声没有作用。

与图 9(b)比较,图 9(d)中黑色斑点明显减少,但是没有完全消除,说明本文算法比双边滤波具有更好的第二种噪声的降噪效果。比较图 9(d)与图 4,场景中的黑色块状基本消除,整个场景的深度信息被补全,说明本文算法能够很好地去除第

一种噪声。

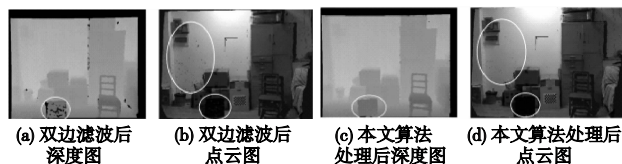


图9 两种滤波算法性能比较

分别对上述两种算法程序进行计时,得到两者运行时间如表 1 所示。相对于双边滤波算法需要的 1 750 ms,本文算法将运算时间降到了 294 ms,为原来的 0.168,大大提高了程序运算效率。但是,考虑到三维重建的实时性,需要进一步使用多核技术或者 GPU 加速技术对程序进行优化,将时间减少到 30 ms 以下。

表 1 双边滤波及本文算法运行时间

算法	运行时间/ms
双边滤波	1 750
本文算法	294

5 结束语

本文利用 Kinect 技术实现了实时场景三维重建。由于 Kinect 的技术原因,获取的深度图像存在大量噪声。之后分析噪声,根据来源将深度图像噪声分为第一种缺失信息噪声和第二种常规图像噪声。针对现有的双边滤波,提出了一种改进算法。新算法根据深度图像噪声范围为 3~4 mm 的特点,将双边滤波中的深度差部分的权值改为二值函数,还结合 RGB 图像弥补了缺失的深度信息。实验表明,本文提出的算法针对第二种噪声无论在降噪性能还是计算效率上,都远远优于双边滤波。另外,对于双边滤波无法处理的第一种噪声也有很好的降噪效果。但是,本文算法在实现上还需要进一步优化,提高计算效率以达到实时效果。

参考文献:

- [1] DELLART F, SEITZ S, THORPE C, et al. Structure from motion without correspondence[C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2000:557-564.
- [2] SEITZ S M, CURLESS B, DIEBEL J, et al. A comparison and evaluation of multi-view stereo reconstruction algorithms[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2006: 519-526.
- [3] Introducing Kinect for Xbox 360[EB/OL]. (2012) [2012-08-10]. <http://www.xbox.com/en-US/kinect>.
- [4] PMD[vision] CamCube 3.0[EB/OL]. (2012) [2012-08-10]. <http://www.pmdtec.com/products-services/pmdvisionr-cameras/pmd-visionr-camcube-30/>.
- [5] NEWCOMBE R, DAVISION A, IZADI S. KinectFusion: real-time dense surface mapping and tracking[C]//Proc of IEEE/ACM International Symposium on Mixed and Augmented Reality. 2011:127-136.
- [6] KONOLIGE K, MIHELICH P. Technical description of Kinect calibration [EB/OL]. (2011) [2012-08-10]. http://www.ros.org/wiki/kinect_calibration/technical.
- [7] TOMASI C, MANDUCHI R. Bilateral filtering for gray and color images[C]//Proc of the 6th International Conference on Computer Vision. 1998:839-846.
- [8] 杨学志,徐勇,方静,等. 结合区域分割和双边滤波的图像去噪新算法[J]. 中国图象图形学报,2012,17(1):40-48.
- [9] 谢勤岚. 结合双边滤波和多帧均值滤波的图像降噪[J]. 计算机工程与应用,2009,45(27):154-156.
- [10] KOPF J, COHEN M F, LISCHINSKI D, et al. Joint bilateral upsampling[J]. ACM Trans on Graphics,2007,26(3):96.