论文

基于语义约束与 Graph Cuts 的稠密三维场景重建

王伟, 高伟*, 胡占义

中国科学院自动化研究所模式识别国家重点实验室, 北京 100190

* 通信作者. E-mail: wgao@nlpr.ia.ac.cn

收稿日期: 2013-11-26; 接受日期: 2014-02-11; 网络出版日期: 2014-05-09

国家高技术研究发展计划 (批准号: 2013AA12A202) 和国家自然科学基金 (批准号: 61273280) 资助项目

摘要 本文提出一种新颖、有效的稠密三维场景重建算法. 在城市建筑场景的重建中, 为了快速恢复稠密、准确的深度信息, 本文算法首先在视图中对建筑区域进行了语义分割以降低非重建区域 (如天空、地面等) 的干扰, 在提高整体重建速度的同时也增强了采用平面模型对其进行重建的可靠性; 然后, 在通过基于 DAISY 特征的空间点扩散方法获取的初始深度图的基础上, 针对传统算法难以重建的弱纹理、倾斜表面等区域, 本文算法依据场景分段平滑的假设, 在超像素级 MRF 能量优化框架中对其相应的空间平面进行了推断. 由于能量函数融合了初始深度图的约束、空间平面先验及空间平面间的几何关系等信息, 而且候选平面集通过平面拟合和已知平面约束下的多方向平面扫描两种方法构造, 使得相应的两阶段迭代 Graph Cuts 对能量函数的求解更快速和精确. 在标准数据集和真实数据上的实验表明, 本文算法能有效克服光照变化、透视畸变、弱纹理区域等因素的影响, 快速恢复建筑区域完整的深度图.

关键词 语义标注 平面拟合 匹配扩散 深度图 能量优化

1 引言

利用从同一场景获取的多幅视图恢复出高精度、稠密的深度信息对三维场景重建的完整性有着重要的意义. 近年来, 尽管许多相关工作已针对此问题进行了深入的研究, 但由于光照变化、透视畸变、弱纹理区域、遮挡、视图分辨率较低等因素的影响, 恢复高质量的深度图始终是很多算法面临的难题, 特别对于存在较多弱纹理、倾斜表面等区域的城市建筑, 重建完整性问题尤为突出, 如基于特征点匹配或扩散的方法^[1,2], 由于在弱纹理区域无法检测到特征及特征匹配多义性问题的存在, 通常很难获得完整的深度图; 而根据场景分段平滑假设的基于视图过分割的方法^[3~7]以及基于特定场景模型 (如 Manhattan-world 模型) 的方法^[8~10] 尽管在一定程度上可以解决弱纹理区域的重建问题, 但在场景结构比较复杂的情况下, 则会由于所采用的求解模型或场景模型过于简单而产生较大的错误; 最新报道的融入语义信息并利用机器学习的方法^[11,12] 虽然在弱纹理、倾斜表面等区域的重建中表现出相对较好的优势, 但机器学习方法固有的精度低的缺点导致此类算法很难获取高精度的重建结果, 而且计算速度也难以得到保证.

引用格式: 王伟, 高伟, 胡占义. 基于语义约束与 Graph Cuts 的稠密三维场景重建. 中国科学: 信息科学, 2014, 44: 774-792, doi: 10.1360/N112013-00211

为了克服以上问题, 本文提出一种基于语义约束与 Graph Cuts^[13] 的快速、稠密的场景重建算法. 对于场景重建中主要感兴趣的建筑区域, 本文算法首先在视图中对其进行快速语义分割, 排除了天空、地面等非重建区域的干扰, 不但提高了整体重建速度, 而且增强了采用平面结构模型对建筑区域进行重建的可靠性, 进而使得后续在 MRF 能量优化框架下对弱纹理、倾斜表面等区域对应空间平面的推断更可靠. 在获取建筑区域的初始深度图时, 本文利用扩散的方式同时完成 DAISY 特征在多幅视图中的匹配及相应空间点的求取. 由于 DAISY 特征匹配的搜索空间被有效缩小, 使得深度图的求取过程更快速, 结果更稠密, 从而进一步增强了对弱纹理区域和倾斜表面所对应的空间平面进行全局优化求解的可靠性. 整体上, 本文算法克服了利用 Manhattan-world 场景模型重建复杂场景结构时的误差较大、利用机器学习方法推断表面法向量速度慢和精度低等缺点, 能够快速重建出包括弱纹理区域、透视畸变严重的倾斜区域等在内的稠密、准确的城市建筑深度图.

本文结构组织如下: 第 2 节、第 3 节分别对相关工作及本文算法流程进行概述; 第 4 节对本文采用的语义标注算法进行简要介绍; 第 5 节、第 6 节则分别对基于 DAISY 特征的空间点扩散算法及两阶段空间平面全局优化过程做详细介绍; 第 7 节进行实验与分析; 第 8 节为总结与展望.

2 相关工作

在利用两幅或多幅视图恢复场景稠密视差或深度图的过程中, 许多算法^[3~7] 通常根据场景分段平滑的假设, 将颜色相近、位置相邻像素的集合与相应的视差或空间平面进行关联, 然后在利用传统匹配度量如 ZNCC(zero-mean normalized cross-correlation) 等获取的初始视图或深度图的基础上, 利用平面拟合的方法求取初始平面集, 最后采用 Graph Cuts, Cooperative Optimization 等算法求取弱纹理、倾斜表面等区域对应的视差或空间平面以获取完整的视差或深度图. 此类算法的最大问题在于, 由于像素匹配的多义性、噪声等问题的存在, 由此求取的初始平面集通常并不能完整地表达场景结构, 如倾斜表面区域, 在像素匹配及平面拟合阶段一般难以获取较好的匹配结果及相应的平面, 因而其对应的真实平面与初始平面集中的平面并不一定存在相关性, 此时如果用初始平面集中的平面推断其视差或深度往往会产生较大的错误. Bleyer 等^[14] 为了克服此问题, 首先为每个像素随机地分配一个视差平面, 然后根据所定义的视差平面下的像素累积匹配代价, 利用迭代扫描的方法对每个像素对应的视差平面进行优化. 事实上, 由于此算法采用的像素累积匹配代价仍然存在匹配多义性问题, 因而对弱纹理区域视差的推断并不理想, 而且迭代扫描方式导致整体重建速度也较低. Shen^[15] 将此算法推广至多视场景重建中深度图的推断中, 尽管可利用更丰富的视图信息以构造具有强约束性的像素累积匹配代价, 但类似问题仍然存在.

根本上而言, 由于诸多干扰因素的存在, 仅仅依赖像素级匹配的算法很难获取场景稠密的重建结果. 从视图中挖掘更多可用于重建的信息 (如城市建筑的平面结构特征及与地面垂直等先验) 以构造对弱纹理、倾斜表面等区域视差或深度求解的更严格的约束是一种解决场景稠密重建的有效途径. 在此类算法中, Çigla 等^[7] 根据空间平面诱导下弱纹理区域在多个视图中映射区域的颜色差异, 通过深度与法向量扫描获取其对应的空间平面. 然而, 此类穷举式方法不但速度较慢, 而且由于仅依赖于颜色差异的像素累积匹配代价的约束性通常较低, 导致算法的可靠性也较差. Gallup 等^[16] 则根据城市建筑平面结构的先验, 利用平面模型解决弱纹理区域的重建问题, 而且为了增强平面模型的可靠性, 首先采用机器学习的方法将场景粗略地分为平面结构与非平面结构等语义区域, 然后以初始稀疏或准稠密深度图为基础, 对建筑区域采用平面拟合、Graph Cuts 等算法进行稠密重建. 然而, 由于算法过多

图 1 本文算法基本流程图

Figure 1 Flow diagram of the proposed algorithm

地依赖于平面结构与非平面结构区域的分类精度, 而且只从初始深度图中拟合出非常少的平面来近似场景结构, 使得精度与可靠性在复杂场景中都难以得到保证. 根据 Manhattan-world 场景模型的先验, Furukawa 等 [8]、Mičušik 等 [10] 则分别利用 PMVS 算法 [2] 获取带方向的空间点、城市建筑中消影点检测等方法首先确定场景模型的主方向, 然后在每个主方向上求取所有可能的空间平面以构成对弱纹理区域进行深度推断的候选平面集, 最后利用 Graph Cuts 等算法推断弱纹理区域对应的空间平面以获得稠密的重建结果. 正如前述, 由于场景模型过于简单, 当真实空间平面法向量与模型主方向不一致时, 如本文实验 (参见 7.3 小节) 中的“清华物理楼”、“生命科学院”数据集, 场景结构并不符合 Manhattan-world 场景模型的特征, 此类算法往往会产生较大的重建错误. Häne 等 [11] 将视图语义标注与场景重建问题融合到体素多类标注框架下进行求解, 不但使得语义标注结果为场景中特定类别的重建提供了具有几何意义的先验信息 (如建筑物是垂直的、地面是水平的), 而且比较好地克服了传统算法难以解决的弱纹理、倾斜表面等区域的重建问题. 然而, 利用体素进行场景或对象重建时, 不但时间与空间复杂度较高, 而且体素尺寸也会对重建精度产生较大的影响; 此外, 利用机器学习方法推断场景中特定类别的表面方向, 其可靠性、速度也很难得以保证. Bao 等 [12] 利用机器学习方法获取的特定语义类别的形状先验解决物体重建的完整性问题, 但对于城市场景, 由于城市建筑结构、形状复杂, 算法并不适用.

与以上算法不同, 尽管本文算法也采用平面模型、语义信息及全局优化等方法解决建筑区域弱纹理、倾斜表面等区域的重建问题, 但在语义信息的利用、初始深度图的求取及空间平面优化求解模型的构造等环节中采用了不同的、更为鲁棒的方法. 本文算法具有以下一些特点: (1) 利用语义约束极大地提高了整体重建的速度, 增强了采用平面模型对建筑区域进行重建的可靠性; (2) 利用扩散的方式同时完成 DIASY 特征在多幅视图中的匹配与相应空间点的求取, 可以快速获取更稠密、更准确的初始深度图, 克服了传统算法在求取深度图时由于 DIASY 特征的匹配搜索空间较大 [17]、单独进行深度图外点剔除等过程导致的结果稀疏、速度慢的问题; (3) 利用空间点的平面拟合及已知平面约束下的多方向平面扫描两种方法获取全局优化阶段所用的候选平面集及先验, 可以获得更准确的场景结构, 克服了 Manhattan world 场景模型等过于简单的缺点; (4) 在能量函数中融入了区域颜色差异、初始深度图约束、空间平面先验及空间平面间的几何关系等信息, 并采用两阶段的 Graph Cuts 迭代求解, 保证了空间平面全局优化求解的可靠性, 同时降低了整体算法对语义标注精度的依赖性.

3 算法流程与主要模块

本文算法根据场景分段平滑的假设, 从已标定的视图序列中快速恢复相应的稠密深度图. 其基本流程如图 1 所示. 其中, 3 个主要模块可简要描述为:

(1) 视图语义分割

对视图中的建筑区域进行分割, 剔除天空、地面等非重建区域的干扰, 并增强后续环节采用平面

模型对弱纹理、倾斜表面等区域对应空间平面进行推断的可靠性.

(2) 基于 DAISY 特征的空间点两阶段扩散

以 DAISY 特征为基础, 利用扩散的方式快速获取建筑区域初始稠密的深度图, 并为建筑区域先验信息的获取、弱纹理与倾斜表面等区域所对应的空间平面的推断提供可靠的基础.

(3) MRF 能量优化框架下的空间平面推断与深度图获取

对弱纹理、倾斜表面等区域对应的空间平面进行推断, 进而获取完整的深度图.

以下各节将对以上模块分别进行介绍.

4 视图的语义分割

鉴于当前语义分割算法对基本类别 (如建筑、天空、地面等) 的分割精度都比较高, 同时考虑到语义分割的速度, 本文采用了基于区域的语义分割算法^[18]对视图中的建筑区域先进行分割. 此处仅对其进行概述, 算法具体细节参见文献^[18].

该算法首先利用过分割算法将视图分割成超像素集合, 然后, 针对每个超像素, 在融合其外观特征及几何特征的能量函数最小化框架下利用机器学习的方式对其进行语义标注. 由于过分割算法不精确性问题的存在, 超像素可能会被标记为错误的语义类别, 为克服此问题, 算法采用不同的过分割参数将视图分割成多个超像素集合 (即“过分割字典”), 而对于当前超像素, 则根据“过分割字典”与能量函数最小化准则对其进行语义标注优化操作, 如为其分配新的语义类别、与相邻的相同语义类别的超像素进行合并等, 进而获得最优的标注结果. 由于算法以超像素为语义标注基元, 因而速度比较快.

5 基于 DAISY 特征的空间点两阶段扩散

对于建筑区域中包含的弱纹理、倾斜表面等区域, 由于像素匹配的多义性、严重的透视畸变等因素的影响, 其对应的深度信息一般很难获取. 为了获取完整的深度图, 本文首先求取纹理丰富区域的初始深度图, 进而提取建筑结构的先验信息以对弱纹理、倾斜表面区域的深度进行推断. 为了快速获取更稠密、更可靠的初始深度图, 本文采用了基于 DAISY 特征^[19]的空间点扩散的方式. 采取 DAISY 特征的主要原因在于: (1) 视图中全部像素的 DAISY 特征可以同时被快速提取; (2) DAISY 特征所采用的菊花状邻域结构使得其即使在视图基线距离较宽的情况下往往也能表现出高解析力的性能.

5.1 空间点扩散原理

在传统匹配扩散算法^[20]中, 对于已知种子匹配 (x, x') 及相应邻域 $N(x)$, $N(x')$ 中可能相互匹配的两点 $\mu \in N(x)$, $\mu' \in N(x')$, 在视差梯度约束条件下, (u, u') 存在下述关系, 即

$$\|(u' - u) - (x' - x)\|_{\infty} \leq \epsilon, \quad (1)$$

其中 ϵ 为预先指定的阈值. 式 (1) 表明, 对于 x 邻域内的任意像素 $\mu \in N(x)$, 在 x' 的邻域中总可以找到与种子匹配 (x, x') 具有相同偏移的点 u' , 而以 u' 为中心的指定邻域尺寸内的点即为 u 的候选匹配集 (也在 x' 的邻域内). 随着视图间基线距离的增大, 在确定视差梯度约束条件时应同时考虑水平与垂直两个方向的视差约束, 而在极限已知的情况下, 极限的约束将进一步缩小种子匹配邻域像素的匹配搜索范围.

图 2 空间点扩散原理 (相邻视图数 $k = 2$)Figure 2 The principle of spatial points propagation (the number of neighboring views $k = 2$)

如图 2 为参考视图 I_r 与 $k = 2$ 个相邻视图 $\{N_i\} (i = 1, \dots, k)$ 之间的匹配情况. 设参考视图 I_r 中的像素 p_r 在视图 N_i 中的匹配为 p_i , 而由匹配 $(p_r, p_1, \dots, p_k)$ 确定的种子空间点为 $P(p_r, p_1, \dots, p_k)$ (黑色点所示). 根据水平与垂直方向的视差梯度约束, 像素 p_r 邻域内的像素 m_r 在视图 N_i 中的候选匹配区域 (尺寸设为 w) 也在 p_i 的邻域内 (粗线框所示), 而极线的约束则进一步缩小了像素 m_r 的候选匹配范围 (短虚线所示为 m_r 的极线在候选匹配区域中的部分).

在确定像素 m_r 的匹配时, 为了保证匹配精度, 本文将极线在候选匹配区域内的部分进行等间隔采样 (采样间隔设为 ϑ), 而将采样点作为其候选匹配点, 其中采样点的特征根据相邻像素的 DAISY 特征并采用双线性插值方法求取. 此外, 由于 DAISY 特征并不具有旋转不变性, 为了获得更准确的匹配度量, 在对像素 m_r 进行匹配时采用了根据相应极线方向校正后的 DAISY 特征.

最后, 利用欧氏距离度量标准可在视图 N_i 中确定与像素 m_r 最相近的匹配 m_i (白色点所示). 为了验证匹配 $\{m_i\} (i = 1, \dots, k)$ 的可靠性以确定其是否对应于场景中真实的空间点 $M(m_r, m_1, \dots, m_k)$ 本文根据像素 m_r 在视图 N_i 中的匹配 m_i , 利用三角化方法^[21] 求取像素 m_r 相应的深度值 $d_{r,i}$. 显然, 如果匹配 m_i 是可靠的, 则深度值集合 $\{d_{r,i}\} (i = 1, \dots, k)$ 中任意两元素之差应很小, 因此, 本文将度量 $M(m_r, m_1, \dots, m_k)$ 是否可靠的标准 $D(M)$ 定义如下:

$$D(M) = \frac{1}{k} \sum_{i \neq j} \|d_{r,i} - d_{r,j}\| (i, j = 1, \dots, k). \quad (2)$$

如果 $D(M)$ 值小于指定阈值 ς , 则表明空间点 $M(m_r, m_1, \dots, m_k)$ 是可靠的, 可以作为种子空间点继续进行空间点的扩散; 而为了增强扩散过程的可靠性, 本文算法也采用了“最优最先”的扩散原则^[20], 即 $D(M)$ 值最小的种子空间点总是优先进行扩散. 需要强调的是, 在基线距离较窄的视图对中检测到的匹配, 相应深度值的求取通常会产生较大的偏差, 为了保证空间点扩散的可靠性, 根据文献^[17] 的建议, 本文选择基线距离相对较宽的多幅视图进行匹配的检测与扩散.

事实上, 本文基于空间点扩散的深度图求取方法不但度量了空间点在多幅视图中的映射点之间的 DAISY 特征相似性, 而且也考察了相邻空间点之间的平滑性约束关系, 因而更易于获取稠密的深度图; 而且, 由于在扩散过程中极大地缩小了 DAISY 特征的匹配搜索空间, 也有效地克服了传统算法^[17] 在场景深度范围未知时在较大的深度范围进行匹配搜索导致的计算复杂性, 所以具有更快的速度. 此外, 扩散过程结果后, 相邻视图建筑区域中的大部分像素也获得了可靠的深度值, 因而在相邻视图作为参考视图时, 仅对未获得可靠深度值的像素进行了匹配检测与空间点的扩散, 从而极大地节约了初始深度图的求取时间.

5.2 初始种子空间点的选取

对于空间点扩散开始时的初始种子空间点, 通常可采用 2 种方法获取, 即: (1) 对于参考视图中像素 p_r , 直接与视图 N_i 中相应极线上的所有采样点分别进行匹配并从中确定最优匹配 p_i , 然后对全部匹配结果 $\{p_i\} (i = 1, \dots, k)$ 对应空间点 $P(p_r, p_1, \dots, p_k)$ 的可靠性进行验证, 如果 P 是可靠的, 则作为种子空间点进行扩散; (2) 借助其他特征检测算法 (如 SIFT^[22] 特征) 在多幅视图中进行特征检测并匹配, 然后将由匹配求取的相应空间点作为种子空间点.

实验中发现, 由于极线上较多的候选匹配会消耗大量的时间, 方法 (1) 更适于扩散过程中种子空间点的检测 (参见 5.3 小节), 而方法 (2) 可以以更快的速度获取更多的种子空间点, 适于扩散开始时种子空间点的选取.

5.3 空间点两阶段扩散过程

在理想情况下, 1 个种子空间点即可获得稠密的扩散结果, 然而, 由于室外场景中存在的遮挡、光照变化等诸多因素的影响, 初始种子空间点往往不能得以充分的扩散. 为了解决此问题, 本文采用了种子空间点扩散与检测两阶段的循环过程, 即: 在当前种子空间点扩散结束后, 对视图中建筑区域中所有未参与扩散的像素进行了检测. 对于每个未扩散像素, 在其相邻视图相应极线上尚未匹配的采样点中确定其最优 DAISY 特征匹配并对相应的空间点进行可靠性验证, 如果该空间点是可靠的, 则将其作为种子空间点继续进行扩散, 如算法 1 所示. 实验中发现, 随着第一阶段扩散过程的继续, 由扩散得到的空间点为第二阶段种子空间点的检测提供了越来越丰富的先验信息, 极大地降低了第二阶段进行种子空间点检测的计算复杂性, 同时, 本文也采用了 KD-tree 算法以提高匹配速度.

由以上步骤可以发现, 空间点在扩散的同时也完成了相应深度值的求取与外点剔除, 避免了传统深度图求取方法^[17]在像素匹配结束后单独对深度图进行外点剔除所导致的计算复杂性. 在实验中, 本文对标准数据集及实际场景数据集的参考视图分别采用 2 个及 4 个相邻视图进行空间点的扩散 (具体结果参见第 7 小节), 总体上, 两种情况均可获得较好的结果, 表明本文扩散算法对相邻视图数量具有较好的适应性.

然而, 由于弱纹理区域像素匹配的多义性 (如“清华物理楼”的玻璃区域)、倾斜区域严重的透视畸变 (如“生命科学院”的倾斜区域) 以及遮挡区域 (如“清华物理楼”中的自遮挡及“生命科学院”中参考视图左边区域遮挡) 等问题的存在, 本文的扩散方法仍无法获得建筑区域完整的匹配结果及深度图, 因而需要进一步利用场景的平面先验对这些区域进行深度推断.

6 MRF 能量优化框架下的空间平面推断与深度图获取

为了获取弱纹理、倾斜表面等区域的深度信息, 类似于其他重建算法, 本文利用 Mean-shift 算法^[23]将当前视图过分割成超像素集合. 根据场景分段平滑的假设, 每个超像素对应的空间点位于空间某一平面上; 因而, 对弱纹理区域、倾斜表面等区域的深度推断问题可转化为求其内部超像素对应空间平面的问题.

而为了克服简单场景模型 (如 Manhattan-world 模型) 的局限性, 本文通过将超像素的颜色信息、空间平面之间的位置关系以及场景的先验信息融合在统一的 MRF 能量优化框架下, 分别对建筑区域主方向及非主方向上的空间平面进行了推断 (参见 6.4 小节), 进而获取完整的深度图. 具体流程如图 3 所示.

算法 1: 基于 DAISY 特征的空间点两阶段扩散

输入: 种子空间点集合 Seed、已扩散像素列表 Mark;

输出: 深度图 Depth;

初始化: 根据 Seed 更新 Mark, Depth;

% 种子空间点的扩散

1: **while** Seed $\neq \emptyset$ **do**

2: 从 Seed 中取出 $D(P_i)$ 最小的种子空间点 $P_i(p_r, p_1, \dots, p_k)$;

3: **for each** $m_r \in N(p_r)$ **do**

4: 确定视图 N_i 中的匹配 m_i ;

5: **if** $D(M) < \varsigma$

6: 根据 $D(M)$ 值将 M 插入至 Seed 并分别将 m_r, M 的深度值保存于 Mark, Depth;

7: **end-if**

8: **end-for**

9:**end-while**

% 种子空间点的检测

10: **if** Seed = $\emptyset$

11: **for each** $q_r \notin \text{Mark}$ **do**

12: 确定视图 N_i 中的匹配 q_i ;

13: **if** $D(Q) < \varsigma$

14: 根据 $D(Q)$ 值将 Q 加入至 Seed 并分别将 q_r, Q 的深度值保存于 Mark, Depth;

15: goto 1;

16: **end-if**

17: **end-for**

18: **end-if**

图 3 空间平面推断流程

Figure 3 Flow diagram of inferring spatial planes

下面对其中的关键步骤进行介绍.

6.1 空间平面拟合与外点剔除

如前所述, 在求取建筑区域的初始深度图时, 空间点在扩散的同时也剔除了不可靠的深度值, 因而最终获得的深度值绝大部分是可靠的, 而其中极少量的外点可在空间平面推断时进行剔除, 此处将它们统称为“可靠深度值”. 根据初始深度图, 当超像素内部包含的已求取可靠深度值的像素数量大于

3 时, 本文利用 RANSAC 方法进行了平面拟合, 然后通过求取超像素内部所有像素的反投影线与此空间平面交点的方式获得更稠密的空间点或深度值. 对于少量超像素, 如细长结构超像素, 以上过程可能会产生错误的空间平面及深度值. 为了检查错误空间平面, 本文根据相邻视图对应深度图的拟合结果, 采用文献 [17] 所述方法对当前深度图的拟合结果进行错误深度值检查与剔除. 如果超像素内部错误深度值数量与像素总数之比大于给定阈值 K_1 , 则表明该超像素对应的空间平面是不可靠的.

对于利用上述拟合方法求取的空间平面不可靠或内部已确定深度值的像素数量少于 3 而无法拟合平面的超像素, 如弱纹理、倾斜表面等区域内的超像素, 本文则在 MRF 能量优化框架下利用已拟合求取的空间平面及相应空间点、建筑区域特定平面结构及方向先验 (如垂直于地面、主方向的数量较少等) 等信息对其相应的空间平面进行推断. 为便于下文表述, 特将此类超像素统称为不可靠超像素, 而空间平面能够被正确拟合的超像素, 则称为可靠超像素, 所有可靠超像素所对应的空间平面的集合称为初始空间平面集.

6.2 能量函数的定义

为了在 MRF 能量优化框架下推断不可靠超像素对应的空间平面, 类似于其他算法 [9,10], 对于超像素 s 及候选空间平面 H , 本文定义如下能量函数:

$$E(s, H) = E_{\text{data}}(s, H) + E_{\text{smooth}}(s), \quad (3)$$

其中, $E_{\text{data}}(s, H)$, $E_{\text{smooth}}(s)$ 分别为数据项和平滑项.

(1) 数据项

为了定义数据项, 首先定义超像素 s 内部像素的匹配代价.

设参考视图 I_r 的相邻视图及深度图分别为 N_i , D_i , 而在 D_i 中, 已确定可靠深度值的像素集合为 M , 其余像素集合为 $\bar{M}$. 此外, 如果超像素 s 的候选空间平面为 H , 则对于像素 $p \in s$, 其在 H 诱导下在 D_i 中的映射点及相应深度值分别表示为 $H(p)$, $d(H(p))$.

由于 $H(p)$ 可能位于区域 $\bar{M}$, M , 同时考虑到遮挡情况, 则定义 p 与 $H(p)$ 之间的匹配代价为

$$C(p, H, N_i) = \begin{cases} m(p, H, N_i), & H(p) \in \bar{M}, \\ \lambda_{\text{occ}}, & D_i(H(p)) \leq d(H(p)), \\ m(p, H, N_i) + \lambda_{\text{occ}}, & D_i(H(p)) > d(H(p)), \end{cases} \quad (4)$$

上式中 λ_{occ} 为遮挡惩罚常数, 而 $m(p, H, N_i)$ 则度量了 p 与 $H(p)$ 之间颜色差异, 即

$$m(p, H, N_i) = \min(\|I_r(p) - N_i(H(p))\|, \tau), \quad (5)$$

其中, τ 为截断阈值 (实验设置为 0.25), 以增强遮挡情况下像素匹配度量的鲁棒性.

在式 (4) 中, 第 1 种情况表示映射点 $H(p)$ 位于区域 $\bar{M}$, 此时的 H 更可能为超像素 s 对应的真实空间平面, 则采用 p 与 $H(p)$ 间的颜色差异作为匹配度量; 第 2 种情况表示映射点 $H(p)$ 位于区域 M 且被 D_i 中的相应点所遮挡, 所以给予较大的惩罚量; 第 3 种情况表示映射点 $H(p)$ 位于区域 M 且遮挡了 D_i 中的相应点, 由于 D_i 中的深度值均是可靠的, 不可能被遮挡, 因而给予更大的惩罚量.

最后, 数据项 $E_{\text{data}}(s, H)$ 定义为超像素 s 内部全部像素匹配代价之和, 即

$$E_{\text{data}}(s, H) = \frac{1}{k} \sum_{i=1}^k \sum_{p \in s} C(p, H, N_i), \quad (6)$$

其中 k 为相邻视图的数量.

为了提高全局优化的可靠性与速度, 本文也将空间平面 H 的先验 $p(H)$ (参见 6.3 小节) 融入到能量函数. 假设在空间平面 H 已知时, 超像素 s 的数据项 $E_{\text{data}}(s, H)$ 在诸如噪声、摄像机参数偏差等因素影响下服从高斯分布, 即

$$p(E_{\text{data}}(s, H)|H) = e^{-\frac{E_{\text{data}}(s, H)}{2\sigma^2}}. \quad (7)$$

为了使概率 $p(H|E_{\text{data}}(s, H))$ 最大化, 则等同于最小化 $-\log(p(H|E_{\text{data}}(s, H)))$, 根据 Bayes 公式可得融合先验概率 $p(H)$ 的数量项 $E_{\text{data}}^{\text{prior}}(s, H)$, 即

$$E_{\text{data}}^{\text{prior}}(s, H) = E_{\text{data}}(s, H) - 2\sigma^2 \log(p(H)), \quad (8)$$

式 (8) 中, 空间平面的先验 $p(H)$ 越高, 则数据项 $E_{\text{data}}^{\text{prior}}(s, H)$ 越小, 超像素 s 对应的空间平面越偏向于空间平面 H . 此外, 如果超像素 s 内部部分像素对应的空间点集 P_s 已知, 则距离点集 P_s 越近的空间平面 H 更有可能为其对应的空间平面, 因而, 本文在空间平面先验 $p(H)$ 的基础上增加了由点集 P_s 与当前所选空间平面 H 所确定的偏置项 θ , 即

$$\theta = \begin{cases} 0, & |P_s| = 0, \\ \omega \cdot e^{-d(P_s, H)}, & |P_s| > 0, \end{cases} \quad (9)$$

其中, $d(P_s, H)$ 为空间点集 P_s 中的所有元素到空间平面 H 的平均距离, $|P_s|$ 表示集合 P_s 中的元素总数, ω 是偏置项的作用权值 (实验中设置为 0.5).

最终, 融合空间平面先验概率及偏置项的数据项为

$$E_{\text{data}}^{\text{prior}}(s, H) \propto E_{\text{data}}(s, H) - 2\sigma^2 \log(p(H) + \theta). \quad (10)$$

式 (10) 表明, 在空间平面 H 的先验不变的情况下, 如果其距离超像素 s 对应的空间点集 P_s 越近, 则相应的偏置项值越大, 数据项 $E_{\text{data}}^{\text{prior}}(s, H)$ 越小, 因而, 其成为超像素 s 对应空间平面的可能性也越大.

(2) 平滑项

在确定平滑项时, 本文采用了 Potts 模型^[13], 即

$$E_{\text{smooth}}(s) = \lambda_{\text{smo}} \cdot \sum_{t \in N(s)} \omega_{st} \cdot \delta(l_s \neq l_t), \quad (11)$$

其中, $N(s)$ 表示超像素 s 的所有相邻超像素集合, $\delta(\cdot)$ 函数在括号内的条件为真时值为 1, 否则为 0 (下文采用相同标记), l_s 为超像素 s 当前标记值, 实际与特定空间平面相对应, ω_{st} 则为超像素 s 与其相邻超像素 t 分配不同标记值的惩罚量, λ_{smo} 为平滑项权重.

在定义式 (11) 中的惩罚量 ω_{st} 时, 为使其可以根据相邻超像素间的属性差异而自适应地调整, 本文综合考虑了相邻超像素之间相邻接的像素数、颜色相似性以及对应空间平面之间的几何关系, 即

$$\omega_{st} = b_{st} \cdot (1 - c(s) - c(t)) \cdot \|n(s) - n(t)\|, \quad (12)$$

其中, $c(s)$ 表示超像素 s 的平均颜色, $n(s)$ 表示超像素 s 对应空间平面的单位法向量, b_{st} 则为超像素 s 与 t 相邻接的像素数, 即

$$b_{st} = |\{(p, q) | p \in s \wedge q \in t \wedge (p, q) \in N_4\}|, \quad (13)$$

上式中的 N_4 表示像素 4 邻域关系.

式 (12) 表明, 超像素 s 与相邻超像素之间的颜色越相似、相邻接的像素数越多, 则超像素之间的平滑性越好, 如果两者对应的空间平面法向量差异较大, 则应给予较大的惩罚量.

6.3 候选空间平面及先验

建筑区域通常具有规则的几何结构, 其内部超像素对应的空间平面通常分布在特定的方向与深度上, 因此, 本文首先根据初始深度图确定建筑区域的主方向, 然后推断主方向上存在的所有空间平面, 最后在主方向空间平面的约束下进一步完成非主方向上空间平面的推断.

6.3.1 主方向候选空间平面及先验

由于建筑区域 DAISY 特征匹配及相应空间点的充分扩散, 求取出的初始深度图及空间平面集已包含建筑区域大部分结构与先验, 为利用这些信息进一步对弱纹理、倾斜表面等区域的深度进行推断, 本文算法首先利用 AP (affinity propagation) 聚类算法^[24] 对初始空间平面集对应的法向量集合进行了聚类. 由于 AP 聚类算法并不需要事先指定类别数量, 而是将每个法向量均视为潜在的聚类中心, 因而在建筑区域主方向未知的情况下, 往往可以获得更可靠的聚类结果. 本文将与类别对应的所有法向量作为建筑区域的初始主方向, 标记为 $\{n_i\} (i = 1, \dots, m)$, 而主方向上相应的空间平面集则标记为 $\{P_i\} (i = 1, \dots, m)$. 需要强调的是, $\{n_i\}$ 并不一定是建筑区域完整的法向量集合, 如倾斜表面区域, 可能由于无法获得任何匹配 (空间点), 其对应的空间平面无法通过拟合的方法获取, 因而集合 $\{n_i\}$ 中并不包含该区域的法向量. 针对该问题, 本文采用两阶段全局优化算法 (参见 6.4 小节) 进行处理.

为了进一步提取建筑区域包含的先验信息, 对于集合 $\{n_i\}$, 本文沿每个主方向 n_i 将相应空间平面集合 P_i 中的元素根据位置的相近性进行 AP 聚类 $\{D_i\} (k = 1, \dots, n)$, 然后统计每个类别 D_k 中所有空间平面对应的空间点总数 M_k . 显然, M_k 值越高, 则由方向 n_i 与相应位置 D_k 确定的空间平面 $\Pi_{i,k}$ 包含的先验信息越丰富, 其成为弱纹理、倾斜表面等区域对应的空间平面的可能性也就越大; 所以, 本文用 M_k 与主方向 n_i 上全部空间点数之比度量空间平面 $\Pi_{i,k}$ 所包含的先验信息. 所有候选空间平面 $\{\Pi_{i,k}\}$ 将用于在全局优化阶段对弱纹理、倾斜表面等区域对应的空间平面进行求解.

事实上, 尽管在匹配及平面拟合过程中可能导致极少数超像素对应的空间平面不正确, 但与可靠空间平面相比, 这些错误空间平面包含的先验信息非常少, 往往对后续基于全局能量优化的空间平面推断影响不大, 另外, 本文在利用 Graph Cuts 对能量函数的迭代求解中, 不断地剔除错误空间平面并更新空间平面的先验信息, 使得这些错误空间平面能够很快地得以校正 (参见 6.4 小节).

6.3.2 非主方向候选空间平面及先验

在主方向上空间平面的推断中, 由于能量函数的约束性、空间平面先验的融入及在每次迭代中的更新, 弱纹理、倾斜表面等区域中越来越多的超像素获得了正确的空间平面, 而这些可靠的空间平面对法向量与主方向不一致的空间平面的正确求取构成的约束也越来越强. 事实上, 不在主方向上的空间平面法向量通常具有较高的随机性, 传统多方向空间平面扫描算法^[7] 一般要在较大的深度与多方向上进行扫描, 然后从中选择使特定能量函数值最小的深度与法向量构成的空间平面作为最优空间平面. 其中的法向量通常采用球面坐标系表达式, 即

$$n = \begin{bmatrix} \cos \theta \sin \phi \\ \sin \theta \sin \phi \\ \cos \phi \end{bmatrix}. \quad (14)$$

考虑到空间平面在各个视点的可见性, 通常假设其法向量与相应摄像机主轴的夹角不大于某一阈值^[15](如 60°), 即 $\phi \in [0^\circ, 60^\circ]$, $\theta \in [0^\circ, 360^\circ]$.

然而, 由于深度与法向量的扫描空间较大、能量函数约束性较低等原因, 此类算法的速度与精度都较差. 而在本文中, 由于第一阶段沿主方向对空间平面的确定将法向量与主方向不一致的空间平面的扫描范围约束到非常小的空间内, 使得多方向空间平面扫描过程更可靠、更快速. 具体而言, 对于当前超像素 s , 本文首先采用多视图区域匹配算法^[10] 确定其对应空间平面所在的 K_2 个最有可能的深度值, 与算法^[10] 略有不同的是, 算法^[10] 虽然指定了超像素 s 的主方向, 但仍需要在较大的深度范围内进行匹配区域的扫描; 而本文算法虽然没有指定超像素 s 的主方向, 但在每幅相邻视图中, 则将其候选匹配区域限制在未获得深度值的区域内, 极大地缩小了候选匹配区域的数量, 因而具有更高的可靠性与速度. 此外, 对于每个可能的深度值下的法向量扫描, 类似于融入空间平面先验的数据项, 本文也将建筑区域垂直结构的先验融入到空间平面扫描时的可靠性度量 $E(s, H_{d,n})$ 中, 即

$$E(s, H_{d,n}) = E_{\text{data}}(s, H_{d,n}) - 2\sigma^2 \log(p(H_{d,n})), \quad (15)$$

式 (15) 中, $H_{d,n}$ 为由当前深度与法向量构成的空间平面, $p(H_{d,n})$ 为 $H_{d,n}$ 的垂直方向先验, 可用球面坐标系下法向量参数 ϕ 表达为

$$p(H_{d,n}) = e^{-d(\phi_{d,n}, \phi_o)}, \quad (16)$$

其中, $\phi_{d,n}$ 为当前空间平面 $H_{d,n}$ 的法向量 ϕ 值, ϕ_o 为主方向 $\{n_i\}$ 中与水平法向量的 ϕ 值最近的主方向 n_i 的 ϕ 值, 度量了建筑区域当前的垂直特性. 显然, 空间平面 $H_{d,n}$ 的方向与建筑区域主体方向越一致, 则成为超像素 s 对应的空间平面的可能性越大.

最后, 从所有深度与法向量的组合中选择使 $E(s, H_{d,n})$ 值最小的前 K_3 个可能的空间平面作为超像素 s 的初始空间平面集. 同样, 通过对所有超像素的初始空间平面集进行聚类则可确定非主方向上空间平面推断时的候选空间平面集 $\{\Pi_{i,k}\}$ 及空间平面 $\Pi_{i,k}$ 的先验 (参见 6.3.1 小节).

6.4 两阶段空间平面的推断

在确定了候选空间平面集 $\{\Pi_{i,k}\}$ 及空间平面 $\Pi_{i,k}$ 的先验后, 本文在 MRF 能量优化框架中推断弱纹理、倾斜表面等区域对应的空间平面; 其中, MRF 的位置集对应于视图过分割得到的超像素, 而标记集 L_s 则由候选空间平面集 $\{\Pi_{i,k}\}$ 中的元素序号构成. 最终通过标记问题的求解, 为每个超像素 s 确定最优标记 $l_s \in L_s$, 使得 $\{l_s\}$ 对应的空间平面集相应的能量函数最小.

式 (3) 形式的能量函数的求解一般属于 NP-hard 问题, 通常无法确定全局最优解, 本文采用 Graph Cuts 获取其近似解. 为了解决主方向集合不完备的问题, 本文采用两阶段的 Graph Cuts 求解弱纹理、倾斜表面等区域对应的空间平面, 即: (1) 根据主方向候选平面集, 利用 Graph Cuts 迭代求取主方向上的所有空间平面; (2) 在主方向空间平面的约束下, 确定非主方向候选空间平面集, 然后利用 Graph Cuts 迭代求取非主方向上的所有空间平面. 此外, 为了提高 Graph Cuts 的求解速度, 本文将 MRF 中已正确求取空间平面的超像素视为常量处理.

综上所述, 对于参考视图 I_r 及其相邻视图 $\{N_i\}$, 以参考视图 I_r 为例, 设其初始不可靠超像素集合为 U_0 , 分别以 U_1, U_2 表示两阶段 Graph Cuts 迭代求解中产生的不可靠超像素集合, $|U_1^i|, |U_2^i|$ 表示第 i 次迭代求解时 U_1, U_2 包含的元素数, 则两阶段空间平面优化过程如算法 2 所示.

算法 2: MRF 能量优化框架下的空间平面推断与深度图获取

输入: 初始深度图 Depth、不可靠超像素集合 U_0 ;
 输出: 稠密化的深度图 Depth;
 初始化: $U_1 = \emptyset, U_2 = \emptyset$;
 % 主方向的空间平面的推断
 1: 利用主方向候选空间平面集确定 MRF 标记集;
 2: **while** $U_1 = \emptyset$ or $|U_1^{i+1}| - |U_1^i| > 0$ **do**
 3: Graph Cuts 对能量函数 (3) 进行求解;
 4: 检测不可靠超像素 (参见 6.1 小节) 并将其加入至 U_1 ;
 5: 更新空间平面先验;
 6: **end-while**
 7: 根据 $(U_0 - U_1)$ 内超像素的标记更新深度图 Depth;
 % 非主方向的空间平面的推断
 8: 求取非主方向候选空间平面集 (参见 6.3.2 小节) 及 MRF 标记集;
 9: **while** $U_2 = \emptyset$ or $|U_2^{i+1}| - |U_2^i| > 0$ **do**
 10: Graph Cuts 对能量函数 (3) 进行求解;
 11: 检测不可靠超像素 (参见 6.1 小节) 并将其加入至 U_2 ;
 12: 更新空间平面先验;
 13: **end-while**
 14: 根据 $(U_1 - U_2)$ 内超像素的标记更新深度图 Depth;

需要强调的是, 以上过程同时在参考视图及其相邻视图中进行, 因而, 最终可同时获得多幅相应的深度图. 实验中, 相邻视图数至少应为 2. 随着相邻视图数量的增加, 算法可靠性增强, 但速度随之降低.

7 实验与分析

从本文算法原理上而言, 相对于普通分辨率视图, 采用高分辨率视图可以获得更好的效果. 考虑到利用普通分辨率进行稠密场景重建是传统算法的难点, 因而, 为了验证算法的可行性及对视图分辨率的鲁棒性, 本文采用牛津大学提供的标准数据集¹⁾ Wadham, Valbonne 及自行拍摄的数据集“清华物理楼”、“生命科学院”(视图分辨率分别为 1024×768 、 512×768 、 819×546 及 1092×728) 对其进行测试. 此外, 两类数据集参考视图分别采用 2 个和 4 个相邻视图进行实验以验证算法对视图数量的依赖性. 本文实验环境为 32 位 Win7 系统, 处理器为 Intel 2.33 GHz 双核 CPU、内存为 4 GB. 算法采用 Matlab 2012 实现.

1) VGG: <http://www.robots.ox.ac.uk/vgg/data/data-mview.html>.

7.1 参数设置

本文所有实验采用相同的参数设置, 以下对其中的关键参数设置问题进行讨论.

对于空间点扩散阈值的选择, 我们实验发现, 在相邻视图数 k 为 2 时, 将候选匹配区域尺寸 w 设置为 5×5 , 极线采样间隔 ϑ 及空间点可靠性验证阈值 ς 分别设置为 0.5, 0.1 时, 扩散结果和速度能得到比较好的均衡. 随着相邻视图数的增加, 扩散结果中包含的外点减少, 计算复杂度也相应增大.

在空间平面全局优化过程中, 实验发现, 本文算法对参数的设置并不敏感.

空间平面先验常数 σ 设置过小, 则先验信息的约束力较小, 不利于正确空间平面的求取, 反之, 如果 σ 设置过大, 则趋向于强制为所有超像素分配先验较大的空间平面, 可能会导致真实空间平面中先验较小的超像素无法获得正确的空间平面. 实验发现, σ 设置在区间 $[0.8, 1.5]$, 性能影响不大.

遮挡惩罚常数 λ_{occ} 如果设置过小, 则即使在错误空间平面的诱导下, 能量函数中的数据项仍然可能获得较小值, 不利于正确空间平面的求取; 如果设置过大, 则即使在正确空间平面的诱导下, 能量函数中的数据项仍然可能获得较大值, 也不利于正确空间平面的求取. 实验发现, 遮挡惩罚常数 λ_{occ} 可选区间为 $[0.4, 0.8]$.

平滑项权重 λ_{smo} 度量了当前超像素的相邻超像素对其空间平面求取的约束能力, 实际中应根据数据项值进行设置. 在本文所有数据集的实验中, 平滑项权重统一设置为 0.5, 表示数据项与平滑项对空间平面的求取具有相同的贡献.

空间平面外点剔除时的比例阈值 K_1 在实验中设置为 0.1, 而深度候选阈值 K_2 及每个超像素对应的初始空间平面数 K_3 均设置为 10. 此三值设置越高, 则全局优化时的可靠性越高, 而求解速度则越慢.

7.2 标准数据集

从实验结果可以发现, 两组数据集都存在诸如天空、地面等非重建区域, 如果采用传统算法, 这些区域的像素匹配及外点剔除过程将会对整体重建速度造成很大的影响, 尤其在 valbonne 数据集中, 天空区域也表现出比较丰富的纹理特征, 匹配时往往会产生大量难以剔除的外点. 而在本文中, 通过视图的语义约束直接对建筑区域进行重建, 不但极大地提高了整体重建速度 (表 1 所示), 同时也保证了利用平面场景模型对弱纹理、倾斜表面等区域的深度进行推断的可靠性. 而在求取初始深度图时, 利用基于 DAISY 特征的空间点的扩散通常可以获得更稠密、更可靠的结果, 使得后续的平面拟合、建筑区域主方向及空间平面先验的求解更可靠. 如图 4(f) 所示的 Wadham 数据集建筑区域的主方向显然比传统 Manhattan-world 模型的 3 个主方向更准确 (如屋顶区域). 对于弱纹理、倾斜表面等区域, 如图 4(g) 中的白色矩形区域, 两阶段的空间平面全局优化准确地求取了其对应的空间平面, 最终获得了稠密的重建结果.

7.3 实拍数据集

对于自行拍摄的两组数据集, 从图 5 可以发现, 尽管建筑区域具有平面结构特征, 但其中的平面分布比较复杂, 如果采用传统的 Manhattan-world 模型进行重建, 必然会产生较大的错误; 而且, 各平面之间也存在着自遮挡情况, 两平面邻接区域的像素很难得以正确匹配. 此外, 两组数据集集中的视图间的透视畸变及光照变化也比较大, 而且存在相对较多的弱纹理与倾斜表面区域, 如“清华物理楼”数据集集中的玻璃区域及在“生命科学院”数据集集中的倾斜表面区域 (图 6(d) 中的白色矩形区域). 最终实

表 1 阈值 $\varepsilon = 0.01$ 重建正确率及消耗时间的对比Table 1 Comparison of accuracy rate and elapsed time result when $\varepsilon = 0.01$

Data set	Elapsed time (minute)		Accuracy rate	
	Our algorithm	Algorithm ^[17]	Our algorithm	Algorithm ^[17]
Wadham	14.9034	58.7612	0.8867	0.6838
Valbonne	9.6669	34.1797	0.9189	0.7939
Tsinghua physics building	10.9781	41.0841	0.9057	0.7195
Life science building	18.0301	61.3377	0.9209	0.7632

图 4 标准数据集实验结果

Figure 4 Experiment result of standard data set. (a) Sample views; (b) semantic constraint of building; (c) over-segmented views; (d) initial depth map; (e) initial plane set; (f) planes along directions which contain the top four (two) most planes; (g) the result of global planes optimization; (h) close-up of rectangular region of reconstruction result

验表明, 以上问题利用本文算法均可得到比较好的解决, 特别对于一些严重遮挡区域, 如“生命科学学院”数据集集中的视图左侧的被遮挡区域, 本文算法仍然可以得到正确的重建结果, 如图 6(d), 图 7 所示.

图 5 实拍数据集及其中两相邻视图

Figure 5 Two neighboring views of our own data set. (a) Tsinghua physics building; (b) life science building

图 6 实拍数据集实验结果

Figure 6 Experiment result of our own data set. (a) Semantic constraint of building; (b) initial depth map; (c) planes along principal direction; (d) reconstruction result

图 7 重建结果的顶部视图与图 6(d) 中矩形区域放大显示

Figure 7 Top view of reconstruction result and close-up of rectangular region in Figure 6(d). (a) Top view of tsinghua physics building; (b) close-up of tsinghua physics building; (c) top view of life science building; (d) close-up of life science building

7.4 算法比较

为了进一步验证本文算法的性能, 本文也与其他优秀的重建算法进行了对比实验. 鉴于许多算法 (如 Bundler+PMVS^[2]) 从数量较少的普通分辨率视图中很难获得较好的稠密重建结果, 所以我们仅与稠密重建算法^[17]进行了对比实验. 此外, 由于本文算法首先利用语义标注算法对建筑区域进行了约束, 为了保证对比结果的合理性, 文献^[17]的算法也仅针对建筑区域进行, 算法实现的环境也与本文算法相同.

图 8 不同阈值 ε 时正确率的对比

Figure 8 Comparison of accuracy rate with different ε . (a) Wadham; (b) valbonne; (c) Tsinghua physics building; (d) life science building

为了保证对比结果的公平性, 类似于文献 [17] 中剔除深度图外点的方法, 本文定义了重建完整性的度量标准, 即重建正确率 T . 设参考视图及 k 个相邻视图对应的深度图分别为 $D_r, \{D_i\}(i = 1, \dots, k)$, D_r 中建筑区域全部像素集合为 A . 对于 $m \in A$, 其对应的空间点 X_m 在深度图 D_i 中的已知深度值为 $d(X_m, D_i)$, 而根据 X_m 与 D_i 的摄像机参数求取的 X_m 相对于 D_i 摄像机中心的深度值为 $\lambda(X_m, D_i)$; 显然, 如果 X_m 相对于 D_i 是可靠的, 则 $d(X_m, D_i)$ 与 $\lambda(X_m, D_i)$ 的差别应当非常小, 如果 X_m 在至少 N 个相邻深度图中均是可靠的, 则认为 X_m 即是可靠的空间点, 相应深度值也是可靠的. 因而, 重建正确率 T 可定义为

$$T = \frac{1}{|A|} \sum_{m \in A} \delta \left(\sum_{i=1}^k \delta \left(\frac{\|\lambda(X_m, D_i) - d(X_m, D_i)\|}{d(X_m, D_i)} < \varepsilon \right) > N \right), \quad (17)$$

其中, ε 为深度值差异阈值, 函数 $\delta(\cdot)$ 的定义同式 (11), 阈值 N 度量了空间点 X_m 相对多个视图的可靠性, 实验统一设置为 4.

表 1 为 $\varepsilon = 0.01$ 时的重建正确率及消耗时间的对比. 由于本文算法同时求取多幅稠密深度图, 表 1 中的本文算法消耗时间为平均消耗时间, 正确率则为参考视图对应深度图重建的正确率. 为了更全面地考察两种算法性能的差异, 对于每组数据集, 本文也对不同阈值 ε 时的两种算法获得的正确率 T 进行了对比, 如图 8 所示. 图 9 所示为算法 [17] 相应的重建结果.

图 9 算法^[17]重建结果

Figure 9 Reconstruction result of algorithm^[17] ($\varepsilon = 0.01$). (a) Wadham; (b) valbonne; (c) Tsinghua physics building; (d) life science building

实验结果可以发现, 算法^[17]并不能解决弱纹理、倾斜表面等区域的重建问题 (如图 9 中的白色矩形区域), 相对而言, 本文算法的整体性能明显更高. 分析表明, 基于 DAISY 特征的空间点扩散算法不但节约了大量的 DAISY 特征计算与比对时间, 而且在一定程度上缓解了 DAISY 特征的匹配多义性问题, 从而可以获得更稠密、更准确的深度图; 而在对弱纹理、倾斜表面等区域所对应的空间平面进行推断时, 由于视图过分割获取的超像素数量较少及空间平面先验的融入, 使得法向量聚类及两阶段 Graph Cuts 对能量函数的迭代求解等过程均耗时不多, 如数据集“生命科学院”建筑区域包含 5522 个超像素, 法向量聚类及全局优化过程仅用时 1 分钟左右. 显然, 如果将场景中的天空、地面等非重建区域也考虑在内, 两种算法消耗时间的差异将会更大, 本文算法将表现出更强的优势. 其实, 在阈值 ε 较小 (如图 8 所示) 时, 本文算法的重建正确率并未表现出较好的优势, 据我们分析, 这应该是算法所依赖的场景分段平滑假设所致, 这是我们下一步研究中需要改进的地方.

8 总结与展望

本文提出了一种快速有效的基于多视图的场景稠密深度图重建算法. 为了快速地获取稠密、准确的重建结果, 本文算法首先对视图中的建筑区域进行语义分割剔除了非重建区域的干扰, 而为了有效解决弱纹理、倾斜表面等区域的深度推断问题, 本文根据场景分段平滑的假设, 首先采用基于 DAISY 特征的两阶段空间点稠密扩散技术及平面拟合方法获取比较多的可靠初始空间平面, 然后在此基础上, 将其他空间平面的推断问题转化为 MRF 能量优化框架下的能量优化问题, 最后采用 Graph Cuts 对融合颜色、空间平面之间的位置关系、场景先验等信息的能量函数进行了有效求解. 实验结果表明, 本文算法能有效克服光照变化、透视畸变、弱纹理区域等因素的影响, 快速恢复建筑区域完整的深度图. 目前, 本文算法的缺点主要体现在以下两个方面: 首先, 在有些情况下, 场景分段平滑假设可能会产生较大的误差, 如建筑中的弧面区域, 即使使用很小的空间平面去近似, 最终深度值的推断仍会存在偏差; 其次, 算法在一定程度上会受到视图过分割算法精度的影响, 如当超像素内部像素对应的深度值跳变较大时, 其对应的空间区域并不能用空间平面近似. 针对这些问题, 在下一步的工作中, 本文计划根据超像素内部像素颜色分布及相应深度值的变化等信息对其进行子分割以进一步增强平面模型对其近似的可靠性, 同时在能量优化时融入高阶能量项以获取更准确的空间平面优化结果, 进而获取更稠密、更准确的深度图, 也是我们下一步准备进行的工作.

参考文献

- 1 Snavely N, Seitz S M, Szeliski R. Modeling the world from internet photo collections. *Int J Comput Vision*, 2008, 80:

- 189–210
- 2 Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereopsis. *IEEE Trans Patt Anal Mach Intel*, 2010, 32: 1362–1376
- 3 Hong L, Chen G. Segment-based stereo matching using graph cuts. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, Washington, 2004. 1-74–1-81
- 4 Zitnick C L, Kang S B. Stereo for image-based rendering using image over-segmentation. *IEEE Trans Patt Anal Mach Intel*, 2007, 75: 49–65
- 5 Wang Z F, Zheng Z G. A region based stereo matching algorithm using cooperative optimization. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, Anchorage, 2008. 1–8
- 6 Klaus A, Sormann M, Karner K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: *Proceedings of IEEE 18th International Conference on Pattern Recognition*, Hong Kong, 2006. 15–18
- 7 Çiğla C, Zabulis X, Aydin Alatan A. Segment-based stereo-matching via plane and angle sweeping. In: *Proceedings of 3DTV Conference*, Kos Island, 2007. 1–4
- 8 Furukawa Y, Curless B, Seitz S M, et al. Manhattan-world stereo. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, Miami, 2009. 1422–1429
- 9 Sinha S N, Steedly D, Szeliski R. Piecewise planar stereo for image-based rendering. In: *Proceedings of the 12th IEEE International Conference on Computer Vision(ICCV)*, Kyoto, 2009. 1881–1888
- 10 Mičušík B, Košecká J. Multi-view superpixel stereo in urban environments. *Int J Comput Vision*, 2010, 89: 106–119
- 11 Häne C, Zach C, Cohen A, et al. Joint 3D scene reconstruction and class segmentation. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, Portland, 2013. 97–104
- 12 Bao Y, Chandraker M, Lin Y, et al. Dense object reconstruction with semantic priors. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, Portland, 2013. 1264–1271
- 13 Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Trans Patt Anal Mach Intel*, 2001, 23: 1222–1239
- 14 Bleyer M, Rhemann C, Rother C. Patchmatch stereo-stereo matching with slanted support windows. In: *Proceedings of the British Machine Vision Conference(BMVC)*, Dundee, 2011. 1–11
- 15 Shen S. Accurate multiple view 3D reconstruction using patch based stereo for large-scale scenes. *IEEE Trans Image Process*, 2013, 22: 1901–1914
- 16 Gallup D, Frahm J M, Pollefeys M. Piecewise planar and non-planar stereo for urban scene reconstruction. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, San Francisco, 2010. 1418–1425
- 17 Tola E, Strecha C, Fua P. Efficient large-scale multi-view stereo for ultra high-resolution image sets. *Mach Vision Appl*, 2012, 23: 903–920
- 18 Gould S, Fulton R, Koller D. Decomposing a scene into geometric and semantically consistent regions. In: *Proceedings of the 12th IEEE International Conference on Computer Vision(ICCV)*, Kyoto, 2009. 1–8
- 19 Tola E, Lepetit V, Fua P. Daisy: an efficient dense descriptor applied to wide-baseline stereo. *IEEE Trans Patt Anal Mach Intel*, 2010, 32: 815–830
- 20 Lhuillier M, Quan L. Match propagation for image-based modeling and rendering. *IEEE Trans Patt Anal Mach Intel*, 2002, 24: 1140–1146
- 21 Hartley R, Zisserman A. *Multiple View Geometry in Computer Vision*. 2nd ed. Cambridge: Cambridge University Press, 2004. 1–672
- 22 Lowe D. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis*, 2004, 60: 91–110
- 23 Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. *IEEE Trans Patt Anal Mach Intel*, 2002, 24: 603–619
- 24 Frey B J, Dueck D. Clustering by passing messages between data points. *Science*, 2007, 315: 972–976

Dense 3D scene reconstruction based on semantic constraint and graph cuts

WANG Wei, GAO Wei* & HU ZhanYi

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

*E-mail: wgao@nlpr.ia.ac.cn

Abstract A new and effective dense 3D scene reconstruction algorithm is proposed in this work. At first, in order to speedup the reconstruction process and reduce possible unnecessary interference by unrelated regions, such as sky, ground, etc., to the desired architectural scene, a semantic image segmentation step is carried out at first to mask those uninterested parts. After that, an initial depth map is obtained by Daisy-feature based 3D point propagation. To tackle the difficult dense reconstruction problem of poorly textured regions and slanted surfaces, the problem is formulated as a plane inference problem under the MRF energy minimization framework to assign a possible planar patch to each image superpixel. More specifically, at first, the plane principal directions and possible non-principal are clustered by the affinity propagation from the reconstructed point clouds, then the plane inference is carried out along the principal directions first, followed by the non-principal directions. In each of the above two plane inference steps, a Graph Cuts based algorithm is employed to minimize a properly designed cost function which takes into account a variety of factors, such as appearance similarity, the constraints from the initial depth map, spatial planes priors, as well as geometric relations between spatial planes etc. Experiments on the standard data set as well as our own data set show that the our proposed algorithm can satisfactorily handle many challenging factors, e.g. illumination variation, perspective distortion, poorly textured regions, etc., to efficiently recover complete dense depth map of buildings.

Keywords semantic annotation, plane-fitting, match propagation, depth map, energy optimization

WANG Wei received his master degree in 2011 from the SouthWest Jiao-Tong University. Currently, he is a Ph.D. candidate at the Institute of Automation, the Chinese Academy of Sciences. His research interest covers computer vision and machine learning.

GAO Wei was born in 1980. He received his Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences at 2008. Currently, he is an Associate Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers computer vision and 3D reconstruction.

HU ZhanYi was born in 1961. He received his B.S. degree in automation from the North China University of Technology in 1985, the Ph.D. degree (Docteur d'Etat) in computer vision from the University of Liege, Belgium, 1993. He is now a Professor at the Institute of Automation, Chinese Academy of Sciences.