

基于物体面对应的 RGB-D 图像拼接优化方法^{*}于广旺¹ 杨家志^{1,2} 陈梦强¹ 沈 洁¹

(1. 桂林理工大学信息科学与工程学院 桂林 541006; 2. 广西嵌入式技术与智能系统重点实验室 桂林 541006)

摘 要: 针对 RGB-D 中深度图像分辨率低、范围小、噪声大而不利于三维重建的问题,研究了一种基于物体面对应的 RGB-D 图像拼接优化方法。先对 RGB-D 图像进行预处理对齐,使用基于特征匹配算法对特征点提取和粗匹配,其次通过本文研究的不同视角下同一物体面对应关系来剔除误匹配,最后根据单应矩阵得到宽视角的 RGB-D 图像以及三维模型。本文使用了尺度不变特征变换(SIFT)、加速稳健特征(SURF)和定向 FAST 和旋转 BRIEF (ORB)3 种算法来进行对比实验。实验结果表明,添加本文方法后的算法在有形变、旋转的图像上分别剔除 41%、29% 和 52% 的误匹配,均方根误差减少了 5%、27% 和 33%。在缩放的图像上分别剔除 53%、57% 和 51% 的误匹配,均方根误差减少了 14%、17% 和 28%,提高了匹配精度,验证了本文方法的可行性。

关键词: RGB-D 图像;图像拼接;三维重建;深度信息

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6040

RGB-D image mosaic optimization method based on
object-face correspondenceYu Guangwang¹ Yang Jiazhi^{1,2} Chen Mengqiang¹ Shen Jie¹

(1. College of Information Science and Engineering, Guilin University of Technology, Guilin 541006, China;

2. Guangxi Key Laboratory of Embedded Technology and Intelligent System, Guilin 541006, China)

Abstract: An object-face based RGB-D image stitching optimization method is studied to solve the problem of low resolution, small range, and high noise of RGB-D depth images, which is not conducive to three-dimensional reconstruction. First, the RGB-D images are pre-processed and aligned; the feature points are extracted and roughly matched using the algorithm. Then, the mismatching is eliminated by the corresponding relationship of the same object face under different perspectives studied in this paper. Finally, the RGB-D images with wide viewing angles and three-dimensional models are obtained based on the homography matrix. Three algorithms, Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), and Oriented FAST and Rotated BRIEF (ORB), are used for comparison experiments. The experimental results show that 41%, 29%, and 52% erroneous matches are removed on distorted and revolved images, and the Root Mean Square Error is reduced by 5%, 27%, and 33% respectively. In the scaled image, 53%, 57%, and 51% erroneous matches are removed, and the Root Mean Square Error is reduced by 14%, 17%, and 28%, which improves the matching accuracy and verifies the feasibility of this method.

Keywords: RGB-D image; image mosaic; 3D reconstruction; depth information

0 引 言

在计算机图形学、机器视觉等领域中,基于图像的三维重建^[1]一直是研究的重点。广泛应用于医疗领域上器官模型^[2]、科学领域上化合物模型^[3]以及地球科学领域上三维

地质模型的构建等等。获取物体的三维模型可以通过接触式扫描、激光扫描的方式,但是这些方式存在操作复杂、设备昂贵的缺点^[4-6]。Kinect 深度相机^[7]则避开了这样的缺点,它操作简单、价格便宜,原理是通过红外传感器感知环境形成一幅深度图像,再通过多幅深度图像拼接完成三维

收稿日期:2022-04-26

^{*} 基金项目:国家自然科学基金(41961065)、广西创新驱动发展专项资金项目(桂科 AA18118038)、广西科技基地和人才专项(桂科 AD19254002)资助

重建。图像拼接的原理是同一场景下,从不同视角下拍摄的存在重叠区域的图像,对其进行配准、投影和融合,得到宽视场的图像。但深度图像存在分辨率低、噪声大和范围小的缺点,直接拼接会导致图像的畸变甚至失败。

图像配准是图像拼接的关键,深度图像配准更是三维重建的关键。目前的图像拼接算法分为基于变换域、灰度信息、特征匹配以及深度学习四类算法,其中基于特征匹配的算法由于准确度高,运算量少等优点成为目前最常用的图像拼接算法。SIFT^[8-9]、SURF 和 ORB 都是其中较为成熟的算法。但上述算法在深度图像拼接过程中会出现误匹配的情况,降低匹配精度,极大影响拼接效果。为此,国内外研究人员做了大量的相关工作。

Elank 等^[10]指出了上述算法影响拼接效果的原因是缺乏对特征点质量和分布的关注,无法避免冗余特征点和误匹配的出现。因此,剔除冗余特征点和误匹配成为了提高匹配精度的两种主要方式。Zhao 等^[11]提出了一种改进的图像配准方法,通过欧氏距离分组匹配点对,选择距离最小比率组的匹配,确保匹配的均匀分布。Hosseini 等^[12]提出了一种自适应的冗余特征点消除法,根据特征点的数量和距离分布度量出自适应阈值。刘佳等^[13]通过修改欧氏距离和积分图像的阈值来剔除误匹配,但匹配精度不高。沈悦等^[14]提出了一种基于深度和彩色双信息特征源的 Kinect 植物图像拼接方法,其中通过彩色图像引导深度图像拼接的方式,能够防止拼接出现畸变。接着通过深度信息有效剔除误匹配,但只比较了深度信息,却没有注意特征点周围区域。陈伟等^[15]提出了一种基于改进的 FAST-SURF 快速图像拼接算法,其中通过几何约束法剔除匹配连线不在合理范围内的误匹配,匹配效果有待进一步验证。

本文结合深度图像的特点,研究了一种基于物体面对应的图像拼接优化方法。在欧式距离比值法得到初始匹配后,利用匹配点对斜率阈值和深度值比值剔除不合理的匹配点对,最后计算出单应矩阵完成图像拼接。

1 方法原理

1.1 本文方法

基于特征匹配的算法在图像拼接中分为 5 步,包括特征点提取、描述符计算、特征点匹配和单应矩阵计算。在输入 RGB-D 彩色图像后,使用特征匹配算法提取两幅彩色图像的特征点,并通过描述符为特征点进行粗匹配。由于传统算法在特征点匹配中缺乏对物体识别的能力,所以会出现不在同一物体上的特征点依旧匹配的情况。深度图像中的深度值信息(相机镜头距物体的距离)可以简单判断是否为同一物体。本文提出的方法主要分为两部分,一是剔除斜率不在合理区间的匹配,原因是彩色图像中物体特征点连线应该与物体面连线斜率近似。二是深度图像中,特征点的深度信息应该接近深度图像的缩放倍率。本文方法可以有效减少误匹配的数量,提高匹配的精度。本文拼接整

体流程如图 1 所示。

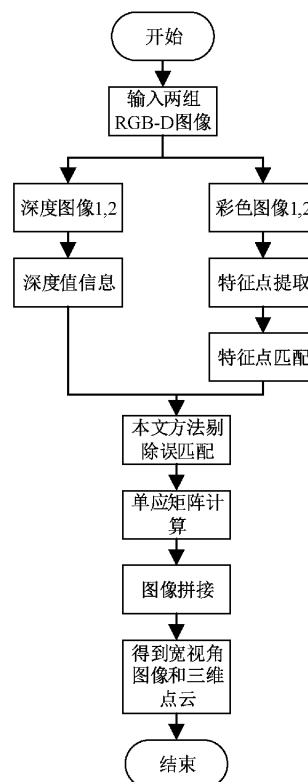


图 1 拼接整体流程

1.2 彩色图像对应关系

图 2 展示了实际场景中会遇到的情况。可以看到同一物体在不同视角中存在着对应关系,比如左侧泰迪熊与右侧泰迪熊对应,而在左侧泰迪熊身上提取的特征点也近似对应于右侧泰迪熊身上的特征点。所以当同一物体在不同视角下对应时,这个物体面上提取的特征点也满足相应的关系,即特征点连线斜率应该接近整个物体面连线斜率,同一物体所有的像素集合正是本文定义的物体面。

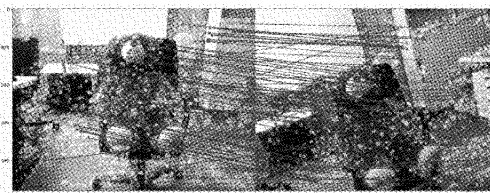


图 2 彩色图像示例

彩色图像特征点坐标分别记为 (x_i^l, y_i^l) 、 (x_i^r, y_i^r) , 按式(1)计算出左侧图像和右侧图像中特征点连线斜率, 式(2)为特征点连线与水平线夹角。当这个夹角在某个区间时才能构成合理的匹配, 正确角度应该在物体连线平均值 θ_{mean} 附近, 式(3)为两幅图像中所有特征点匹配连线夹角平均值。本文方法设置的角度范围为 $[\theta_{\text{mean}} - 5^\circ, \theta_{\text{mean}} + 5^\circ]$, 在这个角度范围外的被认为是误匹配, 需要从初始匹配中剔除。

$$k_i = \frac{y_i^r - y_i^l}{x_i^r - x_i^l + W} \quad (1)$$

式中: k_i 表示第 i 对匹配连线斜率, i 为第 i 对匹配, l 为 left 左侧图像, r 为 right 右侧图像, $(x_i^l, y_i^l), (x_i^r, y_i^r)$ 为匹配的特征点坐标, W 为图像宽度。

$$\theta_i = \arctan \frac{y_i^r - y_i^l}{x_i^r - x_i^l + W} \quad (2)$$

式(2)为式(1)中斜率 k_i 所在匹配连线夹角。

$$\theta_{\text{mean}} = \frac{1}{n} \sum_{i=1}^n \theta_i = \frac{1}{n} \sum_{i=1}^n \arctan \frac{y_i^r - y_i^l}{x_i^r - x_i^l + W} \quad (3)$$

式(3)为两幅图像中所有特征点匹配连线夹角平均值, n 为匹配数量。

1.3 深度图像对应关系

可视化后的深度图像如图 3 所示, 展示了缩放时深度图像的关系。左侧深度图像离物体较远, 特征点深度值大于右侧深度图像。当两幅图片无缩放即缩放倍率为 1 时, 匹配的特征点邻域深度值应该接近。当两幅图片有缩放时, 放大的物体面深度值总是小于小的物体面, 所以可以通过深度值来剔除误匹配。而基于特征匹配算法计算出的特征点坐标不同于像素值坐标, 所以需要特征点坐标进行取整和转置处理, 如式(4)所示, 单应矩阵各元素含义如式(5)和表 1 所示。

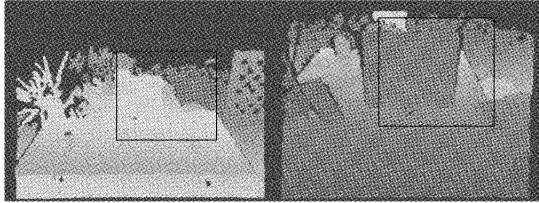


图 3 深度图像示例

本文选择的是向下取整, 处理后彩色图像中的 $(x_i^l, y_i^l), (x_i^r, y_i^r)$ 分别对应深度图像矩阵中的 $(\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor), (\lfloor y_i^r \rfloor, \lfloor x_i^r \rfloor)$, 从而得到两个深度矩阵中的 $(\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor)$ 和 $(\lfloor y_i^r \rfloor, \lfloor x_i^r \rfloor)$ 处深度值, 为了防止对齐中出现误差, 分别计算 $(\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor)$ 、 $(\lfloor y_i^r \rfloor, \lfloor x_i^r \rfloor)$ 和以 $(\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor), (\lfloor y_i^r \rfloor, \lfloor x_i^r \rfloor)$ 为中心的周围 8 个点的平均深度值 d_{mean}^l 和 d_{mean}^r , 该值可表示为特征点邻域内的平均深度。当其中较大与较小的比值不在缩放比率区间内, 则认为特征点的深度值误差过大, 需要从初始匹配中剔除。 d_{mean}^l 和 d_{mean}^r 如式(6)所示。

$$\begin{bmatrix} x_i^l \\ y_i^l \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} x_i^r \\ y_i^r \\ 1 \end{bmatrix} \quad (4)$$

式中: 右侧图像特征点坐标 (x_i^r, y_i^r) 通过单应矩阵变换到左侧图像的 (x_i^l, y_i^l) , $\mathbf{H}$ 代表单应矩阵^[16]。

$$\mathbf{H} = \begin{pmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{pmatrix} \quad (5)$$

表 1 单应矩阵参数表

单应矩阵参数				含义
m_0	m_1	m_3	m_4	尺度变换和旋转
m_2				水平位移
m_5				垂直位移
m_6 m_7				水平和垂直变形

$$\begin{cases} d_{\text{mean}}^l = \frac{1}{9} \sum_{\delta=0}^1 f((\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor) - \delta, (\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor) + \delta), \mathbf{l}_{\text{matrix}} \\ d_{\text{mean}}^r = \frac{1}{9} \sum_{\delta=0}^1 f((\lfloor y_i^r \rfloor, \lfloor x_i^r \rfloor) - \delta, (\lfloor y_i^r \rfloor, \lfloor x_i^r \rfloor) + \delta), \mathbf{r}_{\text{matrix}} \end{cases} \quad (6)$$

式中: δ 为特征点邻域半径, $(\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor)$ 为 (x_i^l, y_i^l) 向下取整后、转置后的坐标。

$f((\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor) - \delta, (\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor) + \delta), \mathbf{l}_{\text{matrix}}$ 表示 $((\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor) - \delta, (\lfloor y_i^l \rfloor, \lfloor x_i^l \rfloor) + \delta)$ 点在左侧深度图像矩阵 $\mathbf{l}_{\text{matrix}}$ 中的深度值, l 为 left 左侧图像, matrix 表示为深度图像的矩阵形式, d_{mean}^r 同 d_{mean}^l 一样。

2 实验方法

2.1 实验环境

为了验证本文方法在误匹配剔除上的效果, 进行了原算法与添加本文方法下原算法的对比实验, 实验分为有旋转、有形变下拼接和有缩放下的拼接。本文实验平台为 Windows10 操作系统, CPU i5-10600kf, 内存 16GB, CPU GTX2060, 编程环境为 Python3.7 和 3.8。所选取的数据来源是慕尼黑工业大学的 3D 对象重建数据集^[17]。选取其中的 Freiburg1_plant 和 Freiburg2_xyz 数据集。此类数据集包含有形变、旋转、缩放的 640×480 的 RGB-D 图像, 能满足实验要求。

2.2 评价指标

为了验证本文方法的有效性, 选取了匹配数量、均方根误差(RMSE)2 个指标进行比较, 均方根误差公式如式(7)所示。

$$R_{\text{mse}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \| (m_0 x_i^r + m_1 y_i^r + m_2, m_3 x_i^r + m_4 y_i^r + m_5) - (x_i^l, y_i^l) \|^2} \quad (7)$$

式中: $(m_0 x_i^r + m_1 y_i^r + m_2, m_3 x_i^r + m_4 y_i^r + m_5)$ 坐标为 (x_i^l, y_i^l) 通过式(4)和(5)投影到左侧图像的预测值坐标, $RMSE$ 为预测值坐标与真实值坐标 (x_i^l, y_i^l) 偏差的平方与匹配数量 n 比值的平方根, 公式中用 R_{mse} 表示。

3 实验结果与分析

实验分为有旋转、有形变、无缩放下图像拼接和仅缩放下图片拼接两种。SIFT 和 SURF 算法使用最近邻近匹配器, ORB 使用暴力匹配器匹配特征点, 欧式距离比值都设置为 0.7。在求解单应矩阵方法的选择上, 实验 4.1 使用最小

二乘法求解单应矩阵,能明显展示本文方法下和原算法之间的对比。而实验 4.2 需要使用随机抽样一致性算法(RANSAC)来精确优化模型参数,原因是 ORB 算法只能提取 500 个特征点,过少过于紧密的特征点会导致拼接失败。

3.1 有旋转、有形变、无缩放图像拼接效果分析

Freiburg1_plant 数据集中选取的图像在本文方法下与原算法的对比结果如表 2 和 3 所示,由表 2 匹配数量可得,添加本文方法能剔除 41%、29%和 52%的误匹配。在表 3 均方根误差上,可见本文方法较原算法的均方根误差减少了 5%、27%和 33%。

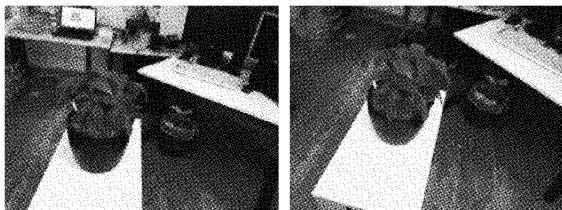
表 2 基于特征匹配算法和本文方法下匹配对比结果

算法	特征点数量		匹配数量
	左图	右图	
SIFT	1 180	878	130
SIFT+Ours	1 180	878	76
SURF	1 737	1 452	223
SURF+Ours	1 737	1 452	158
ORB	500	500	134
ORB+Ours	500	500	64

表 3 基于特征匹配算法和本文方法下均方根误差对比结果

算法	SIFT	SIFT+Ours	SURF	SURF+Ours	ORB	ORB+Ours
RMSE/pixel	259.76	245.15	242.68	176.08	506.11	338.41

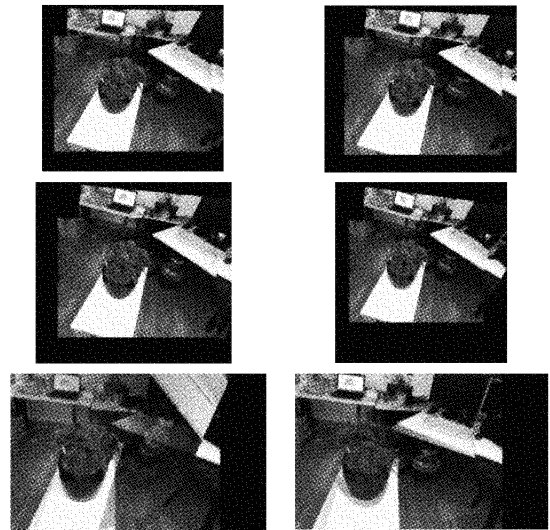
图 4 为实验选取的彩色图像,左侧彩色图像如图 4(a)所示,右侧彩色图像如图 4(b)所示。由图 5(a)和(b)可知,原算法会出现明显的误匹配,特征是斜率异常,不满足不同视角下物体对应的条件。由图 5(c)和(d)可知,本文方法拼接效果明显优于 SURF 和 ORB 算法,表现为右下角的桌角拼接更为准确。但本文方法跟 SIFT 算法仅在左上角的桶有细小区别,这也与表 3 中两者均方根误差很小吻合。不管是实验数据还是拼接效果,都验证了本文方法在有旋转、有形变、无缩放图像上的有效性。



(a) 左侧彩色图像

(b) 右侧彩色图像

图 4 彩色图像



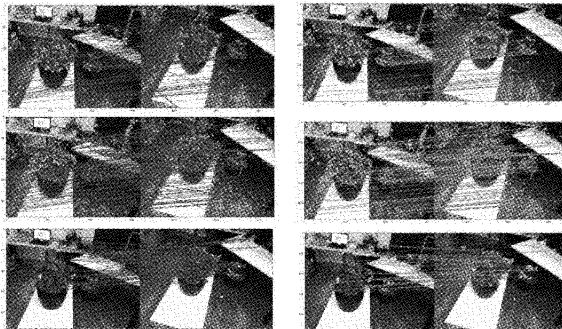
(c) SIFT、SURF、ORB
算法拼接效果

(d) 添加本文方法拼

图 5 特征点匹配和彩色图像拼接

3.2 仅缩放下图像拼接效果分析

Freiburg2_xyz 数据集中选取的图像在使用本文方法和不使用本文方法的对比结果如表 4 和 5 所示。从表 4 匹配数量可见,添加本文方法较原算法在仅有缩放下能剔除



(a) SIFT、SURF、ORB 算法匹配

(b) 添加本文方法后匹配

表 4 基于特征匹配算法和本文方法下匹配对比结果

算法	特征点数量		匹配数量
	左图	右图	
SIFT	1 672	1 574	449
SIFT+Ours	1 672	1 574	209
SURF	2 048	1 868	614
SURF+Ours	2 048	1 868	261
ORB	500	500	252
ORB+Ours	500	500	122

53%、57%和 51%的误匹配。在表 5 均方根误差上,本文方法较原算法的均方根误差减少了 14%、17%和 28%。

图 6 为选取的彩色和深度图像,彩色图像如图 6(a)、(b)所示,可视化深度图像如图 6(c)、(d)所示。从图 7(a)

表 5 基于特征匹配算法和本文方法下方根误差对比结果

算法	SIFT	SIFT+Ours	SURF	SURF+Ours	ORB	ORB+Ours
RMSE/pixel	323.63	274.79	291.98	239.01	509.37	362.70

和(b)可知,原算法在远处地面提取的特征点和桌面上的特征点组成错误的匹配,究其原因是原算法只关注是否为最佳匹配,却不关心特征点是不是来自一个物体。从图 7(c)和(d)来看, SIFT 与 SURF 算法因为使用了 RANSAC,所以在拼接效果上看不出明显的差异。但原 ORB 算法较本文方法有很大的差别,原因是原 ORB 算法提取特征点数量本身就少,加之缩放的图像特征点过于紧密,最终求解出异常的单应矩阵。从另一个角度看,本文方法能防止这类情况发生。

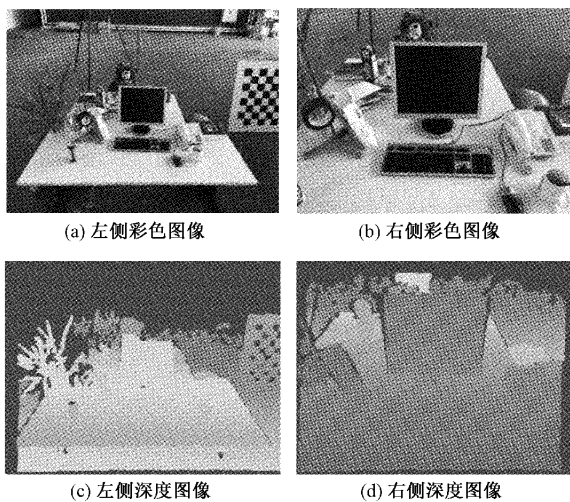


图 6 彩色图像和可视化深度图像

3.3 三维重建效果分析

为了更好展示本文方法在三维重建上的效果,利用 RGB-D 图像与三维点云相互转换的特点,再加上相机内参来重建三维模型。图 8 为数据集中选取的 RGB-D 图像,彩色图像如图 8(a)、(b)所示,可视化深度图像如图 8(c)、(d)所示。图 9(a)是基于特征匹配算法下不添加本文方法拼接

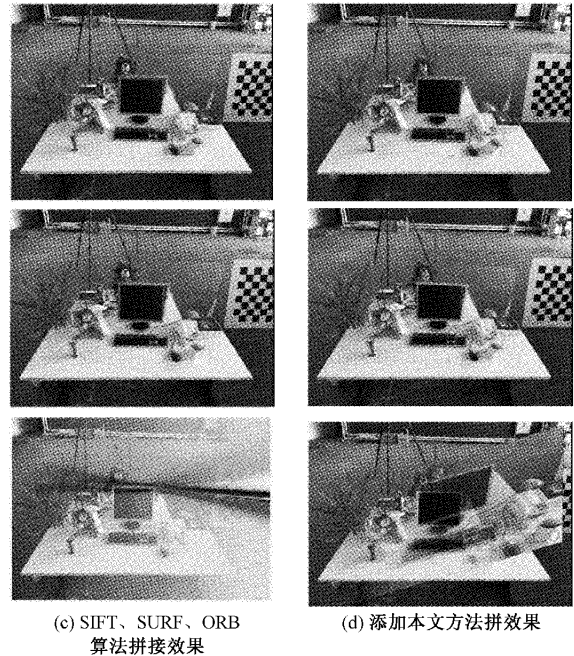
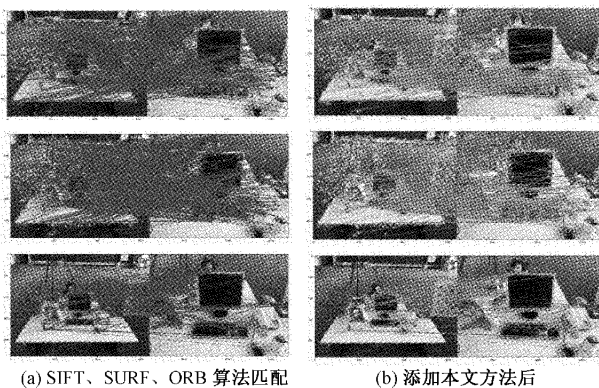


图 7 特征点匹配和图像拼接

的 RGB-D 图像三维重建后的效果,图 9(b)是添加本文方法后的效果。可以明显看出,在本文方法剔除误匹配后,电脑、键盘和可乐都较原算法配准更好。所以无论是从指标对比还是实验效果来看,本文方法都是可行的,通过剔除误匹配这一方法在单应矩阵匹配精度、图像拼接效果和三维重建上均有提升。

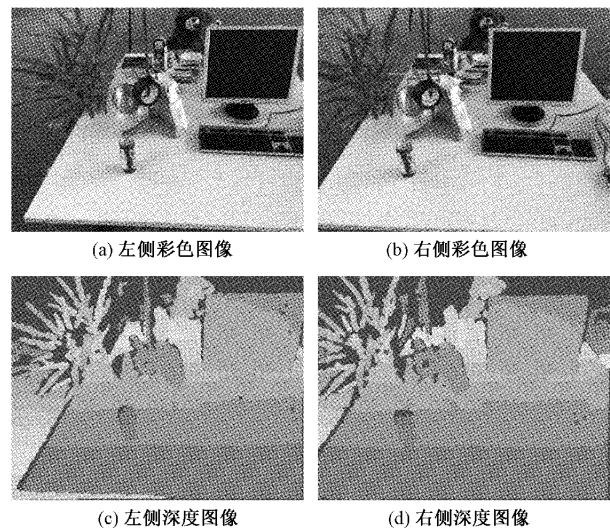
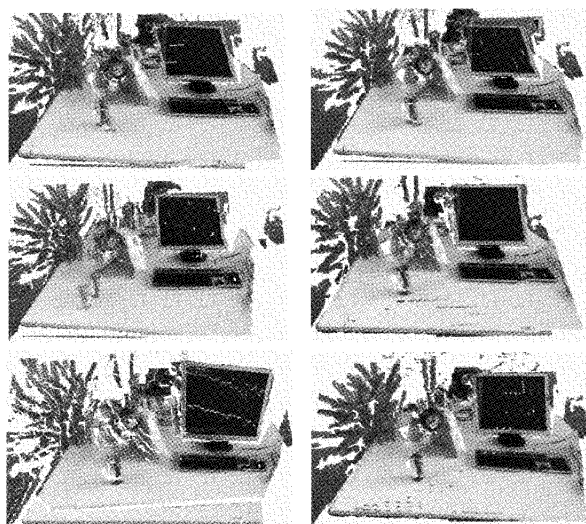


图 8 RGB-D 图像

(a) SIFT、SURF、ORB
算法下三维重建

(b) 本文方法下三维重建

图 9 三维重建对比

4 结 论

本文针对深度图像难以直接拼接的问题,研究了一种基于物体面对应的 RGB-D 图像拼接优化方法。添加本文方法后的算法,可以根据不同视角同一物体对应的关系来剔除误匹配。通过实验验证该方法能获取更为准确的单应矩阵,使图像拼接更为准确。最终通过实验仿真得到拼接后的图像和点云数据,通过多组实验证明了添加本文方法可以得到更好的三维重建效果。接下来将结合传统算法和神经网络进行展开与研究,研究如何通过神经网络赋予传统算法物体识别能力。

参考文献

- [1] 郑太雄,黄帅,李永福,等. 基于视觉的三维重建关键技术综述[J]. 自动化学报,2020,46(4):631-652.
- [2] 李秀艳,马会梅,王琦,等. 基于轮廓形状的肺部电阻抗成像三维插值方法[J]. 电子测量与仪器学报,2021,35(10):89-97.
- [3] 李双双,赵全堂,曹树春,等. 基于代数重建算法的高能电子三维成像研究[J]. 原子核物理评论,2021,38(1):73-79.
- [4] 郑立华,王露寒,王敏娟,等. 基于 Kinect 相机的油菜菜自动化三维点云重建[J]. 农业机械学报,2021,52(7):159-168.
- [5] 原志路,李游,汤圣君,等. 基于 RGB-D 传感器的室内三维建模及应用研究综述(英文)[J]. Frontiers of Information Technology & Electronic Engineering, 2021, 22(6): 815-827.
- [6] 谢亮亮,屠大维,张旭,等. 深海原位激光扫描双目立体视觉成像系统[J]. 仪器仪表学报,2020,41(6):106-114.
- [7] 耿璇,王召巴,金永,等. 基于单个深度相机的人体三维重建[J]. 国外电子测量技术,2020,39(9):143-146.
- [8] 孙艳丽,周伟,张亚周,等. 基于 SIFT 的航空侦查图像拼接算法[J]. 电子测量技术,2020,43(19):112-116.
- [9] 胡涛,茅健. 基于降采样和改进 Shi-Tomasi 角点检测算法的 PCB 图像拼接[J]. 电子测量技术,2021,44(22):134-140.
- [10] ELANK, ALWAN A H S, HUSSEIN A A. A review over panoramic image stitching techniques[J]. Journal of Physics: Conference Series, 2021,1999(1),DOI:10.1088/1742-6596/1999/1/012115.
- [11] ZHAO Y, GAO F, YU J, et al. Underwater image mosaic algorithm based on improved image registration [J]. Applied Sciences, 2021, 11(13): 5986.
- [12] HOSSEIN N Z, AGAHI H, MAHMOODZADEH A. Image matching based on the adaptive redundant keypoint elimination method in the SIFT algorithm[J]. Pattern Analysis and Applications,2021, 24(2): 669-683.
- [13] 刘佳,傅卫平,王雯,等. 基于改进 SIFT 算法的图像匹配[J]. 仪器仪表学报,2013,34(5):1107-1112.
- [14] 沈跃,朱嘉慧,刘慧,等. 基于深度和彩色双信息特征源的 Kinect 植物图像拼接[J]. 农业工程学报,2018,34(5):176-182.
- [15] 陈伟,刘宇,王亚伟,等. 基于改进的 FAST-SURF 快速图像拼接算法[J]. 应用光学,2021,42(4):636-642.
- [16] 李小雨. 全景图像拼接的关键技术研究及实现[D]. 上海:华东师范大学,2017.
- [17] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]. Intelligent Robots and Systems(IROS), 2012 IEEE/RSJ International Conference on,2012.

作者简介

于广旺,硕士研究生,主要研究方向为深度图像处理。

E-mail:1244182951@qq.com

杨家志,博士,教授,主要研究方向为信号处理与检测 and 智能信息处理。

E-mail:jiaazhi_yang@126.com

陈梦强,硕士研究生,主要研究方向为机器学习、激光雷达。

E-mail:1312681533@qq.com

沈洁,硕士研究生,主要研究方向为点云数据处理。

E-mail:1215924848@qq.com