

基于稀疏点云的多平面场景稠密重建

缪君^{1,2} 储珺² 张桂梅² 王璐²

摘要 多平面场景是生活中常见的一种场景,然而由于该类场景中常常存在物体表面纹理缺乏和纹理重复的现象,导致从多视图图像重建获得的三维点云数据中存在点云过于稀疏甚至孔洞等问题,进而导致以微面片拟合三维点云所得到的重建表面出现平面颠簸现象.针对这些问题,本文提出了一种基于稀疏点云的分段平面场景重建方法.首先,利用分层抽样代替随机抽样,改进了 J-Linkage 多模型估计算法;然后,利用该方法对稀疏点云进行多平面拟合,来获得场景的多平面模型;最后,将多平面模型和无监督的图像分割相结合,提取并重建场景中的平面区域.场景中的非平面部分用 CMVS/PMVS (Clustering views for multi-view stereo/patch-based multi-view stereo) 算法重建.多平面模型估计的实验表明,改进的 J-Linkage 算法提高了模型估计的准确度.三维重建的实验证实,提出的重建方法在有效地克服孔洞和平面颠簸问题的同时,还能重建出完整平面区域.

关键词 多视重建, 三维重建, 多模型估计, 图像分割

引用格式 缪君, 储珺, 张桂梅, 王璐. 基于稀疏点云的多平面场景稠密重建. 自动化学报, 2015, 41(4): 813–822

DOI 10.16383/j.aas.2015.c140279

Dense Multi-planar Scene Reconstruction from Sparse Point Cloud

MIAO Jun^{1,2} CHU Jun² ZHANG Gui-Mei² WANG Lu²

Abstract There are multi-planar scenes everywhere in our daily life. However, given its lack and self-repeat of the texture, there would be problems of over scarcity and holes on the reconstructed point cloud by the method of multi-view reconstruction. Further, there would be vacillation over the reconstructed facades using the method of fitting the reconstructed point cloud with miniature facets. To address these problems, we propose a method of piecewise reconstruction of each plane from the sparse point cloud. The proposed method first improves the J-linkage algorithm, with the stratified sampling instead of the random sampling. We then fit the point cloud with planes using the improved J-linkage algorithm, to obtain the multi-planar model of the scene. Finally, we extract and reconstruct the planar regions with the multi-planar model as well as an unsupervised segmentation algorithm. Besides, the non-planar areas are reconstructed by using the clustering views for multi-view stereo/patch-based multi-view stereo (CMVS/PMVS) algorithm. Experimental results of the multi-planar model demonstrate that the improved J-linkage algorithm can enhance the accuracy of the multi-planar model. Also, the experimental results of 3D reconstruction show that our method not only can effectively overcome holes and jaggies problems, but also can model the complete planar regions.

Key words Multi-view reconstruction, 3D reconstruction, multiple structures estimation, image segmentation

Citation Miao Jun, Chu Jun, Zhang Gui-Mei, Wang Lu. Dense multi-planar scene reconstruction from sparse point cloud. *Acta Automatica Sinica*, 2015, 41(4): 813–822

多平面场景是周围环境中常见的一种场景,例如城市建筑、办公室环境、街道等.基于图像或视频重建此类场景在诸如数字地球、虚拟城市以及三维影视与娱乐等方面都有广泛的应用^[1–3].

多平面场景三维重建的方法可分为两大类:过程建模^[4–5]和多目立体视觉建模^[6–7].过程建模能简练、快速地生成大场景模型,但是由于重建结果是各种基本体素的结合,因此重建结果只是真实场景的近似.多目立体视觉的方法通常根据输入的一组图像或视频,通过求解相机的参数几何结构,来恢复场景表面的三维模型.这类方法的核心内容包含两种技术:从运动信息中恢复场景结构 (Structure from motion, SFM)^[6, 8–10]和多视立体 (Multi-view stereo, MVS)^[11–13].SFM 技术通常基于多视点图像的低层特征 (点特征)^[6, 8]、基本几何结构约束^[9] (如平面、长方体等) 或二者结合^[10] 来恢复相机的运动参数和场景结构. Snavely 等^[8] 提出的针对互联网无序照片集的 SFM 方法能够准确得到相机内、

收稿日期 2014-04-22 录用日期 2014-10-13
Manuscript received April 22, 2014; accepted October 13, 2014
国家自然科学基金 (61263046, 61462065), 江西省自然科学基金 (20122BAB201037) 资助
Supported by National Natural Science Foundation of China (61263046, 61462065) and Natural Science Foundation of Jiangxi Province (20122BAB201037)
本文责任编辑 贾云得
Recommended by Associate Editor JIA Yun-De
1. 南昌大学机电工程学院 南昌 330031 2. 南昌航空大学计算机视觉研究所 南昌 330063
1. School of Mechanical and Electrical Engineering, Nanchang University, Nanchang 330031 2. Institute of Computer Vision, Nanchang Hangkong University, Nanchang 330063

外参数, 并获得稀疏的三维场景重建数据. Zhou 等^[10] 通过检测和跟踪场景中的主平面构建了一个完整的 SFM 系统. Furukawa 等^[13] 提出的 CMVS/PMVS (Clustering views for multi-view stereo/patch-based multi-view stereo) 算法可将 SFM 技术获得的稀疏点云作为种子点输入, 利用 CMVS 算法根据视角对图像序列进行聚类, 再用基于微面片模型的 PMVS 算法将种子点向周围扩散得到空间有向点云或面片, 并在局部光度一致性和全局可见性约束下完成稠密重建. 文献 [8, 13] 的算法通用性强, 能适用于多种场景重建, 也代表了当前多视图重建所达到的水平^[14]. 然而, 多平面场景的图像经常表现出单一、重复的纹理. 这些方法在特征匹配过程中都依赖于图像纹理信息, 在重复纹理或无纹理的表面缺乏匹配特征, 特别是在图像灰度值一致的区域, 估计不出深度值, 因此重建的模型会出现很多孔洞. 另外, 在重建平面时, 文献 [13] 的算法由于采用大量微面片组合表示整个平面, 重建的平面显的颠簸不平.

近年来, 常采用分段平面 (Piecewise plane) 模型拟合多平面场景来解决重建中的孔洞和平面颠簸的问题. 这种方法的关键在于提取场景的平面区域. 分段平面的重建方法大致可分为两类: 1) 只利用二维图像信息重建^[15-17]. 这类方法首先要提取图像中的点、线、灭点等特征, 然后通过分析它们的拓扑关系提取平面区域. 因此, 重建的结果依赖于图像特征检测的成功性和精度. 2) 基于多源信息 (包括图像、深度图和三维稀疏点云) 重建^[18-21]. 如 Sinha 等^[18] 基于三维稀疏点云和稀疏直线段, 首先, 估计出场景中存在的平面模型, 然后用图割技术从预先给定

的深度图中分割出各平面区域. Gallup 等^[19] 将分段平面重建方法和 CMVS/PMVS 方法相结合重建场景, 解决实际场景中都会存在的非平面结构. 该方法首先对场景中的平面和非平面场景分类, 然后使用 Multi-RANSAC (Random sample consensus) 算法找出给定的深度图中的平面, 再用图割技术分割出平面区域, 非平面部分仍采用 CMVS/PMVS 方法重建. Hane 等^[20] 提出的方法也利用了深度信息和二维图像分割相结合的方法稠密重建场景. 由于深度图对应了图像中完整的平面区域, 因此文献 [18-20] 的算法克服使用点或微面片构建场景模型时出现的孔洞和平面颠簸的现象, 但实际应用中获取较大场景的精确的深度图需要很高的成本和计算代价. Toldo 等^[21] 利用多平面模型估计技术 (J-Linkage 算法) 从稀疏三维点中估计平面, 由于稀疏点云并不能像深度图那样提供完整的平面区域信息, 因此用凸集面片来拟合场景模型, 但凸集面片不能排除场景真实的空洞区域, 而且也容易包含错误区域.

本文提出一种将三维稀疏点云的多平面模型估计和无监督的二维图像分割相结合, 在不借助深度图的条件下, 提取完整的平面区域, 分段重建三维多平面场景的方法. 提出算法的目的在于克服点或微面片重建时出现的孔洞和平面颠簸现象的同时, 还能重建出完整的平面场景. 本文算法的流程如图 1 所示, 首先, 用 SFM 技术^[8] 获得 3D 稀疏点云; 然后, 基于稀疏点云进行多平面估计. 在此步骤中, 本文用分层抽样代替随机抽样, 对 J-Linkage 算法的假设模型的采样做了改进, 使估计的结果更准确; 接着, 将属于同一平面的点作为顶点生成相关邻域图

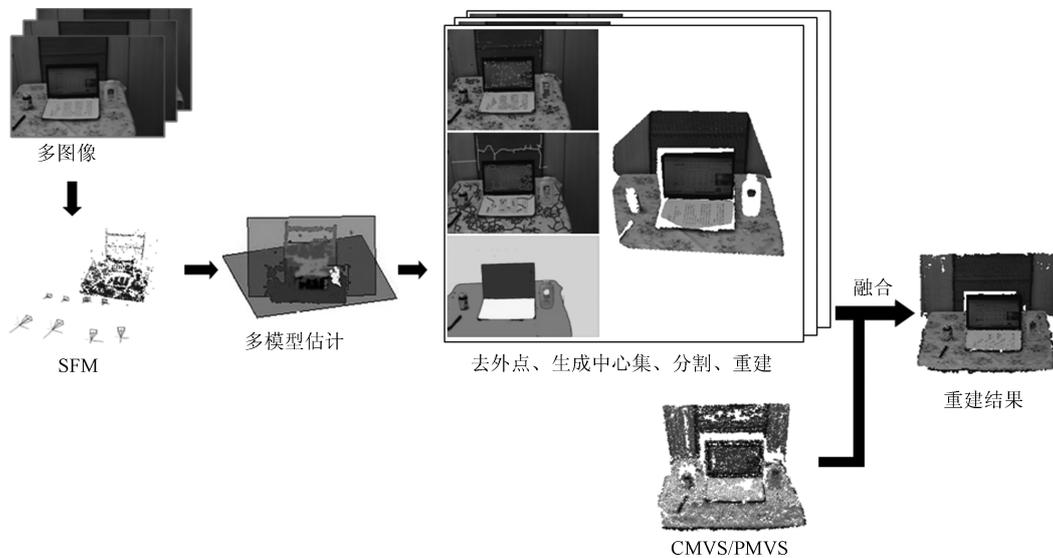


图 1 重建算法的流程

Fig. 1 Process of our reconstruction approach

(Relative neighborhood graph, RNG), 并自动生成分割标记, 利用测地星形凸性 (Geodesic star convexity) 图像分割算法提取场景中的平面区域; 最后, 对平面区域进行重建和融合, 为保证场景重建的完整性, 场景非平面部分用 CMVS/PMVS 算法重建.

1 基于改进的 J-Linkage 算法估计三维平面模型

从输入的多视图像中使用 SFM 技术^[8] 获得的三维稀疏点云只包含三维坐标信息和颜色向量, 若要利用它们进行分段平面重建, 则首先要将这些点分配到它们实际对应的场景中的平面之上, 即对它们进行多平面模型估计.

J-Linkage 算法是由 Toldo 等^[22] 提出的一种多实例模型估计方法. 该算法的基本思想是: 属于同一模型实例的数据点会聚类于相似的概念空间. 该算法不像 Multi-RANSAC^[23] 算法那样需要预先指定模型数量. J-Linkage 算法首先从数据点集合中随机抽取的 n 个数据点计算假设模型, 假设模型的数目即为相似概念空间的维度 M , 确定一个模型所需的最小数据点个数 n 即组成一个最小采样集 (Minimal sample set, MSS). 每个 MSS 生成之后, 可以利用其包含的数据点计算出它的模型实例的假设; 然后通过一致集构造用于描述数据点对于每个模型倾向情况的倾向向量; 最后对倾向向量自底向上地累计聚类, 每次扫描都会将两个具有最小距离的类别合并, 最终剩下的每个聚类即为对应的不同模型实例.

可见, 由抽样点构成的 MSS 是 J-Linkage 算法后期聚类的基础. J-Linkage 算法对每个模型假设的种子点采取的是随机抽样, 这种抽样方式使当属于每个模型实例的点的总数差别很大的时候, 假设模型总是在点总数大的模型实例中形成, 而总数小的模型实例可能就此丢失. 例如, 用 SFM 获得初始点云时, 纹理丰富的区域数据点多且密, 而纹理稀疏的区域数据点少且稀疏, 若 J-Linkage 运用在此类的点云中估计模型实例时, 点分布密度小的模型实例就会丢失, 最终得到的是局部最优解. 因此, 本文采用了一种先聚类后分层抽样的方法使初始点空间变得更均匀, 当 J-Linkage 在这样的点空间运行时, 能获得更好的解. 具体操作如下:

步骤 1 (K-Means 聚类). 设一共有 N 个 3D 数据点, 每个数据点由归一化后的空间坐标和归一化的颜色值构成向量 $P_i(x, y, z, r, g, b)$, 我们采用经典的 K -Means 聚类算法将数据点分为 K 个类, 即最小化函数:

$$J = \sum_{i=1}^N \sum_{k=1}^K r_{ik} \|P_i - \mu_k\|^2 \quad (1)$$

其中, r_{ik} 在数据点 P_i 被归类到第 k 类的时候为 1; 否则为 0. μ_k 为聚类 k 中的数据点的平均值.

步骤 2 (奈曼法分层抽样). 将步骤 1 中获得的 K 个聚类视为待抽样的 K 个分层, 从各层应抽样本数与该层总体数及其标准差的积成正比. 即第 h 层的应抽样本数为

$$n_h = \frac{n(N_h \times \sigma_h)}{\sum_{i=1}^K N_i \times \sigma_i} \quad (2)$$

其中, n 为样本总数, N_h 为第 h 层样本总数, σ_h 为第 h 层样本方差.

由式 (2) 可知, 当某层中样本密度稀疏时, 样本方差很大, 则被抽取的样本数量会较直接随机抽样大, 反之亦然. 可见, 分层抽样后, 每个 MSS 的空间密度差异减小, 那些少而稀疏的模型更易被纳入假设模型.

2 二维图像的平面区域的分割

当用初始点集完成多模型实例的估计, 即从点云中提取多个三维平面之后, 并不能直接进行三维平面的重建. 原因有二: 1) 平面模型内的三维点间的拓扑关系和场景的平面区域无直接关联, 很难构成和真实平面场景一致的空间平面区域; 2) 每个平面模型实例中只包含该平面区域的部分点, 并没有包含实际属于该三维平面对应的二维图像平面区域的所有点, 若只用这些少数点去重建平面, 则重建的平面区域不完整. 本文以属于已检测出的各三维平面包含的数据点为基础, 基于测地星形凸集技术自动地分割出每一副图像中的二维平面区域.

2.1 测地星形凸集

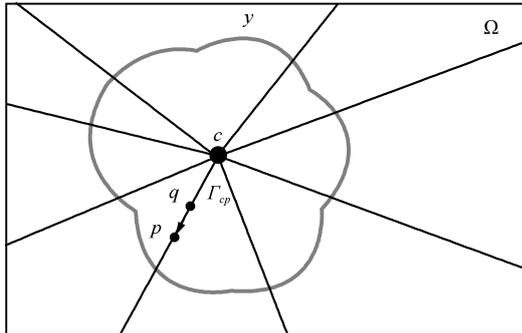
测地星形凸集源于 Veksler^[24] 提出的星形凸集. 星形凸集的定义为: 如果存在 y 中的点 c , 使得对于 y 中的所有 p , 从 c 到 p 的线段也位于 y 内, 则集合 y 称为星形凸集, c 被称作 y 的中心点, 这样的集合 y 被称为单星形状, 如图 2(a) 所示. 如果将 1 和 0 分别标记为前景和背景, 则可知, $\forall p \in \Omega : p \in S \Leftrightarrow S_p = 1$, Ω 为图像域.

在文献 [24] 方法的基础上, Gulshan 等^[25] 将单星形状扩展到多星形状 (图 2(b) 所示), 即将单个中心点 c 扩展为中心点集合 Θ . 连接点 c 与 p 的线段扩展为连接集合 Θ 和 p 的线段的集合 $\Gamma_{\Theta, p}$. 定义集合 Θ 到点 p 的最短距离 (欧氏距离) $d(c, p)$, 其可通

过求解函数:

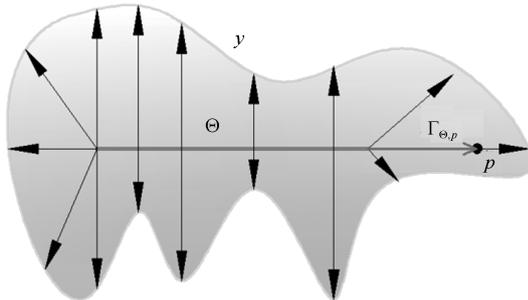
$$c(p) = \arg \min_{c \in \Theta} d(c, p), \Gamma_{\Theta, p} = \Gamma_{c(p), p} \quad (3)$$

获得.



(a) 单星凸集

(a) Single star-convexity



(b) 多星凸集

(b) Multiple star-convexities

图 2 星形凸集

Fig. 2 Star-convexity

2.2 测地星形凸集中心集的自动生成

文献 [24–25] 的方法将星形凸集中心集作为形状先验应用于图像分割, 但需要人工标记, 包括标记前景 (中心集) 和背景, 因此不能实现无监督的分割. 另外, 利用他们的方法每次只能分割出一个前景目标. 为了实现无监督的分割, 我们针对单一图像, 围绕测地星形中心集合的自动生成和测地星形凸集区域的分割展开研究, 并同时分割出多个前景目标. 星形凸集中心的自动生成包括前景和背景标记的生成. 这些标记以已获得的三维平面包含的数据点为基础来构建.

首先, 去除数据点集中的外点. J-Linkage 算法只利用了数据点的三维坐标信息估计模型实例, 所以每个模型实例包含的三维数据点只是属于确定的 3D 平面, 并不确定属于图像中的同一个 2D 面片区域, 这些点可能分布在不同的面片. 为使 3D 平面内包含的三维点只对应于图像的一个平面区域, 本文将输入的图像从 RGB 颜色空间转换到 $L^*a^*b^*$ 颜色空间, 在 $L^*a^*b^*$ 空间利用 K -Means 算法对图像进

行分割, 并对图像像素标记类别:

$$M(X, c) = \sum_{i=1}^{N_p} \arg \min_j \|X_i - c_j\|^2 \quad (4)$$

其中, X 表示图像像素, c 为聚类中心, N_p 为数据点总数. 当指定 k 为类别总数时, 则有 $M(X, c) = 1, 2, \dots, k$.

然后, 对每个估计出的 3D 平面, 将属于它的数据点用 SFM 所获得的投影矩阵 P_c 投影到二维图像, 即 $u(x, y, 1) = P_c U(x', y', z', 1)$, 并将 $u(x, y)$ 代入式 (4), 找出其所属类别:

$$S_j(u, c) = \begin{cases} 1, & M(u, c) = j \\ 0, & \text{其他} \end{cases} \quad (5)$$

接着, 累计每个类别中的点个数 $\delta = \sum S_j(u, c)$, 并将 δ_{\max} 对应的类别中包含的点作为内点, 其他的点都作为外点. 需要指出的是, δ 值会受到类别总数 k 的影响. 当 k 值指定过大, 则图像分割的区域过多, 属于同一区域的点总数便减少, 因此 δ 值变小, 这使得后续分割标记过少, 最终导致分割不完整, 甚至分割错误. 同理, k 值过小也容易包含分割错误区域. 大量的实验表明, $k = 4 \sim 6$ 是适合的选择.

上述得到的离散内点数量过少, 并不适合直接作为图像上的前景或背景标记, 但注意到这些点已经分别分布在图像的各个平面区域内. 因此, 遵守两个原则来构建标记: 1) 同一性, 即凸集中心集 (前景) 只包含属于同一平面区域的点; 2) 排它性, 即当属于某一平面区域的点已用来构建前景, 则其他点只用来组成背景标记. 假设已估计得到的平面模型数目为 m , 围绕这两个原则, 首先在每个平面区域内建立相关邻近图 (Relative neighborhood graph)

$$G_i = (V, E), \quad i = 1, 2, \dots, m \quad (6)$$

其中, V 为该平面区域内的所有内点, 该图为一个包含点集 V 和边集 E 的无向图, 某无向边 (u, v) 属于该图当且仅当点 v 和 u 位于对方的邻域内.

在整个图像内, 可以获得和之前估计的 3D 平面数量相等的 m 个相关邻近图. 将每个图的顶点和边所经过的图像像素定义为一个集合 O_i , 生成的这些在分布每一个平面区域的像素点集合 O_i 便作为下一步图像分割所用的中心集.

2.3 平面区域分割

平面区域分割使用和文献 [25] 类似的方法, 利用 Graph-cut 模型进行分割. 和文献 [25] 不同的是, 本文算法可以分割出多个前景区域.

在每幅图像分割过程中, 依次将像素点集合 O_i 作为前景标记, 与此同时将除 O_i 外的其他集合

O_j ($j = 1, 2, \dots, i-1, i+1, \dots, m$) 标记为背景. 定义能量最小化函数

$$E(L) = \sum_{p \in \Omega} E_{\text{data}}(L(p)) + \sum_{p, q \in Ne} \lambda_{\text{smooth}} E_{\text{smooth}}(L(p), L(q)) + \sum_{i=1}^m E_i^*(L) \quad (7)$$

式中, L 是星凸集合 (前景), Ω 是图像像素集合, Ne 是相邻接的像素, E_{data} 为数据项, E_{smooth} 为平滑项, $E_i^*(L)$ 表示定义在某平面区域上的星凸集合的能量. 能量函数

$$L = \arg \min_{L \in S^*(\Theta)} \sum_{i=1}^m E_i^*(L) \quad (8)$$

的最小值便对应着图像的最优分割. 这里, $S^*(\Theta)$ 表示以集合 Θ 作为中心的星凸形状.

平面区域的分割虽然是对每幅图像分别进行的, 但每幅图像中的相同平面区域在不同图像中的体现是一致的, 所以分割后的结果应是全局一致的. 但当某一平面模型上实际包含多个平面区域时, 平面区域之间的部分也会被纳入分割结果. 为了提高分割的精度, 本文采取类似文献 [13] 的方法, 利用多视图之间的几何关系, 对分割的区域进行一致性检测.

首先, 将每一个平面划分成一系列小的元胞 (Cell), $C = \{\rho_1, \dots, \rho_c, \dots, \rho_r\}$, 每一个元胞大小为 $t \times t$ 的像素子块, 其中, $\rho_c = \{\rho_{c11}, \dots, \rho_{c1t}\}$, $\rho_{c1j} = (u \ v \ 1)^T$ 为像素的齐次坐标. 然后, 将元胞 ρ_c 用单应性矩阵 H_{π, I_r} 投影到参考图像 I_r ; 接着, 利用图像间的单应性矩阵 H_{I_j, I_r} , 将其他每幅图像 I_j 上对应的分割区域投影到参考图像 I_r 上. 如果平面模型的元胞真实地属于场景平面区域, 那么所有投影在图像上的元胞区域应是匹配的. 一致性的判断从包含已匹配特征点的元胞开始, 逐步向其周围扩散. 元胞采用以特征点为中心的 8 邻域区域. 本文通过中心点的归一化互相关 (Normalized cross correlation, NCC) 条件来判断元胞的匹配, 删除不匹配的元胞区域. 由于 SFM 的结果中已获得匹配的特征点对应的图像, 因此, 围绕元胞的一致性判断并不在所有图像间进行, 而只在包含匹配特征的几副图像间进行.

3 平面与非平面区域的重建

本节主要是对场景中平面重建的研究, 其中包括两个内容: 1) 利用前述步骤得到的稀疏重建和场景分割的结果, 从单一视点重建三维平面区域; 2) 利用图像一致性 (Photo consistent) 对点云进行融合.

对于多视重建来说, 场景内容通常会在多个视图中都可见, 因此需要将不同视点得到的三维数据进行融合. 考虑到长序列图像通常背景复杂、视点杂乱无序, 利用传统的融合方法耗时较长, 本文采取了一种单向链式的融合算法对不同视点获得的三维数据进行了融合.

在多图像的 SFM 计算中, 我们已获得多图像的匹配特征, 而且这些二维匹配特征点和已重建的三维数据点是一一对应的. 将三维数据点重新投影回图像中, 找到对应在各图像平面区域的已被重建的三维数据点. 将每个平面区域的三维点构建为凸包 (Convex hull). 在构建凸包后, 只取出其投影在上节已分割出的 2D 平面区域内的部分, 以便保证三维平面区域和二维平面区域的一致性.

由于每一副图像可能包含多个平面区域, 若不经融合而直接将重建出的平面都作为最后重建的结果, 则在映射纹理后, 整个场景纹理会多处重叠而不能达到真实纹理 (Realistic texture) 的效果. 为了简化计算, 本文不采用多视点匹配的方法将所有像素进行匹配, 而是依据 SFM 计算稀疏重建的视点顺序, 将位置邻近图像的平面区域融合: 假设 π_{i-1} 、 π_i 为两个待融合的重建平面, 任取一点 $a \in \pi_{i-1}$, 如果至少存在一点 $b \in \pi_i$, 使

$$\begin{cases} |a - b| < \varepsilon \\ \langle \vec{N}_{i-1, a}, \vec{N}_{i, b} \rangle \leq \theta \end{cases} \quad (9)$$

成立, 则 π_i 中的点保留. 接下来再重复上述过程, 融合平面 π_i 和 π_{i+1} .

式 (9) 中, ε 为两点的欧氏距离阈值, $\vec{N}_{i-1, a}$ 和 $\vec{N}_{i, b}$ 为 a 和 b 法向量, $\langle \cdot, \cdot \rangle$ 为两法向量夹角, θ 为角度阈值.

为重建的完整性, 场景中的非平面部分 (包括场景中未提取出的平面部分) 采用 CMVS/PMVS 重建结果.

4 实验结果

4.1 多模型估计实验

多平面模型不但包含场景中的平面参数信息, 而且也包含了属于场景各平面上的三维稀疏点的聚类信息, 因此准确的多平面模型对场景中平面区域的分割与重建有重要意义. 本文对改进的 J-Linkage 多模型估计算法进行了多平面场景的平面模型估计实验, 并将实验结果和 Toldo 等 [22] 的算法进行了比较. 为保证比较的客观性, 两种算法都分别对同一组三维稀疏点云进行平面估计. 由于本文改进算法的模型假设和聚类原理与原始文献 [22] 算法是一致的, 因此在它们的运算过程中都采用相同的参数 (例

如, 假设模型、MSS).

图 3 为 “Dunster” 序列图像的三维点的平面模

型估计结果. 图 3 的第一列为输入数据, 第二列为文献 [22] 算法结果, 第三列为本文算法结果. 从文献

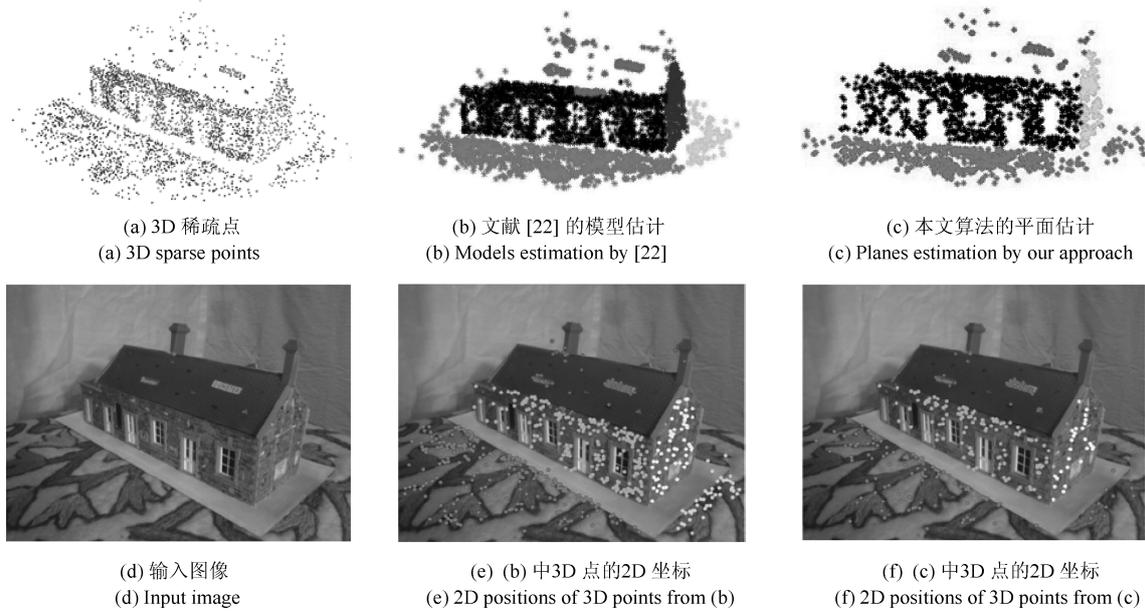


图 3 “Dunster” 场景平面模型估计

Fig. 3 Plane structures estimation of “Dunster”

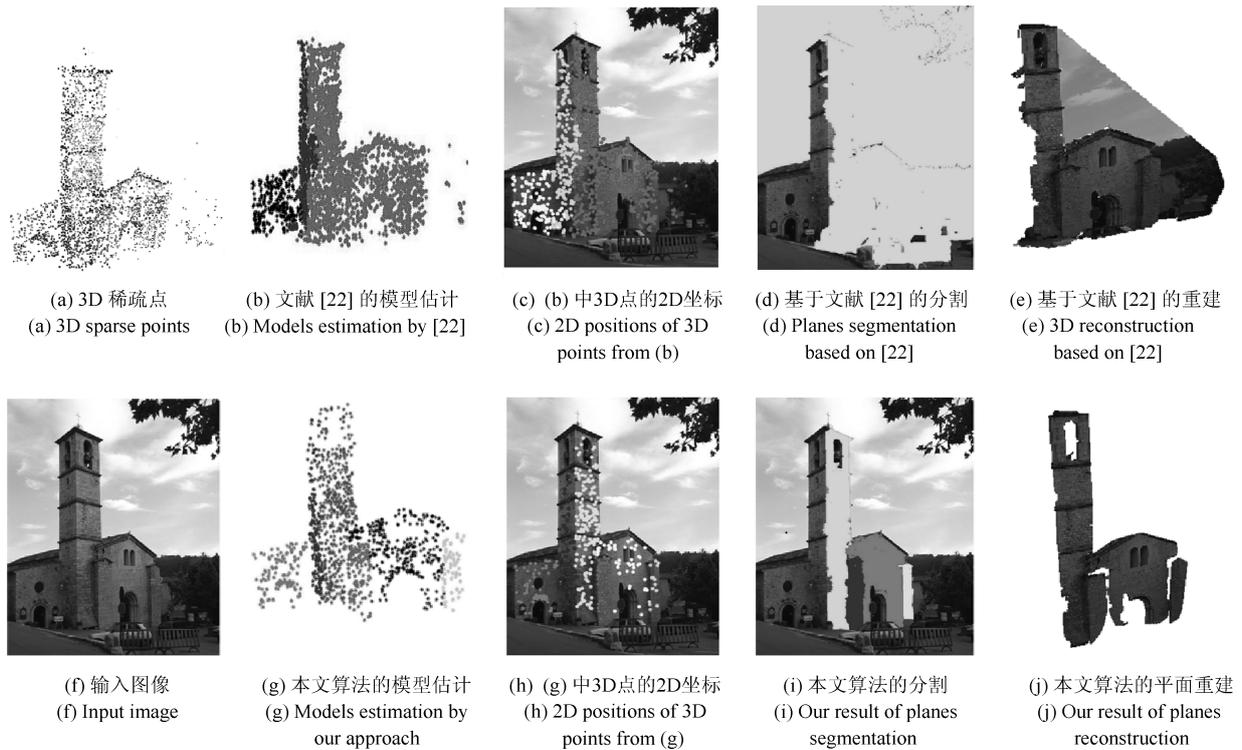


图 4 “Valbonne” 场景平面模型估计

Fig. 4 Plane structures estimation of “Valbonne”

[22] 算法结果 (图 3(b)、(e)) 可看出, 场景中的“地面”在右前方的转角处被“断开”, 文献 [22] 算法便将“地面”视为两个平面, 而本文算法则视为同一平面. 同时也可注意到, 深灰色点对应的“屋顶”平面, 本文的算法结果更正确, 这是因为分层抽样可以将一些外点排除在样本之外.

图 4 是对“Valbonne”序列图像的实验. 图 4(a) 和 (f) 为输入数据: “Valbonne”序列图像的三维稀疏点和该序列中的一幅图像. 图 4(b)~(e) 为文献 [22] 算法结果: 图 4(b) 为平面模型估计结果, 图 4(c) 为三维点对应在图像上的二维点, 图 4(d) 和 (e) 分别为基于文献 [22] 算法的平面分割和最终重建结果. 图 4(g)~(j) 为本文算法结果. 注意到图 4 中的建筑物“教堂”前端存在两个小平面 (图 4(f) 中所示的 #1 和 #2), 文献 [22] 算法并未估计出这两个平面. 从本文算法的结果 (图 4(h)) 可看出, 对输入点采用分层抽样后, 使点云密度变得相对均匀, 从而使包含小样本点的两个小平面被检测出 (见图 4(h) 中右边门洞所处的平面). 另外, 从图 4(c) 注意到, 文献 [22] 算法的深灰色点对应的平面中都包含错误的内点, 而本文算法在分层抽样中已将那些点排除在样本之外, 这样既可使估计的平面参数更准确, 还可以减少后续图像分割的错误, 最终得到更准确的重建结果. 如图 4(c) 和 (d), 在图中深灰色点覆盖区域, 平面模型估计的不准确造成平面区域分割错误. 虽然在后续重建中生成的凸集面片可以对错误进行弥补, 但当凸集面片中也包含错误区域时 (如图 4(e) 中所示的天空等区域), 最终的重建结果仍然不准确. 本文算法结果中, 如图 4(i) 和 (j) 所

示, 好的平面估计模型能提供好的分割结果, 最终得到精度更高的重建模型.

4.2 多平面场景重建实验

为了验证本文重建算法的实验效果, 我们使用了多组不同环境的图像集合进行测试, 包括室内场景和室外建筑物图像.

图 5 为“桌面”图像序列实验结果, 该序列为包含 10 幅图像的室内场景. 图像都用佳能相机采集, 图像的分辨率为 720×480 . “桌面”场景大部分由平面区域组成, 包含少量非平面区域 (瓶子、笔). 图 5(c) 为多平面模型估计后, 稀疏点云对应在该图像上的二维点. 当不去除外点时, 从图 5(d) 的分割结果可以看出, 由于少量外点的存在, 而导致每个平面区域都包含了错误. 当去除外点 (图 5(e) 中黑色点) 后, 如图 5(f) 所示, 除了平面 3 部分缺失, 其他平面区域都被正确的分割. 从 SFM 的结果 (图 5(b)) 可以看出, 由于图像数量少, 初始点云的数量稀少, 特别是处于后方的木质平面 (平面 4), 由于缺乏纹理信息, 几乎无重建点生成. CMVS/PMVS 算法 (图 5(g)) 比 SFM 的结果有很大改进, 但区域增长法对于特征稀疏的区域仍存在孔洞, 而本文算法 (图 5(h)) 在 SFM 点云覆盖的平面区域采用分段平面重建, 不会存在孔洞, 因此效果最好.

为了对各种算法的重建结果的平面颠簸程度做出定量评估, 本文对多模型估计中获得的平面到包含于该平面上的三维点的距离进行了统计. 统计包括两方面: 距离平均偏差 (D_{mean}) 和距离均方差 (D_{std}), 分别定义为

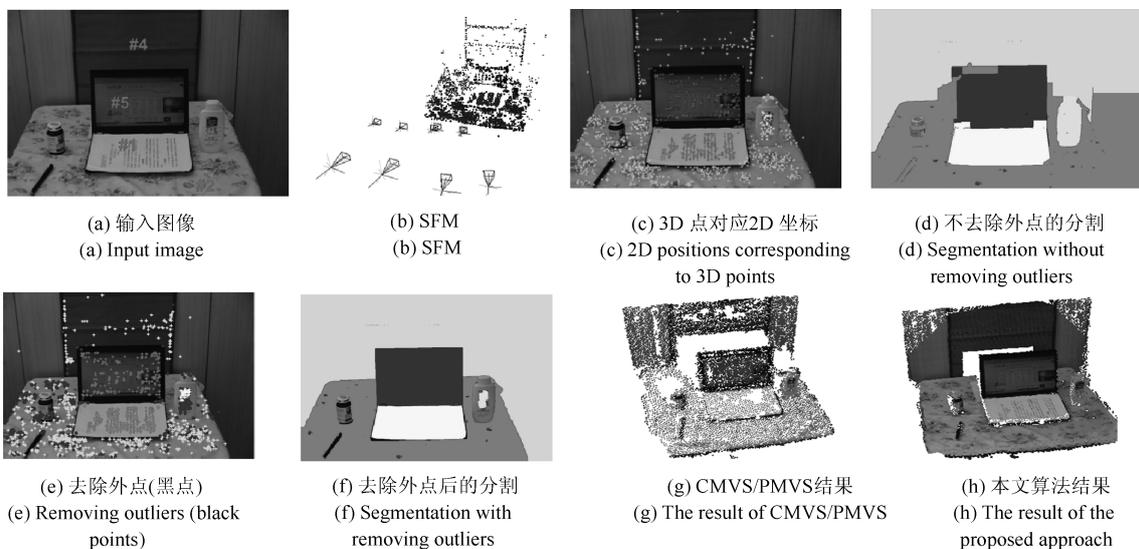


图 5 “桌面”图像序列重建.

Fig. 5 Reconstruction of “Desktop” sequence

表 1 SFM、CMVS/PMVS 和本文算法的平面颠簸比较
Table 1 Comparisons of planar jaggies of SFM, CMVS/PMVS and our results

图像序列	SFM	CMVS/PMVS	本文算法
#1 纸面	428/0.093/0.074	713/0.16/0.137	14 836/2.4777E-04/1.463E-04
#2 桌子	2 312/0.128/0.136	2 954/0.111/0.11	25 354/2.497E-04/1.4477E-04
#4 墙面	246/0.172/0.168	4 533/0.169/0.139	22 264/6.784E-03/3.2288E-03
#5 显示器	365/0.085/0.116	2 417/0.125/0.189	37 367/0.012/7.2390E-03

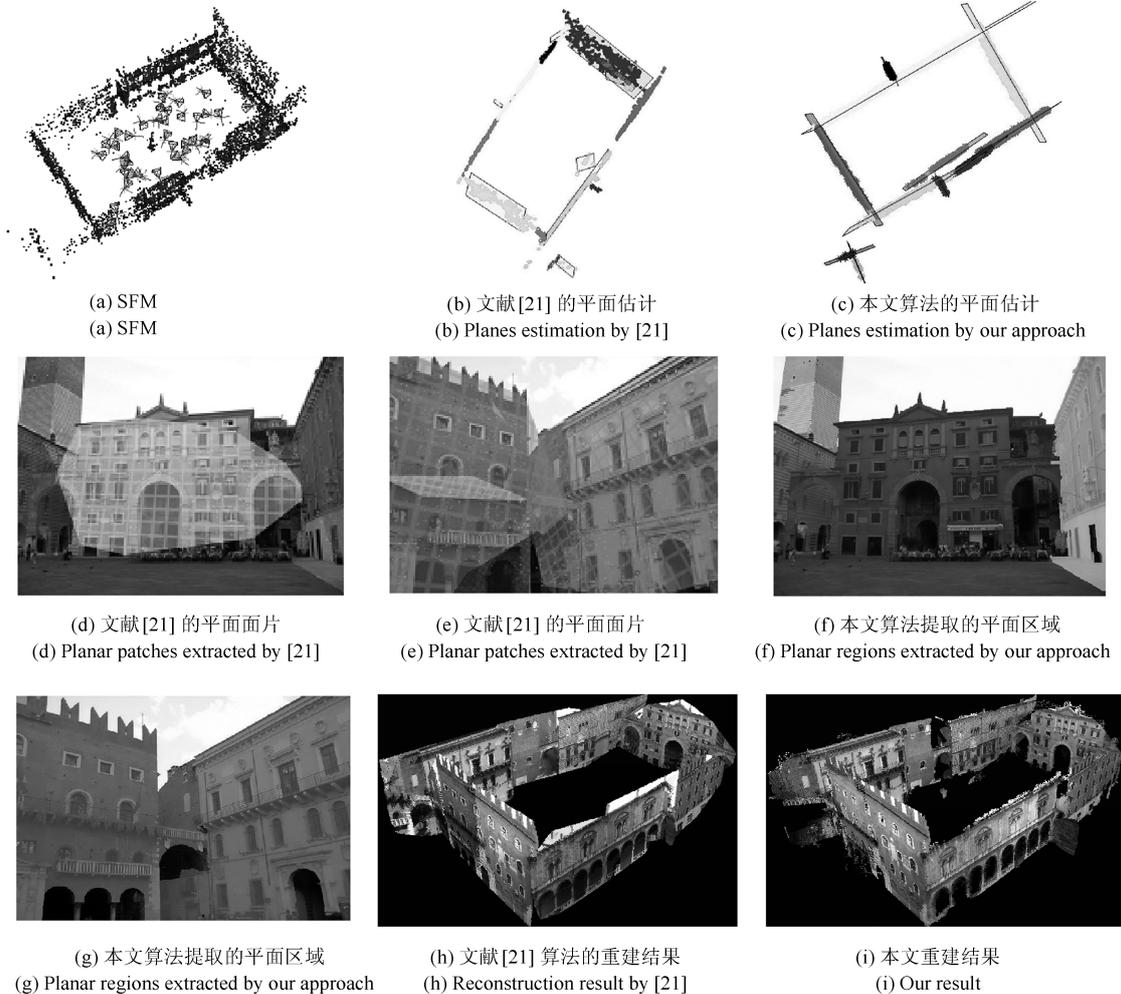


图 6 “Dante” 图像序列重建

Fig. 6 Reconstruction of “Dante” sequence

$$D_{\text{mean}} = \frac{\sum_{i=1}^N D_i(P, X_i)}{N} \quad (10)$$

$$D_{\text{std}} = \frac{\sum_{i=1}^N (D_i(P, X_i) - D_{\text{mean}})^2}{N} \quad (11)$$

此处, P 为平面参数, X_i 为三维点坐标, $D_i(P, X_i)$ 表示点 X_i 到平面 P 距离, N 为该平面包含的点的总数. 从式 (10) 和式 (11) 可看出, 若 D_{mean} 的值越

大, 则点云和平面偏差越大, 若 D_{std} 的值越大, 则平面颠簸程度越大. 反之亦然.

表 1 为“桌面”场景中的 4 个平面的颠簸程度的比较结果. 表 1 的第一行为用来比较的方法名称, 第一列为平面名称, 它和图 5(a) 中的标号一致. 表中每一单元格的数据表示“点数目/ $D_{\text{mean}}/D_{\text{std}}$ ”. 表 1 显示, SFM 算法和 CMVS/PMVS 算法的 D_{mean} 和 D_{std} 值相差不多, 这是因为 CMVS/PMVS 算法是以 SFM 算法获得稀疏点云作为种子点向周围扩散而得到的空间点云. 本文算法的 D_{mean} 和 D_{std} 值

比其他两种方法要小于 100 倍以上, 这表明本文算法的平面颠簸程度远小于其他两种方法。

图 6 是文献 [26] 提供的数据集 “Dante” 图像序列测试的结果, 并且, 本文算法的结果和文献 [21] 的结果进行了比较. 该图像序列包含 39 副图像, 图像大小为 1144×856 . 图 6 (a) 是利用 SFM 技术获得的稀疏点云和相机结构, 图 6 (b) 和 (c) 分别为文献 [21] 和本文算法的多平面模型估计结果, 估计的平面个数分别为 16 和 11 个, 主要的平面都被找出. 图 6 (d) 和 (e) 是文献 [21] 算法得到的平面面片, 图 6 (f) 和 (g) 是本文算法得到的平面区域. 从图 6 (d) 和 (e) 可看出, 由于图像的特征点是 SIFT 特征点, 凸集面片形成的平面面片边界不会和真实平面区域边界接近, 而且, 这种面片容易包含错误区域, 例如图中所示的门洞, 凸集面片不能避免这种真实存在的空洞. 本文算法提取的平面区域 (如图 6 (f) 和 (g) 所示) 更接近真实的区域边界, 有效地减少了错误区域. 对比文献 [21] 的重建结果 (图 6 (h)), 本文结果 (图 6 (i)) 的区域边缘更接近场景的真实边界, 包含更少的错误区域.

5 结论

本文提出了一种基于稀疏点云的分段平面重建多平面三维场景的方法. 该方法将稀疏点云的多平面模型估计和无监督的图像分割相结合, 在不依赖场景深度图的条件下, 对多平面场景进行重建. 三维重建的实验表明, 本文提出的分段平面重建算法能克服场景重建因纹理缺乏和纹理重复而导致的孔洞问题, 也能避免微面片近似大平面出现的颠簸现象, 恢复的平面场景更接近真实的结构.

References

- 1 Vanegas C A, Aliaga D G, Wonka P, Müller P, Waddell P, Watson B. Modelling the appearance and behaviour of urban spaces. *Computer Graphics Forum*, 2010, **29**(1): 25–42
- 2 Sheppard S R J, Cizek P. The ethics of Google Earth: crossing thresholds from spatial data to landscape visualisation. *Journal of Environmental Management*, 2009, **90**(6): 2012–2117
- 3 Shi Li-Min, Guo Fu-Sheng, Hu Zhan-Yi. An improved PMVS through scene geometric information. *Acta Automatica Sinica*, 2011, **37**(5): 560–568
(史利民, 郭复胜, 胡占义. 利用空间几何信息的改进 PMVS 算法. *自动化学报*, 2011, **37**(5): 560–568)
- 4 Simon L, Teboul O, Koutsourakis P, Van Gool L, Paragios N. Parameter-free/Pareto-driven procedural 3D reconstruction of buildings from ground-level sequences. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Rhode Island, USA: IEEE, 2012. 518–525
- 5 Vanegas C A, Aliaga D G, Benes B. Automatic extraction of manhattan-world building masses from 3D laser range scans. *IEEE Transactions on Visualization and Computer Graphics*, 2012, **18**(10): 1627–1637
- 6 Agarwal S, Furukawa Y, Snavely N, Simon I, Curless B, Seitz S M, Szeliski R. Building rome in a day. In: Proceedings of the 12th International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009. 72–79
- 7 Mičušík B, Košecká J. Multi-view superpixel stereo in urban environments. *International Journal of Computer Vision*, 2010, **89**(1): 106–119
- 8 Snavely N, Simon I, Goesele M, Szeliski R, Seitz S M. Scene reconstruction and visualization from community photo collections. Proceedings of the IEEE, 2010, **98**(8): 1370–1390
- 9 Bartoli A, Sturm P. Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. *International Journal of Computer Vision*, 2003, **52**(1): 45–64
- 10 Zhou Z H, Jin H L, Ma Y. Robust plane-based structure from motion. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Rhode Island, USA: IEEE, 2012. 1482–1489
- 11 Goesele M, Snavely N, Curless B, Hoppe H, Seitz S M. Multi-view stereo for community photo collections. In: Proceedings of the 11th International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1–8
- 12 Seitz S M, Curless B, Diebel J, Scharstein D, Szeliski R. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2006. 519–528
- 13 Furukawa Y, Ponce J. Accurate, dense, and robust multi-view stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(8): 1362–1376
- 14 Musialski P, Wonka P, Aliaga D G, Wimmer M, Gool L, Purgathofer W. A survey of urban reconstruction. In *Computer Graphics Forum*, 2013, **32**(6): 146–177
- 15 Bartoli A. A random sampling strategy for piecewise planar scene segmentation. *Computer Vision and Image Understanding*, 2007, **105**(1): 42–59
- 16 Yu W C, Kai L Y, Chung P C. 3D reconstruction of piecewise planar models from multiple views utilizing coplanar and region constraints. *Journal of Information Science and Engineering*, 2013, **29**(2): 361–378

- 17 Micusik B, Kosecka J. Piecewise planar city 3D modeling from street view panoramic sequences. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 2906–2912
- 18 Sinha S N, Steedly D, Szeliski R. Piecewise planar stereo for image-based rendering. In: Proceedings of the 12th IEEE Conference on Computer Vision. Rio de Kyoto, Japan: IEEE, 2009. 1881–1888
- 19 Gallup D, Frahm J M, Pollefeys M. Piecewise planar and non-planar stereo for urban scene reconstruction. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 1418–1425
- 20 Hane C, Zach C, Cohen A, Angst R, Pollefeys M. Joint 3D scene reconstruction and class segmentation. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE, 2013. 97–104
- 21 Toldo R, Fusiello A. Photo-consistent planar patches from unstructured cloud of points. In: Proceedings of the 2010 European Conference on Computer Vision. Crete, Greece: Springer, 2010. 589–602
- 22 Toldo R, Fusiello A. Robust multiple structures estimation with J-linkage. In: Proceedings of the 2008 European Conference on Computer Vision. Marseille, France: Springer, 2008. 537–547
- 23 Zuliani M, Kenney C S, Manjunath B S. The multiransac algorithm and its application to detect planar homographies. In: Proceedings of the 2005 IEEE Conference on Image Processing. Genoa, Italy: IEEE, 2005. 153–156
- 24 Veksler O. Star shape prior for graph-cut image segmentation. In: Proceedings of the 2008 European Conference on Computer Vision. Marseille, France: Springer, 2008. 454–467
- 25 Gulshan V, Rother C, Criminisi A, Blake A, Zisserman A. Geodesic star convexity for interactive image segmentation. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 3129–3136

- 26 Farenzena M, Fusiello A, Gherardi R. Structure-and-motion pipeline on a hierarchical cluster tree. In: Proceedings of the 12th IEEE Conference on Computer Vision Workshops. Miami, USA: IEEE, 2009. 1489–1496



缪 君 南昌大学机电工程学院博士研究生. 南昌航空大学航空制造工程学院讲师. 主要研究方向为计算机视觉与图像处理. E-mail: miaojun@nchu.edu.cn
(**MIAO Jun** Ph.D. candidate at the School of Mechanical and Electrical Engineering, Nanchang University, lecturer at the School of Aeronautical Manufacturing Engineering, Nanchang Hangkong University. His research interest covers computer vision and image processing.)



储 珺 南昌航空大学软件学院教授. 主要研究方向为图像处理与计算机视觉. 本文通信作者.
E-mail: chujun@nchu.edu.cn
(**CHU Jun** Professor at the School of Software, Nanchang Hangkong University. Her research interest covers image processing and computer vision. Corresponding author of this paper.)



张桂梅 南昌航空大学航空制造工程学院教授. 主要研究方向为图像处理, 计算机视觉与模式识别.
E-mail: zhangguimei@nchu.edu.cn
(**ZHANG Gui-Mei** Professor at the School of Aeronautical Manufacturing Engineering, Nanchang Hangkong University. Her research interest covers image processing, computer vision, and pattern recognition.)



王 璐 南昌航空大学软件学院讲师. 主要研究方向为计算机视觉与图像处理.
E-mail: wanglu@nchu.edu.cn
(**WANG Lu** Lecturer at the School of Software, Nanchang Hangkong University. Her research interest covers computer vision and image processing.)