

摘要：在虚拟现实（VR）技术中，为获得更宽的视场需要将多幅图片通过拼接构建出全景。目前大多数基于特征的图像拼接算法采用单应性矩阵变换或保持内容的变形来实现图像对准。基于单应性矩阵变换方法很难处理视差，而保持内容变形方法不能保持图像的结构特性。本文提出了一种精确的单应性矩阵的求解方法，由匹配点和特征点求得全局单应性矩阵，并通过非线性优化方法对该结果进行优化。进一步针对局部变形带来的误差，结合几何和光度约束来设计代价函数，通过最小化该函数获得更好的图像对准结果，从而产生精确的拼接图像。通过各种开放数据集的实验结果表明，本文提出的图像配准方法效果最好。

关键词：虚拟现实；图像拼接；图像匹配；非线性优化

中图分类号：TP391.41

文献标识码：A

文章编号：1006-883X(2020)05-0013-05

收稿日期：2020-03-15

基于非线性优化的 VR 场景拼接方法研究

张雁鹏¹ 高建勇¹ 周志杰²

1. 苏州热工研究院有限公司设备管理部深圳分部, 广东深圳 518028;

2. 华能山东石岛湾核电有限公司生产管理部, 山东荣成 264312

一、引言

图像拼接是视觉领域中一个重要研究方向^[1-2]。David L. Milgram^[3]第一个提出了这个领域的工作。文献[4]首先提出了一个全自动全景创建框架，采用对象识别技术，根据不变的局部特征计算两个图像之间的重叠区域。文献[5]根据不变的局部特征来实现图像之间的匹配，使用直接线性变换（DLT）来计算基于 SIFT 特征匹配的全局单应性矩阵，但这些基于单应性的方法无法处理大视差并且经常产生鬼影。为了减少鬼影，需要对局部翘曲、变形进行修复。文献[6]提出了类似的工作，首先根据全局单应性对齐图像，然后使用局部运动引导的局部变形进行重新定义。文献[7]提出了一种基于局部变形的运动模型，它结合了点和线，以保存场景的几何和结构信息。这些方法的问题在于它强烈依赖于特征匹配的准确性，因此不能处理特征对应中存在的噪声。光度约束是一种广泛使用的技术，可以克服这种情况。文献[8]提出了一种结合单应性和基于网格的局部变形的类似方法。作者使

用线和点特征来计算全局单应性。进一步使用局部翘曲方法，包括光度和几何约束，并提出了一种基于点、线、强度值、网格和边缘的复杂冗余成本函数。

在虚拟现实领域中，为了获得更大的视野范围或者生成 360° 全景图像，经常要用到图像拼接技术，通过拼接两幅或者多幅图像，并对拼接结果进行深度信息抽取，恢复场景的三维信息，进而建立三维模型。图像拼接主要有三个步骤：空间校准、图像对齐和拼接。空间校准的目的是减少光学缺陷和校正增益；图像对齐为了解校准图像对之间的转换，首先找到待配准图像与参考图像的模板或特征点的对应位置，然后根据对应关系建立参考图像与待配准图像之间的转换数学模型，将待配准图像转换到参考图像的坐标系中，确定两图像之间的重叠区域。在最后一步中，确定了两幅图像之间的转换关系模型，即重叠区域后，需要根据重叠区域的信息将待拼接图像拼接成全景图。

图像拼接技术主要包括两个关键环节，即图像配准和图像融合。对于图像融合部分，由于其耗时不太大，

且现有的主流方法效果差别也不多，所以总体来说算法上比较成熟。而图像配准部分是整个图像拼接技术的核心部分，它直接关系到图像拼接算法的成功率和运行速度，因此配准算法的研究是多年来研究的重点。对于图像配准，一般都需要首先建立起两幅待拼接图像间的变换模型，然后采用非线性最优化方法计算出模型的变换参数，从而确定图像的配准位置。当模型方程指数项增高，或者当变量数量增加（即维度增加，尤其是图像处理中，图像作为一个高维输入）时，这时我们会发现出现大量的局部最优解，如何求解就会变得非常困难。而非线性优化利用迭代的方式特别适合来求解这些高维的非线性方程。目前基于时域的图像配准方法存在两种方式：基于区域的图像配准方法和基于特征的图像配准方法^[9]。基于区域方法直接利用图像的灰度信息计算图像的相似度，这就需要根据待配准图像的特点选择相应的度量函数，通过一种或几种最优化算法寻找一个最佳空间变换。这类方法一般不需要对图像进行复杂的预处理，实现简单，但是该方法执行效率低，收敛范围有限，拼接效果不好^[10]。基于特征的方法使用特征点匹配来计算图像对之间的单应性矩阵，能够保持全局图像结构，避免局部扭曲。但是大多数基于单应性矩阵的图像拼接方法是基于线性变换，忽略了诸如镜头失真之类因素导致不正确的图像拼接，并且基于单应性变换的方法很难处理非平面图像拼接。本文中，在使用基于特征的图像配准基础上，提出了一种新的图像配准方法来克服上述问题：

- 1、提出了一种非线性优化方法来精确估计单应性矩阵；
- 2、提出一种多约束的修复方法，解决图像中局部形变问题，提高图像对准的鲁棒性。

二、配准方法

本文使用特征点来估计初始全局单应性矩阵。借助 vlfeat 对 SIFT 特征的提取获得一组匹配的特征点 ψ_1 ，并参考文献 [11] 得到另一组匹配的边点关系 ψ_2 。使用 $\psi_{noisy} = \psi_1 \cup \psi_2$ 的并集来进行全局单应性矩阵估计。

目前常使用直接线性变换（DLT）和 RANSAC 算

法来估计单应性矩阵。通过这种方法估计出来的单应性矩阵准确度不高，只能粗略地缝合图像。针对该问题，本文引入非线性最小二乘法方法对单应性矩阵估计结果进行优化。

设 I_s 和 I_d 为待拼接的一对图像，并且存在一组类似的图像对。使用 DLT 和 RANSAC 估计成对图像的单应性矩阵，生成对应的单应矩阵 H_{sd} 和一组特征点之间的对应关系 φ 。本文提出的非线性优化方法是指进一步利用单应性矩阵和特征点对应关系来估计全局单应性矩阵，这里设代价函数为：

$$C_w = \sum_{I_s, I_d \in \theta} \sum_{i \in \varphi} \left| (H_d^{-1} H_s) x_i - x'_i \right|^2 + \lambda * F_R(H_{sd}, (H_d^{-1} H_s)) \quad (1)$$

其中， x_i, x'_i —图像 I_s 和 I_d 中的第 i 对匹配特征点；

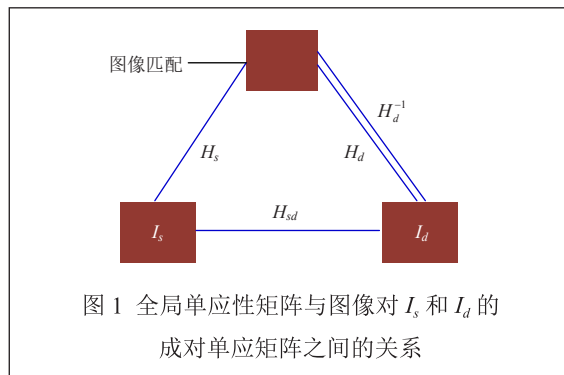
H_s, H_d —图像 I_s 和 I_d 的全局单应性矩阵；

$F_R(A, B)$ —矩阵 A 和 B 之间的 Frobenius 范数。

式（1）的第一部分表示特征点通过矩阵 H_s 从图像 I_s 变换到拼接图像，随后通过矩阵 H_d 从拼接图像变换到图像 I_d 的误差。图 1 以图形方式解释了这个变换过程；式（1）的第二部分约束全局单应性矩阵。由于 DLT 估计单应性矩阵不准确，所以在实现过程中把式（1）的权重 λ 设置为 0.2。

把这种全局单应性矩阵计算和对齐称为初始对齐。由于特征点对应集 φ 和成对单应矩阵 H_{sd} 存在误差，图像可能存在未对准，下面提出局部变形修复，进一步纠正了重叠区域和边缘边界上的误差。

局部变形修复主要处理图像之间的重叠区域。在局部变形修复中引入了三种不同的约束，式（2）表示局部变形修复的成本函数。



$$C_L = C_D + \delta_1 C_P + \delta_2 C_G \quad (2)$$

其中, C_D 、 C_P 、 C_G 分别代表数据项、光度项和几何平滑项的成本函数。数据项和光度项修复局部形变的错位, 几何平滑项确保物体几何形状的平滑。 δ_1 和 δ_2 为权重, 在实现中将它们分别设置为 0.7 和 0.9。下面将这个二次成本函数最小化, 使得在像素移动的平均变化低于单个像素时收敛。

通过以上步骤, 已经获得所有对齐图像对之间的特征对应关系。由于全局对齐中存在错误, 这些特征点对应的位置不完全相同。设输入图像对为 I_s^{al} 和 I_d^{al} , I_s^o 和 I_d^o 为相应的输出存在变形的图像。将 x_i^m 作为输入图像对上的第 i 个点对应的中心, 并将变形与相应的中点匹配。

式 (2) 中对应的数据项函数为:

$$C_D = \sum_{i \in \varphi} \left| x_i^{os} - x_i^m \right|^2 + \left| x_i^{od} - x_i^m \right|^2 \quad (3)$$

其中, x_i^{os} 和 x_i^{od} 是图像 I_s^o 、 I_d^o 中的变形特征点。在输入图像上的第 i 个特征周围的 12×12 窗口内取 9 个角点, 并且用九角围城的封闭区域进行双三次插值来表示第 i 个特征点 P_{ij} ($j=1,2,\dots,9$)。使用式 (4) 计算变形的特征点 x_i^{os} 和 x_i^{od} :

$$x_i^o = \sum_{j=1}^9 \omega_{i,j}^T P_{ij}^o \quad (4)$$

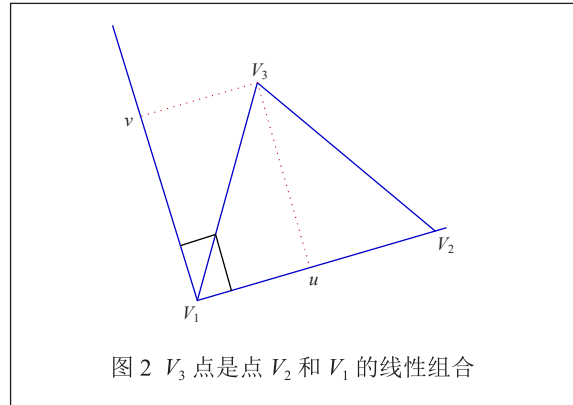
其中, 向量 $\omega_{i,j}^T$ 包含双三次插值系数; P_{ij}^o 是变形图像上的九个角点。

引入图像对之间的光度对局部变形修复进行约束, 光度项定义为:

$$C_D = \sum_k \left| I_s(x_k^o) - I_d(x_k) \right|^2 \quad (5)$$

其中, x_k 是第 k 个点, 其中 $k \in \beta$, 点集 β 来自于重叠部分的边缘和重叠区域中的特征点。 x_k^o 是对应的使用双三次插值的变形点, 是 $I_s(\cdot)$ 和 $I_d(\cdot)$ 表示源图像和目标图像的光强度。

添加了几何约束, 引入了独特的网格模型。通过在重叠区域的边缘边界上均匀采样点以及对应匹配点来创建点集 φ 。对于 φ 中的点, 通过 delaunay 三角剖分算法进一步创建一个三角形网格, 表征重叠区域的几何结构。这里定义了几何平滑度, 其中网格的三角形表示可以为 $\Delta V_1 V_2 V_3$, V_3 与 V_1 和 V_2 是线性关系。



V_3 可以表示为:

$$V_3 = V_1 + u(V_2 - V_1) + v \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} (V_2 - V_1) \quad (6)$$

其中, u 和 v 是两个标量, 可以从 V_1 、 V_2 和 V_3 计算出, 如图 2 所示。

三角形 $\Delta V_1 V_2 V_3$ 应使用 u 和 v 确保相似的关系, 以实现用平滑网格修复变形。因此, 几何平滑度项定义为:

$$C_G = \sum_{i=1}^{\Delta n} \left| V_3^0 - (V_1^0 + u(V_2^0 - V_1^0) + v \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} (V_2^0 - V_1^0)) \right|^2 \quad (7)$$

其中, Δn 一网格中存在的三角形的总数。

四、实验结果

在配置为 Intel i7-8700 (6 核 @ 3.7-4.7GHz) 的电脑上用 C++ 来编程实现本文提出的方法。1024×720 的一对图像计算时间约为 3~4 秒, 其中大部分时间花费在特征检测和匹配上。

1、定量评估

在文献 [10] 中使用的公开数据集上测试本文中提出的方法。将本文的方法与文献 [9-11] 中三种基于特征的方法进行了比较。最终表明了本文提出的方法比现有技术效率和效果上表现更好。

因为本文中提出的方法是基于特征点匹配的, 所以从数据集中剔除了低纹理数据, 保证每个图像对都至少有 30% 重叠。对图像对重叠区域中的所有像素构建 3×3 邻域匹配窗口, 建立目标函数来对匹配窗口进

行度量相关性 (NCC)，并求解 NCC 的均方根误差 (RMSE)。

$$RMSE(I_s, I_d) = \sqrt{\frac{1}{\pi} \sum (1 - NCC(x_s, x_d))^2} \quad (8)$$

其中， π —图像 I_s 和 I_d 之间的重叠区域中的像素总数；

x_s 、 x_d —图像 I_s 和 I_d 中的像素。

表 1 显示了 RMSE 与以前方法的比较。可以看出，本文的方法比特征点估计出的单应性矩阵更精确，对于室外数据集，因为特征比较明显，基于特征点的单应性矩阵估计也非常准确。这表明，全局单应性矩阵估计的准确性与匹配特征点的数量呈正相关。本文使用两个边缘和 SIFT 特征点进行全局单应性估计，从而产生更好的单应性矩阵估计。

表 1 误差比较

数据集	APAP	DFW	MCC	本文方法
学校	12.20	9.89	10.85	9.736
室外	11.90	9.52	6.75	7.433
铁路	14.80	10.58	9.81	8.317
路	2.28	4.59	1.67	1.917
院子	38.30	36.23	29.17	30.258
马	19.80	19.57	14.57	13.113
公园	11.07	8.18	5.85	7.528

注：APAP^[12]：作为可能的投影方法；DFW^[9]：双特征方法；MCC^[10]：多组合约束方法

2、定性评估

图 3 显示了文献 [7] 提出的 DFW 与本文提出的方法之间的定性比较。通过对铁路图片进行了比较，对比铁路轨道的合并情况。图 3 第一行：使用 DFW 的拼接图像，其中未对准出现在铁轨中；图 3 第二行：使用本文提出的方法缝合图像，其中导轨钉完全对齐。通过在边缘和重叠边界上添加更多匹配的点要素，本文提出的方法在合并轨道方面优于 DFW。

五、结论

本文提出了一种新颖的图像拼接方法，其中图像最初使用全局单应性估计进行对齐，并使用多约束局部变形方法进一步纠正对准。在局部翘曲中使用光度和几何约束来实现在重叠图像之间的平滑结构保持拼接。对不同开放数据集的本文提出的方法的评估显示出比最先进的方法更好的准确性。

参考文献

- [1] 钱鸣静. 基于特征点匹配的全景图像拼接技术研究 [D]. 2018.
- [2] 汤强. SIFT 算法在 VR 场景拼接中的应用 %SIFT algorithm in VR scene mosaicking[J]. 长春工业大学学报 (自然科学版), 2018, 039(002): 156-162.
- [3] 盛明伟, 唐松奇, 万磊, 等. 二维图像拼接技术研究综述 [J]. 导航与控制, 2019, 18(01): 32-39+101.
- [4] 罗群明, 施霖. 图像拼接方法综述 [J]. 传感器与微系统, 2017, 36(12):4-6.
- [5] Milgram, D. L. Computer Methods for Creating Photomosaics[J]. IEEE Transactions on Computers, 1975, C-24(11):1113-1119.
- [6] Brown M, Lowe D G. Recognising Panoramas.[C]// IEEE International Conference on Computer Vision, 2003.
- [7] Brown M, Lowe D G. Automatic Panoramic Image Stitching using Invariant Features[J]. International Journal

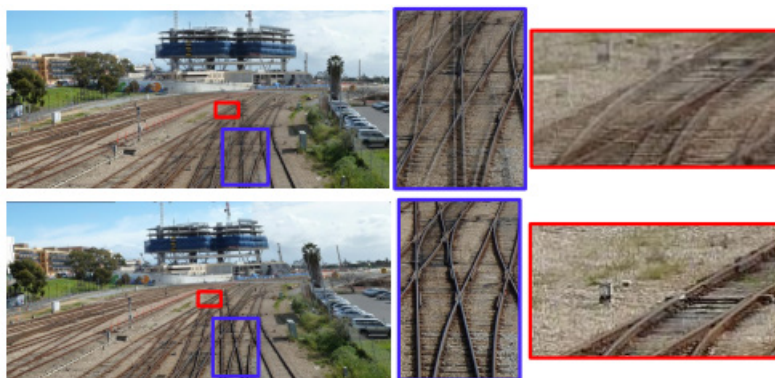


图 3 DFW 与本文方法之间的定性比较

of Computer Vision, 2007, 74(1): 59-73.

[8] Shum H Y, Szeliski R. Construction of Panoramic Image Mosaics with Global and Local Alignment[J]. 2001: 227-268.

[9] Li S, Yuan L, Sun J, et al. Dual-Feature Warping-Based Motion Model Estimation[C]// International Conference on Computer Vision 2015. IEEE, 2015.

[10] K. Chen, J. Tu, B. Xiang, L. Li, J. Yao. "Multiple combined constraints for image stitching" in IEEE International Conference on Image Processing (ICIP), 2018, pp. 1253-1257.

[11] Maity S, Saha A, Bhowmick B. Edge SLAM: Edge Points Based Monocular Visual SLAM[C]// 2017 IEEE International Conference on Computer Vision Workshop (ICCVW). IEEE, 2017.

[12] Zaragoza J, Chin T J, Tran Q H, et al. As-Projective-As-Possible Image Stitching with Moving DLT[J]. 2013: 2339-2346.

Research on VR Scene Stitching Based on Nonlinear Optimization

ZHANG Yan-peng¹, Gao Jian-yong¹, ZHOU Zhi-jie²

(1. Suzhou Thermal Power Research Institute Co., Ltd., Equipment Management Department, Shenzhen, Guangdong 518028, China; 2. Production Management Department, Huaneng Shandong Shidaowan Nuclear Power Co., Ltd., Rongcheng, Shandong 264312, China)

Abstract: In virtual reality (VR), in order to obtain a wider field of view, multiple images need to be stitched to construct a panorama. Despite significant advances in image stitching in recent years, this problem still lacks a reliable solution. Most feature-based image stitching algorithms perform image alignment based on homography matrix transformations or by maintaining distortion of the content. It is difficult to deal with parallax based on the homography matrix transformation method, and the content modification algorithm cannot maintain the structural characteristics of the image. In this paper,

an accurate method for solving the homography matrix is proposed. The global homography matrix is obtained from the matching points and the feature points, and the result is optimized by the nonlinear optimization method. Further errors caused by local deformation, combined with geometric and photometric constraints to design our cost function, minimize this function, and obtain better image alignment, resulting in accurate image stitching. Experimental results through various open datasets show that our proposed method is superior to the most advanced image stitching algorithm.

Key words: VR; Image Stitching; image matching; nonlinear optimization

作者简介

张雁鹏：苏州热工研究院有限公司设备管理部，高级工程师，工程硕士，研究方向为设备可靠性工程。

通信地址：深圳市福田区深南大道 2002 号中广核大厦南楼 12-14 层

邮编：518028

邮箱：18611291962@163.com

高建勇：苏州热工研究院有限公司设备管理部，工程师，学士，研究方向为核电可靠性分析与虚拟仿真技术应用。

周志杰：华能山东石岛湾核电有限公司生产管理部，高级工程师，学士，研究方向为核电站设备管理。