

基于稀疏特征点的单视点深度图像校准

郭庆慧 梁秀霞 张 锐

(山东财经大学计算机科学技术学院 济南 250014) (山东省数字媒体技术重点实验室 济南 250014)

摘 要 高效而精确的校准是基于深度图像三维重建的关键问题。针对 Kinect 摄像头获得同步的彩色图像和深度图像,提出基于稀疏特征点的校准方法。将其中的彩色图像作为参考图像,根据深度图像和彩色图像的几何对应关系计算相应矩阵转换参数,然后利用稀疏特征点确定校准参数,由此得到彩色图像和深度图像间的映射关系,最后实现深度信息的校准。实验结果表明,深度图像相对彩色图像的变换近似满足线性关系,计算复杂度低,精确度高,易于实现,适合在相关领域中得到进一步推广。

关键词 深度图像,图像校准,稀疏特征点,特征点匹配,机器视觉

中图法分类号 TP37 文献标识码 A

Calibration for Single Viewpoint Depth Image Based on Sparse Feature Points

GUO Qing-hui LIANG Xiu-xia ZHANG Rui

(School of Computer Science & Technology, Shandong University of Finance and Economics, Jinan 250014, China)

(Shandong Province Key Lab of Digital Media Technology, Jinan 250014, China)

Abstract Efficient and accurate calibration is the key issue for 3D model reconstruction based on the depth image. Based on the synchronous color image and depth image obtained from the Kinect camera, a calibration method based on the sparse feature point was presented. We took the color image as a reference image, used the geometric relation between the depth image and the color image to calculate the corresponding matrix transformation parameters. Then we extracted corresponding sparse feature points of the two types of data to determine the calibration parameters, and thus obtain a mapping relationship between the two data. Finally, we achieved the calibration of depth information according to the determined parameters. The results show that the relative transformation between the depth image and the color image meets a linear approximation conversion. This method has low computational complexity, high accuracy and easy to realize. So it is suitable for further promotion in related fields.

Keywords Depth image, Image calibration, Sparse feature points, Feature points matching, Machine vision

1 引言

机器视觉在国民经济、科学研究及国防建设等领域有着越来越广泛的应用,如工业检测、医学影像分析等。机器视觉系统中常用的数据是通过摄像头获取的图像。以往基于可见光的二维图像常常遭遇光照变化、阴影、物体遮挡以及环境变化等因素的干扰,不能对三维物体进行准确的识别与定位,深度信息就能很好地克服以上困难,因此在计算机视觉中得到越来越广泛的应用。如何快速实现深度图像的精确校准是三维模型重建的关键问题,特别是近几年,越来越多的应用如虚拟现实、医学诊断以及碰撞检测等都依赖于对三维场景进行准确且快速的分析^[1]。

对于深度信息的获取,早期在某些工业应用中,使用摄像

机阵列形成双目或多目视觉系统,对同一场景从不同位置成像,进而从视差中得到深度信息。这种采用传统计算机视觉的方法来获得深度信息的方法称为被动式方法,又称计算机立体视觉。相对于获得深度几何信息的被动式方法,通过三维视觉传感器获得深度信息的方法成为主动式方法。在计算机视觉中,一般是通过激光或红外光扫描来获得物体表面的三维点云数据,这些设备虽然采集精度高,但要求的条件非常苛刻,再加上价格和操作复杂等原因,并没有得到广泛的应用。作为距离测量中一种可替代的设备,深度传感器相机拥有传统三维测量系统所不具备的一些优点,能够以较低的价格、较高的拍摄速度等直接获取深度图像,特别是微软最新开发的深度摄像设备 Kinect 体积小、价格低廉、使用简单,在虚拟现实、远程电子商务等领域,有着更为广泛的市场推广前

本文受国家自然科学基金(61020106001,61272431,61103117,61202364),教育部人文社会科学研究项目(13YJC860023),中青年科学家奖励基金(BS2011DX024,BS2012DX028)资助。

郭庆慧(1988—),女,硕士生,主要研究方向为多媒体与虚拟现实、图像处理与计算机视觉、计算机游戏与动画,E-mail:guoqinghui621@163.com;梁秀霞(1976—),女,博士,副教授,主要研究方向为计算机图形学、计算机游戏与动画、多媒体与虚拟现实等;张锐(1979—),男,博士,讲师,主要研究方向为信息可视化、图像处理、三维游戏等。

景。

然而通过摄像头获取的深度图像由于受镜头畸变、摄像机质量、景物对象与摄像机三维空间相对位置关系等内外参数的影响,图像像素之间的空间关系会发生变化,也就是说,原始场景中各部分之间的空间关系与图像中各对应像素间的关系不一致了。Kinect 的彩色摄像头和深度摄像头都是对同一场景进行取图,而且 Kinect 的彩色摄像头和深度摄像头的位置不同,使得深度图像相对彩色图像产生畸变。图 1 为 Kinect for xbox360 同步获取彩色图像和深度图像。如果不对得到的深度图像的畸变加以合理、有效的校准,不但图像的视觉效果不佳,而且会影响到目标识别与跟踪的质量和效率等。



(a) Kinect 摄像头获取的彩色图像 (b) Kinect 摄像头获取的深度图像

图 1

本文利用基于图像特征的方法对 Kinect 摄像头得到的深度图像进行校准。该方法通过比较 Kinect 摄像头获得的同步彩色图像和深度图像,获得深度图像相对彩色图像的变换关系,并利用特征点对变换关系进行进一步优化。

2 国内外研究现状

图像校准通常通过一些已知的参考点,即无畸变图像的某些像素点和畸变图像相对应像素点的对应关系,计算出映射关系中的未知系数,以此确定各像素的深度值。传统的二维图像的校准方法是将所有引起图像畸变的因素综合考虑在内,通过分析摄像机的成像过程,改变透射投影矩阵对图像进行校准^[2,3]。这样不仅增加了求解过程的复杂性,同时由于各未知参数的互相牵制,进行计算时会造成图像对应特征的消失、灰度畸变(可视性差)等问题,很难获得满意的精度。郭永刚^[4-6]等人提出的控制对象的校准方法虽然精度高,但计算复杂,运算量大,很难满足实时性的要求;刘堂友^[7-9]等人在以往文献的基础上提出基于标准网格的图像校准方法,有效解决了图像校准的实时性问题。

对图像进行校准,应在保证图像信息不丢失且满足精度要求的基础上使运算量尽量小,以满足实时性的要求^[6]。然而,由于深度图像与二维图像存在一定的差异,采用传统二维图像校准方法直接对深度摄像头进行修正,或者对得到的深度图像进行校准,不仅计算复杂,还会降低校准精度。因此许多学者提出了针对三维图像的校准方法。综合现有的图像校准方法,即可分为基于立体配准的校准方法和基于图像特征的校准方法。

基于立体配准的图像校准就是利用图像的对极几何进行转换。该方法大多应用于双目视觉或者多目视觉中。在双视

点标定方面,Yun-Suk Kang^[10]等人提出利用水平基线对得到的深度图像进行校准,利用计算出的水平基线估计相机参数和进行相应的校准变换。然而该方法只适用于双视点的立体图像校准,针对三视点的图像校准却不能达到很好的效果。Changming Sun^[11]等人利用三目图像获得的三线性张量、射影不变量和基本矩阵,通过旋转、斜交和放射变换等方法对无标定的三目图像进行校准。以上两种方法是对不同视点获得的图像进行校准,进而得到精确的深度图像,如果对深度图像直接研究,通常提取深度图像的点云数据进行分析。Ali Khosravani^[12]等人提出基于对象空间观测的方法,即根据图像数据范围提供的几何信息,结合迭代最近点(ICP)算法确定 Kinect 传感器的位置,进而确定点云数据的位置。

基于立体配准的图像校准方法不需要知道相机的初始位置和方向,使得数据获取变得十分方便且稳定性好,并能消除其他校正方法产生的视觉失真,极大地提高了图像校准的精确性。然而,该方法利用同一个摄像头在不同时间对同一位置获得的两幅图像的对极几何进行同时校准,计算复杂,且改变了原始数据,影响校准的精确度。

为了降低计算量,简化图像处理过程,提高校准精度,有学者提出对三维图像基于图像特征的校准。Peter Henry^[13]等人提出利用场景中外观和形状匹配的联合优化算法对得到的深度图像进行校准。该方法从彩色图像中提取稀疏特征点,然后结合致密迭代最近点(ICP)匹配得到校准后的深度图像。Ali Khosravani^[12]等人还提出基于 Kinect 彩色图像的方法,利用运动中恢复结构的方法对获得的三维点云数据进行处理。根据图像中的特征点确定传感器的位置,进而确定深度图像和彩色图像之间的对应的信息。基于图像的校准方法不仅能得到很好的校准效果,而且简单易行,大大降低了计算量。

综上所述,现有的基于立体配准的校准方法计算量大,计算复杂,而基于图像特征的图像校准方法计算简单,复杂度低,且不用考虑摄像机的内外参数,大大降低了图像处理的难度。目前针对传统摄像头的校准技术已非常成熟,利用传统相机进行场景拍摄得到的彩色图像产生的畸变会越来越小,在某些应用下可以将获得的图像看作无畸变图像。对于 Kinect 摄像机的彩色摄像头采用传统相机来获取彩色图像,得到的彩色图像几乎接近真实场景。而对于深度摄像头修正的技术^[14]却甚少,使得深度图像相对真实场景的畸变很明显。

Kinect 摄像头中,深度图像是通过红外摄像头获得的深度数据,与二维图像不同的是,深度图像中每个像素代表了一个相对的深度信息,但是,在某些应用中仍可按照二维图像进行处理,可以根据二维图像的校准方法对深度图像进行校准。因此,本文将同时获得的彩色图像看成原始无畸变图像,将其作为参考图像,提出基于特征点的校准方法,并由此得到彩色图像和深度图像之间的映射关系。

3 深度图像校准

对 Kinect 摄像头而言,它包含 3 个镜头,中间的镜头是

RGB 彩色摄像机,左右两边镜头则分别为红外线发射器和红外线 CMOS 摄影机所构成的 3D 深度感应器,对于 Kinect 视野范围中的一个点,由于镜头的位置和性质不同,结合摄像机成像原理^[15],使得到的深度图像与彩色图像出现畸变。即使同样的分辨率,深度图像帧的像素也无法一一对应到彩色图像帧中去,如图 2 所示。在以后的应用中若能得到更好的视觉效果,特别是在三维重构应用中能与真实场景一一对应起来,需要对得到的深度图像进行校准。



图 2 未校准的深度-彩色图像合并

本文首先通过深度图像和彩色图像的几何对应关系对图像进行粗略校准,然后通过提取特征点来确定转换矩阵的校准参数,进而进行基于灰度的深度信息校准,以恢复原位置的深度信息,以此完成对深度图像的校准,其校准操作流程如图 3 所示。

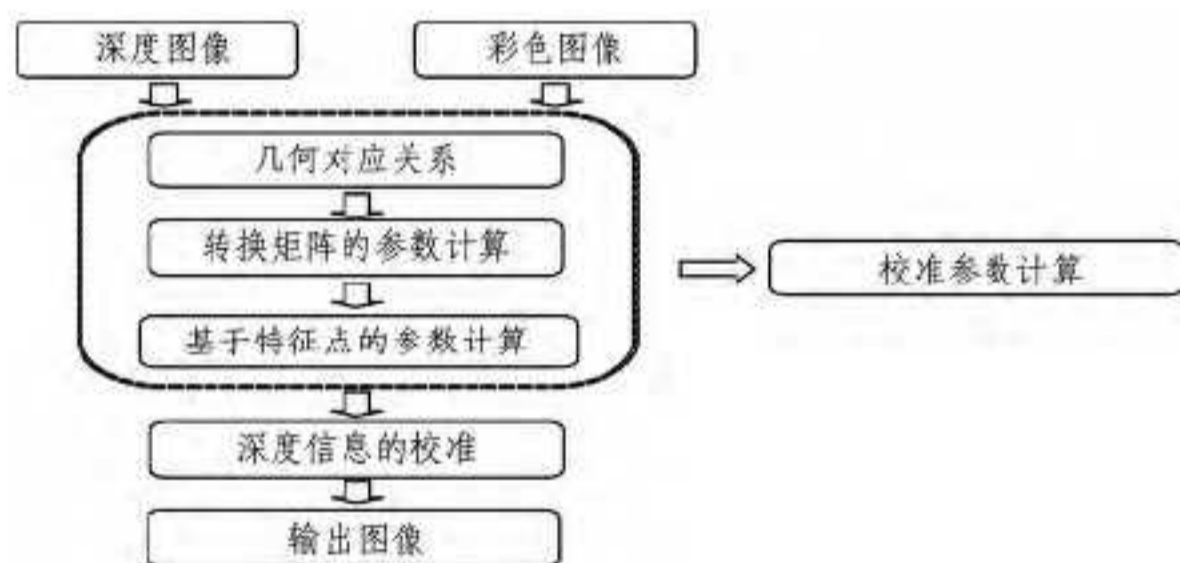


图 3 校准流程

3.1 校准参数计算

3.1.1 转换矩阵参数计算

假设作为参照图像的彩色图像的像素位置为 (x, y) , 其相应的深度信息为 $D(x, y)$, 畸变图像的像素位置为 (x', y') , 相应的深度信息为 $D'(x', y')$, 如图 4 所示。则像素位置之间的变换为,

$$x' = s(x, y)$$

$$y' = t(x, y)$$

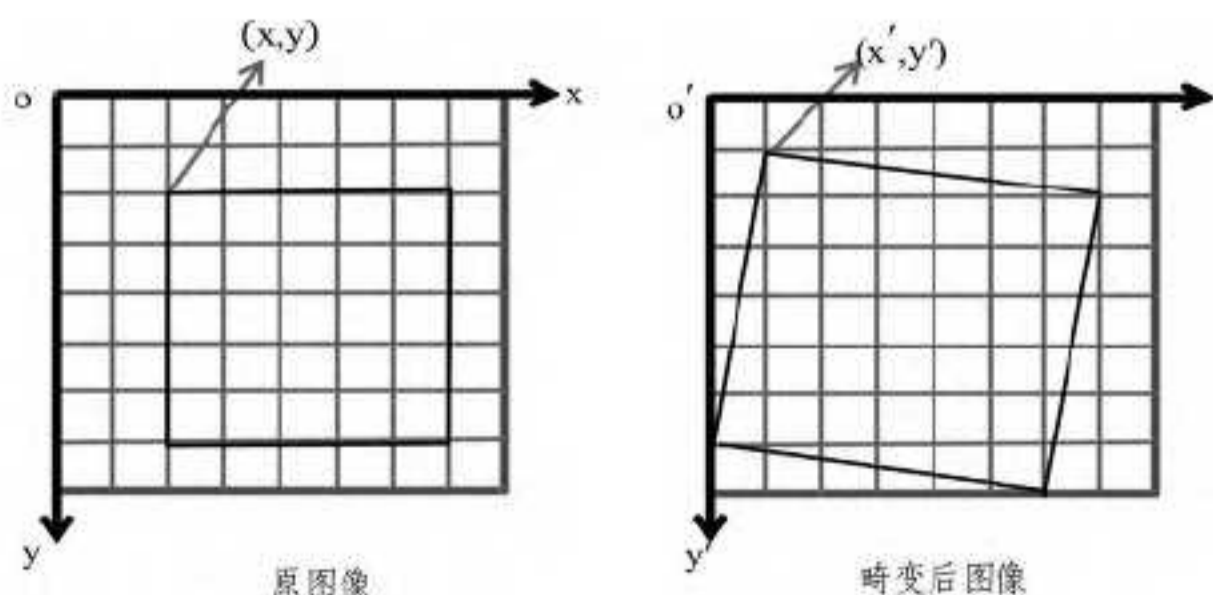


图 4 畸变图像和原图像的对应点

如果能确定 $s(x, y)$ 和 $t(x, y)$ 的解析表达式,就可以通过逆变换来恢复图像。在实际中,通常无法得到畸变情况的解

析表达,为此我们首先需要在原始图像和畸变图像上找一些其位置确切的点,即约束对应点。本文通过两种数据的几何对应关系,得到对应的 4 个坐标点作为计算转换矩阵参数的约束对应点。由于 Kinect SDK NUI API 中包含彩色图像二维坐标和深度图像空间坐标,对于 Kinect 视野范围中的一个点,这两个坐标系的坐标和度量并不一致,无法通过获取的约束点坐标直接进行计算,因此在转换矩阵参数确定时,需要将求取的约束对应点相应的坐标进行转换,将彩色图像和深度图像映射到同一个坐标系中去。最后利用确定的约束对应点计算出转换矩阵中的各个参数,从而建立两幅图像间各个像素空间位置的大体对应关系。

一般情况,我们可用双线性等式来表示 $s(x, y)$ (见式(1)) 和 $t(x, y)$ (见式(2)):

$$s(x, y) = s_1 x + s_2 y + s_3 xy + s_4 \quad (1)$$

$$t(x, y) = t_1 x + t_2 y + t_3 xy + t_4 \quad (2)$$

将得到的 4 个点分别代入上式中,求出矩阵转换参数 $s_1, s_2, s_3, s_4, t_1, t_2, t_3, t_4$ 的值,得到相应的变换,如图 5 所示。



图 5 基于转换矩阵的对应关系

3.1.2 基于特征点的参数

通过几何对应关系的点来计算转换矩阵的参数能够建立起两组数据间空间位置的大概对应关系,但不能确保两组数据间其他点都满足该对应关系。为了确保图像校准的精确性,需要对图像上的其他点进行验证。因此,我们提取彩色图像上的特征点以及深度图像上的匹配点来求取精确的转换参数。

图像特征点是指那些灰度信号在二维方向上都有明显变化的点,如角点、圆点等。图像特征点的提取和匹配技术广泛应用于图像融合、图像校正、三维重建等领域中^[16]。在获取不同图像之间的匹配特征点时,往往面临尺度变化的难题,即需要分析在拍摄时摄像机与目标物体的距离。本文利用检测尺度不变的 SURF 算法提取数据图像的特征点及其描述。首先对每个像素计算 Hessian 矩阵以得到特征,尺度为 σ 的图像点 $I(x, y)$ 的 Hessian 定义如下:

$$H(x, y) = \begin{bmatrix} I_{xx}(x, \sigma) & I_{xy}(x, \sigma) \\ I_{xy}(x, \sigma) & I_{yy}(x, \sigma) \end{bmatrix} \quad (3)$$

其中, $I_{xx}(x, \sigma), I_{xy}(x, \sigma)$ 和 $I_{yy}(x, \sigma)$ 是 x 点处高斯函数二阶偏导数和图像的二维卷积。为了提高高斯卷积的计算速度, SURF 算子采用方框滤波器来近似二阶高斯滤波,构造了一种 Fast-Hessian 矩阵。在原图像上,通过扩大方框滤波器尺寸形成不同尺度的图像金字塔,并使用积分图像对图像卷积进行加速,进一步求解 Fast-Hessian 矩阵的行列表达式为:

$$\Delta H(x, y) = D_{xx}(x)D_{yy}(x) - (0.9D_{xy})^2(x) \quad (4)$$

其中, Δ 为在点 $I(x, y)$ 周围区域的方框滤波器响应值, 用 Δ 进行极值点的检测。方框滤波模板与图像进行卷积运算后的值分别为 $D_{xx}(x), D_{yy}(x), D_{xy}(x)$ 。通过矩阵的行列式和特征值来判别其是否为极值点, 若 Fast-Hessian 矩阵的行列式为正, 且两个特征值不同时为正或为负, 则认定其为极值点。根据 Fast-Hessian 矩阵求出的每个像素点与其三维邻域的 26 个点进行大小比较, 如果它是这 26 个点中的最大值或者最小值, 则保留下来, 当作初步的特征点, 然后采用三维线性插值法得到精确的特征点位置及其所在的尺度值, 并用向量描述作为其描述子。

确定了数据图像的特征点后, 对特征点进行匹配, 为了提高匹配的准确率, 首先采用 FLANN 算法找到具有最近距离的匹配点对, 即将第一幅图像中的每个特征描述子向量与第二幅图像中的描述子进行比较, 得分最高的一对描述子 (即两个向量的欧式距离最近) 将被视为“最佳匹配点对”, 如图 6 所示。



图 6 匹配特征点提取

仔细观察图 6, 会看到里面存在着误匹配, 有对应点的位置并不一致, 如右图头顶上的点却对应到左图大腿的位置, 这就出现了误匹配, 此时, 采用 RANSAC (随机采样一致) 算法去除误匹配, 进一步提高匹配准确度。最后根据确定的“高匹配点对”手动确定更新所需要的 4 对坐标点, 如图 7 所示。



图 7 手动确定的匹配特征点

在 3.1 节中通过几何对应关系的点来计算转换矩阵的参数, 建立起两组数据间空间位置的大概对应关系, 其相应变换公式曲线如图 8 所示。

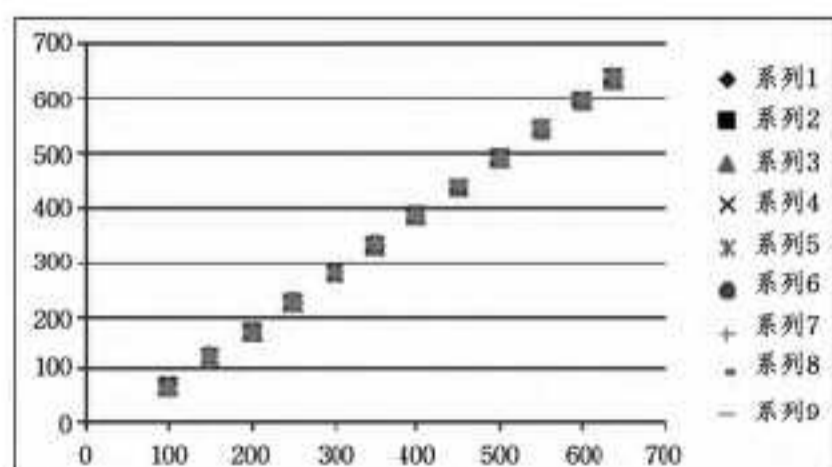


图 8 X 方向校准函数图

图 8 为 X 方向的校准曲线图, 图 9 表示 Y 方向的校准曲线图, 其中系列 1 到系列 9 分别表示人体到摄像头的不同距离, 即 180cm~270cm 的变换公式曲线图, 从上面两图可以看出, 在误差范围内不同的距离的深度图像相对彩色图像都满足同一个线性变换关系。因此, 其修正因子也满足线性变换。

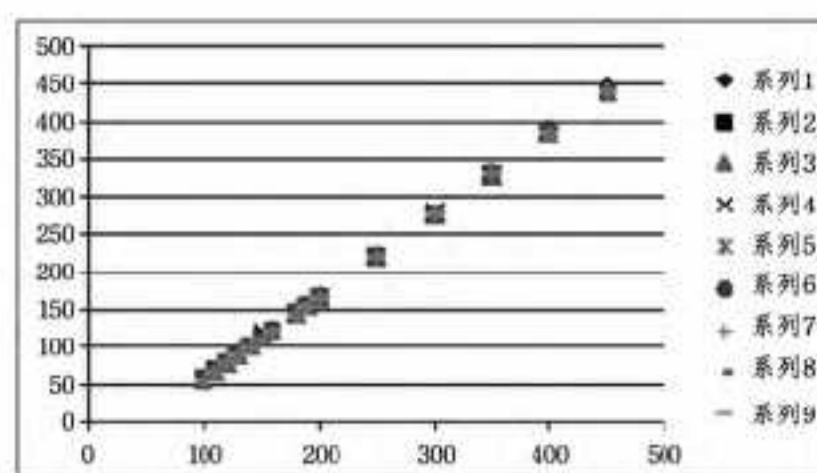


图 9 Y 方向校准函数图

设修正因子:

$$P(x, y) = a_1 x + b_1 y + c_1$$

$$Q(x, y) = a_2 x + b_2 y + c_2$$

则最终校准公式为:

$$s'(x, y) = s(x, y) + P(x, y) \quad (5)$$

$$t'(x, y) = t(x, y) + Q(x, y) \quad (6)$$

将获得的相匹配的特征点对代入上述的校准公式中, 根据产生的误差得到修正因子为一个常量, 即 $P(x, y) = c_1, Q(x, y) = c_2$ 。

3.2 深度信息的校准

根据确定的映射关系对图像中像素的各个空间位置完成校准以后, 需要对深度信息进行插值。实际数字图像中 (x, y) 的值总是整数, 但通过映射关系计算出的 (x', y') 的值一般不是整数。然而深度图像的深度信息仅在坐标为整数处有定义, 所以在非整数处的深度信息就要用其周围一些整数处的深度信息来计算, 以此确定各个像素位置处的深度信息能够与原图像一一对应起来。本文中, 根据上述获得的校准公式, 取某一像素点的 4-邻域内的深度信息, 取其平均值作为该点的深度信息赋给相应的像素点的坐标。将深度图像的各个像素点根据计算得到的校准公式进行计算。

$$x' = s_1 x + s_2' \quad (7)$$

$$y' = s_3 x - s_4' \quad (8)$$

其中, s_2', s_4' 分别为计算得到的校准公式的常量。然后求 (x, y) 取点四邻域的像素值, 并求其平均值, 将得到的平均值作为校准后 (x', y') 点处的像素值。

4 实验结果

本实验采用 Visual Studio2010, OpenCV 对 Kinect SDK 得到的深度图像进行校准, 提出基于稀疏特征点的校准方法。实验中, 我们垂直于 Kinect 摄像头方向每隔 10cm 取一张深度图像和彩色图像, 取彩色图像作为参考图像, 利用两种数据的匹配特征点对深度图像进行校准, 在没有丢失原来信息的基础上, 深度图像得到有效校准, 并得到相应的变换公式主要实验步骤结果如图 10 所示。



图 10 校准后的深度—彩色图像

为了获得校准精度,需要对畸变校准结果进行分析,可通过计算对应像素点的坐标残差的均方差 δ_x, δ_y 来衡量校正精度^[17];

$$\delta_x = \sqrt{\frac{1}{n-1} \sum_{ij} (x_{ij} - x_{cij})^2} \quad (9)$$

$$\delta_y = \sqrt{\frac{1}{n-1} \sum_{ij} (y_{ij} - y_{cij})^2} \quad (10)$$

其中, (x_{ij}, y_{ij}) 和 (x_{cij}, y_{cij}) 分别为修正后的图像上点的坐标和理想成像状态下图像(彩色图像)点的坐标。

本文在摄像头的识别距离内,每隔 10cm 取 6 组数据,利用本文提出的畸变校准算法对获得的 10 幅深度图像进行校准,计算每幅图像校准后的均方误差,其中最大值为 0.278,最小值为 0.251,6 组数据的均方误差的平均值在 0.262 个像素以内,与基于控制点的图像校准算法^[6]和基于 Kinect 彩色图像的方法^[12]比较,该校准算法有很好的校准精度。其均方差如表 1 所列。

表 1 均方差比较

算法	均方差
控制点算法	0.31
基于彩色图像	0.289
本文算法	0.262

结束语 本文中,我们在摄像头的识别距离 80cm-400cm 内,每隔 10cm 取一组数据,将彩色图像作为参考图像,按照彩色图像和深度图像像素点之间的映射关系对深度图像进行校准,得到不同距离的校准曲线。通过对比分析校准曲线的变化,我们发现在一定的误差范围内,不同距离的深度图像相对彩色图像的畸变都近似满足同一个线性关系。

基于立体配准的图像校准方法需要对图像的全部像素进行处理,且需要计算两幅图像的对极几何,计算复杂。本文针对基于立体配准的图像校准方法计算复杂等问题,提出基于图像的校准方法来对深度图像进行校准。在校准过程中,将彩色图像看作没有畸变的参考图像,分析摄像机成像模型及 Kinect 摄像头中各个坐标系之间的关系,得出彩色图像和深度图像之间的几何变换关系,然后通过得到几何变换来校准深度图像中各个像素位置,使得校准后深度图像中像素间的空间关系跟彩色图像中像素间的空间关系一致,最后对空间变换后的像素赋予相应的深度值以恢复原位置的深度,使得畸变的图像恢复到原始图像。

本方法直接将 Kinect 获得的彩色图像作为参考图像,从而引入一定的误差。基于图像特征的校准方法的首要任务是要确认一个无畸变的参考图像。Kinect 摄像头能同时获得彩色图像和深度图像,将彩色图像看作无畸变的图像,这就为深度图像基于图像的方法进行校准提供了可能。本文利用 Flann 算法进行特征点匹配,在确定计算参数时手动确定 4

对坐标,因此会引起一定的误差,在以后的实验中,对特征点的匹配需要进一步优化。

参考文献

- [1] 向学勤,潘志庚,童晶. Depth camera in computer vision and computer graphics: An overview[J]. Journal of Frontiers of Computer Science and Technology, 2011, 33(6): 5-16
- [2] Fusiello A E, Trucco A V. Rectification with unconstrained stereo geometry[C] // Proceedings of the British Machine Vision Conference. University of Essex, UK, 1997: 400-409
- [3] Fusiello A E, Trucco A V. A compact algorithm for rectification of stereo pairs[J]. Machine Vision and Application, 2000, 22(12): 16-12
- [4] 郭永刚,葛庆平,冯平,等. 一种基于控制点自动提取的图像畸变校正算法[J]. 计算机工程与应用, 2007, 43(2): 214-216
- [5] 刘军,胡振伟,张达,等. 基于控制点鲁棒提取的视觉成像畸变校准方法[J]. 东北大学学报(自然科学版), 2011, 31(3): 331-333
- [6] 赵小川,罗庆生,韩宝玲. 四边形分片逼近控制点的图像畸变校正算法[J]. 光电工程, 2009, 36(5): 78-81
- [7] 刘堂友,董爱华. 精确校正图像的径向畸变和倾斜失真[J]. 中国图象图形学报, 2007, 12(10): 1935-1938
- [8] 李久芳. 基于标准棋盘格的图像校准方法[J]. 电子工业专用设备, 2010, 190: 4-6
- [9] 林明汉,胡贤龙. 大视场 CMOS 相机图像的畸变校正[J]. 科学技术与工程, 2009, 9(18): 1671-1819
- [10] Kang Y-S, Ho Y-S. Efficient stereo image rectification method using horizontal baseline[M]. Advances in image and video technology, Springer Berlin Heidelberg, 2012: 301-310
- [11] Sun Chang-ming. Uncalibrated three-view imagerecification[J]. Image and Vision Computing, 2003, 21(3): 259-269
- [12] Khosravani A, Lingenfelder M, Wenzel K, et al. Co-registration of Kinect point clouds based on image and object space observations[OL]. <http://www.ifp.uni-stuttgart.de/publications/2012/LC3D-2012-khosraianiEtAl.pdf>
- [13] Henry P, Krainin M, Herbst E, Xiao Feng Ren, Dieter Fox. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments[J]. The International Journal of Robotics Research, 2012, 31(5): 647-663
- [14] 王学伟,徐庆九,葛先军. 红外光学系统几何畸变效应模拟[J]. 激光与红外, 2012, 42(5): 544-546
- [15] 苏晋. 摄像机标定方法研究[D]. 沈阳, 东北大学, 2010: 7-14
- [16] 徐丽燕,王静,邱军,等. 基于特征点的多光谱遥感图像配准[J]. 计算机科学, 2011, 38(7): 280-282
- [17] 王权岱. 压印对正系统中标记图像几何畸变的校正[J]. 光学精密工程, 2007, 15(3): 422-427