arXiv:2601.14208v1 [cs.CV] 20 Jan 2026

使用高斯泼溅进行刚性架感知的车辆底盘三维重建

Nitin Kulkarni*†、Akhil Devarashetti*、Charlie Cluss*、Livio Forte*、Dan Buckmaster*、Philip Schneider*、Chunming Qiao†、Alina Vereshchaka†
†布法罗大学，美国纽约州布法罗
*ACV Auctions，美国纽约州布法罗{nitinvis, qiao, avereshc}@buffalo.edu，{adevarashetti, ccluss, lforte, dbuckmaster, pschneider}@acvauctions.com

摘要—检查二手车辆的底盘是一项劳动密集型工作，检验员必须蹲下或爬到每辆车底下进行全面检查。此外，在线买家很少能看到车辆底盘的照片。我们提出了一套端到端流程，利用三摄像头设备架在车辆驶过时拍摄底盘视频，并生成底盘的交互式三维模型。该三维模型支持检验员和客户旋转、缩放并剖切车辆底盘，使他们能够在几秒钟内发现锈蚀、泄漏或撞击损伤，从而提升工作场所安全和买家信心。我们的主要贡献是一套刚性架感知的运动恢复结构（SfM）流程，专为克服广角镜头畸变和低视差场景带来的挑战而设计。我们通过整合精确的相机标定、同步视频流以及来自相机支架的强几何先验，克服了广角镜头畸变和低视差场景带来的挑战。我们采用结合学习组件的受约束匹配策略，使用 DISK 特征提取器和基于注意力机制的 LightGlue 匹配器，生成高质量的稀疏点云；使用标准运动恢复结构流程通常难以获得此类点云。这些点云为高斯泼溅过程提供初始数据，进而生成可实时渲染的照片级真实感底盘模型。我们的实验和消融研究表明，这些设计选择对于实现最先进的质量至关重要。

索引词—高斯泼溅、运动恢复结构、LightGlue、神经辐射场、三维重建、摄影测量、车辆检查

I. 引言

二手车辆构成了一个规模达数十亿美元的全球市场 [1], ，数字拍卖所占份额也日益扩大。然而，评估车辆底部状况——这是决定车辆价值的关键因素之一——仍然颇具挑战。锈蚀、液体泄漏或车架损坏的迹象，往往难以通过标准成像和检查流程发现。由于工作环境狭窄，车辆检查人员可能会漏检问题；而买家，尤其是居住在农村或偏远地区的买家，通常难以获得可信赖的检查服务。对于这些地区的个人而言，这一挑战尤为突出，他们往往无法获得专业检查服务。由于无法清楚查看车辆底部，买家只能在不确定的情况下做出决定，既可能面临高昂的维修费用，也会降低对在线交易的信任。

C. Qiao 获得美国国家科学基金会部分资助，资助项目编号为 CNS-2413876 和 CNS-2120369。作者感谢 ACV Auctions 提供计算资源、车辆数据和相机支架。

为弥合这一信息差距，我们提出了一种解决方案，可提供高分辨率、照片级真实感且可交互的车辆底盘三维模型。这些三维模型使买家能够远程细致地检查车辆。对于卖家和检验员而言，该流程提高了透明度和问责制，最终增强信任，并为二手车市场树立了新标准。

我们的系统通过三台横向间隔布置的广角相机，在车辆驶过定制支架时采集高分辨率视频。由于相机与车辆底盘之间的距离很短（12 – 30厘米），必须使用广角镜头（160°视场角）才能拍摄到底盘。然而，这类广角镜头会产生明显畸变，使特征匹配变得困难。当车辆沿直线驶过相机时，单个相机镜头只能提供沿车辆长度方向的视差。为生成用于深度估计的宽度方向视差，我们使用三台横向间隔布置的广角相机。传统的运动恢复结构（SfM）流程在这些条件下往往难以奏效，而神经辐射场（NeRF） [2] 的速度又无法满足生产级车辆检测流程的要求。

我们提出了一种新颖的模块化流程，将传统多视图几何与 3D 高斯泼溅（3D-GS）相结合， [3] 以高效重建精细且可交互的车辆底盘模型。首先，我们使用 ChAr Uco 标定板进行精确的相机标定，并挑选多样且清晰的图像，以建模镜头畸变并尽量减小投影误差（第 III-A节）。接下来，我们通过相位相关估计全局垂直位移，以亚帧精度同步多相机视频流，然后利用自适应互相关进一步优化时间偏移（第 III-B节）。这种两阶段对齐实现了毫秒级同步，这是后续处理所必需的。

随后，我们采用基于先进学习组件的考虑设备约束的运动恢复结构方法。选出最清晰的图像三元组后，我们使用标定得到的相机内参对图像进行去畸变。接着，我们从这些去畸变图像中提取稠密局部特征，使用的是 DISK [4], ，这是一种专为应对复杂条件而设计的学习型关键点检测器。我们使用基于注意力的匹配器 LightGlue [5] ，在一组受限的时空邻近图像对上匹配特征。随后，在 COLMAP [6] 中对得到的原始匹配进行几何验证。接着，我们使用 GLOMAP [7] 运行全局光束法平差

© 2025 IEEE。允许将本材料用于个人用途。对于任何其他用途，无论是现有媒体还是未来媒体，均须获得 IEEE 的许可，包括为广告或宣传目的重新印刷/再版本材料、创作新的合集作品、转售或再分发至服务器或列表，或在其他作品中重复使用本作品的任何受版权保护部分。

，并利用已知相机架几何结构提供的先验信息，优化相机位姿，从而生成干净、精确的车辆底盘点云（第III-C节）。该点云和优化后的相机位姿将作为高斯泼溅的种子，用于生成交互式三维模型（第III-D节）。

我们提出的流程能够高效生成达到检测级精度的三维重建，使车辆检查员和买家能够实时交互式地检查汽车底盘。我们证明，通过精确的标定、同步和几何感知的运动恢复结构，即使面对具有挑战性的广角、低视差序列，也能达到神经辐射场级别的质量，使其成为适用于真实场景部署的实用且可扩展的解决方案。这项技术有望规范在线交易平台上的车辆检测，增强信任，并支持保险理赔和车队维护等应用。

II. 背景

从二维图像进行精确的新视角合成，是虚拟现实、文化遗产数字化和工业检测等应用的基础 [8]。我们的工作建立在三维重建的几个关键领域之上：经典多视图几何、用于特征提取与匹配的深度学习模型，以及辐射场领域的最新进展。

A. 运动恢复结构

运动恢复结构（SfM）通过检测并匹配不同视图中的关键点，根据重叠图像估计三维场景几何结构和相机位姿。从初始图像对开始，COLMAP 等系统通过后方交会和三角测量逐步添加新视图，随后利用光束法平差（BA）联合优化相机位姿和三维点位置。最终得到稀疏三维点云和全局一致的相机轨迹。

然而，标准的运动恢复结构流程在处理畸变和低视差场景时面临挑战，重建质量高度依赖于特征匹配的质量及其分布。在我们的车辆底盘检测中，必须使用广角镜头拍摄场景，这会引入严重的非线性畸变；如果未能准确建模，这些畸变会干扰特征匹配 [9]。此外，当相机配置的视差较低时，运动恢复结构也难以应对；而视图之间的差异对于深度估计至关重要 [10]。这种由我们的驶越式相机支架固有带来的情况，可能导致几何退化以及估计的相机位姿发生漂移，最终降低生成的三维点云质量。

B. 学习型局部特征与匹配器

传统的运动恢复结构流程一直依赖尺度不变特征变换等手工设计的特征 [11], ，但深度学习的最新进展带来了更强大的替代方案。学习型局部特征（如 DISK）通过大型图像数据集进行训练，以检测关键点并生成描述子；与经典方法相比，它们对极端视点变化和光照变化具有更强的鲁棒性。

同样，基于深度学习的匹配器增强了跨图像匹配特征的能力。它们不再依赖最近邻匹配，而是采用 LoFTR [12],SuperGlue [13], 和 LightGlue [5]，等现代方法，利用基于注意力机制的图神经网络，同时考虑两幅图像中所有特征的全局上下文。这使它们即使在传统方法难以处理的低纹理或重复纹理场景中，也能找到准确的匹配结果。我们的工作将这些学习得到的特征提取器和匹配器整合到一个受几何约束的运动恢复结构流程中，以克服我们设置中的挑战。

C. 辐射场与新视角合成

虽然运动恢复结构能够捕捉稀疏几何结构，但无法提供照片级真实感的渲染。早期流程通过多视图立体视觉（MVS） [14], 表面重建和纹理映射弥补了这一不足。近年来，神经辐射场（NeRF） [2]展现出卓越的新视角合成能力，能够通过神经网络学习体积场景表示。

神经辐射场能够生成逼真的三维场景，并捕捉反射和半透明性等精细细节。然而，其计算成本高昂，通常需要数小时进行训练，而且渲染速度仅为每分钟几帧。后续的神经辐射场变体缩短了单场景训练时间；例如，Instant-NGP 使用多分辨率哈希编码和经过优化的计算内核，实现更快的训练和交互式渲染[15]。相比之下，3D 高斯泼溅（3D-GS）是一种由运动恢复结构初始化的显式表示方法，将场景建模为各向异性高斯，并采用可见性感知泼溅，实现快速优化和实时渲染；这一权衡非常适合我们的车辆底盘驶越式工作流程 [3], [16]。

D. 高斯泼溅

3D 高斯泼溅（3D-GS） [3]为神经辐射场提供了一种实时替代方案，训练和渲染速度更快，同时达到相当甚至更高的视觉保真度。3D-GS 不采用隐式神经表示，而是将场景显式建模为大量 3D 高斯的集合。

每个高斯分布都由一组可优化属性定义：其位置（均值 μ）、形状和方向（ 3×3 协方差矩阵 Σ）、颜色（由球谐函数表示）以及不透明度（α）。这些高斯分布由运动恢复结构（SfM）生成的稀疏点云初始化。训练过程中，会优化其属性以重建训练图像。

与神经辐射场不同，3D-GS 无需在场景中追踪光线。相反，它会将每个高斯分布直接投影到图像上，形成类似柔和二维圆点的“泼溅”，再通过阿尔法混合将它们合成。借助快速光栅化器，这一过程能够高效完成。因此，训练速度很快，并且能够实时渲染。

由于三维高斯泼溅依赖初始运动恢复结构点云，因此初始重建的质量至关重要。这直接促使我们设计了支架感知运动恢复结构流程，以处理车辆底盘这类低视差广角场景。

III. 方法

我们的流程如图1所示，可将来自多个相机视角的原始视频转换为高质量的交互式三维

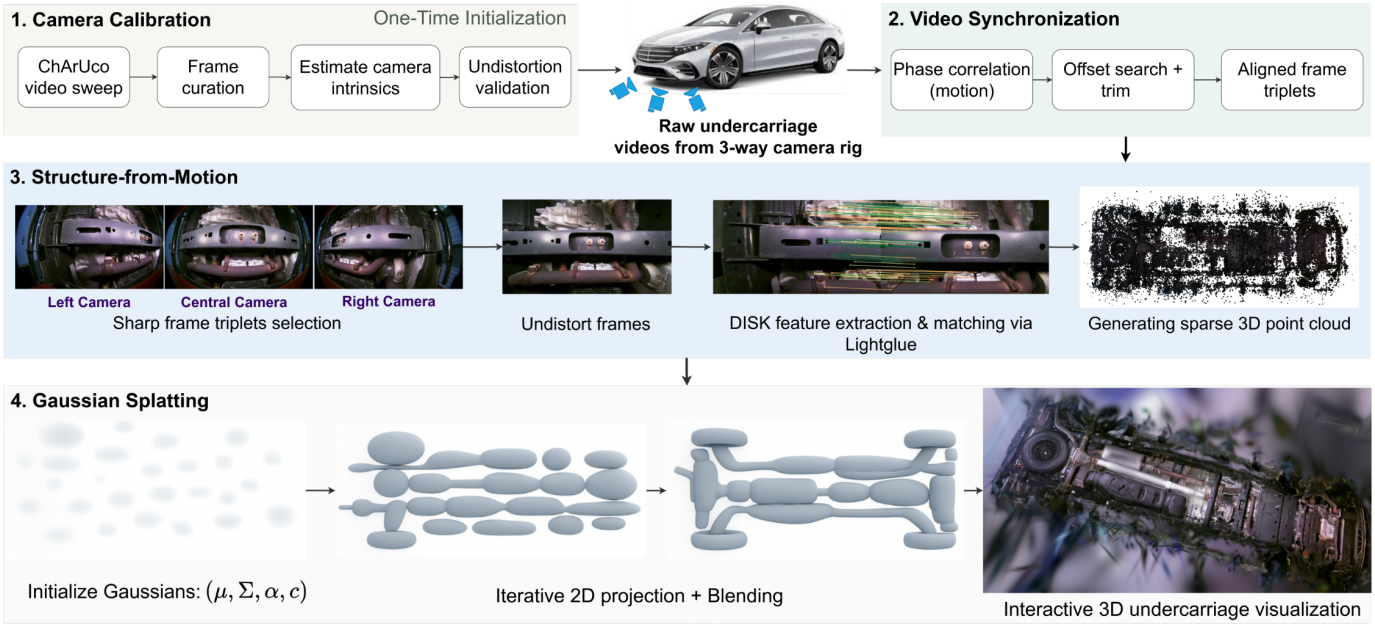


图 1：三维重建流程概览。(1) 我们使用 ChArUco 标定板进行一次性相机标定，以建模并校正严重的广角镜头畸变。(2) 对于每辆车，我们同步采集设备上三台相机拍摄的原始视频，以确保时空对齐。(3) 从同步后的视频中，均匀采样最清晰的帧三元组，对帧进行去畸变，提取 DISK 特征，并通过 LightGlue 应用约束特征匹配策略，以查找不同帧之间的匹配点，随后通过光束法平差生成车辆底盘的稀疏点云。(4) 最后，使用点云初始化 3D 高斯，并对其进行优化，以生成可交互的车辆底盘三维模型。

车辆底盘三维模型。我们的流程包含四个主要步骤。第一步（第III-A节）是进行精确的相机标定，以校正严重的镜头畸变。此标定只需执行一次；标定后的相机参数可用于重建任意车辆底盘的三维模型。第二步（第III-B节）是对来自三台相机的视频进行同步，以确保三台相机之间的时空对应关系。第三步（第III-C节）是采用刚性架感知的运动恢复结构（SfM），以估计三维几何和相机位姿。最后一步（第III-D节）是使用 3D 高斯泼溅生成车辆底盘的照片级真实感、交互式三维模型。每一步都旨在应对广角镜头或低视差视图带来的挑战。

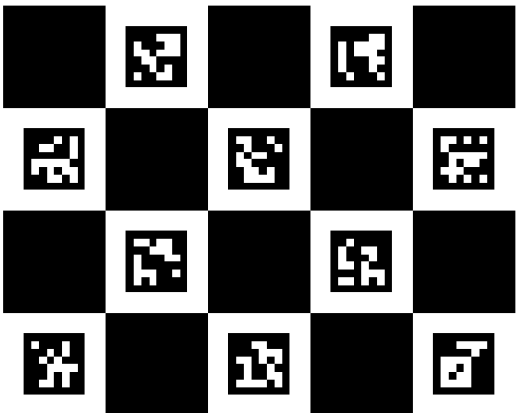


图 2：ChArUco 标定板示例 [18]。

A. 相机标定

广角镜头对于近距离拍摄车辆底盘必不可少，但会引入严重的径向畸变和切向畸变；如果没有准确的相机内参，这些非线性畸变会传播到运动恢复结构（SfM）阶段，并使稀疏点云发生扭曲。尽管 COLMAP 等 SfM 框架能够估计相机内参，但使用专用标定流程可显著降低投影误差，并减少光束法平差漂移。

1）ChArUco 标定板：我们使用 ChArUco 标定板（图2），这是一种每隔一个方格便以 ArUco 标记替换的棋盘格图案。ChArUco 标定板能够对棋盘格角点进行精确的亚像素角点检测，并通过 ID 实现稳健的 ArUco 标签检测，与仅使用标记或普通棋盘格图案相比，投影误差更低 [17]。我们的标定板包含 53×37 个边长为 22mm 的内部方格，每个 ArUco 标记的宽度为 16 mm。我们将标定板依次扫过三台广角相机，拍摄标定视频，同时改变滚转角、俯仰角、偏航角和距离（$\approx 30-120$cm），确保每台相机都能观测到较大的基线变化。

*2)*帧筛选：我们从每段视频中连续取十张图像作为一个滑动窗口，从中采样出一组多样且清晰的标定图像。对于每个滑动

窗口，先将图像转换为灰度图，再根据拉普拉斯算子的方差（公式1）进行评分。拉普拉斯算子能够突出强度快速变化的区域，因此其方差可作为衡量边缘清晰度和整体图像锐度的有效指标。

$$\mathrm{Var}(\nabla^2 I) = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\left(\nabla^2 I(x,y) - \mu_{\nabla^2}\right)^2 \quad (1)$$

$$\text{where, } \mu_{\nabla^2} = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\nabla^2 I(x,y)$$

数值越高表示边缘越清晰、模糊越少。在每个十帧窗口中，我们采样拉普拉斯方差最高的帧，从而确保选取最清晰且运动模糊最小的帧。

3) 相机模型与优化：三维点 $P = [X, Y, Z]^\top$ 到二维像素坐标 $p = [u, v]^\top$的投影分为两个步骤。首先，将三维点投影到归一化、无畸变的图像坐标 $p' = [x', y']^\top$，其中 $x' = X/Z$ 且 $y' = Y/Z$。其次，对 p' 应用畸变函数，得到畸变坐标 $P_d = [x_d, y_d]^\top$，然后使用相机内参矩阵将其映射到像素坐标 K （式2）。

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

我们使用 Levenberg-Marquardt 优化器 [19]拟合八参数 OpenCV 完整相机模型（式3），以求得镜片的畸变参数。该模型结合了径向畸变（包含 k_i 的项）和切向畸变（包含 p_i 的项），从而考虑偏心和非平行镜片元件的影响。该模型在标准 OpenCV 模型的基础上增加了三个额外的径向畸变系数（k_4、k_5、k_6），以更好地拟合极端广角镜头的复杂畸变特征。

$$\begin{aligned} x_d &= x' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + 2p_1 x' y' + p_2(r^2 + 2x'^2) \\ y_d &= y' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + p_1(r^2 + 2y'^2) + 2p_2 x' y' \end{aligned} \quad (3)$$

4) 误差指标： 优化过程通过最小化均方根（RMS）重投影误差来进行，该误差衡量 ChArUco 标定板上观测到的角点位置 p_i 与其对应的投影位置 $\hat{}p_i$之间的欧几里得距离（以像素为单位），计算公式如下：RMS $= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{p}_i - \hat{\mathbf{p}}_i\|^2}$

除这一量化指标外，我们还进行了定性评估，通过目视检查去畸变图像，确认场景中的直线看起来仍是直线（图4）。随后固定得到的内参矩阵 K 和畸变系数，并将其传入我们的运动恢复结构流程（第III-C节）。

B. 视频同步

我们的相机支架经过编程，可同时触发三台相机开始录制视频。然而，由于信号和编码延迟，生成的视频可能会略有不同步。我们分两个阶段解决这一问题：首先，我们估计全局垂直运动，然后通过最小化运动曲线之间的 L1 损失来对齐这些运动曲线。

1) 全局运动估计： 对于每一对连续帧，我们使用相位相关计算沿垂直方向移动的像素数。由此得到垂直位移序列，即偏移量，对于包含 N 帧的视频记为 $S \in \mathbb{Z}^{N-1}$。我们对每一帧进行预处理，依次应用高斯模糊、限制对比度自适应直方图均衡化（CLAHE）和拉普拉斯滤波器，以突出其关键特征。

对于由左侧、中央和右侧摄像机 $\{l, c, r\}$拍摄的三个视频，我们分别获得三个对应的位移序列 $\{S_l, S_c, S_r\}$，其长度为 $\{N_l - 1, N_c - 1, N_r - 1\}$。这些位移构成压缩信号，用于概括随时间变化的全局垂直运动，并据此进行时间对齐。

2) 通过 *L1* 损失最小化搜索偏移量：我们通过找出使位移信号之间的平均绝对差最小的时间偏移量，来同步视频。对于任意两个视频 a和 b，我们将 L1 对齐损失定义为：

$$\mathcal{L}(a,b) = \frac{1}{\overline{N}}\sum_{i=1}^{\overline{N}}|S_i^a - S_i^b|, \quad \text{where } \overline{N} = \min(N^a - 1, N^b - 1) \quad (4)$$

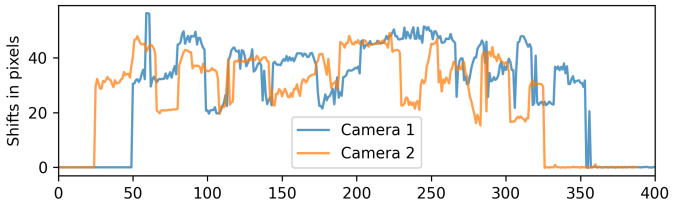
随着偏移量增大或减小，误差会单调趋近最小值（图3c），因此我们将其视为凸优化问题。为了优化误差，我们采用迭代方法。首先，在图中蓝点所示的固定偏移量范围内进行均匀搜索。3c。然后，我们选取误差最小的偏移量，并在橙点所示的更小范围内进行搜索。重复这一过程，直到达到最小值，且该轮迭代的范围内没有尚未探索的偏移量。

对于三段视频，我们以中央摄像机作为时间参考，并分别计算左侧和右侧视频的时间偏移。对齐后，我们裁剪多余帧，使视频长度一致，从而得到长度为 N_{synced} 的同步序列。

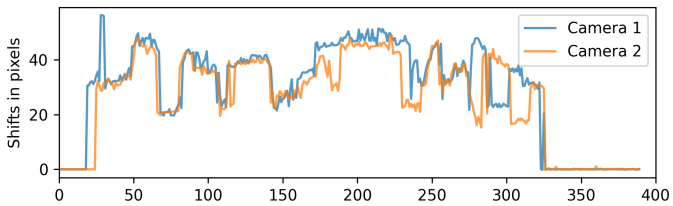
C. 运动恢复结构（*SfM*）

我们将同步视频和相机内参输入运动恢复结构流程，以生成表示车辆底盘的点云，并计算精确的相机位姿。随后，这个稀疏点云将作为高斯泼溅的种子（第III-D节）。我们的运动恢复结构流程专为应对低视差、广角拍摄的挑战而设计，将稳健的学习式特征匹配流程与来自相机支架的强几何先验相结合。下文将详细介绍关键算法选择：

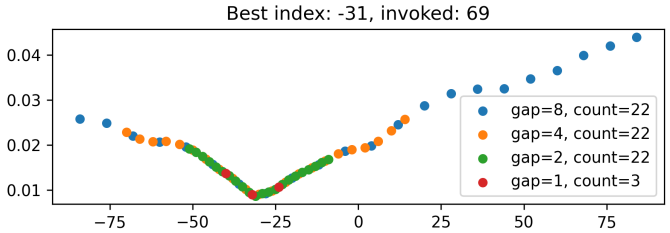
1）帧选择和图像去畸变：考虑到车辆的速度以及与摄像机的距离，运动模糊可能会降低图像质量。为减少计算量并提高质量，我们从同步的左侧、中央和右侧摄像机视频中均匀采样最清晰的 $k = 250$ 帧三元组。通过计算帧三元组的平均拉普拉斯方差来评估清晰度（公式5）。



(a) 同步前两个视频的偏移量。



(b) 同步后的偏移对齐。



(c) 不同偏移量下的 L1 损失。最低点表示最佳同步。

图 3：使用相位相关和 L1 损失最小化对两个视频进行同步。

$$\mathrm{Score}(I) = \frac{1}{3}\big(\mathrm{Var}(\nabla^2 I_l) + \mathrm{Var}(\nabla^2 I_c) + \mathrm{Var}(\nabla^2 I_r)\big) \quad (5)$$

其中，I_l、I_c和 I_r 分别是左侧、中央和右侧摄像机拍摄的灰度图像，$\nabla^2 I$ 是灰度图像 I的离散拉普拉斯算子。

我们流程中的一个关键步骤是首先校正严重的镜头畸变。我们使用通过一次性 ChArUco 标定流程（第III-A节）获得的相机参数，对每个选定帧进行去畸变。此预处理步骤可校正图像，确保在几何形状正确的空间中检测特征。

2) 特征提取：为增强局部对比度并显现车辆底盘较暗、曝光不足区域中的特征，我们应用 CLAHE，使特征数量增加 $15-25\%$。

我们采用学习型局部特征描述符 DISK，因其在严苛条件下具有较强的鲁棒性。它经过海量图像数据集的训练，能够检测并描述在视点和光照发生显著变化时仍高度可重复且可靠的特征；这类变化在我们的车辆底盘采集环境中十分常见。我们从每个预处理帧中提取最多 8192 个 DISK 特征。

3）约束特征匹配：标准的运动恢复结构（SfM）方法通常依赖穷举匹配或序贯匹配；对于具有重复结构或因多相机架带来复杂运动的场景，这些方法计算开销较大且容易出错。我们利用已知的时空结构的捕获过程，实现了一种高效且稳健的受约束的匹配策略。

对于索引为 i 的每个三图像组，我们考虑一个时间窗口 $\mathcal{W}_5(i) = \{i-5, \ldots, i+5\}$，并创建三类图像对：

$$\begin{aligned}
&\text{(i) intra-camera} : (c,i) \leftrightarrow (c,j), \quad c \in \{\mathrm{L,C,R}\}, \quad j \in \mathcal{W}_5(i),\\
&\text{(ii) cross L}\rightarrow\text{C} : (\mathrm{L},i) \leftrightarrow (\mathrm{C},j), \quad j \in \mathcal{W}_5(i),\\
&\text{(iii) cross C}\rightarrow\text{R} : (\mathrm{C},i) \leftrightarrow (\mathrm{R},j), \quad j \in \mathcal{W}_5(i).
\end{aligned}$$

我们有意省略左侧（L）与右侧（R）相机视图之间的直接匹配，因为这两台相机之间相对较大的水平距离（基线为 62 cm），与相机到车辆底盘的较短垂直距离（$12-30$cm）相比，会造成极大的视角差异，使得准确的特征匹配难以实现。相机内匹配有助于追踪相机随时间的运动，而跨相机匹配则提供实现稳健三角测量所需的立体基线。我们使用 LightGlue 对这些受约束图像对中的特征描述符进行匹配。LightGlue 利用基于注意力的图神经网络寻找对应关系，并综合考虑两幅图像中所有特征的全局上下文，从而生成精度很高、离群点很少的原始匹配。随后将这些原始匹配导入 COLMAP，并使用 RANSAC [20]进行最终几何验证，以筛除任何残余离群点，并确保结果与我们标定的相机模型一致。这种两阶段方法先在时空上约束搜索空间，再采用高保真学习型匹配器，两者协同加速匹配过程，同时最大限度地减少错误对应关系。

4) 感知相机架的稀疏点云生成：我们采用增量式运动恢复结构方法，从一组成像效果良好的图像对开始，逐步加入更多视图。我们方法中的一个关键环节，是在光束法平差（BA）优化中将已知的相机架几何结构作为先验加以整合。

三维点三角测量后，光束法平差会通过最小化总重投影误差（式6）来优化相机位姿和三维结构。目标函数为：

$$J = \min_{\mathbf{C}_i,\,\mathbf{P}_j} \sum_{i,j} w_{ij} \left\| \pi(\mathbf{C}_i, \mathbf{P}_j) - \mathbf{p}_{ij} \right\|^2 \qquad (6)$$

其中，$\mathbf{C}_i$ 是相机 i的姿态，$\mathbf{P}j$ 是点 j的三维位置，$\mathbf{p}ij$ 是相机 i中点 j 的观测投影，而 w_{ij} 是表示可见性的二值项。

为确保符合已知的相机架几何结构，我们在目标函数中加入一个基于相机架上相机相对位置和方向的正则化项。根据相机架示意图，我们为左、右相机相对于中央摄像机设置姿态先验。

$$\mathbf{t}_{\text{L-C}} = [-0.31, 0, 0]^\top, \quad \mathbf{t}_{\text{C-R}} = [+0.31, 0, 0]^\top$$

这种刚性架感知的光束法平差可防止相机间漂移，从而生成准确的稀疏点云。

D. 高斯泼溅

我们将运动恢复结构流程生成的稀疏点云（第 III-C 节）与去畸变图像一同输入 Kerbl 等人提出的 3D 高斯泼溅框架。 [3]该框架可将离散的三维点集转换为稠密的辐射度表示，并可从任意视点进行渲染。

稀疏点云中的每个点都被初始化为一个高斯分布：

$$G_i(\mathbf{x}) = \alpha_i \exp\left[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathsf{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right] \tag{7}$$

其中， $\boldsymbol{\mu}_i \in \mathbb{R}^3$ 是均值位置， α_i 是其初始不透明度， $\boldsymbol{\Sigma}_i \in \mathbb{R}^{3\times3}$ 是根据局部点密度缩放的对角协方差矩阵。

随后对 3D 高斯参数进行优化，从而获得场景的稠密辐射度表示。渲染时，每个 3D 高斯分布 G_i 都会被投影到图像空间，成为一个 2D 高斯分布 $G_i(\mathbf{u})$ ，其中， $\mathbf{u} = (u, v)$ 。可见性和颜色通过由前向后的阿尔法混合进行合成：

$$\mathbf{C}(\mathbf{u}) = \sum_{i=1}^{N} w_i(\mathbf{u})\,\mathbf{c}_i, \quad w_i(\mathbf{u}) = \tilde{\alpha}_i(\mathbf{u}) \prod_{j<i}\left[1-\tilde{\alpha}_j(\mathbf{u})\right] \tag{8}$$

运动恢复结构点用于初始化均值 $\boldsymbol{\mu}_i$，颜色 c_i根据关联图像像素的颜色进行初始化，协方差 $\boldsymbol{\Sigma}_i$ 则初始化为各向同性的 $\sigma^2\mathbf{I}$。

IV. 实验

本节评估了流程的各个阶段，并重点介绍了实现高保真 3D 高斯泼溅的设计选择。

我们使用十辆品牌、型号和底盘状况各异的车辆（例如全新、锈蚀、受损）测试了系统，并用广角相机架进行了录制。每辆车的驶入过程会生成三段视频（左、中、右），分辨率为 1920×1080 ，帧率为每秒 120 帧（帧每秒）。视频时长为 ≈ 8 秒，因此每台相机可获得 ≈ 960 帧。

所有实验均在一台配备 AMD EPYC 7763 CPU、2 TB 随机存取存储器和八块 NVIDIA RTX A6000 GPU 的工作站上进行。

A. 相机标定

校正广角镜头造成的极端畸变，需要精确的相机标定。不准确的相机模型会使误差贯穿整个重建流程，导致点云扭曲、渲染结果失真。我们使用精心筛选的清晰 ChArUco 标定板图像，对 OpenCV 完整相机模型进行了拟合评估。该八参数完整 OpenCV 模型考虑了高阶径向畸变项和切向畸变，取得了较低的均方根投影误差：0.74 像素。

图 4 展示了 OpenCV 完整相机模型如何校正标定图像中的严重畸变。图 5 展示了对真实底盘图像进行相同校正的效果。准确的内参可将运动恢复结构流程中的光束法平差漂移降至最低，从而生成几何精确的点云。

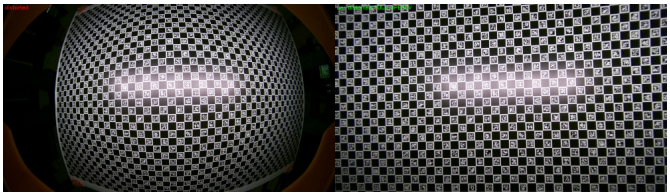


图 4：原始标定图像（左）与使用完整 OpenCV 模型进行去畸变后的图像（右）的定性比较。

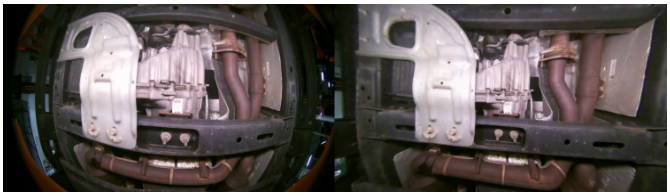


图 5：原始车辆底盘图像（左）与使用完整 OpenCV 模型进行去畸变后的图像（右）的定性比较。

B. 视频同步

即使使用硬件触发，轻微的延迟也可能导致三个视频流之间出现时间错位。我们的两阶段同步算法旨在以亚帧精度校正这些偏移。

三台摄像机的视频最多可能相差 ± 35 帧。同步后，摄像机对之间垂直运动的平均差异从 774 像素降至 ≈ 22 像素。

该算法采用凸损失函数，有助于可靠地找到最佳对齐方式。精确同步可确保三元组选择所需的时空对应关系，从而避免高斯泼溅中出现重影伪影。

C. 运动恢复结构

运动恢复结构阶段为我们的模型提供几何基础。它输出稀疏三维点云和相机位姿，这些结果决定了高斯泼溅质量的上限。为验证我们的方法，我们进行了全面评估，将所提出的流程与两个关键基线进行比较：原始运动恢复结构方法，以及使用经典尺度不变特征变换特征和 COLMAP 匹配器、而非 DISK+LightGlue 的版本（该版本采用了考虑设备结构的流程）。随后，我们对流程进行了一系列消融研究，以展示各项核心贡献的影响。

定量结果详见表 I。不包含我们的标定、同步、自定义匹配策略和基于刚性架的姿态先验的原始基线运动恢复结构（SfM），无法生成连贯的重建结果。我们采用经典尺度不变特征变换（SIFT）特征的刚性架感知流程构建了强有力的基线，显著提升了所有指标。通过集成学习型特征（DISK）和基于注意力的匹配器（LightGlue），我们提出的方法取得了整体最佳表现，生成了最稠密、最准确的稀疏点云。消融研究凸显了最终流程中各个组件的重要性。移除专用的



图 6：车辆底盘的稀疏点云。

相机标定（因此也移除了预去畸变）、自定义匹配策略或基于刚性架的姿态先验，都会导致生成无法表示车辆底盘的退化点云，进而产生退化的高斯泼溅，如图7所示。

表 I：运动恢复结构基线比较与消融研究。每项消融均移除我们运动恢复结构流程中的一个特定组件。

指标	基线 SfM*	刚性架感知 运动恢复结构（只度不变特征变换）	- 相机 标定*	- 视频 Sync	- 自定义 匹配*	- 姿态 先验*	我们的运动恢复结构 DISK+LG
已注册图像 ↑	749	**750**	**750**	**750**	745	673	**750**
稀疏三维点数 ↑	126,496	323,005	189,121	339,209	368,751	313,395	**427,299**
平均轨迹长度 ↑	**10.0868**	8.6509	6.8651	8.1906	**14.0779**	8.0244	**8.6674**
重投影误差（像素）↓	0.7817	0.5376	0.7058	0.4995	0.5279	0.5478	**0.4909**

* 这些运动恢复结构方法生成了退化的稀疏点云，无法表示车辆底盘。因此，这些稀疏点云会产生退化的高斯泼溅。

如图 6所示，我们的最终方法生成了车辆底盘的稠密、一致表示，为高斯泼溅提供了坚实的几何基础。

D. 高斯泼溅

最后，我们评估三维车辆底盘的高斯泼溅结果。为了评估结果，我们采用标准图像质量指标：峰值信噪比（PSNR）、结构相似性指数（SSIM）以及感知指标学习感知图像块相似度（LPIPS）。

如表 II所述，我们的流程取得了良好的结果，平均 PSNR 为 30。66 dB，平均 SSIM 为 0。92，平均 LPIPS 为 0。19，表明我们的渲染视图在视觉和感知上都更接近真实相机图像。这证明，与使用经典 SIFT 特征和匹配器的流程版本相比，我们的方法有了明显改进，表明 DISK+LightGlue 方法生成的更稠密、更准确的点云能够为高斯泼溅优化提供更优的种子。如图 7所示，基线方法和若干消融运动恢复结构初始化会产生退化、噪声较多或不完整的渲染结果，证实我们的流程对于实现高质量三维模型至关重要。

我们的方法实现了实时渲染，平均帧率超过 130 帧每秒（fps），因此适用于



图 7：基线运动恢复结构，以及针对相机标定、自定义匹配和姿态先验的消融实验，均产生了退化、噪声较多或不完整的渲染结果。这些结果直观地印证了表 II中的定量发现，并表明我们提出的流程中每个组件对于生成连贯且高保真的三维模型都至关重要。

交互式检查。图9 展示了我们的模型具有照片级真实感，且几何结构连贯。高质量的运动恢复结构初始化使三维高斯泼溅优化能够收敛为清晰的表示，捕捉螺栓头、锈蚀纹理、液体痕迹以及排气系统金属光泽等精细细节（图8）。值得注意的是，这些模型基本没有“漂浮物”伪影和朦胧区域，而这些问题常见于使用精度较低的相机位姿训练的辐射场。

在 RTX A6000 GPU 上训练耗时 $\approx 8-10$ 分钟。高效的训练和高保真的三维模型使我们能够从不同角度查看车辆底盘，识别腐蚀、泄漏或损伤等潜在问题，从而使其适用于生产环境。

表 II：高斯泼溅基线比较与消融研究。每项消融实验都会从我们的运动恢复结构流程中移除一个特定组件。

指标	基线 SfM*	刚性架感知 运动恢复结构（只度不变特征变换）	- 相机 标定*	- 视频 Sync	- 自定义 匹配*	- 姿态 先验*	我们的运动恢复结构 DISK+LG
峰值信噪比（dB）↑	23.46	29.68	21.40	29.47	20.78	16.96	**30.66**
SSIM ↑	0.78	0.91	0.75	0.89	0.76	0.68	**0.92**
LPIPS ↓	0.42	0.21	0.57	0.26	0.43	0.66	**0.19**

五、结论

本文提出了一套完整的流程，利用多视角视频生成高质量、交互式三维车辆底盘模型。我们的系统旨在应对车辆底部检查中的实际挑战，而这类检查通常十分繁琐、耗费体力，



图8：高斯泼溅渲染展示了我们的流程捕捉精细纹理和诊断细节的能力。渲染图显示车架上有明显锈蚀。这种细节水平使检验员能够准确评估车辆状况，并增强买家在在线市场上的信心。

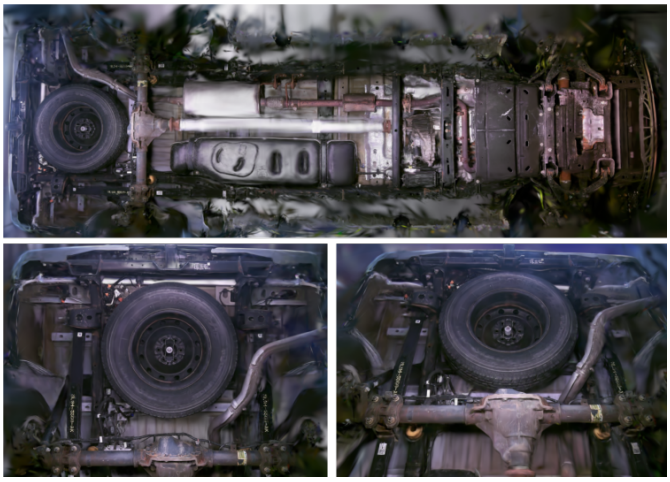


图9：我们的流程生成的照片级真实感3D底盘模型。上图展示了车辆底盘的全方位俯视图，下方图像则从不同角度展示了车辆后部。

并且对于在线交易平台而言尤其难以规模化。通过将原始视频转换为细致、具有照片级真实感的三维模型，我们的方法能够实现更快速、更安全且更透明的车辆评估。

我们方法的核心是支架感知运动恢复结构流程，专为支持3D高斯泼溅而构建。我们发现，在低视差和线性运动等具有挑战性的场景中，准确处理广角镜头畸变并利用相机支架的物理先验，是恢复可靠几何结构的关键。这一坚实的几何基础使我们能够生成清晰、无伪影的辐射场。

我们的实验表明，该流程表现良好，能够生成高质量的渲染结果（峰值信噪比 > 30，SSIM > 0.90）实时速度可达（> 130 帧每秒）。生成的三维模型能够呈现锈蚀、泄漏和表面磨损等诊断特征，使该工具对车辆检查员和消费者都很有价值。我们认为，这项工作为未来开展自动化3D损伤检测研究奠定了基础，并有望应用于其他工业检测场景。

参考文献

[1]S. Chauhan 和 L. J. Katare，“二手车市场规模、份额、竞争格局及趋势分析报告：按车辆类型、燃料类型和分销渠道划分：全球机遇分析与行业预测，2023–2033年”，Allied Market Research，技术报告 A06429，2月。2025, [在线；访问日期：19-Jul-2025]。[在线]。获取地址：https://www.alliedmarketresearch.com/used-cars-market-A06429

[2] B. Mildenhall、P. P. Srinivasan、M. Tancik、J. T. Barron、R. Ramamoorthi 和 R. Ng，“NeRF：将场景表示为用于视图合成的神经辐射场”，《ACM通讯》，第65卷，2021年，第1期，第99–106页。[3] B. Kerbl、G. Kopanas、T. Leimk¨uhler 和 G. Drettakis，“用于实时辐射场渲染的3D高斯泼溅”。《ACM图形学汇刊》，第42卷，第4期，第139页，2023年。[4] M. Tyszkiewicz、P. Fua 和 E. Trulls，“DISK：使用策略梯度学习局部特征”，《神经信息处理系统进展》，第33卷，第14 254–14 265页，2020年。[5] P. Lindenberger、P.-E. Sarlin 和 M. Pollefeys，“LightGlue：以光速进行局部特征匹配”，载于 IEEE/CVF国际计算机视觉大会论文集，2023年，第17 627–17 638页。[6] J. L. Sch¨onberger 和 J.-M. Frahm，“重新审视运动恢复结构”，载于计算机视觉与模式识别会议（CVPR），2016年。[7] L. Pan、D. Bar´ath、M. Pollefeys 和 J. L. Sch¨onberger，“重新审视全局运动恢复结构”，载于欧洲计算机视觉会议。施普林格，2024年，第58–77页。[8] N. Snavely、S. M. Seitz 和 R. Szeliski，“照片旅游：探索3D照片集”，载于 ACM SIGGRAPH 2006论文集，2006年，第835–846页。[9] C. Mei 和 P. Rives，“基于平面网格的单视点全向相机标定”，载于2007年IEEE国际机器人与自动化会议论文集。IEEE，2007年，第3945–3950页。[10] R. Hartley 和 A. Zisserman，《计算机视觉中的多视图几何》。剑桥大学出版社，2003年。[11] D. G. Lowe，“基于尺度不变关键点的独特图像特征”，《国际计算机视觉期刊》，第60卷，2004年，第2期，第91–110页。[12] J. Sun、Z. Shen、Y. Wang、H. Bao 和 X. Zhou，“LoFTR：无需检测器、基于Transformer的局部特征匹配”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021年，第8922–8931页。[13] P.-E. Sarlin、D. DeTone、T. Malisiewicz 和 A. Rabinovich，“SuperGlue：利用图神经网络学习特征匹配”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2020年，第4938–4947页。[14] M. Goesele、B. Curless 和 S. M. Seitz，“重新审视多视图立体视觉”，载于2006年IEEE计算机学会计算机视觉与模式识别会议（CVPR’06），第2卷。IEEE，2006年，第2402–2409页。[15] T. M¨uller、A. Evans、C. Schied 和 A. Keller，“利用多分辨率哈希编码实现即时神经图形基元”，《ACM图形学汇刊》（TOG），第41卷，第4期，第1–15页，2022年。[16] B. Fei、J. Xu、R. Zhang、Q. Zhou、W. Yang 和 Y. He，“作为新时代技术的3D高斯泼溅：综述”，《IEEE可视化与计算机图形学汇刊》，2024年。[17] G. H. An、S. Lee、M.-W. Seo、K. Yun、W.-S. Cheong 和 S.-J. Kang，“基于ChArUco标定板的全向相机标定方法”，《电子学》，第7卷，第12期，421页，2018年。[18] OpenCV开发团队，“ChArUco标定板的检测”，在线。[在线]。获取地址：https://docs.opencv.org/3.4/df/d4a/tutorial charuco detection.html[19] A. Ranganathan，“Levenberg–Marquardt算法”，LM算法教程，第11卷，第1期，第101–110页，2004年。[20] R. Schnabel、R. Wahl 和 R. Klein，“用于点云形状检测的高效随机采样一致性算法”，载于《计算机图形学论坛》，第26卷，第2期。Wiley在线图书馆，2007年，第214–226页。