STANFORD ROBOTICS

Leopard 双目深度软件

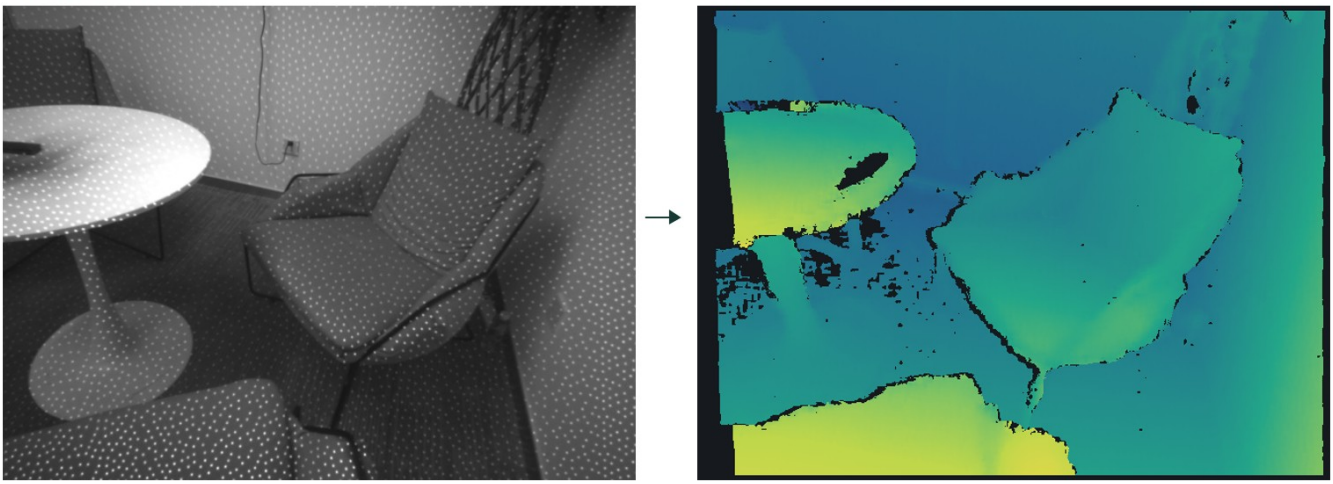
客户评估与源代码授权报告

2026 年 11 月

Leopard 为机器人与嵌入式视觉团队提供可配置的双目匹配引擎，便于查看实现、集成和调优。我们的实现结合了倾斜支持窗口匹配、CPU 与 GPU 后端、亮度校正以及明确的有效性控制。输入须为经过立体校正的双目图像，无需预训练神经网络。

左侧双目输入图像

Leopard 输出视差图



项目样例 frame0408。启用局部亮度校正；显示范围为 0-40 px（超出部分截断），搜索上限为 256 px。该示例没有真实深度参考值。[2]

为什么值得评估 Leopard

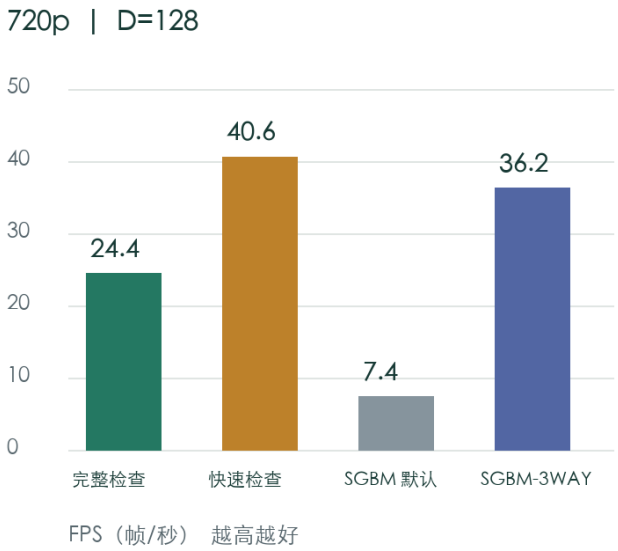
按应用选择速度与质量的平衡。完整反向匹配提供更强的一致性检查；当应用可以接受不同的像素剔除行为时，可选择更快的投影一致性检查来降低延迟。

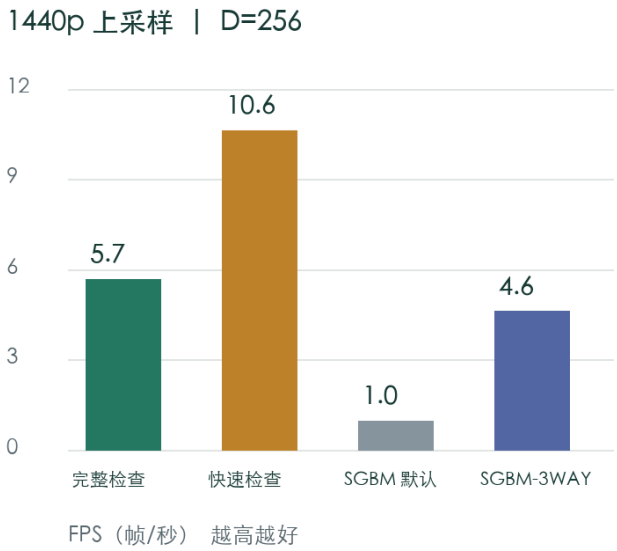
掌握深度处理流程的控制权。源代码访问支持针对具体应用调整阈值、集成相机和开展诊断。无效测量会被明确标记；可选的补洞功能会单独标识推断得到的像素。

基于实测验证开展集成。在最新 CPU 实验中，快速 FP32 配置在 720p 下达到 40.6 FPS；另一项独立的 Apple Metal 测试在 1248 × 688 分辨率、启用亮度校正时达到 67.8 FPS。这些数字对应不同的配置与测试任务，后文分别说明。[1,3]

CPU 性能与规模扩展

当分辨率和视差搜索范围使穷举匹配的计算成本升高时，Leopard 尤其值得评估。在图像块、窗口和迭代次数固定时，其计算量近似随像素数线性增长，无需构建稠密的“像素 × 视差”代价体。相对于 SGBM 的性能交叉点取决于输入与硬件。





同一台 Apple M3 Pro，除注明外为 12 线程预算，OpenCV 4.13.0。“默认”仅指五方向 MODE_SGBM，其他参数与 3WAY 相同。Leopard 使用 FP32；快速模式采用较弱的投影检查。四项配置在本轮一并重测。[1]

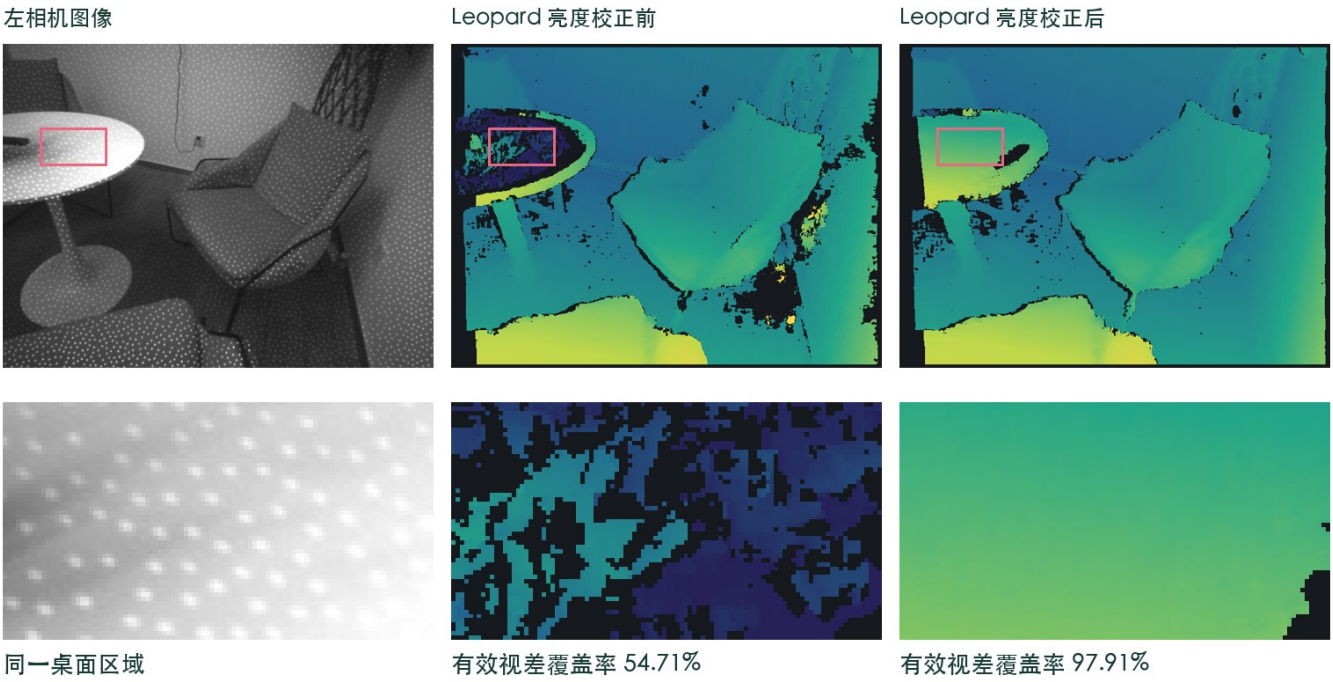
办公室样例	D	Leopard 完整 FPS	Leopard 快速 FPS	SGBM 默认 FPS	SGBM 3WAY FPS
640 × 360	64	86.85	154.48	52.02	230.77
1280 × 720	128	24.41	40.56	7.37	36.25
2560 × 1440 *	256	5.67	10.62	0.97	4.61
1280 × 720 单线程	128	4.02	7.62	7.38	10.16

在 720p 下，Leopard 完整与快速模式的吞吐量分别为默认 SGBM 的 3.31 倍和 5.51 倍。对比更快的 SGBM-3WAY，快速 Leopard 在 720p 和上采样 1440p 下分别达到其 1.12 倍和 2.30 倍。360p 下的 SGBM-3WAY 仍更快；720p 下也快于完整检查的 Leopard。选用哪种基准模式会显著影响加速比。单线程行显示了并行执行的重要性；相同线程预算并不代表实际利用相同数量的核心。

*1440p 由 720p 上采样，仅用于计时；分辨率与视差范围 D 同时增加。两个 OpenCV 模式均为 blockSize=5、P1/P2=200/800、uniqueness=10、preFilterCap=31、disp12MaxDiff=1，并关闭散斑过滤。三轮随机顺序测试，每轮七帧，均经预热；含逐帧准备、匹配、过滤与输出，不含初始化、解码、缩放、采集、立体校正与显示。其他桌面应用仍在运行。[1]

恢复高亮表面的视差细节

曝光和局部亮度差异可能使左右对应区域的外观不同，即使其纹理仍然可见。Leopard 支持在匹配前启用局部亮度归一化。在此高亮桌面示例中，固定桌面区域的有效视差覆盖率从 54.71% 提升至 97.91%，且未使用补洞。[2]



各面板标出的区域完全相同。覆盖率表示通过有效性检查的视差像素比例，并非基于真实参考值计算的准确率。该示例没有真实深度参考值。[2]

将预处理保留在 GPU 上

Metal 实现在 Apple GPU 上先执行所选的高斯亮度校正，再进行匹配。在另一项包含 27 组实拍数据、分辨率为 1248 × 688 的评估中，将相同的校正从 CPU 移到 GPU 后，处理帧率中位数从 42.8 提升至 67.8 FPS，加速比为 1.59 倍。[3]

1248 × 688 下的 Metal 处理流程	帧率中位数 FPS
CPU 亮度校正 + Metal 匹配	42.8
Metal 亮度校正 + Metal 匹配	67.8
Metal 亮度校正、匹配及补洞	62.8

补洞是可选的保守处理：保留已接受的测量值，并单独记录推断得到的像素。亮度归一化和补洞可以提高有效覆盖率，但也可能引入更多异常值。应结合应用自身的图像和误差容限选择参数。

GPU 数字包含主机端准备、数据传输、求解和同步，表示对图像重复处理的帧率，不是实时相机采集与输出的帧率，也不是 CPU 与 SGBM 的比较。每组数据在预热后按交替顺序重复 15 次；汇总值为各组帧率的中位数。[3]

精度与有效覆盖率

实用的深度引擎需要兼顾测量误差与场景覆盖程度，因此我们同时报告两者。下表汇总了 Skoltech3D 的 27 组实拍数据：三个物体、三个视角和三种照明条件；输入为 RealSense D435 红外双目图像，真实参考值来自独立扫描仪。[1,5]

方法	MAE 像素	RMSE 像素	有效 覆盖率	有效且误差 ≤1 像素 占全部 GT 的比例
Leopard FP32 完整	0.402	1.022	85.0%	80.0%
Leopard FP32 快速	0.455	1.356	87.3%	81.0%
OpenCV SGBM 默认	0.502	1.469	98.3%	90.3%
OpenCV SGBM 3WAY	0.435	1.330	97.8%	91.1%

MAE：平均绝对误差；RMSE：均方根误差；GT：真实参考值。表中误差单位为视差像素，不是毫米。

在各自接受的像素上，Leopard 完整检查的平均绝对误差和均方根误差均低于两个 OpenCV 模式。两个 SGBM 模式返回更多有效像素，且全部评估像素中误差不超过一个像素的有效结果比例更高。本数据集上，默认 SGBM 的覆盖率略高于 3WAY，但误差更大、正确像素占比更低。仅比较已接受像素的误差，不等于同覆盖率比较，也不能证明整体精度更优。

如何选择配置

如果应用宁愿保留空缺，也不希望接受不确定的测量，可从 FP32 与完整反向一致性检查开始。如果吞吐量优先，且应用已有自身的校验或后续过滤机制，可以比较快速投影检查。无论选择哪种配置，都应同时评估误差与覆盖率。

16 位定点路径仍为实验选项。在我们的测试中，FP32 的速度收益更稳定，单个像素的变化也更小。双精度采样参考模式仍然保留，便于复现与对照。

支持产品集成的验证

CPU 验证包括 24 组已知视差的合成对照数据，使用三个随机种子，覆盖整数与亚像素位移、倾斜表面、噪声、深度阶跃和稀疏光点。在这些对照数据上，FP32 的合并有效像素 MAE 为 0.01597 像素，覆盖率为 99.619%。合成测试用于验证数值行为，不能据此推断真实场景的深度精度。[1]

11 项回归测试、标量与 SIMD 内核对照、内存安全检查及线程确定性检查均已通过。保持不变的参考模式在 24 组合成对照与 27 组实拍数据上，均与已保存结果逐位一致。实验精度模式可能改变个别像素的获胜候选及有效性掩码。[1]

真实参考值仅覆盖部分物体表面，距离约为 0.63-0.88 米。这些数据是三个场景的重复拍摄，并非 27 个独立场景，也不是 4 米测试。MAE 与 RMSE 分别在各方法接受的像素上计算；覆盖率和误差不超过一个像素的有效结果比例以固定评估区域为分母。上表 CPU 比较未启用补洞或亮度归一化。[1,5]

集成到您的深度处理流程

Leopard 源代码授权的价值在于提供可调整的实现及配套评估工具。您的团队可以查看匹配阶段、选择目标后端，并将视差与有效性输出接入感知系统的其他模块。商业授权范围可围绕实际需要的后端与平台确定。

实现	现有能力	实测平台
C++17 CPU 与 Python 适配器	复用缓冲区与工作线程；支持参考、FP32 和 fixed16 选项	Apple M3 Pro CPU
Apple Metal	GPU 匹配、局部亮度归一化及可选补洞	Apple 芯片 macOS
CUDA 与 Colab 笔记本	独立 GPU 移植及可复现实例	NVIDIA Tesla T4
Python 与 Numba 参考实现	便于阅读的参考代码、诊断与评估工具	桌面 CPU

各后端的封装与性能特征有所不同，应针对所选版本核验功能和数值一致性。报告中的 CPU 计时不代表 CUDA 或 Metal 性能。其他处理器需单独开展目标平台评估。[1,3,6]

明确的集成接口

输入：尺寸一致、经过立体校正的左右灰度图像，以及适合该相机的视差搜索范围。标定与立体校正须在匹配前完成；普通单摄像头视频无法满足这一输入要求。

输出：以像素为单位的浮点视差、明确的有效性掩码，以及可选的匹配代价和平面诊断数据。无效视差与合法的零视差分别表示。可选补洞会区分实测、推断和仍未解决的像素。

物理深度：使用标定后的焦距与双目基线，将有效的正视差换算为深度。实际测距能力取决于镜头、基线、分辨率、纹理和曝光，不能脱离具体相机给出统一的毫米级精度保证。

为什么选择源代码授权

团队可以针对复杂光照和像素剔除策略进行调整，而不必只接受不可解释的深度输出。通过复用预先创建的处理对象完成逐帧处理，查看各阶段耗时，并使用保存的配置与原始计时记录复现对照结果。

源代码集成也允许您的团队针对自身产品调整内存管理和平台接口。已有测试与参考输出可作为验证这些改动的起点。具体交付范围、集成工作及支持条款应结合目标应用约定。

在您的相机上评估 Leopard

下一步是在具有代表性的客户图像上开展实测。我们建议在相同输入与硬件上，对照 Leopard 的不同配置和您当前的双目匹配器，再根据结果确定源代码授权范围。

明确应用需求

提供经过标定的双目图像、图像尺寸、所需工作距离、目标处理器及延迟预算。样本应覆盖运动、低纹理、强反光、曝光不一致和深度边界等困难情况，并在条件允许时提供真实参考值。

约定验收指标

测量端到端延迟及其波动、已接受像素的误差、无效像素比例、正确像素占比和时间稳定性。选择 CPU 或 GPU 后端，对比完整检查与投影检查，并将归一化、补洞作为独立选项测试。同时明确相机采集与后续处理的时间开销。

确定商业合作范围

围绕选定的后端、目标平台和部署需求讨论源代码授权。在约定商业条款前，明确集成交付物、文档、支持内容和验收测试。价格与授权条款可通过直接沟通确定。

Stanford Robotics 联系人 Leo Pang

leo.pang@gmail.com

测试依据与来源

[1] Stanford Robotics，CPU 同条件对比，2026 年 11 月 6 日。Apple M3 Pro；OpenCV 4.13.0 默认 MODE_SGBM 与 SGBM-3WAY，除模式外的参数和线程预算一致。四项配置一并重测，每项 21 次预热后计时及 27 组原生分辨率 GT 数据。24 组合成对照和回归验证沿用原混合精度评估；保留原始计时与配置记录。

[2] Stanford Robotics，frame0408 高亮桌面评估。项目双目图像为 640 × 480，视差搜索上限为 256；固定 ROI 为 x=55:155、y=125:180。原始与 Metal 归一化视差图使用同一显示尺度。覆盖率比较未使用补洞，也没有真实深度参考值。

[3] Stanford Robotics，Metal 高斯亮度归一化评估。Apple M3 Pro；27 组数据，分辨率 1248 × 688，视差上限 128。CPU 与 GPU 预处理比较使用相同归一化方法和匹配阈值。覆盖率提高可能伴随均方根误差增大；该实验旨在加速所选校正方法，而非消除这一取舍。

[4] Tankovich 等，SOS: Stereo Matching in O(1) with Slanted Support Windows，IROS 2018。Leopard 为独立实现，并增加了工程扩展，不是 Google 的源代码。本报告未将论文中的 GPU 性能用作 Leopard 产品性能声明。

Google Research 上的 SOS 论文介绍

[5] Voynov 等，Multi-sensor large-scale dataset for multi-view 3D reconstruction，CVPR 2023。Skoltech3D 为实拍评估提供 RealSense 双目图像与独立扫描仪参考值。本评估使用近距离子集，并非对 RealSense 出厂精度指标的验证。

Skoltech3D 数据集与论文

[6] Stanford Robotics，Leopard CUDA 基准测试及自包含 Colab 笔记本。独立 CUDA 移植已在 NVIDIA Tesla T4 上测试，其硬件、封装及计时范围与本报告中的 CPU、Metal 对照不同。