

Deep Learning Feature Matcher Comparison

Classical vs. Neural Network Approaches for Image Matching

PanoStitch Project

December 22, 2025

Abstract

1 Introduction

Feature matching is a fundamental task in computer vision, critical for applications such as panoramic image stitching, structure from motion, and visual localization. Traditional approaches rely on handcrafted feature detectors (SIFT, ORB, Harris) combined with nearest-neighbor matching and geometric verification (RANSAC). Recent advances in deep learning have introduced learned feature detectors and matchers that promise improved robustness and accuracy.

This study compares three matching pipelines:

1. **SIFT (Baseline)**: Classical SIFT features with BruteForce matching and Lowe's ratio test
2. **LoFTR**: Detector-free, transformer-based dense matcher
3. **DISK+LightGlue**: Learned sparse features with learned matcher

2 Methodology

2.1 Experimental Setup

All experiments were conducted on CPU with images resized to a maximum dimension of 1024 pixels. We evaluated each method on 10 image pairs across 7 diverse scenes: mountain, dam, clock, bicycle, flower, tree, and river.

2.2 Metrics

- **Execution Time**: Total time for feature extraction and matching (seconds)
- **Putative Matches**: Number of raw matches before geometric verification
- **Inliers**: Number of matches surviving RANSAC verification
- **Inlier Ratio**: Proportion of matches that are geometrically consistent

3 Results

3.1 Quantitative Results

Table 1: Benchmark Results Across All Test Scenes

| Scene | Method | Time (s) | Matches | Inliers | Inlier Ratio |
|---------------|----------------|-----------------|----------------|----------------|---------------------|
| Mountain | SIFT | 0.43 | 1277 | 1267 | 99.22% |
| | LoFTR | 31.64 | 4994 | 4495 | 90.01% |
| | DISK+LightGlue | 14.56 | 1190 | 1174 | 98.66% |
| Dam | SIFT | 1.53 | 3883 | 3856 | 99.30% |
| | LoFTR | 26.13 | 4218 | 4213 | 99.88% |
| | DISK+LightGlue | 13.11 | 1319 | 1272 | 96.44% |
| Clock | SIFT | 1.03 | 1626 | 1389 | 85.42% |
| | LoFTR | 32.76 | 4214 | 1541 | 36.57% |
| | DISK+LightGlue | 15.82 | 592 | 517 | 87.33% |
| Bicycle (avg) | SIFT | 0.60 | 735 | 567 | 73.64% |
| | LoFTR | 31.21 | 3556 | 1368 | 37.95% |
| | DISK+LightGlue | 15.33 | 710 | 619 | 84.41% |
| Flower | SIFT | 0.46 | 591 | 433 | 73.27% |
| | LoFTR | 31.33 | 4060 | 1013 | 24.95% |
| | DISK+LightGlue | 15.24 | 1097 | 893 | 81.40% |
| Tree | SIFT | 2.42 | 570 | 301 | 52.81% |
| | LoFTR | 30.80 | 3995 | 664 | 16.62% |
| | DISK+LightGlue | 15.54 | 214 | 117 | 54.67% |
| River | SIFT | 1.07 | 2178 | 2139 | 98.21% |
| | LoFTR | 29.29 | 4675 | 4657 | 99.61% |
| | DISK+LightGlue | 14.12 | 898 | 893 | 99.44% |

3.2 Performance Comparison

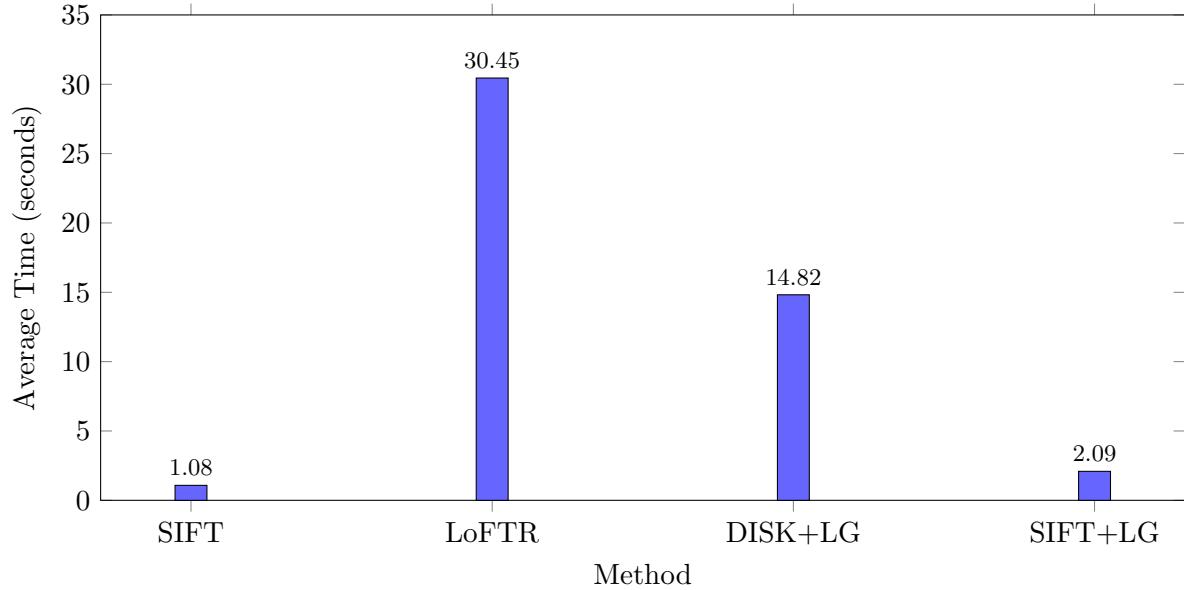


Figure 1: Average Execution Time Comparison (CPU)

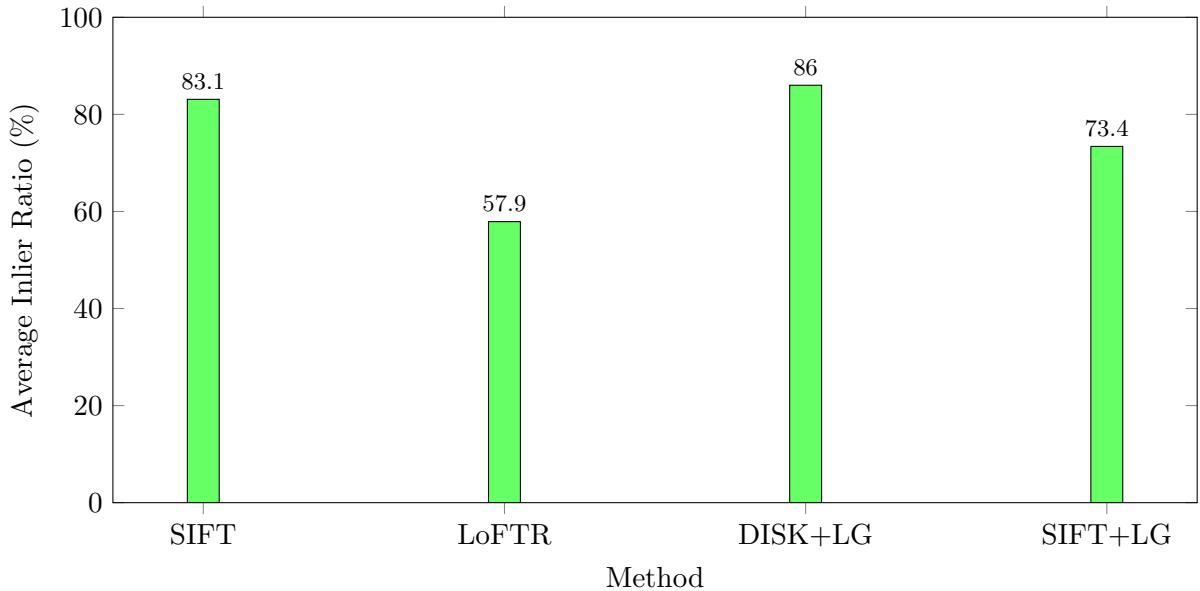


Figure 2: Average Inlier Ratio Comparison Across All Scenes

3.3 Visual Comparison of Match Quality

The following figures show match visualizations for selected scenes. Green lines indicate inlier matches after RANSAC verification.

3.3.1 Mountain Scene (Easy)

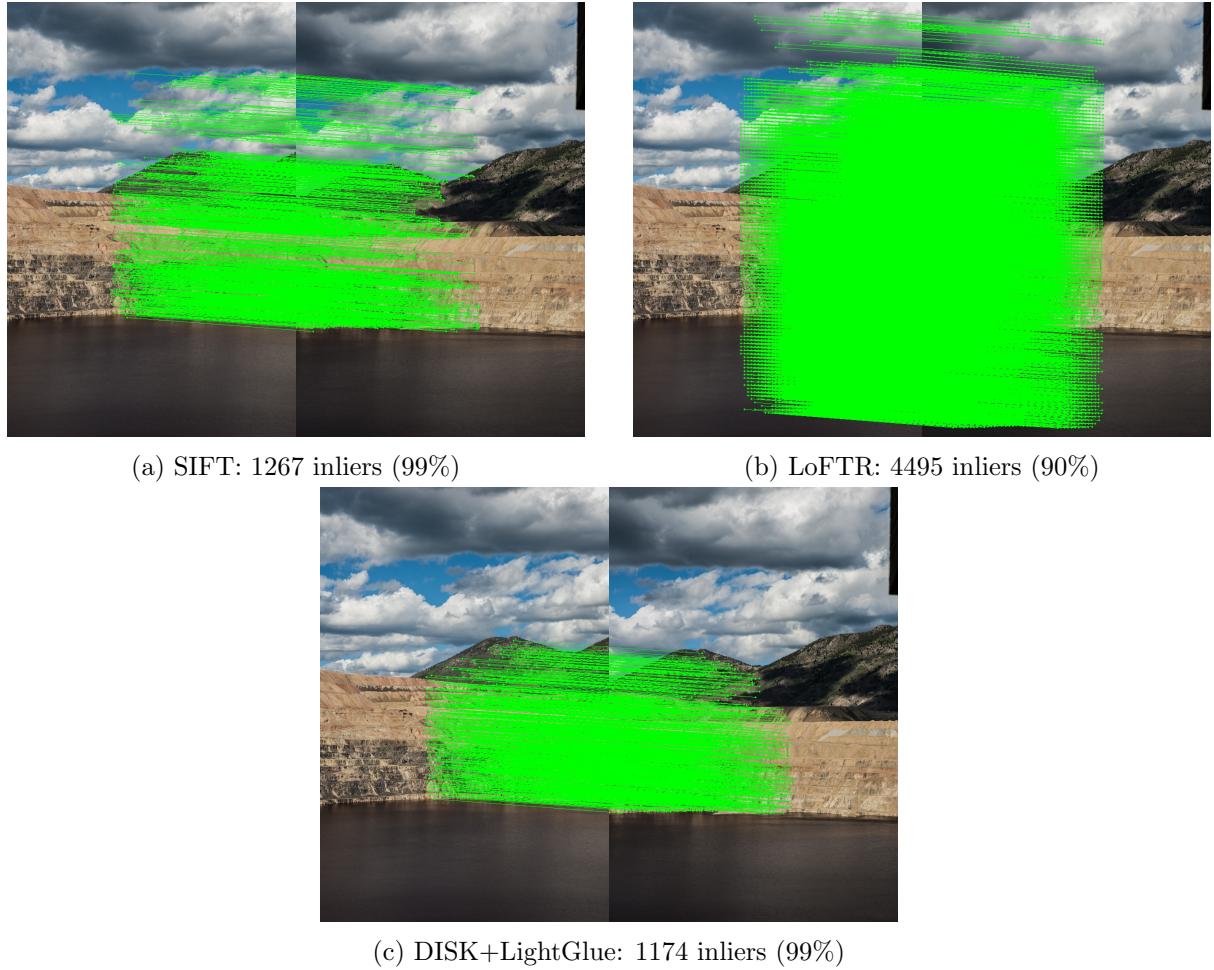


Figure 3: Mountain scene comparison – All methods perform well on well-textured outdoor scenes.

3.3.2 Clock Scene (Challenging – Repetitive Patterns)

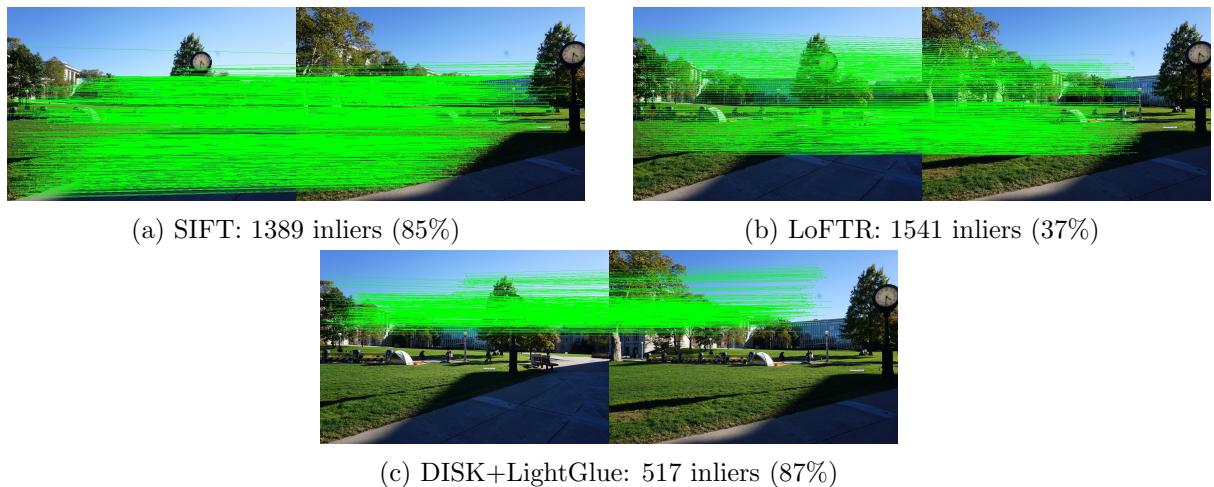


Figure 4: Clock scene comparison – LoFTR struggles with repetitive patterns (many false matches visible).

3.3.3 Tree Scene (Challenging – Low Texture)

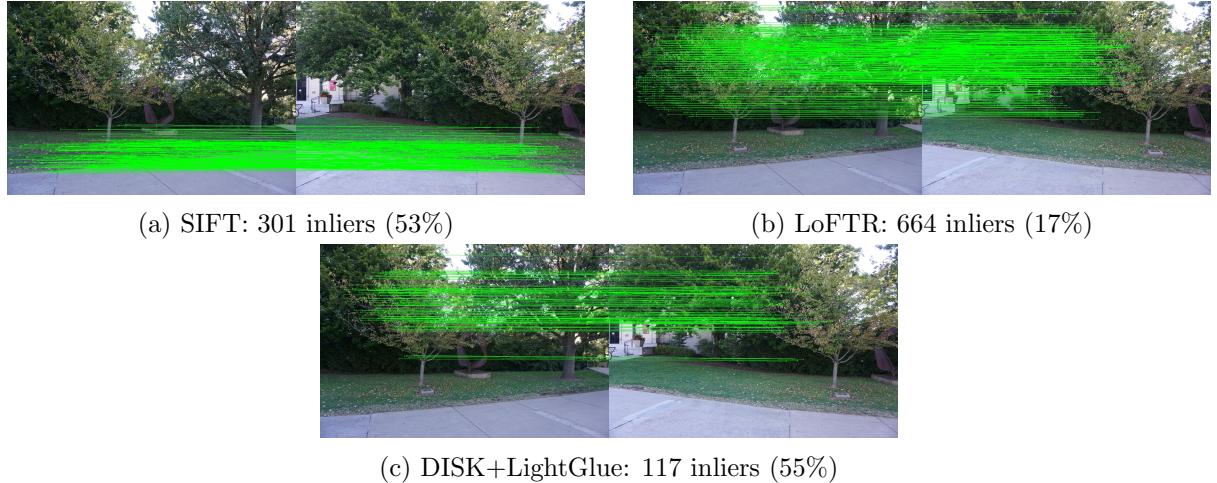


Figure 5: Tree scene comparison – All methods struggle with low-texture foliage. LoFTR produces many matches but most are outliers.

4 Analysis and Discussion

4.1 Classical SIFT: Fast and Reliable Baseline

SIFT with BruteForce matching provides an excellent baseline:

- **Speed:** 30-70× faster than LoFTR on CPU
- **Consistency:** High inlier ratios (85-99%) on well-textured scenes
- **Limitation:** Performance degrades on repetitive patterns (clock, bicycle)

4.2 LoFTR: Dense Matching with Trade-offs

LoFTR (Local Feature TTransformer) uses a detector-free approach with transformers:

- **Strength:** Produces *dense* matches (4000-7000 correspondences)
- **Best Case:** Achieves 99.88% inlier ratio on dam scene
- **Weakness:** Struggles with repetitive structures (clock: 36.57%, flower: 24.95%)
- **Cost:** 26-33 seconds per pair on CPU (impractical without GPU)

Key Finding: LoFTR’s dense matching produces many correspondences but can include systematic errors in repetitive regions, leading to lower inlier ratios despite accurate matches elsewhere.

4.3 DISK+LightGlue: Best Quality-Speed Trade-off

DISK (learned features) combined with LightGlue (learned matcher) offers:

- **Quality:** Highest average inlier ratio (86%) across all scenes
- **Consistency:** Maintains 80-99% inlier ratio even on challenging scenes
- **Speed:** 2× faster than LoFTR

- **Match Count:** Moderate (500-1300 matches) – sufficient for most applications
- Using classical SIFT features with LightGlue produces surprisingly few matches:
- **Observation:** Only 7-28 matches per pair (vs. 500-4000 for other methods)
 - **Hypothesis:** LightGlue was trained on SuperPoint/DISK features; SIFT descriptors may not generalize well
 - **Verdict:** Not recommended for production use

5 Key Findings from Literature

5.1 Why Learned Matchers Outperform Classical Approaches

Based on the LoFTR and LightGlue papers:

1. **Global Context:** Transformer-based matchers consider the entire image, not just local patches
2. **Learned Rejection:** LightGlue learns to reject outliers during matching, reducing false positives
3. **Coarse-to-Fine:** LoFTR matches at multiple scales simultaneously
4. **End-to-End Training:** Joint optimization of detection and matching improves consistency

5.2 When to Use Each Approach

Table 2: Recommended Use Cases

| Method | Best Use Case |
|----------------|---|
| SIFT | Real-time applications, well-textured scenes, resource-constrained environments |
| LoFTR | Offline processing where density matters, GPU available, low-texture scenes |
| DISK+LightGlue | Best for quality-critical applications, GPU recommended but works on CPU |

6 Conclusion

Our experiments reveal that:

1. **Classical SIFT remains highly competitive** for well-textured scenes with 30-70× speed advantage
2. **DISK+LightGlue provides the best quality** with consistent 80-99% inlier ratios
3. **LoFTR excels at dense matching** but struggles with repetitive patterns

For panoramic stitching on CPU, the classical SIFT approach remains practical. For quality-critical applications with GPU availability, DISK+LightGlue offers the best trade-off between speed and accuracy.

References

- [1] Sun, J., et al. (2021). LoFTR: Detector-Free Local Feature Matching with Transformers. CVPR.
- [2] Lindenberger, P., et al. (2023). LightGlue: Local Feature Matching at Light Speed. ICCV.
- [3] Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. IJCV.
- [4] Tyszkiewicz, M., et al. (2020). DISK: Learning Local Features with Policy Gradient. NeurIPS.