

# 基于空间信息聚类的 RGB-D 数据拼接融合

李文月, 何迪\*, 赵爽, 刘畅, 周哲海

北京信息科技大学, 北京 100192

**摘要** 针对单一成像系统获取 RGB-D 数据时视场过小难以满足大视场成像需求的问题, 提出了基于空间信息聚类的 RGB-D 数据拼接融合方法。根据 RGB-D 数据中蕴含的空间信息定义物点距离, 利用 simple linear iterative clustering (SLIC) 方法实现空间信息聚类。将场景分割成若干平面子块, 每个子块具有单应性, 可以进行准确的单应性矩阵计算, 进而实现将小视场 RGB-D 数据准确拼接融合生成大视场 RGB-D 数据。实验结果表明, 所提方法能够改善图像变换过程中产生的扭曲变形及拼接过程中在重叠区域产生的错位。峰值信噪比 (PSNR) 和结构相似性 (SSIM) 两个定量评价指标说明基于空间信息聚类的 RGB-D 数据拼接相较于全局拼接, 结果质量得到了提升。

**关键词** 成像系统; 计算成像; 大视场; RGB-D 数据; 空间信息聚类; 单应性

中图分类号 TN209

文献标志码 A

DOI: 10.3788/LOP202259.1011004

## RGB-D Data Stitching Based on Spatial Information Clustering

Li Wenyue, He Di\*, Zhao Shuang, Liu Chang, Zhou Zhehai

Beijing Information Science & Technology University, Beijing 100192, China

**Abstract** To address the problem of a small field of view of RGB-D data obtained using a single imaging system, where a large field of view is required, an RGB-D data stitching method is proposed based on spatial information clustering. Based on the spatial information present in RGB-D data, the distance between object points is defined to realize spatial information clustering using a simple linear iterative clustering (SLIC) on the RGB-D data. The scene is divided into several planar sub-blocks. Each sub-block shows homography, which can be used to accurately determine the homographic matrix and then realize the accurate splicing and fusion of small-field RGB-D data to generate large-field RGB-D data. Results of a real scene-based experiment shows that the proposed method can decrease the distortion during image warping and reduce the dislocation in overlap regions during stitching. Based on the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) values, the efficiency of RGB-D data stitching based on spatial information clustering is quantitatively shown to improve compared with global stitching.

**Key words** imaging systems; computational imaging; large field of view; RGB-D data; spatial information clustering; homography

## 1 引言

传统 RGB 图像是对三维场景进行二维投影得到的结果, 丢失了深度信息, 难以满足实际应用中

对于深度信息重建、三维场景重构的需求<sup>[1]</sup>, RGB-D 数据包含了场景中物点的深度信息, 能够应用于更广泛的计算成像问题中。获取 RGB-D 数据的方法可以分为直接拍摄和间接计算两种方式, 直接拍

收稿日期: 2021-03-31; 修回日期: 2021-05-14; 录用日期: 2021-05-28

基金项目: 国家自然科学基金(61931003, 62001036, 61871042)、北京长城学者支持计划(CIT&TCD20190323)、北京青年拔尖人才支持计划(Z2019042)、北京信息科技大学校基金(1925026)

通信作者: \*dihe@bistu.edu.cn

摄可以由消费级 RGB-D 相机同时拍摄场景的纹理信息和深度信息(如微软 Kinect 系列、华硕 Xtion 系列等)<sup>[2-5]</sup>,但需要对其 RGB 数据与深度数据进行进一步配准;间接计算可以通过光场成像等计算成像技术<sup>[6]</sup>获取场景准确配准的 RGB-D 数据<sup>[7-8]</sup>。由于硬件限制,单一成像系统获得的 RGB-D 数据视场通常都比较小,难以满足许多重要应用场景对大视场成像的需求<sup>[9]</sup>。为了获取大视场的 RGB-D 数据,一类重要的方法是对小视场数据进行准确的拼接。

2017 年, Li 等<sup>[10]</sup>对 Kinect 相机获得的 RGB-D 数据的深度图进行预处理,对预处理结果和 RGB 图进行配准,对 RGB 图进行特征点的提取和单应性变换矩阵的计算,并对深度图和 RGB 图进行相同的变换。由于单应性描述的是场景中一个平面在成像过程中的变化,只能准确变换对应同一平面场景的图像,当场景不满足平面假设时,此方法会在重叠部分产生重影和错位,在非重叠部分也会产生一定程度的扭曲和变形。通过准确的相机标定可以将 RGB-D 数据转换为 3D 点云数据,然后利用迭代最近点(ICP)或其相关变形算法<sup>[11-14]</sup>实现多组点云数据的配准与拼接,但是此类方法需要数据之间存在较多的重叠部分。2019 年, Ding 等<sup>[15]</sup>基于 2D 特征点匹配进行 3D 点云数据的配准。张元等<sup>[16]</sup>提出了一种将聚类区域分块和凸优化问题相结合的点云配准方法。上述两种方法对低重叠率(低于 60%)的两组三维点云提出了配准方法,分别对重叠率达 30% 和 40% 的点云数据进行配准,但在形成 3D 点云的过程中仍需要很准确的相机参数标定。

本文提出了基于空间信息聚类的 RGB-D 数据拼接方法。根据 RGB-D 数据中蕴含的空间信息,在不利用准确标定的相机参数将 RGB-D 数据转化为点云数据的情况下,定义物点距离,并利用 simple linear iterative clustering (SLIC) 方法<sup>[17]</sup>实现 RGB-D 数据的空间信息聚类,将场景分割成满足平面单应性的子块。对于每一个平面子块,所提方法可以准确计算其单应性变换矩阵,使得数据在坐标变换过程中产生的扭曲变形现象及拼接过程中产生的错位现象能够得到改善,提高拼接结果的质量。

## 2 基于空间信息聚类的场景平面子块分割

单应性<sup>[18]</sup>变换描述了平面场景物点在成像过

程中的坐标变化,由于场景通常不是一个平面,所以全局单应性矩阵无法准确描述场景物点的坐标变化,会在拼接过程中产生扭曲变形和错位。RGB-D 数据蕴含了场景中物点的空间信息,据此定义物点距离,可以对场景物点进行空间信息聚类。将场景分割成若干平面,每个平面在不同成像系统下的坐标变化可以由一个单应性矩阵准确描述。精确计算每个平面的单应性变换矩阵后可以实现准确的 RGB-D 数据拼接。

配准的 RGB-D 数据蕴含了场景中物点的空间信息,由于 RGB 图像中的像素距离和深度图像中的深度值量纲不同,对于场景中任意两个物点  $p_i, p_j$ , 定义包含两项的用于空间聚类的物点距离:

$$D(p_i, p_j) = D_d(p_i, p_j) + \alpha D_p(p_i, p_j), \quad (1)$$

式中:  $D_p(p_i, p_j)$  为两个物点在 RGB 图像中的像素坐标欧氏距离;  $D_d(p_i, p_j)$  为两个物点的深度信息距离;  $\alpha$  是平面参数,控制平面距离影响的权重。由于计算得到的深度图通常在物体边缘处存在误差,而 RGB 图对于边缘的保持有更高的精度,利用 RGB 图的颜色信息对深度信息距离进行校正能够得到更准确的聚类结果,因此定义物点深度信息距离:

$$D_d(p_i, p_j) = d_{is_1}(p_i, p_j) + \beta d_{is_2}(p_i, p_j), \quad (2)$$

式中:  $d_{is_1}(p_i, p_j)$  是两个物点 RGB 图像中的三通道颜色欧氏距离;  $d_{is_2}(p_i, p_j)$  为两个物点的深度欧氏距离;  $\beta$  是深度参数,控制深度距离影响的权重。

式(1)与式(2)中的平面参数  $\alpha$  和深度参数  $\beta$  需要根据不同场景的特点进行调整,以能够较好地保持物体边界的形状,并且能够将场景分割成若干平面。 $\alpha$  取值越大时,聚类结果在 RGB 图像中的形状越接近矩形,当场景中物体形状不规则时,应适当减小  $\alpha$  的取值,以更好地保持物体边界处的形状。 $\beta$  取值越大时,聚类结果中每一类内部物点深度越趋于一致,当深度图计算较准确,且场景中深度信息近似分片常数时,可以适当增大  $\beta$  的取值;当 RGB-D 数据中深度信息计算精度较低,或场景中存在较多深度线性变化的物体时,可以适当减小  $\beta$  的取值,利用物点的颜色信息来对聚类进行校正。

利用 SLIC 算法,根据式(1)所定义的距离对场景物点进行聚类,可以实现 RGB-D 数据的空间信息的聚类,形成对场景的平面子块分割。每一个子块具有平面单应性,可以准确计算单应性变换矩

阵。由于平面子块分割结果在 RGB 图和深度图中可以体现为超像素分割,对于 RGB 图和深度图中的超像素,运用对应子块的单应性变换矩阵进行坐标变换可以实现准确的拼接融合,得到大视场 RGB-D 数据。

### 3 基于平面子块的 RGB-D 数据拼接融合

对于配准的 RGB-D 数据,基于平面子块的 RGB-D 数据拼接融合流程如图 1 所示。

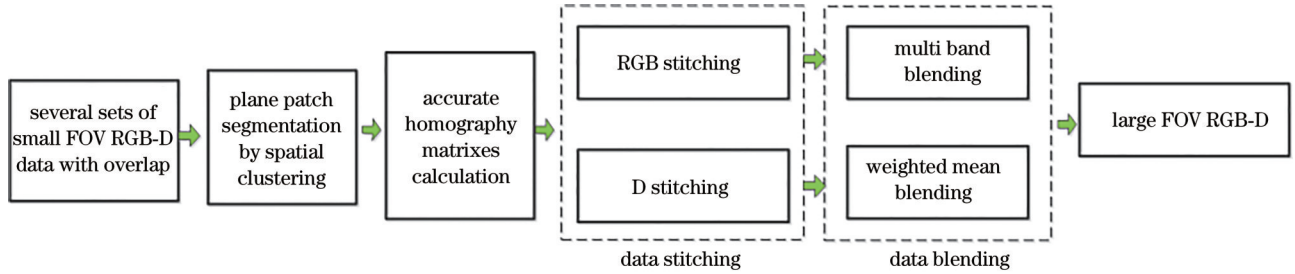


图 1 RGB-D 数据拼接融合流程图

Fig. 1 Flow chart of RGB-D data stitching

#### 3.1 平面子块单应性矩阵的准确计算

利用所提方法对场景进行空间信息聚类,可以形成平面子块,每一个平面子块都满足单应性。可以准确计算单应性矩阵,通过计算 RGB 图像中超像素的单应性矩阵来描述平面子块在不同成像系统下的坐标变换。对多组 RGB-D 数据的 RGB 图像进行 scale-invariant feature transform (SIFT)<sup>[19]</sup> 特征点提取与匹配,并经过 Random Sample Consensus (RANSAC)<sup>[20]</sup> 剔除离群点,获得能够用于计算单应性矩阵的特征点对。对于同一个子块内特征点在不同 RGB 图像中对应的齐次坐标  $[x, y, 1]^T, [x', y', 1]^T$ , 变换关系为

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}, \quad (3)$$

式中:该子块单应性矩阵  $\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$ 。

令  $\mathbf{h} = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]^T$ , 可以根据  $n$  对匹配特征点坐标建立线性方程组:

$$\mathbf{A}_{2n \times 9} \mathbf{h} = 0, \quad (4)$$

通过最小二乘法求解单应性矩阵。

若每一个子块内部存在足够多的准确匹配的特征点对,则利用这些准确匹配的特征点对进行系数矩阵  $\mathbf{A}$  的构造,计算准确的单应性矩阵。若无法找到足够多的准确匹配的特征点,则利用所有匹配的特征点对加权构造系数矩阵  $\mathbf{A}$ 。在基于超像素的传统 2D 图像拼接算法中,利用第  $i$  个特征点  $(x_i, y_i)$  到第  $k$  个超像素中心点  $(x_k, y_k)$  的平面距离进行

加权<sup>[21]</sup>:

$$W_{ik} = \max\left(e^{-\frac{d_{ik}^2}{\sigma^2}}, \eta\right), \quad (5)$$

式中:  $d_{ik} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}$ ;  $\sigma$  和  $\eta$  为两个预设参数。对于平面子块,由于深度信息已知,选择以特征点的深度和子块中心点的深度距离作权重来构造系数矩阵  $\mathbf{A}$ ,即将式(5)中的距离  $d_{ik}$  定义为第  $i$  个特征点和第  $k$  个平面子块中心点的深度距离,  $d_{ik} = \sqrt{(d_i - d_k)^2}$ , 这样距离越近的特征点越近似来自与平面子块相同的平面,权重越大,单应性矩阵的计算越准确。

#### 3.2 RGB-D 数据的拼接

待拼接的 RGB 图和深度图中的每一对超像素都对应了同一个空间中的平面子块,运用相应子块单应性变换矩阵,分别对 RGB 超像素和深度超像素进行坐标变换,可以实现 RGB-D 数据的拼接。不同超像素之间会出现重叠和黑洞两种需要特殊处理的情况。图 2 为超像素变换示意图,左图为空间信息聚类后 RGB 图像中对应的超像素分割结果,右图是经过变换之后的图像,超像素 5 和超像素 6 之间产生了重叠,超像素 1, 2, 4 和 5 之间产生了黑洞。

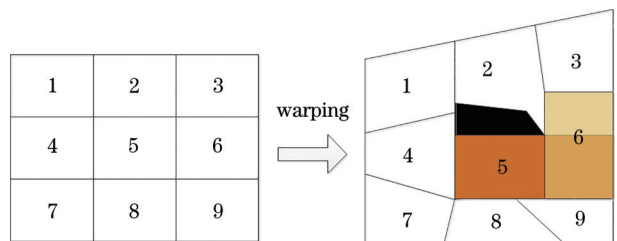


图 2 超像素变换示意图

Fig. 2 Super-pixel warping



1) 重叠部分

对于重叠部分,采用插值的方式进行处理,待变换图像坐标经单应性变换矩阵变换后得到新的坐标,此坐标通常不是整数,该点的像素值将用于生成变换后图像其相邻四个像素的像素值。最后遍历变换后图像的每个像素位置,将与其相邻的非整数坐标点的像素值按照距离加权求和生成该像素的像素值,距离越大权重越小。具体算法如图 3 所示。

对于变换后图像中的一个像素  $O$ ,若有  $a, b, c, d, e$  五个变换得到的非整数坐标点与之相邻,该像素值由这五个非整数坐标点的像素值加权求和生成。 $d_a, d_b, d_c, d_d, d_e$  为  $a, b, c, d, e$  到  $O$  点的欧氏距离,假设距离从小到大的排序为

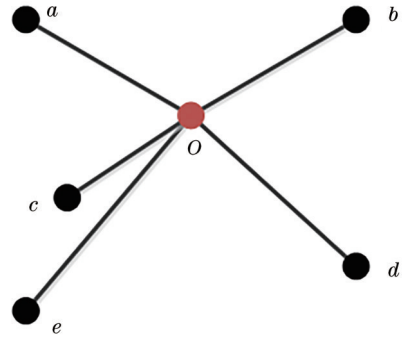


图 3 重叠部分插值示意图

Fig. 3 Schematic of overlap interpolation

$d_c, d_a, d_b, d_d, d_e, p_a, p_b, p_c, p_d, p_e$  为五个非整数坐标点对应的像素值,则  $O$  的像素值为

$$p_o = \frac{(d_d \times p_a) + (d_b \times p_b) + (d_e \times p_c) + (d_a \times p_d) + (d_c \times p_e)}{d_a + d_b + d_c + d_d + d_e} \quad (6)$$

2) 黑洞部分

对于黑洞部分,遍历变换后图像的所有像素,找到像素值为零的位置,即黑洞点。为了提高效率和准确性,对搜索区域进行限制,以该像素为起点,在  $S \times S$  区域向四个方向搜寻,直到找到不为零的像素为止,并以距离为权重对该像素进行赋值。对黑洞部分的处理如图 4 所示, $O$  为检测到的黑洞点, $U, L, D, R$  是沿黑洞点上下左右检测到的第一个不为零的像素, $D_1, R_1$  为黑洞点。

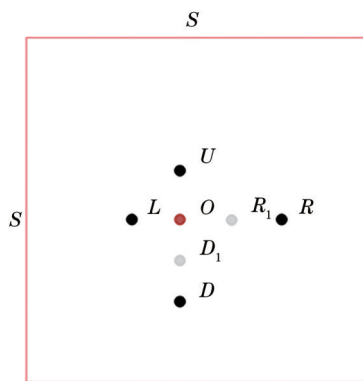


图 4 黑洞部分插值示意图

Fig. 4 Schematic of black hole interpolation

记  $U, L, D, R$  的像素值为  $p_U, p_L, p_D, p_R$ , 与  $O$  的欧氏距离分别为  $d_U, d_L, d_D, d_R$ , 假设距离从小到大排序为  $d_U, d_L, d_D, d_R$ , 则  $O$  点的像素值  $p_o$  为

$$p_o = \frac{(d_R \times p_U) + (d_D \times p_L) + (d_L \times p_D) + (d_U \times p_R)}{d_U + d_L + d_D + d_R} \quad (7)$$

对 RGB 图和深度图中变换后的超像素按照上述方法进行处理后,可以避免重叠部分颜色叠加造成的亮度过高及黑洞部分中的图像不连续。

3.3 RGB 图的多频段融合与深度图的加权平均融合

每个成像系统在获取小视场数据时光照情况可能不同,导致每幅 RGB 图像亮度不同,在 RGB 图像经过子块单应性矩阵进行坐标变换实现拼接后,结果看起来不自然。利用多频段图像融合的方法<sup>[22]</sup>对拼接图像进行融合可以使拼接结果更加自然。由于深度图记录的是场景的几何结构信息,不受光照强弱的影响,所以在融合时仅采用加权平均即可。这样经过准确变换及融合就得到了大视场的 RGB 图和深度图,即大视场 RGB-D 数据。

4 实验结果及分析

4.1 场景搭建和数据采集

为了验证所提方法的可行性和准确性,在光学平台上搭建了深度层不一致的实验场景,深度范围为 80~110 cm。对此场景进行拍摄,获得四个视点下的小视场 RGB-D 数据,如图 5 所示,此 RGB-D 数据的 RGB 图和深度图是经过准确配准的。同时采集一个完整实验场景的大视场 RGB 图像作为参考图像用来对拼接结果进行评价,如图 6 所示。

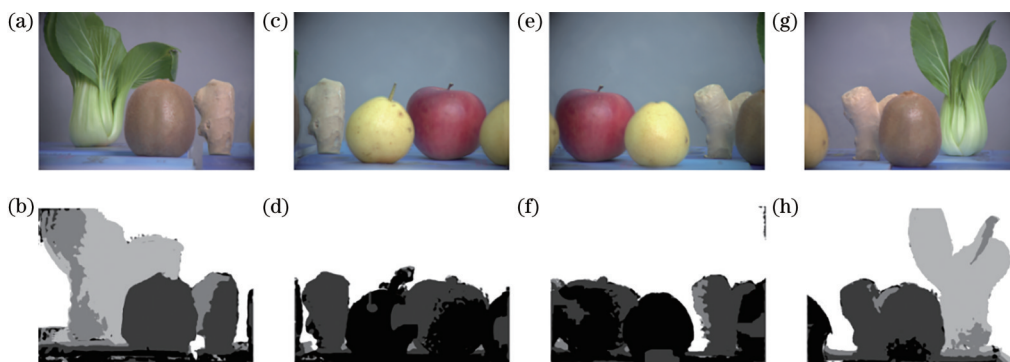


图5 四个视点下小视场RGB-D数据。(a)第一视点下的RGB图;(b)第一视点下的深度图;(c)第二视点下的RGB图;(d)第二视点下的深度图;(e)第三视点下的RGB图;(f)第三视点下的深度图;(g)第四视点下的RGB图;(h)第四视点下的深度图

Fig. 5 Small FOV RGB-D data acquired under four viewpoints. (a) The first view RGB image; (b) the first view depth map; (c) the second view RGB image; (d) the second view depth map; (e) the third view RGB image; (f) the third view depth map; (g) the fourth view RGB image; (h) the fourth view depth map

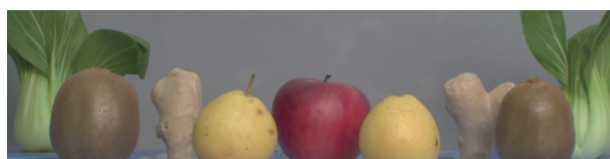


图6 大视场RGB图像

Fig. 6 Large FOV RGB image

#### 4.2 基于空间信息聚类的平面子块分割结果及分析

采用所提方法对四组小视场RGB-D数据进行空间信息聚类。其中,子块的个数设为50,  $\alpha$  设置为0.0001,  $\beta$  设置为8.5。其中第1视点和第4视点RGB-D数据空间信息聚类的结果在RGB图和深度图中体现为超像素分割,如图7所示。

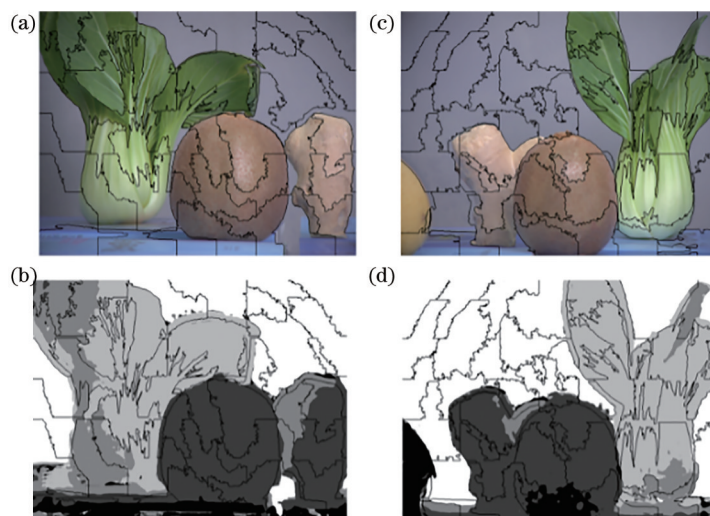


图7 第1个和第4个视点下RGB-D数据在RGB图和深度图中的空间信息聚类结果。(a)第1个视点下RGB图像分割结果;(b)第1个视点下深度图分割结果;(c)第4个视点下RGB图像分割结果;(d)第4个视点下深度图分割结果

Fig. 7 Spatial information clustering results of RGB-D data in RGB images and depth images of the first view and the fourth view. (a) Result of the first view RGB image segmentation; (b) result of the first view depth image segmentation; (c) result of the fourth view RGB image segmentation; (d) result of the fourth view depth image segmentation

子块的个数与权重系数  $\alpha$  和  $\beta$  的选取均对平面子块划分结果产生影响。若保持权重参数  $\alpha$  和  $\beta$  不变,减少或增加子块个数,结果如图8所示。图8(a)为将子块数设为20的深度图分割结果,图8(b)为将

子块数设为100的深度图分割结果。

由图8(a)方框区域可知,当子块个数设置过小时,每个子块内部深度差较大,导致拼接结果不准确。由图8(b)方框区域可知,当子块数设置过大

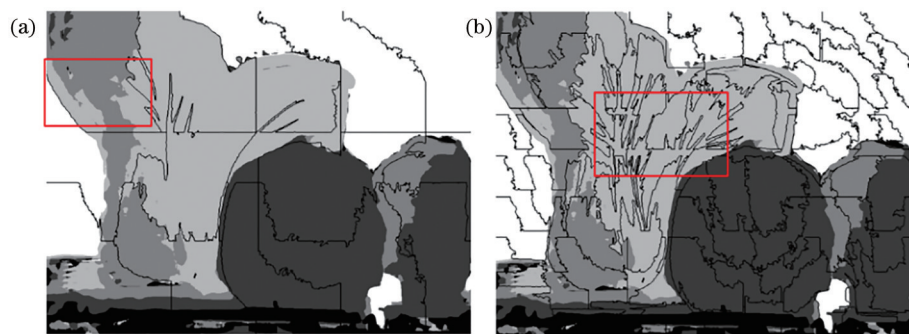


图 8 不同子块数的深度图分割结果。(a)子块数为 20;(b)子块数为 100

Fig 8 Depth map segmentation results with different number of sub-blocks. (a) 20 blocks; (b) 100 blocks

时,相邻子块处于相同深度,增加了计算时长。

若保持子块个数 50 和权重系数  $\beta$  不变,增大  $\alpha$  的大小,来验证  $\alpha$  对平面子块划分的影响, $\alpha=1$  时的深度图分割结果如图 9 所示。

由图 9 可知,当  $\alpha$  增大到 1 时,深度图的超像素形状规整。由于场景中的物体边缘不规则,并没有将 RGB-D 数据划分为平面子块。

若保持子块个数 50 和权重系数  $\alpha$  不变,减小  $\beta$  的大小,来验证  $\beta$  对平面划分的影响。图 10(a) 是  $\beta=0.1$  时的深度图分割结果,图 10(b) 是  $\beta=8.5$  时的深度图分割结果。将图 10(a) 和图 10(b) 的方框部分放大,得到的结果如图 10(c) 和图 10(d) 所示。



图 9  $\alpha=1$  时的深度图分割结果

Fig. 9 Depth map segmentation result with  $\alpha=1$

由图 10(c) 和图 10(d) 可知,相较于  $\beta=0.1$  时的深度图分割结果, $\beta=8.5$  时,所提方法可以将同深度或深度相差较小的像素划分到同一子块。

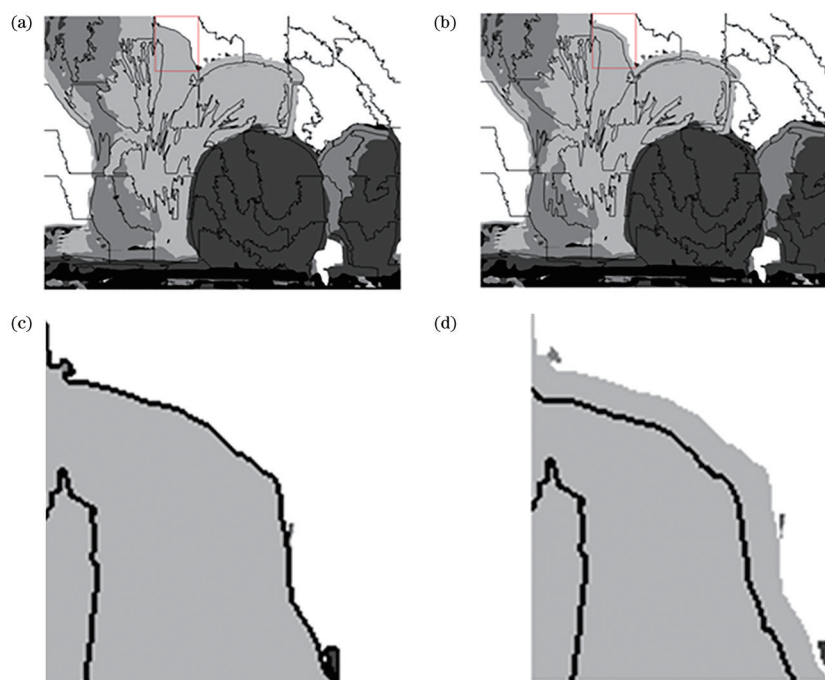


图 10 不同  $\beta$  下的深度图分割结果。(a)  $\beta=0.1$ ; (b)  $\beta=8.5$ ; (c)(d) 对应的局部放大图

Fig 10 Depth map segmentation results with different  $\beta$ . (a)  $\beta=0.1$ ; (b)  $\beta=8.5$ ; (c)(d) corresponding local enlargement



### 4.3 重叠、补黑洞处理结果及分析

得到平面子块后,将对每个子块进行单应性矩阵计算及变换,在变换过程中会产生重叠、黑洞现象。利用所提方法进行处理会得到效果很好的拼接结果。为了更好地展示所提方法在重叠和黑洞方面

的处理结果,只展示第二视点向第一视点的变换结果。图 11(a)为直接对分割后的图像逐子块进行坐标变换的结果,可以看到子块之间存在重叠和黑洞。图 11(b)为进行了重叠区域的处理但未进行黑洞处理的结果。图 11(c)为进行重叠、黑洞处理的结果。

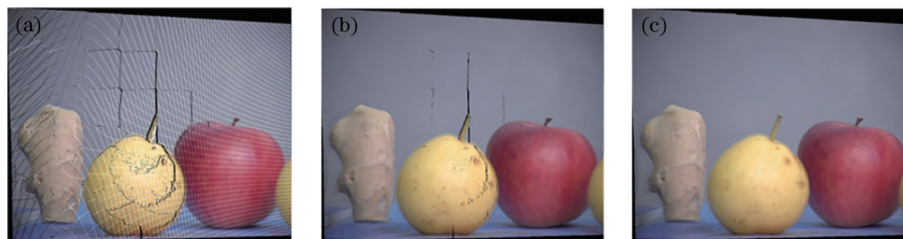


图 11 重叠、黑洞差值结果对比。(a)直接进行坐标变换后存在重叠和黑洞;(b)重叠插值但未处理黑洞的结果;(c)重叠、黑洞插值处理的结果

Fig. 11 Comparison of overlap and black hole interpolations. (a) Overlap and black hole after direct coordinate transformation; (b) result obtained by only overlap interpolation; (c) result obtained by overlap and black hole interpolations

由于每个平面子块的变换矩阵不同,多个像素变换到同一坐标后就会产生像素的重叠,也会产生变换后的子块之间存在黑洞的情形,若不进行处理,直接变换后将会得到如图 11(a)的结果,经过重叠、黑洞插值处理后可以得到如图 11(c)较为理想的结果。

### 4.4 不同拼接融合算法结果对比及分析

基于空间信息聚类,将多组 RGB-D 数据拼接成大视场 RGB-D 数据,结果如图 12 所示,经过重叠、黑



图 13 基于全局单应性变换的 RGB 拼接融合结果  
Fig. 13 Result of RGB stitching based on global homography transformation

洞插值处理后可以得到如图 11(c)较为理想的结果。油菜发生了严重扭曲变形。而在所提空间信息聚类的拼接结果中,如图 12(a)所示,错位和扭曲变形有明显的改善。

对所提方法的拼接结果与基于网格剖分的方法的拼接结果进行比较。在网格数设置为与本文平面子块数相近时,拼接结果如图 14(a)所示。

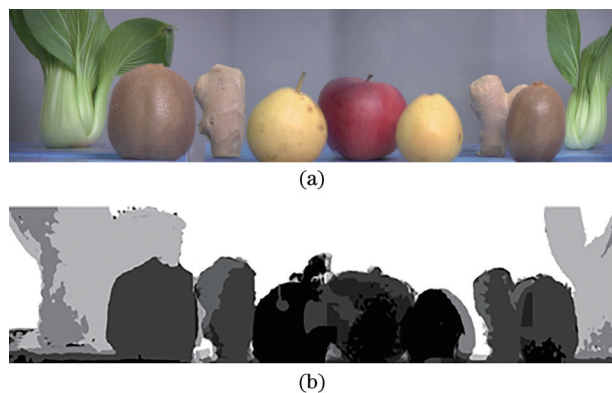


图 12 基于空间信息聚类的 RGB-D 数据拼接融合结果。(a)RGB 图拼接融合结果;(b)深度图拼接融合结果

Fig. 12 Results of RGB-D data stitching based on spatial information clustering. (a) Result of RGB image stitching; (b) result of depth image stitching

洞插值处理后可以得到如图 11(c)较为理想的结果。

对所提方法的拼接结果与基于全局单应性变换矩阵的拼接结果进行比较。图 13 为全局单应性矩阵拼接 RGB 图像的结果,可以明显看出,右侧生姜在两组数据拼接的位置处出现了明显错位,右侧

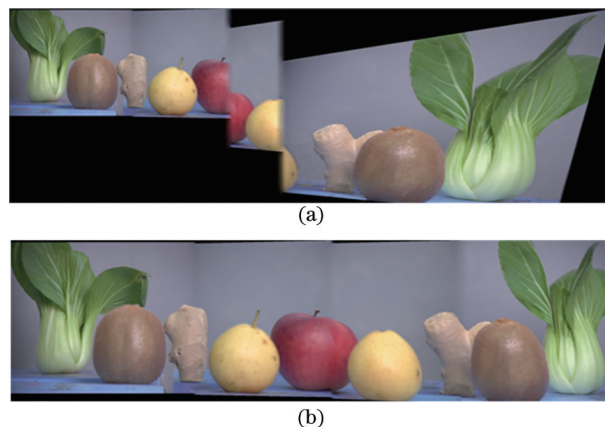


图 14 不同网格数下的拼接结果。(a)网格数为 49;(b)网格数为 1600

Fig. 14 Results of image stitching with different number of grids. (a) 49 grids; (b) 1600 grids

由图 14(a)可知,将网格数设置为 49 块时,拼接结果明显错位。由此可知,与基于网格剖分的图像拼接相比,在分割块数相近的情况下,所提方法拼接结果好。

继续增大网格划分的数量,使每一个网格内部更接近同一个平面,拼接结果会更准确。网格数增大为 1600 时,拼接结果如图 14(b)所示。继续增大网格数目,拼接结果未得到明显改变,定量评价指标峰值信噪比(PSNR)与结构相似性(SSIM)值变小,且由于需要计算更多的单应性矩阵,计算时间会增加。

采用 PSNR 和 SSIM 对所提方法、基于全局的单应性变换和基于网格剖分的图像拼接进行定量评价。PSNR 是以图像的峰值信噪比作为评价图像质量的衡量指标。首先计算待评价图像  $I(x, y)$  和参考图像  $K(x, y)$  的均方误差(MSE),假设两张图像的大小都为  $M \times N$ , PSNR 计算公式为

$$\begin{cases} E_{MS} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(i, j) - K(i, j)]^2 \\ R_{PSN} = 10 \lg \left[ \frac{(2^n - 1)^2}{E_{MS}} \right] \end{cases}, \quad (8)$$

式中:  $n$  为图像位深。PSNR 值越大,则说明待评价图像与参考图像越相似,失真越小,拼接效果越好。由于 PSNR 评价方法的计算复杂度小,易于实现,已经被广泛应用于图像评价领域。

SSIM 方法<sup>[23]</sup>采用图像间的结构相似度评价图像的质量。此方法主要针对图像的结构相似度进行评价。亮度相似度、对比度相似度和结构相似度共同决定 SSIM 值的大小,定义为

$$S_{SSIM}(I, K) = \frac{(2\mu_I\mu_K + C_1)(2\sigma_{IK} + C_2)}{(\mu_I^2 + \mu_K^2 + C_1)(\sigma_I + \sigma_K + C_2)}, \quad (9)$$

式中:  $\mu_I$  和  $\mu_K$  为待评价图像  $I(x, y)$  和参考图像  $K(x, y)$  的平均强度;  $\sigma_I$  和  $\sigma_K$  为标准差;  $\sigma_{IK}$  为两图像的相关系数。SSIM 值越接近 1,代表待评价图像与参考图像之间的结构越相似,图像拼接质量好。

PSNR 和 SSIM 评价结果如表 1 所示。

由表 1 可见,相较于全局单应性变换和网格剖分的拼接结果,所提基于空间信息聚类的拼接融合结果具有更高的 PSNR 和 SSIM 值,说明所提方法得到

的结果更接近大视场 RGB 图,具有更高的拼接质量。

## 5 结 论

针对单一成像系统获取的 RGB-D 数据视场较小的问题,提出了基于空间信息聚类的 RGB-D 数据拼接融合方法,实现了大视场 RGB-D 数据的获取。考虑 RGB-D 数据中蕴含的场景空间信息,定义物点距离,利用 SLIC 算法进行 RGB-D 数据空间信息聚类,将场景分割为若干具有单应性的平面子块。场景中的平面子块在 RGB 图和深度图中可以体现为超像素,RGB-D 数据的拼接可以由 RGB 超像素和深度超像素的拼接实现。在计算子块单应性矩阵时,选择子块内部准确配准的特征点对,或以特征点与子块中心点的深度距离为权重选择全部特征点对构造系数矩阵,使子块的单应性矩阵计算更准确。单应性能够准确描述平面场景在不同成像系统中的坐标变换,但是真实场景中更常见的情况是由分片光滑的曲面组成的,因此设置了物体表面同时存在平面和曲面的实验场景。利用所提方法进行空间信息聚类后,每一个空间子块近似一个平面子块,近似满足单应性。场景曲面曲率越大,就需要越多的子块个数对场景进行分割,来保证每个子块可以近似一个平面。因此对于场景曲面变化剧烈的情形,所提方法的计算时间相应会增加。实验结果表明,利用所提方法对 RGB-D 数据进行拼接,相较于全局单应性变换,能够改善由真实场景不满足单应性平面假设而在变换过程中产生的图像扭曲和变形及在拼接重叠部分的错位现象。相比基于网格剖分的拼接方法,当分割部分数量相同时,所提方法的拼接效果明显更好;网格数量增加到一定程度,拼接结果会有提升,但是需要计算更多的单应性矩阵,增大了计算量。通过对拼接结果的 PSNR 和 SSIM 计算,也定量说明了所提方法的拼接质量更高。所提方法为大视场、高分辨计算成像问题提出了一种准确的 RGB-D 数据拼接融合方法,能够促进计算成像技术在更多重要实际应用中的发展。

## 参 考 文 献

- [1] 卢荣胜, 史艳琼, 胡海兵. 机器人视觉三维成像技术综述[J]. 激光与光电子学进展, 2020, 57(4): 040001. Lu R S, Shi Y Q, Hu H B. Review of three-dimensional imaging techniques for robotic vision[J]. Laser & Optoelectronics Progress, 2020, 57(4): 040001.
- [2] Liu H, Li H Y, Liu X H, et al. A novel method for

表 1 不同方法的定量评价

Table 1 Quantitative evaluation of different methods

Parameter	Proposed method	Global homography	Grid split
PSNR	19.2632	17.6843	17.4831
SSIM	0.8567	0.8483	0.8359



- extrinsic calibration of multiple RGB-D cameras using descriptor-based patterns[J]. *Sensors*, 2019, 19(2): 349.
- [3] Sarbolandi H, Lefloch D, Kolb A. Kinect range sensing: structured-light versus Time-of-Flight Kinect [J]. *Computer Vision and Image Understanding*, 2015, 139: 1-20.
- [4] Hadfield J, Koutras P, Efthymiou N, et al. Object assembly guidance in child-robot interaction using RGB-D based 3D tracking[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid. New York: IEEE Press, 2018.
- [5] Gatesichapakorn S, Takamatsu J, Ruchanurucks M. ROS based autonomous mobile robot navigation using 2D LiDAR and RGB-D Camera[C]//2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP), January 16-18, 2019, Bangkok, Thailand. New York: IEEE Press, 2019.
- [6] Wetzstein G, Ihrke I, Lanman D, et al. Computational plenoptic imaging[J]. *Computer Graphics Forum*, 2011, 30(8): 2397-2426.
- [7] Wu Y C, Rivenson Y, Wang H D, et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning[J]. *Nature Methods*, 2019, 16(12): 1323-1331.
- [8] Tan Z P, Johnson K, Clifford C, et al. Development of a modular, high-speed plenoptic-camera for 3D flow-measurement[J]. *Optics Express*, 2019, 27(9): 13400-13415.
- [9] 李唐薇, 童官军, 李宝清, 等. 大视场域的目标检测与识别算法综述[J]. *激光与光电子学进展*, 2020, 57(12): 120002.  
Li T W, Tong G J, Li B Q, et al. Review on object detection and recognition in large field of view[J]. *Laser & Optoelectronics Progress*, 2020, 57(12): 120002.
- [10] Li H Y, Liu H, Cao N, et al. Real-time RGB-D image stitching using multiple Kinects for improved field of view[J]. *International Journal of Advanced Robotic Systems*, 2017, 14(2): 172988141769556.
- [11] Serafin J, Grisetti G. Using augmented measurements to improve the convergence of ICP[M]//Brugali D, Broenink J K, Kroeger T, et al. Simulation, modeling, and programming for autonomous robots. Lecture notes in computer science. Cham: Springer, 2014, 8810: 566-577.
- [12] Serafin J, Grisetti G. NICE: dense normal based point cloud registration[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 28-October 2, 2015, Hamburg, Germany. New York: IEEE Press, 2015.
- [13] Serafin J, Grisetti G. Using extended measurements and scene merging for efficient and robust point cloud registration[J]. *Robotics and Autonomous Systems*, 2017, 92: 91-106.
- [14] Segal A, Haehnel D, Thrun S. Generalized-ICP [C]//Robotics: Science and Systems V, June 28-July 1, 2009, Seattle, USA. Robotics: Science and Systems Foundation, 2009.
- [15] Ding C Q, Liu H, Li H Y. Stitching of depth and color images from multiple RGB-D sensors for extended field of view[J]. *International Journal of Advanced Robotic Systems*, 2019, 16(3): 172988141985166.
- [16] 张元, 李晓燕, 韩燮. 一种低重叠率的三维点云配准方法[J]. *激光与光电子学进展*, 2021, 58(8): 0810014.  
Zhang Y, Li X Y, Han X. Three-dimensional point cloud registration method with low overlap rate[J]. *Laser & Optoelectronics Progress*, 2021, 58(8): 0810014.
- [17] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282.
- [18] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. Cambridge: Cambridge University Press, 2004.
- [19] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [20] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[M]//Readings in computer vision. Amsterdam: Elsevier, 1987: 726-740.
- [21] 王锦锦. 基于膨胀超像素的快速图像拼接算法研究[D]. 西安: 西安电子科技大学, 2019.  
Wang J J. A fast image stitching algorithm based on expanded super pixel[D]. Xi'an: Xidian University, 2019.
- [22] Allene C, Pons J P, Keriven R. Seamless image-based texture atlases using multi-band blending[C]//2008 19th International Conference on Pattern Recognition, December 8-11, 2008, Tampa, FL, USA. New York: IEEE Press, 2008.
- [23] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2004, 13(4): 600-612.