arXiv:2601.01257v1 [eess.IV] 3 Jan 2026

无缝自然：兼顾自然外观保持的图像拼接

Gaetane Lorna N. TCHANA[1], Damaris Belle M. Fotso[1], Antonio Hendricks[2], Christophe Bobda[2*]

[1]雅温得第一大学，喀麦隆雅温得，812。[2]佛罗里达大学，电子与计算机工程系，美国佛罗里达州盖恩斯维尔，32611。

*通讯作者。电子邮箱：cbobda@ece.ufl.edu；贡献作者：gaetane.tchana@facsciences-uy1.cm；makwate.fotso@facsciences-uy1.cm；a.hendricks1@ufl.edu；

摘要

本文介绍了**SENA（无缝自然）**，这是一种几何驱动的图像拼接方法，旨在应对具有视差和深度变化的复杂真实场景，并优先保证**结构保真度**。传统图像拼接依赖单应性对齐，但这种刚性的平面假设在场景深度较大的双摄像头设置中往往不适用，进而导致明显形变和球形鼓胀等失真。

SENA 通过三项关键贡献解决了这些根本性局限。首先，我们提出一种分层仿射变形策略，将全局仿射初始化、局部仿射细化与平滑的自由形变相结合。这种设计能够保留局部形状、平行性和纵横比，从而避免基于单应性变换的模型通常会引入的结构性失真。其次，我们引入一种由几何驱动的适宜区域检测机制，无需依赖语义分割，直接根据经 RANSAC 筛选的特征对应关系的视差一致性，识别视差最小化区域。第三，在此适当区域的基础上，我们执行基于锚点的接缝线切割与分割，从构造上确保图像对之间一一对应的几何关系，从而有效消除最终全景图中的重影、重复和拖影伪影。

在具有挑战性的数据集上进行的大量实验表明，SENA 的对齐精度可与领先的基于单应性变换的方法相媲美，同时在形状保持、纹理完整性和整体视觉真实感等关键视觉指标上显著优于这些方法。

关键词：图像拼接，自然外观保持，仿射变形，无视差区域

1 引言

数字成像技术的兴起从根本上改变了人们与数字内容及其周围环境互动的方式 [1], ，直接推动了虚拟现实和增强现实等先进沉浸式体验的发展。这些平台的一项关键要求是能够从多个视角捕捉复杂场景，从而为用户呈现连续、无缝的全景视觉效果。这通常通过合成并合并由两台或更多摄像机同时采集的图像数据来实现，这一过程称为图像拼接。大多数图像拼接方法利用显著视觉特征在图像之间建立对应点 [2–6]。这些对应点用于确定几何变换，以便将图像变形到公共坐标系中。随后，通过直接合并重叠区域，或借助接缝选择、融合等额外机制减少视觉伪影，生成最终的全景图。当前最先进方法的改进通常侧重于优化或加速上述一个或多个独立阶段。

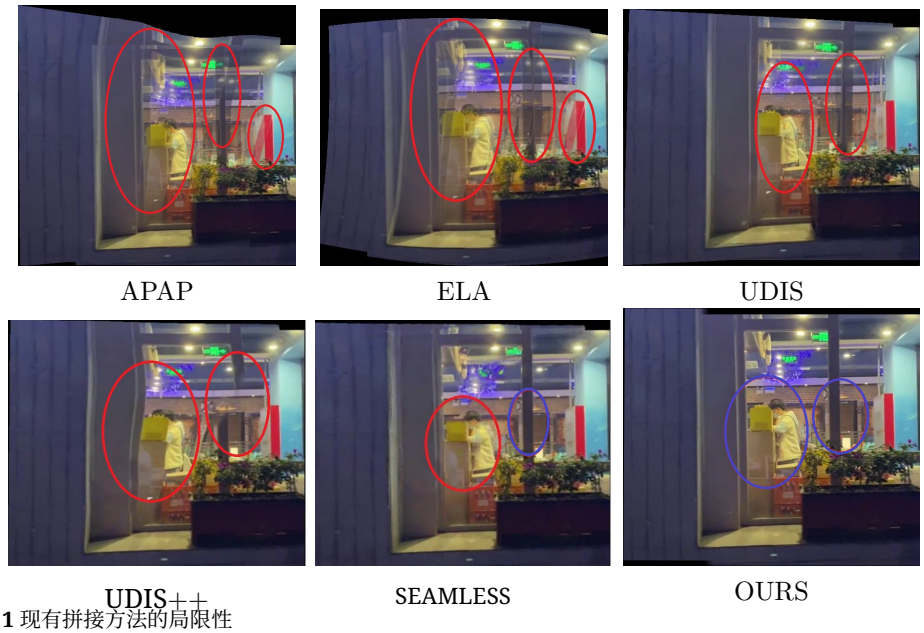


图 1 现有拼接方法的局限性

几何变形在拼接质量中发挥着核心作用，而单应性变换仍是主导的变换模型。由于其具有八个自由度，

单应性变换能够准确建模完整的透视效应，并为平面场景或纯相机旋转提供精确的对齐。然而，现实世界中的场景很少是平面的，且通常包含显著的深度变化。在此类情况下，单应性变换试图使用平面模型协调非平面运动，从而导致过度灵活的几何畸变，例如拉伸、直线弯曲和形状变形。

大量图像拼接研究致力于通过日益灵活的形变模型来改进几何对齐。混合变形模型 [7]针对不同场景区域或深度层估计多个单应性变换，以更好地保持结构一致性。其他方法则用基于局部样条的配准（例如薄板样条（TPS）形变）取代全局单应性变换，从而提高对齐的灵活性 [8, 9]。尽管这些方法通常能改善重叠区域的对齐，但额外的自由度往往会导致过度的几何变形，进而造成纹理拉伸、线条弯曲以及结构真实感缺失。即使结合单应性变换和仿射变换 [10]，也只能减少而无法彻底消除这些伪影，尤其是在视差显著的场景中。图1 展示了具有代表性的最先进方法（如 APAP [11], ELA [12],UDIS [13],UDIS++ [8]和 SEAMLESS [9]）的这些局限，其中可见波纹、重影和信息丢失等视觉伪影（以红色标出）。这些伪影主要源于基于单应性的模型固有的倾向：通过非物理的几何畸变来补偿由深度引起的视差。

与此同时，许多拼接方法并不通过校正几何关系来处理错位伪影，而是仔细选择图像的融合位置。基于接缝的方法会在重叠区域中确定一条拼接路径，并根据预定义的代价函数将视觉差异降至最低。代表性工作 [14–17] 利用强度差异、梯度、语义标签或特征一致性等线索来引导接缝位置。尽管这些方法能有效地掩盖部分错位，但它们通常依赖复杂的预处理流程——包括语义分割、特征分类或深度推断——这会增加计算成本，并使方法对误差更加敏感。此外，仅靠接缝优化并不能保证接缝两侧的几何一致性，因此在许多情况下，重复或重影问题仍未解决。

总体而言，拼接方法仍面临三大挑战。首先，全局或过于灵活的形变（例如仅使用单应性变换或基于样条的模型）往往会造成几何畸变，并损害结构保真度。其次，基于代价函数最小化的接缝选择策略通常依赖语义分割或特征分类等复杂预处理步骤，从而增加了对误差的敏感性。第三，即使局部对齐准确，不同区域间的融合往往缺乏一致性，导致重复、拉伸或重影，降低拼接全景图的视觉真实感。

本研究通过三项贡献解决上述局限。（1）我们提出了一种分层形变策略，将全局仿射初始化（使用经 RANSAC 筛选的对应点）、重叠区域内的局部仿射细化，以及由 SeamGuard 自适应平滑调控的平滑自由形变（FFD）场结合起来。这种多尺度设计在保持结构保真度的同时，避免了全局模型造成的过度失真和基于样条的形变所带来的不稳定性。

（2）在相同的接缝选择范式下，我们提出了一种有别于以往基于语义分割的方法的适宜区域检测策略。我们不采用这类复杂预处理，而是分析特征对应点的视差一致性，以轻量、无模型的准则稳健地识别视差最小化区域。（3）我们依据检测到的适当区域进行基于锚点的分割，确保图像对之间的结构一致性，并实现无缝拼接。不同于以往仅确定接缝线、再依靠融合隐藏伪影的方法，我们的方法以精化关键点为锚点，将两幅图像划分为相互对应的垂直切片。这确保了各图像段之间一一对应的几何对应关系，从而减少最终全景图中的重复和重影。

2 相关工作

图像拼接研究探索了从基于特征的方法到深度学习方法等多种策略，但即使是最先进水平，仍存在三大主要局限。

•**全局变形导致的几何畸变：** 主流的单应性模型 Zheng 等人[18], Yadav 等人 [19], Jia 等人 [20] 经常被采用，因为其 8 个自由度使其能够对完整的透视效应进行建模。然而，当应用于存在视差和深度变化的真实场景时，单应性变换必须协调相互冲突的运动，这往往会导致非均匀畸变，例如纺锤形变形、不自然的拉伸或球形鼓胀。先进的混合变形，例如“尽可能投影”（APAP） [11],和弹性变形虽然提高了灵活性，但也有过拟合或灵活性过高的风险，可能产生局部拉伸伪影。

•**依赖复杂预处理**：许多接缝线优化方法（[14], [15], [16]）会针对重叠区域定义代价函数，并寻找使该代价最小的接缝线，理想情况下，接缝线应穿过视觉上一致的区域。然而，这些方法通常依赖复杂的预处理步骤（如语义分割或深度估计）来引导代价图，从而增加计算成本，并提高对错误和不准确性的敏感度。

- **结构假设与计算开销**:
  - **平面或多单应性** 模型 Jia 等人 [20], [7]尝试通过为每个平面拟合局部投影模型，或融合双单应性，来减少单一单应性变换造成的失真。然而，这些方法依赖于对场景结构的强假设（例如，场景包含两个平面或投影一致区域），在深度几何结构复杂或不规则时也容易变得脆弱。
  - **保结构变形**能够有效减少失真并保留
    显著结构，同时结合复杂的优化框架和
    约束（例如，共线性约束）。然而，这些方法通常
    计算开销较大，并且仍易受特征分布不佳的影响。
  - **基于学习的方法**利用深度学习完成变换器等任务-
    基于变形和光流，并结合图像修复的方法，Nie 等人 [8, 9, 13]。虽然他们

虽然泛化能力良好且性能出色，但这些方法需要大量训练需求、依赖大型数据集，并存在产生内容幻觉或传播错误的风险；与几何模型相比，其可解释性也较低。

总体而言，现有方法仍面临三项主要挑战：(1) 全局或过度灵活的形变会损害结构保真度并引入几何畸变；(2) 基于代价函数最小化的接缝选择策略通常依赖复杂的预处理步骤（如语义分割或特征分类），从而提高了对不准确因素的敏感性，降低了拼接全景图的视觉真实感；(3) 依赖较强的场景假设（平面性、双平面或一致的表面法线）。

3 无缝且结构一致的图像拼接

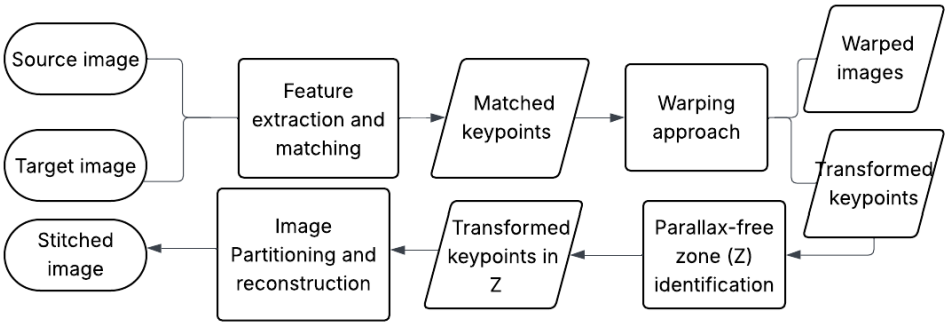


图 2 所提方法的流程图

为克服图像拼接中常见的过度畸变和错位问题，我们提出了一种三阶段框架，强调局部适应性、基于几何的视差处理以及结构一致的重建。该方法将源图像转换到目标图像的域中，同时严格保留几何结构。

3.1 三阶段框架

- 局部图像形变：我们有意不采用全局单应性变换和基于样条的形变。首先，使用全局仿射变换（通过 RANSAC 估算）对源图像进行粗略对齐。随后进入细化阶段，将重叠区域划分为局部网格，并根据局部特征对应关系拟合不同的仿射模型。此过程生成平滑自由形变（FFD）场，再通过接缝保护策略进行融合与自适应平滑；该策略利用渐变掩膜和匹配密度加权

以避免边界处出现不连续性。这样既能保持整体结构保真度，又能减少几何畸变。

- 充分识别视差最小化区域：不同于依赖语义分割的接缝选择方法，我们提出一种无需模型的策略。该方法分析匹配特征之间的视差变化和几何一致性，从而分离出一个能够从根本上减少视差伪影的稳定拼接区域。

- 图像分割与重建：在识别出的稳定区域内，由一系列有序的精化关键点定义最优拼接线。随后，以这些关键点为锚点，将两幅图像分割成相对应的垂直切片。这种基于锚点的分割从构造上确保了两幅图像之间的结构一致性：一幅图像中的每个垂直切片，在另一幅图像中都有直接对齐的对应切片。通过将接缝线与这些几何锚点关联起来，我们的方法消除了融合过程中常见的重复、重影和错位问题。

所提方法的整体流程如图2所示。

3.2 局部自适应图像变形

我们的方法仅利用稀疏特征匹配，即可生成内容感知且受接缝约束的变形场，从而将源图像（I_s）精确对齐到目标图像（I_t）。该流程由六个依次执行的步骤构成，策略性地结合了全局对齐与局部自适应性。通过置信度加权融合和特定的门控机制，该方法兼顾结构保真度与稳健性。

步骤 1：使用 RANSAC 进行全局仿射估计

首先，我们使用 XFeat算法[6], 检测、提取并匹配特征。选择该算法是因为它在计算效率和匹配精度之间取得了有效平衡。根据匹配得到的关键点集合 $\{(p_{s_i},\ p_{t_i})\}$，我们使用 RANSAC算法 $A_{\text{glob}} \in \mathbb{R}^{3\times3}$ 计算稳健的全局仿射变换。此初始步骤有两项关键作用：建立粗对齐并滤除离群点。后续所有局部细化步骤中，仅保留由 RANSAC 判定的内点匹配。

步骤 2：重叠区域与局部网格单元生成

此步骤用于确定对齐区域，并为局部变形做好准备。

使用全局仿射变换 A_{glob}，将源图像 (I_s) 的四个角投影到目标图像 (I_t) 的坐标空间中，生成一个四边形。随后，使用 Sutherland-Hodgman算法将该四边形裁剪至 I_t 的边界范围，以精确确定多边形重叠区域。根据裁剪后的多边形生成二值掩码 M^{overlap} 目标图像。

接下来，将一个均匀的 $G_x \times G_y$ 网格覆盖在该重叠区域的边界框上。对于生成的每个网格单元（c_j），计算以下参数：

- 一个二值掩码 M_{c_j}，由网格单元与 $M_{\text{重叠目标}}$的交集生成。
- 质心 $\mathbf{c}^j = (c_{x_j},\ c_{y_j})$，通过图像矩计算得出。
- 边界框 $((x_{(j)\,0},\ x_{(j)\,1},\ y_{(j)\,0},\ y_{(j)\,1}))$，用于后续的空间加权和诊断。

步骤 3：结合自适应变换选择与置信度评分进行局部仿射细化

对于每个网格单元 c_j，使用落在该单元及其紧邻单元内的特征匹配来拟合局部仿射变换 T_j。拟合采用带正则化 λ_1的岭回归，使局部解偏向初始全局变换 A_{glob}，以确保稳定性。

空间置信度指标

不同于依赖统计残差或协方差的方法，我们根据支持每个局部仿射变换的特征点密度及其空间分布，定义了一种新颖的置信度分数。特征点 i相对于网格单元中心 (c_j) 的高斯权重 $w_i^{(j)}$ 计算如下：

$$w_i^{(j)} = \exp\left(-\frac{\|\mathbf{p}_{t_i} - \mathbf{c}_j\|^2}{2\sigma_j^2}\right)$$

此处，高斯标准差由缩放因子 α 乘以单元格边界框的对角线长度 $\text{diag}(c_j)$ 定义，其中 $\sigma_j = \alpha \cdot \text{diag}(c_j)$。随后，根据加权密度计算总体置信度分数。

置信度分数的计算方式如下：

$$\text{conf}_j = \max\left(\kappa_{\min},\ \min\left(\kappa_{\max},\ \frac{\sum_i w_i^{(j)}}{\beta \cdot \max_i w_i^{(j)}}\right)\right) \tag{1}$$

其中 $\kappa_{\min}$、$\kappa_{\max}$、β 是截断和归一化常数。该分数反映了单元格中匹配项的“权重总量”——这是一种对稀疏性和分布不均具有鲁棒性的启发式指标。

基于综合诊断评分的自适应变换选择：

系统采用自适应策略来维持几何稳定性。如果初始局部仿射变换 T_j 产生不稳定的几何结果——通过均方根误差 (RMSE)、行列式、条件数或相对于全局仿射变换 A_{glob} 的位移等指标进行诊断——则使用更强的正则化参数 $\lambda_2 > \lambda_1$ 重新拟合变换。随后，根据**综合不稳定性评分**选择效果更优的变换版本。

$$\text{score}(T) = \text{RMSE} + \omega_{\text{cond}} \cdot \text{cond} + \omega_{\text{det}} \cdot \max(0, \tau_{\text{det}} - |\det|) + \omega_\delta \cdot \delta_{\text{mean}} \tag{2}$$

其中，ω_{cond}，ω_{det}，ω_δ，τ_{det} 为权重和阈值参数，而 δ_{mean} 是局部仿射变换与全局仿射变换的输出之间的平均位移，该位移是在一个 $N_g \times N_g$ 评估网格上计算得出的，该网格在

单元格的边界框内均匀采样。这种嵌入式选择机制无需人工干预，即可确保几何稳定性。

步骤 4：通过置信度加权局部变换混合构建自由形变场

我们在输出画布上的 $N_y \times N_x$ 格点上构建一个变形场。对于每个格点 {style id='2'>u $= (u, v)$，我们计算其在目标空间中的对应坐标 $\mathbf{p}$ $t = u - o_x$ $v - o_y$ Π A^{-1} Π $=$ $_{\text{base}} =$ (,)，以及通过全局仿射变换得到的基础源图像坐标：$\mathbf{p}(_{\text{glob}} \cdot [\mathbf{p}_t, 1]^T)$，其中 $(\mathbf{x})$ $(x/z, y/z)$ 表示透视除法。对于每个局部变换 T_j，我们计算其映射后的源图像坐标 $\mathbf{p}^{(j)}{}_{\text{local}} = \Pi(T_j^{-1} \cdot [\mathbf{p}_t, 1]^T)$，以及位移 $\Delta\mathbf{p}^{(j)} = \mathbf{p}^{(j)}{}_{\text{local}} - \mathbf{p}_{\text{base}}$。

置信度加权空间混合

最终位移 $\Delta\mathbf{p}$ 是所有局部变换位移的归一化融合，并根据空间邻近度和变换置信度进行加权：

$$\Delta\mathbf{p} = \sum_{j=1}^{J} w_j(\mathbf{p}_t) \cdot \Delta\mathbf{p}^{(j)} \quad \text{where} \quad w_j(\mathbf{p}_t) = \frac{\text{conf}_j \cdot \exp\left(-\frac{\|\mathbf{p}_t - \mathbf{c}_j\|^2}{2\sigma_f^2}\right)}{\sum_{k=1}^{J} \text{conf}_k \cdot \exp\left(-\frac{\|\mathbf{p}_t - \mathbf{c}_k\|^2}{2\sigma_f^2}\right)} \tag{3}$$

其中 $\sigma_f = \alpha_f \cdot$表示平均（单元对角线长度）， α_f 为空间衰减因子。位移会被裁剪到 $[-d_{\max}, d_{\max}]$ ，并在通过双三次插值上采样至完整画布分辨率之前，使用高斯模糊 (σ_l) 进行轻微平滑。

步骤 5：通过双通道门控进行接缝防护

抑制边界附近或稀疏区域伪影的最后一步，是使用由两个主要信号得出的**乘性门控 (G)**，对全分辨率变形场 $\Delta\mathbf{p}$ 进行调制：

- **几何渐变 (R画布)**：将平滑阶跃函数应用于重叠区域的有符号距离场，其带宽与图像对角线 ($b = \rho \cdot \text{diag}_{\text{img}}$) 成正比。
- **匹配密度图 (D画布)**：对内点关键点位置生成高斯模糊热力图，并使用核标准差 (σ_d) 进行归一化。

双门控接缝抑制

最终的门控掩码将两种信号相乘后合并：

$$G(\mathbf{u}) = \underbrace{\left[S\left(R_{\text{canvas}}(\mathbf{u})\right)\right]^{\gamma_p}}_{\text{geometric falloff}} \cdot \underbrace{\left(\gamma_{\min} + (1 - \gamma_{\min}) \cdot S\left(D_{\text{canvas}}(\mathbf{u})\right)\right)}_{\text{density-aware modulation}} \tag{4}$$

其中 $S(t) = 6t^5 - 15t^4 + 10t^3$ 是平滑阶跃函数， γ_p 控制斜坡陡峭度， $\gamma_{\min}$ 是最小门控值。随后对经过门控的变形场进行平滑处理：

$$\Delta\mathbf{p}_{\text{guarded}}(\mathbf{u}) = \text{GaussianBlur}_{\sigma_g}\left(\Delta\mathbf{p}(\mathbf{u}) \cdot G(\mathbf{u})\right)$$

这并非后变形融合，而是一种用于在变形前抑制不可靠变形的机制。最终位移场 Δ_p 由门控 G调制：

$$\Delta\mathbf{p}_{\text{gated}} = \mathbf{G} \odot \Delta\mathbf{p}$$

这种乘性门控确保全分辨率位移仅应用于几何稳定且有特征支持的区域。据我们所知，在图像拼接领域，将几何门控与基于光度信息（匹配密度）的门控结合用于自由形变（FFD）场尚属首创。

步骤 6：最终变形与输出

最终的源图像到画布的映射通过结合全局仿射变换的逆变换与经过门控的变形场计算得出。随后，使用双线性插值将源图像 I_s 变形到最终画布空间。同时，将目标图像 I_t 按照预先计算的偏移量（o_x， o_y）粘贴到输出画布上。

为便于评估，我们计算原始内点匹配的最终变换坐标。具体而言，将目标点简单平移至画布坐标，并对源点应用完整变形（全局仿射变换 + 自由形变 + 乘性门控），同时对位移场进行双线性插值。

这些创新共同使我们的方法即使在特征匹配稀疏、不均匀或存在噪声的情况下，也能实现无缝且无伪影的对齐。

3.3 最优拼接线

3.3.1 充分无视差区域的确定

要确定合适的拼接区域，需要找到一个由占主导地位的几何变换支配、且包含高密度可靠特征对应关系的区域。首先，系统地排除信息量较低的区域；这些区域可通过较低的局部图像方差进行量化，因为它们会产生不稳定的匹配。然而，深度变化显著的场景往往会出现运动视差，导致视差向量呈多模态分布，从而增加寻找一致接缝的难度。

为稳健地处理这种视差，我们的方法着重于分离范围最广、最稳定的表面。具体而言，通过识别主导运动组来实现这一点，该运动组对应于稳定拼接所需的真正内点。算法会定位这样的区域：其中关键点视差在统计上保持一致（局部方差较低），且其局部均值趋近于主要运动模式的全局均值 μ 。这种方法通过分离范围最广、最稳定的表面，使算法能够稳健地处理视差。

给定一组匹配的关键点对 ((${x_S}$, ${y_S}$),(${x_T}$, ${y_T}$))，其中 ${x_S}$ 和 ${y_S}$ 分别表示源图像中的坐标， ${x_T}$ 和 ${y_T}$ 分别表示目标图像中的坐标。为考虑相机几何，首先依据两个主要空间准则对关键点进行排序。

- 首先， ${x_S} > {x_T}$，反映了源图像相机相对靠前的位置。
- （${y_S} > {y_T}$） $\vee$ （${y_T} > {y_S}$），在大多数情况下，一台相机在物理位置上高于另一台。

根据匹配关键点的横坐标值（x 坐标），将其划分为多个类别 ${C_i}$ 。每个类别对应一个空间范围 R，其定义为 R=width/20。对于类别中的每个关键点，我们计算其视差为 (${x_S} - {x_T}$)，随后计算该类别的平均视差。

随后，我们遍历得到的平均视差列表（每个类别对应一个值），并根据这些平均视差值的统计一致性，将相应类别归入不同的簇。此聚类过程可分离出最主要的运动群组，从而识别出视差最小化的“适当区域”。

算法 1 基于阈值的视差聚类

要求： ${D}$：平均视差列表 ${d_1, d_2, \ldots, d_n}$，其中 ${d_i}$ 对应于类别 ${i}$的平均视差。
要求： ${v}$：一个由用户定义的正数阈值，用于衡量视差差异。
确保： ${C}$：一个簇集合 , 其中，每个簇都是由 ${D}$ 中连续类别组成的列表。

初始化一个空的簇列表 ${C}$
初始化当前簇 ${C_{\text{当前}}}$ 与第一个类别 ${d_1}$
对于 ${i = 2}$ 到 ${n}$ **执行**
    如果 ${|d_i - d_{i-1}| \leq v}$ **则**
        将类别 ${i}$ 添加到 ${C_{\text{当前}}}$
    else
        if ${C_{\text{当前}}}$ 至少包含两个类别**然后**
            将当前 添加到 簇列表 ${C}$
            结束条件判断 当前
        使用类别 ${i}$ 创建新簇： ${C_{\text{current}} \leftarrow \{d_i\}}$
    结束 if
结束 for
如果 ${C_{\text{当前}}}$ 至少包含两个类别 **则**
    添加 ${C_{\text{当前}}}$ 添加到簇列表 ${C}$
结束判断
返回 ${C}$

根据评分指标选择最优拼接簇，该指标依据三个关键参数评估每个簇的一致性和可靠性：

- 标准差（σ_k）：用于衡量簇的内部一致性。标准差越小，表示各类别的平均视差越集中于聚类均值周围，这意味着深度平面的一致性越高。

- 基数（C_k）：指簇中所有类别包含的关键点总数。基数较高表示该簇有大量数据作为支撑，从而直接提高其平均视差计算的可靠性和稳健性。

- 视差偏差（$\Delta\mu_k$）：定义为该簇的平均视差（$\mu_{c,\ k}$）与所有类别的全局平均视差（μ_g）之间的绝对差值。

为避免选中可靠性较低（基数较低）、不一致性较高（标准差较高）或与全局均值差异极小（显著性较低）的簇，我们为每个簇 k 定义一个加权得分，w_k，如公式 5 所示。

$$w = \frac{C}{(\sigma + \lambda \cdot \Delta\mu) + \epsilon} \tag{5}$$

其中：

- σ：簇内视差的标准差，
- C：基数（簇中的关键点数量），
- λ：控制视差偏差影响程度的加权因子，
- $\Delta\mu = |\mu_c - \mu_g|$：簇 μ_c的平均视差与全局平均视差 μ_g之间的绝对差值，

- ϵ：为避免除以零而添加的小常数。

计算得分最高的簇 w_{max}将被选为最优聚类。随后，将该选定簇中所有类别的组合定义为“适当区域”。

3.3.2 关键点链细化

一旦确定了适当的重叠区域，就不能直接使用初始关键点对应关系集合，因为这些原始数据本质上并不可靠，并且包含大量离群点或错位匹配。使用这些错误点来指导分区，会在两幅图像之间产生不一致的划分，进而直接导致明显的拼接伪影，例如重影和重复。因此，细化这组初始关键点是至关重要的前提。

采用专用算法生成两组整洁、同步且有序的关键点列表，形成连贯的“关键点链”。为保持亮度一致性，执行筛选步骤，确定每个关键点的强度，仅保留亮度水平相近的关键点。在最接近图像边缘的关键点处建立初始锚点。随后，算法以欧几里得距离为依据，迭代地将第一幅图像中的每个点与第二幅图像中距离最近的对应点进行匹配。在此过程中施加一项关键约束：每个匹配在水平（x）位置上必须具有唯一性

。该筛选步骤会剔除造成结构不一致的模糊或冲突匹配，从而得到一组经过优化的高置信度配对。

每幅图像中的最优拼接线，定义为连接对应关键点对中精化集合的首个元素与最后一个元素的路径。

3.4 分割与重建

3.4.1 图像分割

图像分割过程始于最优拼接线，该拼接线由 n 个可靠关键点对组成的有序集合定义（见图 3）。这些点用作锚点，将每幅图像分割为 $S = n + 1$ 垂直切片。

为确保生成的垂直切片在结构上互补（即两幅图像中的相对顺序和方向保持一致），需对每对连续锚点（A， B）及其在第二幅图像中的对应锚点对（A'， B'）进行方向验证：

- 有效片段（方向一致）：– 如果 A 的 x 坐标大于 B ($x_A > x_B$) 且 A' 大于B' ($x'_A > x'_B$)：这些图像段有效，因为两对点均从右向左移动（见图 4）。– 如果 $x_A < x_B$ 和 $x'_A < x'_B$：这些图像段有效，因为两对点均从左向右移动（见图 5）。

- 无效图像段（方向不一致）：如果两幅图像的方向不一致（例如 $x_A > x_B$ 与 $x'_A < x'_B$，或反之），则这些图像段会被拒绝，因为它们并非互补片段（见图6）。这可防止在最终重建过程中引入结构不一致

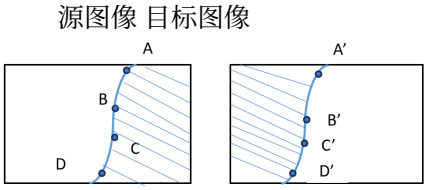


图 3 最优拼接线

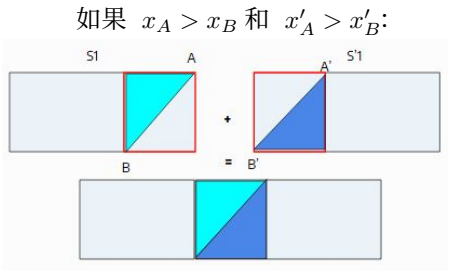


图 4 互补片段

3.4.2 重建

图像段通过验证后，流程进入**拼接阶段**。由于两幅输入图像采用相同的分区模式，每个

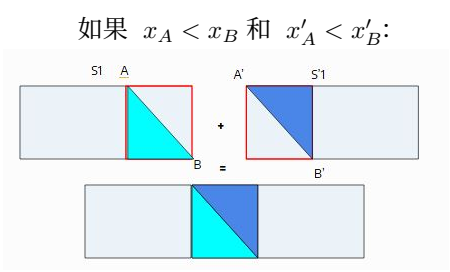


图 5 互补片段

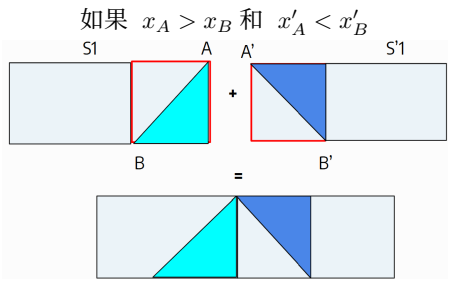


图 6 无效片段

第一幅图像中的图像段都与第二幅图像中的一个图像段直接对应。对应的图像段通过简单的**线性 Alpha 过渡**进行合并：左侧图像段的贡献从左向右线性递减，而右侧图像段的贡献则对称地递增。随后，在接缝区域进行轻微的高斯平滑，以抑制残留的可见伪影。最终，将融合图像水平拼接，再将构建好的各行垂直堆叠，形成最终的拼接图像。

图7 总结了所有描述的过程。

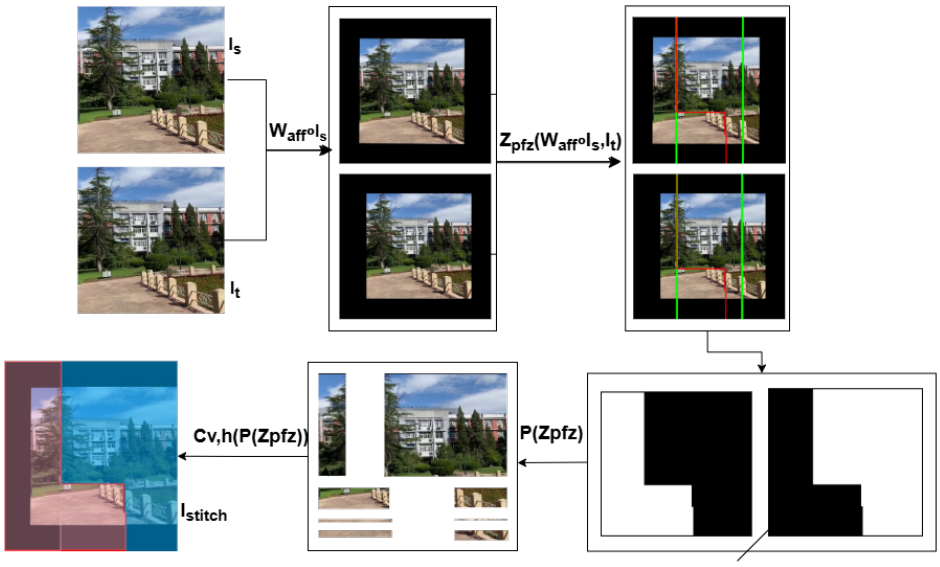


图 7 我们使用局部自适应仿射变形策略（$W_{aff} \circ I_s$），将源图像 I_s 变形到目标图像 I_t 的域中。随后，在重叠区域中确定一个视差最小化区域 Z_{pfz}，并从中提取最优拼接线。根据这条拼接线，通过算子 P将两幅图像划分为对应的垂直切片。最后，将切片水平和垂直拼接（$C_{v,\ h}$），生成最终的拼接图像 I_{stitch}

4 实验与结果

实验在 Google Colaboratory 环境中进行，使用了 12.67GB 的随机存取存储器和 107.72 GB 的存储空间。我们采用了 Nie 等人提供的公开且广受认可的数据集。 [13], Du 等人 [21], 和 Hermann 等人 [22]。代码可在 GitHub 上获取。

4.1 定量评估

表1 本节总结了近期算法在 UDIS-D 数据集上的定量结果（1,106 张图像按难度分为简单、中等和困难三类）。使用**峰值信噪比（PSNR）**和结构**相似性指数（SSIM）**评估拼接图像之间重叠区域的对齐准确度。

指标	PSNR				SSIM			
数据集	Easy	Mod	Hard	Avg	Easy	Mod	Hard	Avg
I_{3x3}	15.87	12.76	10.68	12.86	0.530	0.286	0.146	0.303
SIFT+RANSAC[23]	28.75	24.08	18.55	23.27	0.916	0.833	0.636	0.779
APAP [11]	27.96	24.39	20.21	23.79	0.901	0.837	0.682	0.794
ELA [12]	29.36	25.10	19.19	24.01	0.917	0.855	0.691	0.808
SPW [24]	26.98	22.67	16.77	21.60	0.880	0.758	0.490	0.687
LPC [20]	26.94	22.63	19.31	22.59	0.878	0.764	0.610	0.736
UDIS [13]	25.16	20.96	18.36	21.17	0.834	0.669	0.495	0.648
UDIS++ [8]	**30.19**	25.84	21.57	25.43	**0.933**	**0.875**	0.739	0.838
OURS	24.92	**26.00**	**27.89**	**26.27**	0.823	0.839	**0.882**	**0.848**

表1 使用 PSNR 和 SSIM 的定量结果。

如表 1 所示，我们的方法在所有对比方法中取得了最高的 PSNR 和 SSIM 值，表明其具有更优的拼接质量。这证实了我们的几何驱动策略能够在不同场景中始终实现更好的结构保真和视觉保真度。

4.2 定性评估

我们将我们的方法与多种最先进的方法进行了比较，包括 APAP [11], ELA [12], UDIS [13], UDIS++ [8] 和 SEAMLESS [9]。请随意放大图像，以便更好地观察突出显示的元素。更多数据请参见在线资源，完整的视觉对比结果可在Google Drive上查看。

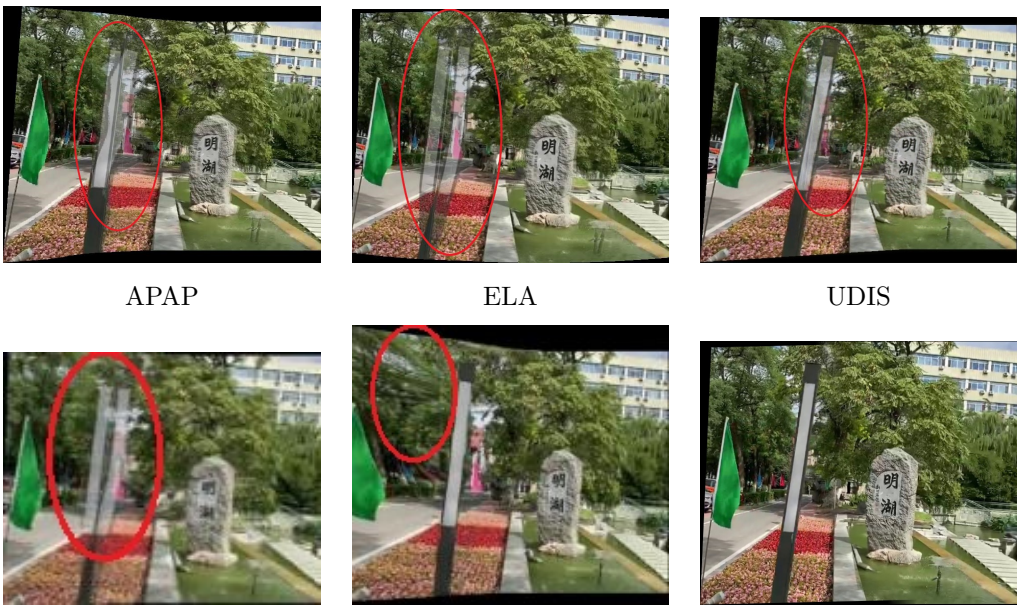


图 8 APAP[11],ELA[12],和 UDIS++ [8] 表现出明显的重影伪影，而 UDIS [13] 会产生显著的模糊现象，SEAMLESS[9] 则会在某些区域引入拉伸。此外，这些方法中的大多数都会使拼接图像整体呈现模糊纹理。相比之下，我们的方法消除了重复，并保留了输入图像原有的清晰纹理。



图 9 我们观察到 APAP、ELA、UDIS、UDIS++ 中存在重影伪影，而 SEAMLESS 中存在模糊纹理。我们的结果没有这些伪影。

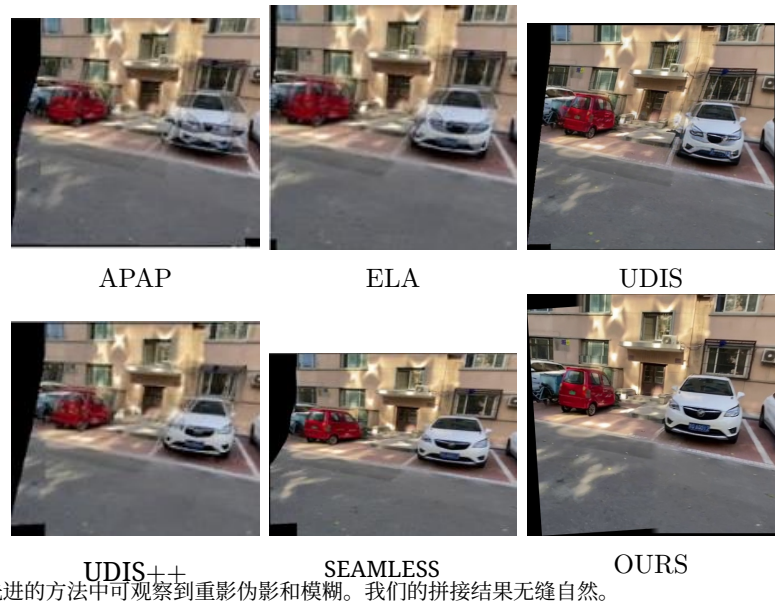
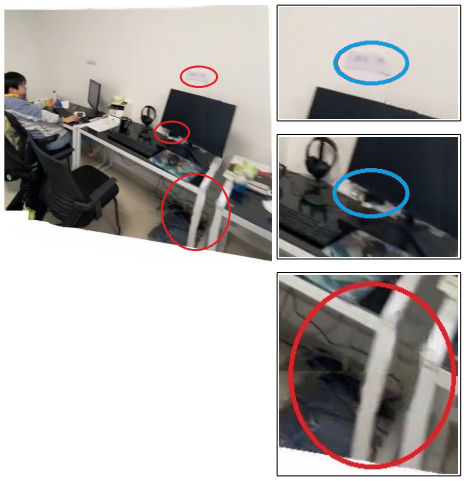


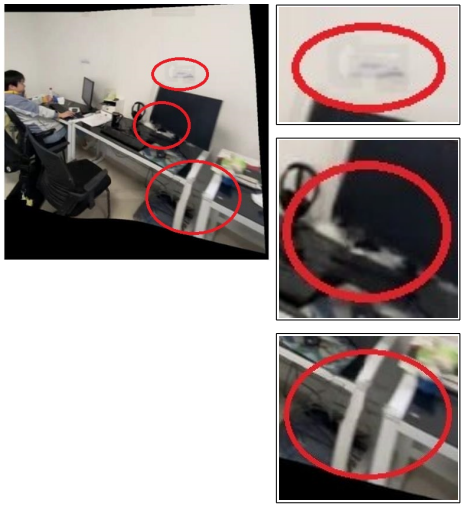
图 10 最先进的方法中可观察到重影伪影和模糊。我们的拼接结果无缝自然。



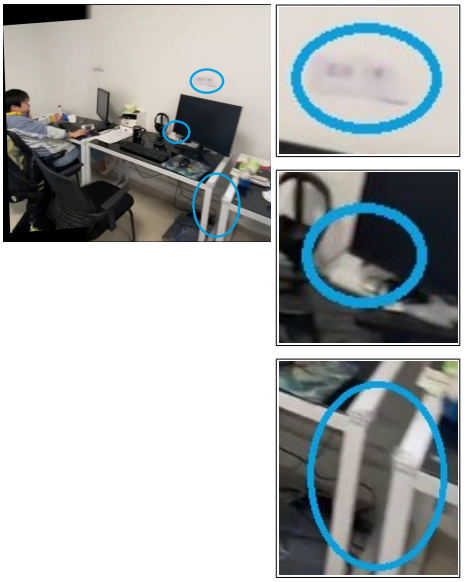
APAP



EPISNET



SEAMLESS

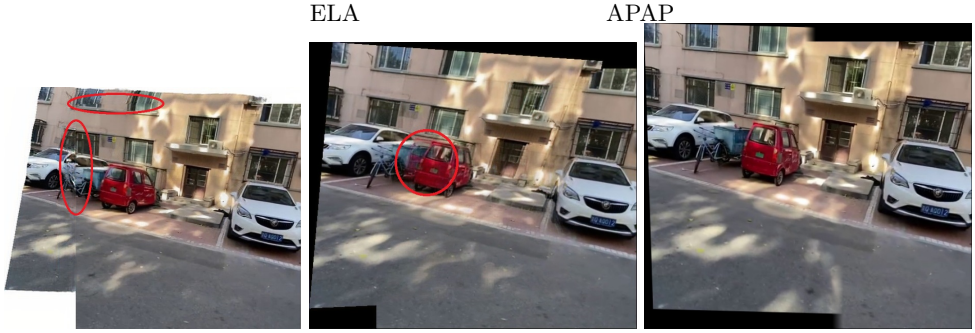


OURS

图 11 与 9中当前最先进水平相似的伪影。



ELA APAP



EPISNET UDIS OURS

图12 与9中当前最先进水平的方法存在类似的伪影。



源图像 ELA APAP



目标图像 UDIS OURS

图13 与9中当前最先进水平的方法存在类似的伪影。

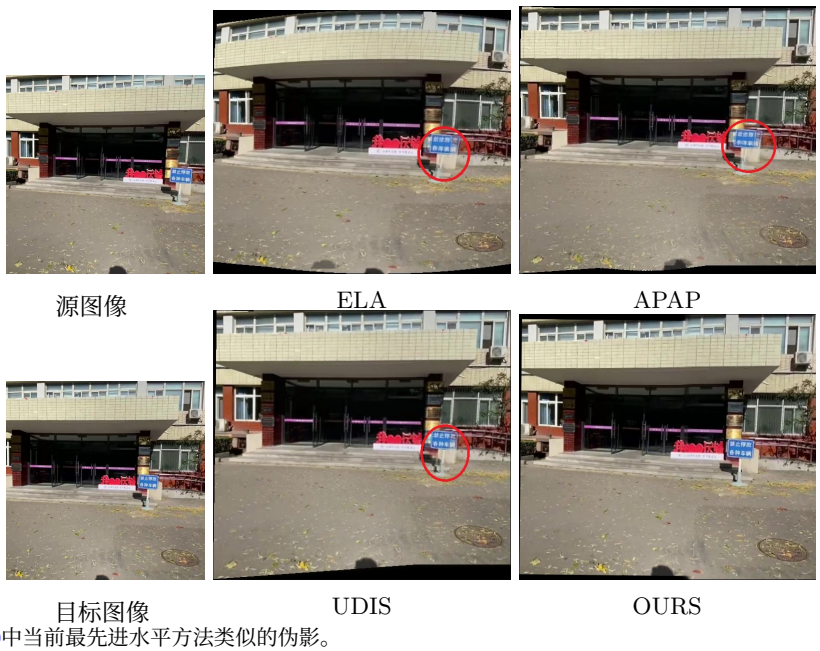


图14 与9中当前最先进水平方法类似的伪影。

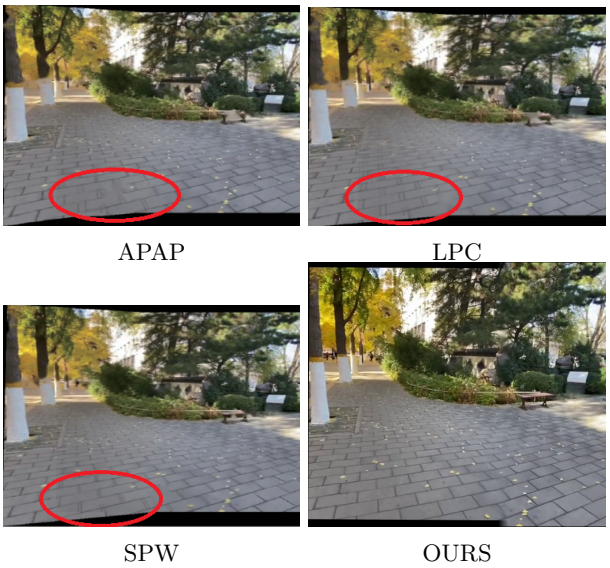


图15 APAP、LPC 和 SPW 中地板地砖的错位，以及 SENA 中的正确对齐。



图16 UDIS++ [8] 和 SEAMLESS [9] 由于图像纹理模糊且存在错位，文字难以辨认。在 SENA 中，文字与输入图像中的一样清晰，且仅观察到轻微错位。

定性结果表明，最先进的方法往往会产生明显的伪影，包括图像结构的重复、纹理丢失和拉伸。相比之下，SENA 生成的拼接图像更加清晰、分辨率更高，同时保持了结构完整性和视觉真实感。

4.3 消融实验

在此，我们评估三个主要组件——局部自适应图像变形、充分的无视差区域识别以及基于分段的重建——在改善对齐和消除伪影方面的效果。



图 17 未采用我们的变形策略。这会导致对齐误差和几何失真。



图 18 采用我们的局部自适应图像变形

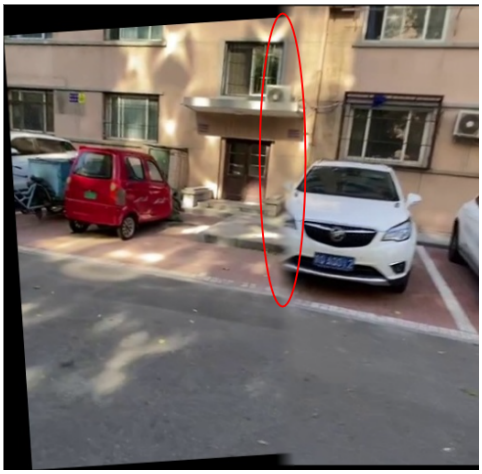


图 19如果未识别无视差区域，拼接可能会在不稳定区域进行，从而导致结果中出现视觉伪影

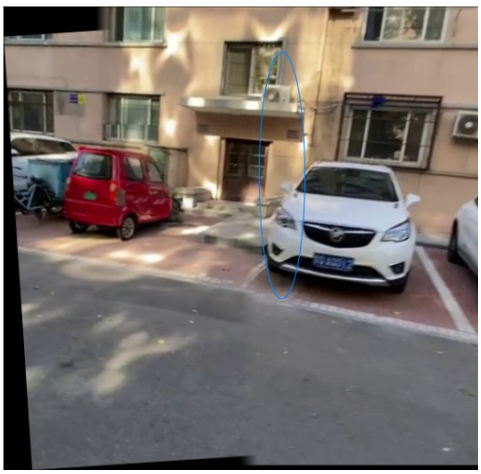


图 20 识别无视差区域后，拼接会在关键点一致的、更稳定的区域进行，从而改善对齐效果。



图 21 不进行融合



图 22 采用我们的融合（线性 alpha 过渡 +轻度高斯平滑）

4.4 局限性

如前所述，SENA 能够有效拼接图像，同时保留其自然外观和几何结构。不过，仍存在一些局限。图像之间的对齐通过多阶段局部变形过程实现：首先利用 RANSAC 估计全局仿射变换，以完成粗对齐；随后在重叠区域内使用局部自适应仿射模型进行细化，并将其插值为平滑的自由形变（FFD）场。这种设计能够实现精准对齐，同时减轻过度失真，并保持形状和纹理的一致性。然而，在深度变化复杂的场景中，局部模型的估计可能变得不稳定，导致局部变形不一致或轻微的几何不连续。此外，检测到的关键点质量会直接影响对视差最小化适当区域的识别，以及后续的拼接线提取。当特征检测或匹配不可靠时，最终全景图中仍可能出现轻微重影、模糊或局部错位等伪影。



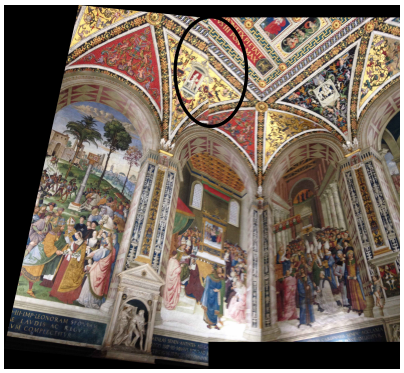


图 23 一些失败案例

5 结论

本研究提出了 SENA，这是一种保结构图像拼接框架，专为解决基于单应性变换的方法和过度灵活的变形模型在几何方面的局限而设计。SENA 不会通过非物理性失真来补偿深度引起的视差，而是采用分层仿射变形策略，将全局仿射初始化、重叠区域内的局部仿射细化以及经过平滑正则化的自由形变场结合起来。这种多尺度设计在提供足够灵活性的同时，能够保留局部形状、平行性和纵横比，从而确保精准对齐。为进一步减轻视差相关伪影，SENA 采用几何驱动拼接机制，直接根据经 RANSAC 筛选的对应点的一致性识别视差最小化适当区域，无需依赖语义分割或基于学习的预处理。在此基础上，系统采用基于锚点的分割策略，将重叠图像划分为由精化关键点对齐的对应垂直切片。这确保了各图像段之间的一一几何对应关系，并维持拼接边界处的结构连续性，从而有效减少重影、重复和纹理拉伸。

大量实验表明，与具有代表性的基于单应性变换和基于样条的方法相比，SENA 始终能够生成视觉上连贯的全景图，同时具有更高的结构保真度和更低的失真。尽管该方法在各种场景下都保持了稳健性，但在极端视点变化或纹理严重缺失的环境中，性能可能会有所下降，因为此时特征提取变得不可靠。未来的工作将探索自适应局部变形模型和混合式学习组件，以进一步提升其在此类挑战性条件下的稳健性。

参考文献

[1] Nghonda, E.：《为现场活动的外向内摄像系统实现360度沉浸式体验》。佛罗里达大学，???（2023）

[2] Lowe, D.G.：基于尺度不变关键点的区别性图像特征。《国际计算机视觉期刊》**60**，91–110（2004）[3] Bay, H., Tuytelaars, T., Van Gool, L.：SURF：加速稳健特征。载于：《计算机视觉–ECCV 2006：第9届欧洲计算机视觉会议》，奥地利格拉茨，2006年5月7–13日。论文集，第一部分9，第404–417页（2006）。Springer[4] Ji, X., Yang, H., Han, C.：基于改进的ORB和拼接线计算的图像拼接方法研究。《电子成像期刊》**31**（5），051404–051404（2022）[5]Liu, W., Zhang, K., Zhang, Y., He, J., Sun, B.：利用归并排序方法提高多场景图像拼接中的拼接效率。《应用科学》**13**（5），2791（2023）[6] Potje, G., Cadar, F., Araujo, A., Martins, R., Nascimento, E.R.：XFeat：用于轻量级图像匹配的加速特征。载于：IEEE/CVF计算机视觉与模式识别会议论文集，第2682–2691页（2024）[7] Wen, S., Wang, X., Zhang, W., Wang, G., Huang, M., Yu, B.：面向视差容忍图像拼接的结构保持与接缝优化。《IEEE Access》**10**，78713–78725（2022）[8] Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.：视差容忍的无监督深度图像拼接。载于：IEEE/CVF国际计算机视觉会议论文集，第7399–7408页（2023）[9] Chen, K., Garg, A., Wang, Y.-S.：Seamless-through-breaking：重新思考图像拼接以实现最优对齐。载于：亚洲计算机视觉会议论文集，第4352–4367页（2024）[10] Li, X., He, L., He, X.：用于图像拼接的组合式区域单应-仿射变形。载于：第十四届图形与图像处理国际会议（ICGIP 2022），第12705卷，第258–264页（2023）。SPIE[11] Zaragoza,J., Chin, T.-J.,Brown, M.S., Suter, D.：采用移动DLT的尽可能投影的图像拼接。载于：IEEE计算机视觉与模式识别会议论文集，第2339–2346页（2013）[12] Li, J., Wang, Z., Lai, S., Zhai, Y., Zhang, M.：基于稳健弹性变形的视差容忍图像拼接。《IEEE多媒体汇刊》**20**（7），1672–1687（2017）[13] Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.：无监督深度图像拼接：将重建的拼接特征还原为图像。《IEEE图像处理汇刊》**30**，6184–6197（2021）

[14]Huang, H., Chen, F., Cheng, H., Li, L., Wang, M.：语义分割引导的特征点分类与接缝融合图像拼接方法。算法与计算技术期刊 **15**，17483026211065399（2021）[15]Qin, Y., Li, J., Jiang, P., Jiang, F.：通过特征定位和接缝消除实现图像拼接。多媒体工具与应用 **80(14)**，20869–20881（2021）[16] Chai, X., Chen, J., Mao, Z., Zhu, Q.：一种用于超大遥感图像镶嵌的升降尺度最优接缝线检测算法。遥感 **15(1)**，89（2022）[17]Garg, A., Dung, L.-R.：用于最优接缝线搜索的拼接条带确定。载于：2020年第4届国际成像、信号处理与通信会议（ICISPC），第29–33页（2020）。IEEE[18] Zheng, J., Wang, Y., Wang, H., Li, B., Hu, H.-M.：一种新颖的基于投影一致平面的图像拼接方法。IEEE多媒体汇刊 **21(10)**，2561–2575（2019）[19] Yadav, S., Choudhary, P., Goel, S., Parameswaran, S., Bajpai, P., Kim, J.：自拍拼接：基于双单应性的广角自拍图像拼接。载于：2018年IEEE国际多媒体与博览会研讨会（ICMEW），第1–4页（2018）。IEEE[20] Jia, Q., Li, Z., Fan, X., Zhao, H., Teng, S., Ye, X., Latecki, L.J.：利用线点一致性，在宽视差图像拼接中保持结构。载于：IEEE/CVF计算机视觉与模式识别会议论文集，第12186–12195页（2021）[21] Du, P., Ning, J., Cui, J., Huang, S., Wang, X., Wang, J.：用于自然图像拼接的几何结构保持变形。载于：IEEE/CVF计算机视觉与模式识别会议论文集，第3688–3696页（2022）[22] Herrmann, C., Wang, C., Bowen, R.S., Keyder, E., Zabih, R.：以对象为中心的图像拼接。载于：欧洲计算机视觉会议（ECCV）论文集，第821–835页（2018）[23] Fischler, M.A., Bolles, R.C.：随机抽样一致性：一种用于模型拟合的范式及其在图像分析和自动制图中的应用。ACM通讯 **24(6)**，381–395（1981）[24] Liao, T., Li, N.：自然图像拼接中的单透视变形。IEEE图像处理汇刊 **29**，724–735（2019）