基于特征点跟踪的视频序列平面提取算法

陶 $\overline{\mathbf{m}}^1$ 王 $\overline{\boldsymbol{\mu}}^1$ 张 $\overline{\mathbf{a}}^2$

(西安交通大学电子与信息工程学院 西安 710049)1 (北京理工大学计算机学院 北京 100081)2

摘 要 平面结构是一种最为常见和重要的几何特征,平面提取在很多领域有着重要的应用。提出了一种毋须恢复 摄像机三维信息、直接从视频序列中提取出平面的算法。先通过在视频相邻帧之间建立极线约束下的射影变换模型, 来计算出由平面诱导的单应,而每一个单应都对应着场景中的一个平面;然后根据计算出的单应对视频的每一帧进行 分割,进一步提取出视频中的平面结构。一系列真实视频序列的实验验证了本算法的有效性和稳定性。 关键词 平面提取,射影变换,Graph Cuts 中图法分类号 TP391.4 文献标识码 A DOI 10.11896/j.issn.1002-137X.2014.10.022

Multiple Plane Extraction Based on Feature Point Tracking over Video Sequences

TAO Lei¹ WANG Ping¹ ZHANG Lei²

(School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)¹ (School of Computer Science, Beijing Institute of Technology, Beijing 100081, China)²

Abstract Planar structures are abundant in both man-made and natural environments which enables the use of planes in various vision tasks. We introduced a new approach for the robust detection of multiple planes over a video sequence without camera calibration or 3D reconstruction. Given a video sequence, we first built a projective geometry model between adjacent frames based on epipolar constraint. Then a homography induced by a plane was computed, with which we could get the planar structure by doing video segmentation. Experimental results on a variety of real video sequences have verified the effectiveness and efficiency of our method.

Keywords Plane detection, Projective transformation homography, Graph Cuts

1 引言

平面提取一直是计算机视觉和图形学领域内一个重要的 研究方向。平面结构是现实场景中最为常见的一类基础几何 特征之一,其中包含了物体的大量二维结构信息,并对位于平 面上的点有着很强的几何约束。在很多实际应用中,如构建 场景空间、障碍物或者目标物体模型等,都需要预先提取出场 景中的平面结构进行处理或者交互。

近年来,已经出现了很多有效的平面提取算法。根据是 否需要提供摄像机参数或者场景的三维信息作为先验知识, 大致可以将它们分为两类。

第一类算法主要是在恢复场景三维信息的基础上,根据 某种准则对三维点云进行聚类,以提取出平面结构。一个经 典代表是 Ogundana 等人^[1]在 2011 年提出的空间三维平面 提取算法。在已知场景三维信息的前提下,该算法将经典 Hough 变换拓展到三维空间,从而实现了立体图像中的平面 目标检测。Zhou^[2]先通过传统的 SFM(structure from motion)得到场景的三维点云信息以及摄像机参数,然后通过一 种被称为 J-Linkage^[3]的点云聚类算法,从三维点云中提取出 空间三维平面,重投影到图像中,得到图像中的平面结构。

然而,作为很多应用的预处理步骤,往往需要在不知道场 景的三维结构信息或者摄像机参数等先验知识的情况下进行 平面提取。由此产生了第二类平面提取算法,这类算法大多 是基于图像间几何特征的提取和匹配来实现的,毋须事先对 场景的三维结构信息或者摄像机参数进行计算恢复。比较经 典的有 Fischler 和 BollesSinclair^[4]提出的基于随机采样一致 性(RANSAC)的平面提取算法,其对两幅图像之间的匹配点 进行平面拟合以提取出平面结构。后续的很多平面提取算法 如 Choi 等人^[5]在 2007 年提出的算法、本文的算法都借鉴了 该算法的思想。Bouchafa 等人^[6]根据相机移动过程中光流的 变化,提出了 c-velocity 空间的概念,文章认为在一个作平移 运动的摄像机中平面结构在 c-velocity 空间中呈现抛物线型, 据此可以提取出场景中的平面结构。Fouhey 和 Scharstein^[7] 在 2010 年提出的平面提取算法借鉴了 J-linkage 聚类算法的 思想,通过提取特征点进行平面单应假设,对假设进行分析拟 合,最终提取出平面。Kim 和 Lee 等人^[8]则通过对两幅图像

到稿日期:2013-06-24 返修日期:2013-08-09 本文受国家自然科学青年基金资助项目(61103159),北京市自然科学基金资助项目 (4132068)资助。

陶 曹(1989一),女,硕士,主要研究方向为三维重建,E-mail;leitao891123@gmail.com;王 萍(1976一),女,博士,讲师,主要研究方向为图像 处理、基于内容的视频分析,E-mail;ping.fu@stu.xjtu.edu.cn(通信作者);张 磊(1981一),男,博士,讲师,主要研究方向为计算机图形学、图像 与视频处理。

进行线匹配,对两条直线的共面性进行假设,进一步进行分析 和合并,最终得到图像中的平面结构。

第一类提取算法得到的提取结果一般是最准确的。但 是,这类算法需要预先恢复出场景的三维信息以及摄像机参 数,而这个过程计算量极大,如果我们仅仅需要提取图像或视 频中的平面结构,这类算法代价过高。第二类算法一般使用 仿射变换来建立图像对之间的变换关系模型,忽略了变换过 程中可能存在的非线性效应,这往往是不够准确的;而光流法 也要求摄像机只做平移运动,这在现实情况下往往是很难做 到的。此外,目前已有的第二类平面提取算法基本都是针对 图像来进行的,完全可以进一步将其拓展到视频空间。图像 的平面提取通常是在两幅视图之间进行,而视频的平面提取 除了要在相邻帧之间提取出平面,还要保证该平面在整个视 频中的全局一致性,即对视频每一帧最终提取出的平面场景 结构应当是一致的。针对这些问题,本文在前人工作的基础 上提出了一种新的毋须恢复摄像机的三维信息而直接可以从 视频序列中提取出平面的算法。该算法主要借鉴了 RANSAC平面提取算法的思想,在此基础上,利用不共线3 点即可确定一个平面这一公理对其进行了改进,同时引入了 平面三角剖分的思想;此外,为了保证视频中平面结构的整体 一致性,在对每一帧进行计算时需要考虑所有视频帧对当前 帧的影响,本文算法选择对特征点的轨迹而非特征点进行操 作。与传统的 RANSAC 算法相比,本文算法的准确性更高, 而且实验次数也更少。

本文第2节对算法涉及的一些基础概念和理论进行简单 介绍;第3节对本文提出的算法进行详细描述;第4节是实验 结果与分析;最后对全文的工作进行总结,并指出下一步研究 的方向。

2 背景知识

在详细描述本文的算法之前,先对一些将要用到的多视 图几何的相关背景知识作一些简要介绍。

2.1 基本矩阵和对极几何

对极几何反映的是同一场景下,两幅不同的投影图像之间存在的内在的相互约束的关系。对极几何与场景环境独立,只和摄像机的外参、内参相关,是在未标定摄像机的情况下,唯一可以从二维图像中得到的信息。

如图 1 所示,在同一场景中,将两个相机的光心设为 O_1 与 O_2 ,将两个相机的成像平面设为 I 和 I',假定存在空间一 点 X,它在两幅图像平面上的投影分别为点 x 和点 x',则点 x和点 x'称为一对匹配点。将光心 O_1 与 O_2 连起来的直线称 为摄像机基线。摄像机基线与两幅图像分别相交于对极点 e和点 e'。空间点 X 与光心 O_1 , O_2 组成的平面被称作极平面, 该平面与两幅图像分别相交于直线 l 和 l'。点 x 可以反向投 影成三维空间的一条射线,它由第一个摄像机光心 O_1 和 x 确 定。这条射线在 I'中的投影就是直线 l'。而空间点 X 必然在 这条射线上,因此 X 在 I'中的投影图像点 x'必然在 l'上,直 线 l'称为图像 I'上对应于 x 的对极线。



图 1 对极几何

基本矩阵是对极几何的代数表示。如前文所述,对于一 幅图像上的任一点 x,在另一幅图像中存在一条对应的对极 线 l[′]。由此存在一个从一幅图像上的点到另一幅图像上与之 对应的对极线的映射,即由称为基本矩阵的矩阵 **F** 表示的从 点到直线的射影映射:

l'=Fx (1)
 基本矩阵是一个秩 2 的 3×3 矩阵,有如下性质^[9]:
 1. 点对应:如果 x 和 x[']是一对对应的图像点,那么有:

$$\boldsymbol{x}^{T}\boldsymbol{F}\boldsymbol{x}=0$$
(2)

2. 对极线:l' = Fx是对应于点 x 的对极线; $l = F^T x'$ 是对应于点 x'的对极线。

3. 对极点:基本矩阵 F 和两幅图像中的对极点 e 和 e'满 足如下关系:

$$\begin{cases} Fe=0\\ F^{\mathsf{T}}e'=0 \end{cases}$$
(3)

2.2 平面诱导的单应

如图 2 所示,存在一空间平面 $\pi = [v^{T}l]^{T}$,平面 $\pi \perp - 点 X_{\pi} \alpha I$ 和 I'两幅视图中的投影分别为点 x和点 x'。从 x到 x'的映射就被称为由平面 π 诱导的单应。对于一般位置的平面,上述单应由该平面唯一确定,反之亦然^[9,10]。



图 2 平面诱导的单应

假设给定两幅图像的投影矩阵为 P = [I|0]和 P' = [A|a],其中 I 是 3×3 单位矩阵,0 是 3 维零矢量,A 是一个 3×3 矩阵,a 是 3 维矢量,则由平面 $\pi = [v^T l]^T$ 诱导的单应 H 可以 表示如下:

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \tag{4}$$

其中,

$$\boldsymbol{H} = \boldsymbol{A} - \boldsymbol{a} \boldsymbol{v}^{\mathrm{T}}$$
(5)

在式(4)中, x 和 x⁷均为以齐次坐标表示的点, 齐次矩阵 H 是一个自由度为 8 的 3×3 矩阵。

2.3 给定基本矩阵 F 和图像对求平面诱导的单应

3D 空间的一张平面可以由 3 个点来确定,而这些 3D 元 素都可以由图像对应来得到。假设 3 个空间非共线点 X_1 , X_2 , X_3 在两幅视图中的影像和基本矩阵 F 都已知,那么由这 3 点所在平面诱导的单应 H 可以通过以下方法来计算:已知 F 和 3 组图像点对应 $x_i \leftrightarrow x_i'(i=1,2,3)$,则 3D 点所在平面诱 导的单应为:

 $\boldsymbol{H} = \boldsymbol{A} - \boldsymbol{e}' (\boldsymbol{M}^{-1} \boldsymbol{b})^{\mathrm{T}}$ (6)

• 96 •

其中, $A = [e']_{\times} F, e'$ 为第二幅图像上的对极点。 $[e']_{\times}$ 表示 e'的反对称矩阵,假设 $e' = (e_1, e_2, e_3)^{T}, f_1$:

$$\begin{bmatrix} e' \end{bmatrix}_{\times} = \begin{bmatrix} 0 & -e_3 & e_2 \\ e_2 & 0 & -e_1 \\ -e_1 & -e_2 & -e_1 \end{bmatrix}$$
(7)

b是一个3维矢量,其元素为:

 $b_i = (\mathbf{x}_i' \times (\mathbf{A}\mathbf{x}_i))^{\mathrm{T}} (\mathbf{x}_i' \times \mathbf{e}') / \| \mathbf{x}_i' \times \mathbf{e}' \|^2$ (8) 而 *M* 是一个行矢量为 $\mathbf{x}_i^{\mathrm{T}}$ 的 3×3 矩阵。

2.4 重投影误差

在计算得到单应矩阵 H 之后,可定义如下重投影误差,以判断一对匹配点对 $x_i \leftrightarrow x_i'$ 是否属于当前单应 H 所对应的 平面:

 $e_{i} = \| \mathbf{x}_{i} - \mathbf{H}\mathbf{x}_{i} \|^{2} + \| \mathbf{x}_{i}' - \mathbf{H}^{-1}\mathbf{x}_{i}' \|^{2}$ (9)

当重投影误差 ei 小于某一给定阈值时,认为匹配点对 xi ↔xi 属于单应矩阵 H 所对应的平面。

3 算法介绍

基于 RANSAC 的平面提取算法通过随机选取两幅图像 之间的匹配点进行实验以拟合出平面,实现对平面结构的提 取。本文算法在该算法的基础上进行了改进,并进一步推广 到视频领域。具体来说,本文算法分为以下 3 个步骤:1)对视 频做特征点跟踪得到轨迹;2)基于轨迹的平面跟踪和提取;3) 根据平面诱导的单应对每一帧图像进行分割,最终得到平面 结构。具体流程如图 3 所示,下面将分别详细介绍。





3.1 特征点跟踪

关于特征点提取和跟踪的理论和方法有很多,比较成熟 的算法有 SIFT 特征匹配算法^[11]和 KLT(Kanade-Lucas-Tomasi)特征点跟踪算法^[12]。在本文中,需要进行提取和跟踪 特征点的对象是视频,如果使用 SIFT 算法,求解速度相对较 慢,比较之下 KLT 算法在保证提取效果的基础上,计算速度 有大幅提升,并且 KLT 算法的前提假设如亮度恒定、时间连 续等,也决定了它更适合于对连续的视频帧进行处理。因此, 本文最终选择了 KLT 特征点跟踪算法。

KLT特征跟踪算法是一种经典的基于特征的跟踪算法, 先通过特征点选择算法在起始图像中选择特征点,然后利用 平移模型进行特征点跟踪,对于第 N 幅图像上跟踪到的特征 点,通过仿射模型进行连续性判断,对跟踪错误的特征点进行 剔除。KLT 算法在研究不同图像之间的匹配问题时,通过计 算两个平移窗口的灰度残差并寻找最小化残差来实现匹配。 在优化过程中,KLT 算法使用泰勒展开直接计算平移矢量, 而不需要通过遍历进行搜索;同时使用 Newton-Raphson 迭 代算法来避免泰勒展开带来的误差轨迹分类。使用 KLT 算 法对视频序列进行特征点跟踪的结果如图 4 所示。



图 4 KLT 特征点跟踪结果

3.2 平面跟踪和提取

平面跟踪和提取部分是本文算法的核心,在这一部分,我 们描述了一种可以从视频序列中跟踪和提取一个平面的新算 法。如果想要得到视频中的所有平面,只需要从轨迹集合中 移除当前平面所包含的轨迹之后,迭代地重复使用这一方法 即可。

本文所提出的平面提取算法主要借鉴了 RANSAC 平面 提取算法的思想。但与传统的 RANSAC 算法每次随机取 4 个特征点进行实验不同,本文利用不共线 3 点即可确定一个 平面这一公理,对跟踪到的特征点进行 Delaunay 三角化,每 次实验随机取一个三角形上的 3 个点结合基本矩阵进行计 算。而一个三角形上的 3 个点属于同一平面的概率明显要高 于随机取的 4 个特征点在同一平面上的概率,因此,本文算法 不仅大大加快了实验收敛的速度,而且准确性更高。同时,与 传统图像平面提取算法不同的是,为了保证提取出的平面结 构在整个视频中的一致性,本文算法直接对特征点的轨迹进 行操作。

下面将对本文的方法进行详细介绍。本文算法是基于抽 样一致性的,因此需要进行多次重复实验,然后从中选取最佳 结果。以下先来介绍一次实验的流程。

对于一个 N 帧的视频序列, 假设已经得到 M 个特征点 轨迹的集合 T = { T_j }^M_j, 对于每一条轨迹 T_j, 令 p_j 和 q_j (1 $\leq p_j < q_j \leq N$)分别表示它的起始和结束帧号, 那么轨迹 T_j 就 可以表示为 T_j = { x_j^i }^A_j p_j 的形式,其中 $x_j^i \in P^2$ 表示第 j 个特 征点在第 i 帧中的齐次坐标。对于跨越第 a 帧和第 b 帧的所 有轨迹的集合可以用 T^{ab} = { T_j \in T: $p_j \leq a, q_j \geq b$ }来表示。 对于一个轨迹点 x_j^i , 如果它位于一个平面上, 那么根据式 (4), 它在第 i 帧和第 k 帧上的坐标就可以由一个该平面所诱 导的单应 H_a来关联,有 $x_j^i = H_a x_j^i$ 。本文使用重投影误差来 衡量一个平面模型对一条轨迹的适应度。对一条轨迹 T_j = { x_j^i } μ_{p_j} ,根据式(9),定义它对第 $k(p_j \leq k \leq q_j)$ 帧的重投影误 差如下:

$$=\sum_{i=1}^{2} (\|x_{j}^{k}-H_{ki}x_{j}^{i}\|^{2}+\|x_{j}^{i}-H_{ki}^{-1}x_{j}^{k}\|^{2})$$
(10)

本文算法的最终目的就是根据重投影误差将所有的轨迹 分为不同的集合,每一个集合对应着一个平面。具体可以分 为两步:

1. 随机采样:给定一个 N 帧的视频序列 $\{F_i\}_{i=1}^N$,可以构造 N-1 对邻接帧对 $C = \{(F_1, F_2), (F_2, F_3), \dots, (F_{N-1}, F_N)\}$ 。首先从 C 中选取任一帧对 (F_{i-1}, F_i) ,取跨越这两帧的轨迹 $T^{(i-1)i}$ 在 F_{i-1} 内的对应点进行 Delaunay 三角化,得到 一个 Delaunay 三角形的集合 T_{ri} 。然后,在 T_{ri} 内取任意一个

三角形 t,根据式(6),使用 t 所对应的 3 条轨迹就可以计算得 到 F_{i-1} 和 F_i 之间的一个单应矩阵 $H_{(i-1)i}$ 。给定一个阈值 ε , 通过判断重投影误差 $e = ||x_i - H_{(i-1)i}x_{i-1}||^2 + ||x_{i-1} - H_{(i-1)i}x_i||^2$ 是否小于 ε ,可以将 $T^{(i-1)i}$ 内的轨迹分为 T_{in} 和 T_{aat} 。

2. 一致性计算:选择与(F_{i-1} , F_i)相邻的帧对(F_i , F_{i+1}), 取 $T^{(i+1)}$ 中已经被标记为 T_m 的轨迹,即 $T_m \cap T^{(i+1)}$ 在 F_i 内 的对应点进行 Delaunay 三角化,得到一个新的 Delaunay 三角 形的集合 T_{ri} 。然后,同上一步,使用 t 所对应的 3 条轨迹计 算得到 F_i 和 F_{i+1} 之间的单应矩阵 $H_{i(i+1)}$,通过判断重投影误 差 e是否小于阈值 ϵ ,可以将 $T^{(i+1)}$ 内的轨迹也分为 T_m 和 T_{out} 。不断重复该步骤直到所有帧均已经过处理或者 $T_m \cap$ $T^{(i+1)}$ 内的轨迹数目小于 3 条。

多次重复步骤 1 和 2,选择 T_{in} 尺寸最大的平面和它所对 应的 $\{H_{(i-1)i}\}_{i=2}^{N}$ 集合,作为本次平面提取的结果输出。

整个流程可以总结为算法 1。本文算法的唯一参数是重 投影误差阈值 ϵ ,这里一般取 ϵ =4。

算法 1

1. 输入: 跨越 N 帧的 M 条轨迹集合 T; 重投影误差阈值 ε 2. repeat for n trials:

- 3. 从 C 中选取任一帧对(F_{i-1},F_i)
- 取跨越这两帧的轨迹 T⁽ⁱ⁻¹⁾ⁱ在 F_{i-1}内的对应点进行 Delaunay 三 角化,得到三角形的集合 T_{tri}
- 5. 在 T_{tri}内取任意一个三角形 t,计算得到单应矩阵 H_{(i-1)i}
- 6. 根据 H_{(i-1)i}将 T⁽ⁱ⁻¹⁾ⁱ内的轨迹分为 T_{in}和 T_{out}
- 7. while 不是 C 中的所有帧对都已处理

8. 选择新的帧对(F_{k-1},F_k)

- 9. if $T_{in} \cap T^{(k-1)k} \leq 3$; break; end if
- 10. 对 T_{in} ∩ T^{(k-1)k}中的轨迹,使用同样的方法计算得到 H_{(k-1)k}
- 11. 根据 H_{(k-1)k}将 T^{(k-1)k}中未分类的轨迹分为 T_{in}和 T_{out}
- 12. end while
- 13. end repeat
- 14. 从实验结果中选择 | T_{in} | 最大的 { H_{(i-1)i} } ≥ 集合
- 15. 输出:平面诱导的单应集合 $\{H_{(i-1)i}\}_{i=2}^{N}$

平面提取的结果如图 5 所示,其中黑色的圆形、十字形和 方形点分别属于地面和左右侧墙壁 3 个不同的平面,白色的 点代表不属于任何平面的噪声点,可以看到本文算法对于视 频序列的划分结果是正确的。而相比传统的基于特征点的平 面提取,本文选择特征点轨迹进行操作,在平面提取的过程中 不单单针对当前帧对,而是综合考虑了整个视频序列的全局 一致性,对平面结构的约束性更强,这使得对于复杂的场景, 本文的算法会更加稳定。具体分析将在实验部分给出。



黑色的不同形状的点分别代表 3 个不同的平面 白色的点表示不属于任何平面的异常点

图 5 平面跟踪提取结果

3.3 基于 Graph Cuts 算法的视频平面分割

在得到了平面模型和它们所对应的 $\{H_{(i-1)i}\}_{i=2}^{N}$ 集合后, 就可以根据 $\{H_{(i-1)i}\}_{i=2}^{N}$ 对视频进行分割以得到最终结果。这 里选择了 Graph Cuts^[13]算法来对视频进行分割。

对一个包含 P 个平面的 N 帧视频序列进行分割的目的 是为每一个像素 $p \in F = \{F_1, F_2, \dots, F_N\}$ 指定一个标签 $f_p \in \{0, 1, \dots, P-1\}$,得到每一帧图像的属于同一平面的像素集 合 $f = \{p; f_p = i, i \in \{0, 1, \dots, P-1\}\}$ 。标签集合 $\{f_p\}$ 通常是 通过最优化一个能量函数得到的。首先,将输入图像用一个 图 G = (N, E)来表示,其中 N 表示图中的结点集合,每一个 结点 $n_i \in N$ 对应于输入图像中的一个像素, E 表示图中结点 之间的边的集合,每一条边对应于一对具有特定关系的结点 $\langle n_p, n_q \rangle$ 。Graph Cuts 算法的最优化目标函数如下:

$$E(f) = \sum_{n \in N} D_n(f_n) + \sum_{n_b < n_q \in N} V_{n_b, n_q}(f_{n_b}, f_{n_q})$$
(11)

其中, $D_n(f_n)$ 表示将结点 n 的标签设为 f_n 时带来的数据项 惩罚。 $n_p < n_q$ 表示图中相邻结点之间的单向组合,这里的 "相邻"可以是四邻域相邻或者八邻域相邻; $V_{n_p,n_q}(f_{n_p}, f_{n_q})$ 表示两相邻结点[n_p, n_q]的标签分别被设为 f_{n_p} 和 f_{n_q} 时带来 的基于空间连续性的平滑项惩罚。各结点的标签{ f_n }通过最 小化上述目标函数得到。

这里,数据项惩罚定义如下:

 $D_n(f_n) = ((R(H[f_n]n) - R(n))^2 + (G(H[f_n]n) -$

 $G(n))^{2} + (B(H[f_{n}]n) - B(n))^{2})^{1/2}$ (12) 其中, $H[f_{n}]$ 为第 f_{n} 个平面所诱导的单应, $R(\cdot), B(\cdot), G$

(•)分别表示当前像素点的 RGB 分量。平滑项惩罚对于节 $点 n_p$ 与其相邻节点 n_q 的连接权值,定义如下:

 $V_{n_p,n_q}(f_{n_p}, f_{n_q}) = \lambda (f_{n_p} - f_{n_q})^2$ (13) 其中, $n_p, n_q \in N, \lambda$ 为常量,用来平衡数据惩罚项和平滑惩罚 项之间的权重,一般取 80~100。

图 6 是对一个视频做 Graph Cuts 所得到的平面分割结果,不同方向的线段代表不同的平面,可以看到,视频中的平面结构已经被很好地分割出来。



图 6 Graph Cuts 分割结果,不同方向的直线分别代表了不同的平面

4 实验结果与分析

为了验证本文算法的有效性,这里分别选择了多个至少 包含了一个平面结构的视频片段进行实验。这些视频序列囊 括了从自然到人工环境、从室内到室外环境等多种可能的情况。

在图 7 中展示了对不同视频序列使用本文算法进行平面 提取的结果,可以看出,不论是在室内还是室外场景中,都提 取出了正确的平面区域,可以说本文算法对于视频序列是有 效的。



图 7 对不同的视频片段进行平面提取的结果

与传统平面提取算法相比,本文所提出的平面提取算法 的最大创新在于:不同于以往通过对图像间的匹配点对进行 分类来提取平面的方法,其是直接对特征点的轨迹进行操作。 在这种情况下,一条轨迹如果在任一帧对内被划分到一个平 面内部,那么在它所跨越的其他所有帧中,该轨迹包含的特征 点也都属于同一个平面。因此,相比传统针对图像的平面提 取算法,本文的算法可以更好地保证平面结构在视频序列中 的整体一致性。如图 8(a)所示,在视角变换的过程中,建筑 物的左侧平面上的特征点数目在逐渐减少。图 8(b)中我们 使用了文献[7]中的平面检测算法,前期虽然可以很好地检测 到左右两个平面,但是到了后期,由于左侧平面上的特征点数 目过少,会将其归为噪声点,认为图像中仅存在右侧平面一个 平面结构,这就与前面结果不相符合,导致平面提取出现错 误。而本文算法洗择特征点轨迹进行计算,进行聚类的对象 也是轨迹,因此即使到了后期左侧平面特征点数目很少的时 候,由于之前已经将包含这些特征点的轨迹分到了左侧平面 上,因此仍然可以检测出左侧平面,如图 8(c)所示,从而保证 了平面结构的前后一致性。



(c) 本文算法提取结果

图 8 平面提取结果对比

而目前已知的对视频进行平面提取的算法仅有 Zhou^[2] 等人在 2013 年提出的方法。该算法先恢复场景的三维点云, 再通过 J-linkage 算法对点云进行聚类,得到空间三维平面结构,图 9(a)所示即为对视频进行三维重建后聚类得到 3 个空间平面。之后将三维空间平面分别重投影到视频的每一帧中去,进而得到视频帧中的平面结构,提取结果如图 9(b)所示。的确,这种方法在正确恢复出场景三维结构的情况下得到的平面提取结果应该是最好的,但是这种方法的唯一缺陷就是三维重建过程的代价过高。而本文算法虽然利用了部分三维信息,但是并没有对整个场景的三维结构进行恢复,提高了算法的整体效率。而且从图 6 与图 9 中的平面提取的结果比较来看可以说效果相差并不大,在保证了提取结果的准确性的前提下,显然本文算法略胜一筹。



图 9 文献[2]中的平面提取算法提取结果

此外,真实的视频中常常存在一些噪声或干扰,例如移动 的物体或者植被等等,这些都会对平面提取造成影响。而本 文算法先对特征点进行跟踪,然后对相邻帧取跨越这两帧的 轨迹进行计算,本身就可以先滤除一部分噪声;在计算过程中 又综合考虑了整个视频序列对平面的一致性约束,尽可能地 降低了噪声和干扰的影响。这里也选择了一些包含动态前景 的视频进行实验,结果如图 10 所示,视频中运动的人物都与 墙壁平面分割开来划分到异常点里。



图 10 对包含动态前景的视频序列进行平面提取的结果

以上结果可以证明,本文算法在兼顾了有效性的同时,也 具有一定的稳定性,对于视频中的噪声和干扰有很高的容忍 度。但是,本文算法由于依赖于特征点的跟踪和提取,因此在 处理一些缺乏纹理信息的区域,例如地面时,往往难以跟踪到 足够多的特征点轨迹,造成平面提取失败,如图 11 所示,地面 作为主要的平面结构之一完全没有被提取出来。



图 11 对于缺乏纹理信息的区域,难以跟踪到足够数量的特征点, 造成提取失败

结束语 本文提出了一种基于视频序列进行平面提取的 新方法。首先对视频序列进行特征点跟踪以提取轨迹;然后 在相邻帧之间构建射影变换模型,利用平面诱导的单应与空 间平面——对应这一原理,引人重投影误差作为衡量平面模 型与轨迹相合程度的标准,对轨迹进行聚类,提取出视频中的 平面结构以及所诱导的单应,最后根据平面诱导的单应使用 Graph Cuts 算法得到平面提取结果。多种环境下的视频序列 实验结果表明,本文算法可以得到较好的效果,且具有稳定性。

但是本文算法也存在着一些缺陷,即对于缺乏纹理信息 的区域,往往难以提取到足够多的特征点,可能会造成部分平 面提取失败或者出现平面分割错误等情况。以后的工作方向 是,考虑加入线匹配等其他特征信息,以适应更多更复杂的情 况。

参考文献

- [1] Ogundana O O, Coggrave C R, Burguete R L, et al. Automated detection of planes in 3-D point clouds using fast Hough transforms[J]. Optical Engineering, 2011, 50(5)
- [2] Zhou Z, Jin H, Ma Y. Plane-Based Content-Preserving Warps for Video Stabilization [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013
- [3] Toldo R, Fusiello A. Robust multiple structures estimation with j-linkage [M]. Computer Vision-ECCV 2008. Springer Berlin Heidelberg, 2008; 537-547
- [4] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automa-

(上接第86页)

实验结果表明,本文提出的基于海陆边界跟踪的海陆分 割方法与改进的二维 otsu 方法相比,在处理效率和准确率上 都有一定的提高。与基于多特征融合的方法相比,在准确率 相当的情况下,时间效率大幅度提高。因此,本文提出的海陆 分割方法能够快速、准确地分割海水和陆地区域。

结束语 本文提出的基于海陆边界跟踪的快速海陆分割 方法,充分利用了海水区域边缘纹理特性,构造了 EBT 纹理 特征,能够准确区分海水与陆地区域的纹理差异。同时,分割 过程只需要对海陆边界的图像区域进行处理,避免了传统分 割方法需要逐像素地对全图像进行处理的劣势。实验证明本 文方法有较好的分割效果,有效减少了时间花费。

虽然本文方法取得了较好的效果,但是对于过于复杂的 状况,如陆地和海面的灰度、纹理过于接近,就不能很好地实 现分割;此外,本文方法是对区域图像块进行处理,因此对海 陆边界的界定无法精确到像素,存在陆地和海洋边界划分粗 糙的现象。我们将在今后的研究中努力解决这些问题。

参考文献

- [1] 高晓亮,王志良,刘冀伟,等.基于灰度统计特征的可变区域图像 分割算法[J].光学学报,2011,31(1):1-5
- [2] 王茜蒨,彭中,刘莉.一种基于自适应阈值的图像分割算法[J]. 北京理工大学学报,2003,23(4):521-524

ted cartography[J]. Communications of the ACM, 1981, 24(6): 381-395

- [5] Choi O, Kim H, Kweon I S. Simultaneous plane extraction and 2D homography estimation using local feature transformations
 [M]. Computer Vision-ACCV 2007. Springer Berlin Heidelberg, 2007: 269-278
- [6] Bouchafa S, Patri A, Zavidovique B. Efficient plane detection from a single moving camera[C]//2009 16th IEEE International Conference on Image Processing (ICIP). IEEE, 2009; 3493-3496
- [7] Fouhey D F, Scharstein D, Briggs A J. Multiple plane detection in image pairs using j-linkage[C]//2010 20th International Conference on Pattern Recognition (ICPR), IEEE, 2010; 336-339
- [8] Kim H, Lee S. Multiple planar region extraction based on the coplanar line pairs[C] // 2011 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2011: 2059-2064
- [9] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. Cambridge, 2000
- [10] Ma Y, Košecká J, Soatto S, et al. An invitation to 3-d vision: from images to geometric models[M]. springer, 2004
- [11] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60 (2):91-110
- [12] Shi J, Tomasi C. Good features to track[C]//1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94. IEEE, 1994: 593-600
- [13] Narayanan P J, Vineet V, Stich T. Fast Graph Cuts for Computer[M]//GPU Computing Gems Emerald Edition, 2011:439
- [3] Zhang Xian-feng, Wang Zhi-yong. Coastline extraction from remote sensing image based on improved minimum filter [C] // 2nd IITA International Conference on Geoscience and Remote Sensing. 2010,2:44-47
- [4] Gonzalez R C, Woods R E. 数字图像处理[M]. 阮秋奇, 阮字智,
 译. 北京:电子工业出版社, 2003
- [5] Roger T, Stamon G, Jean L. Using Color, Texture, and Hierarchial Segmentation for High-Resolution Remote Sensing[J]. IS-PRS Journal of Photogrammetry and Remote Sensing, 2008, 63 (2):156-168
- [6] 陈琪,熊博莅,陆军,等,改进的二维 Otsu 图像分割方法及其快 速实现[J]. 电子与信息学报,2010,32(5):1100-1104
- [7] Gong Jian, Li Li-yuan, Chen Wei-nan, Fast recursive algorithm for two-dimensional thresholding [J]. Pattern Recognition, 1998,31(3):295-300
- [8] 艾国红. 基于多特征动态融合的图像分割研究[D]. 合肥:中国 科技大学,2011
- [9] Boser B E, Guyon I M, Vapnik V N. A training algorithm for optimal margin classifiers [C] // Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory. 1992;144-152
- [10] Shi Jian-bo, Malik J. Normalized Cuts and Image Segmentation
 [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905