Image Matching Challenge 2025: A Computer Vision Semester Project

Ibtesam Hussain, Shaheer Uddin, Safey Ahmed, Sajjad Ali
Department of Artificial Intelligence
FAST National University of Computer and Emerging Sciences
22K4125@nu.edu.pk, 22K8719@nu.edu.pk, 22K4039@nu.edu.pk, 22K8729@nu.edu.pk

December 10, 2025

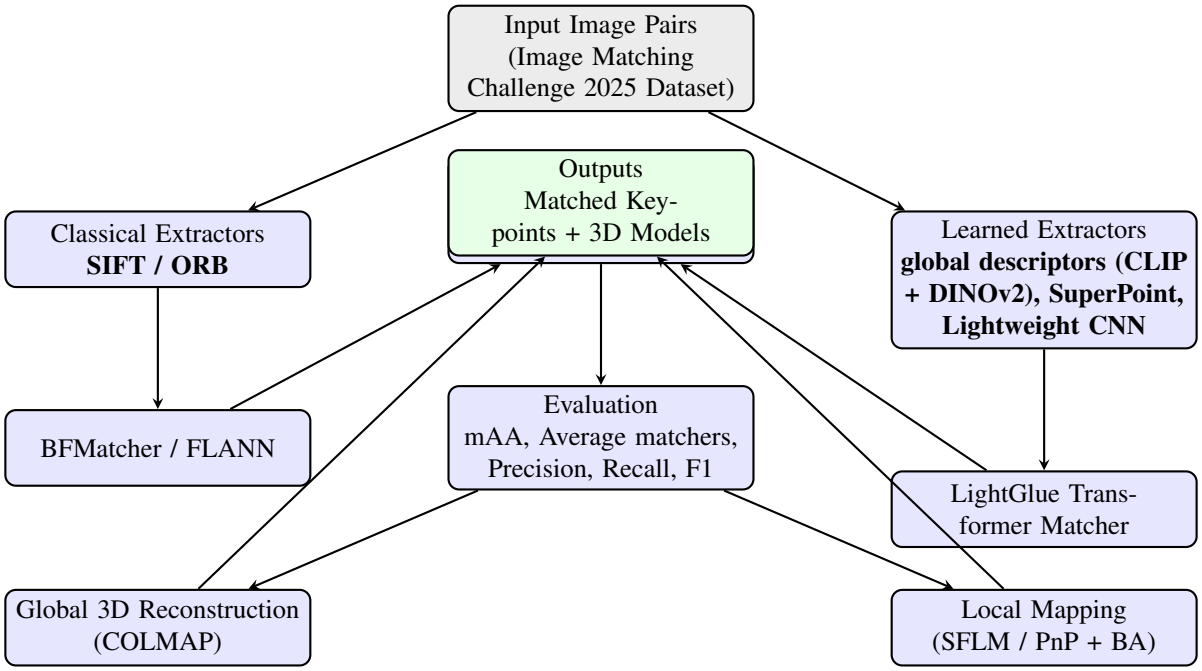

Figure 1: Overall architecture combining classical (SIFT/ORB) and transformer-based (LightGlue/SuperPoint) feature matchers with geometric verification, evaluation, and reconstruction through COLMAP and SFLM modules.

Abstract

Image matching is a fundamental task in computer vision, underpinning applications such as image registration, 3D reconstruction, and SLAM. This paper presents a comprehensive exploration of classical and transformer-based Image matching architectures. The inspiration arose from the latest **Image Matching Challenge 2025** organized by Czech Technical University in Prague. We begin with traditional descriptors like ORB and SIFT integrated with Brute-Force and FLANN matchers. Subsequently, we extend our analysis to learned feature extractors using the LightGlue framework (SuperPoint + Transformer-based matching). We also integrate these pipelines with Structure-from-Motion (COLMAP) and a custom Sparse Feature-based Local Mapping (SFLM) architecture designed for efficient, scalable 3D scene reconstruction. Moreover we also explored the solutions of top 10 scorers in this challenged, in which MAST3R and DUST3R architectures surpassed everything.

1 Introduction

Image matching plays a critical role in modern computer vision pipelines, forming the backbone of 3D reconstruction, visual localization, and augmented reality systems. The goal is to identify reliable keypoint correspondences between image pairs despite variations in viewpoint, illumination, and scale. In this project, we investigated both traditional and modern approaches, building a unified framework for feature extraction, matching, and evaluation.

Our contributions are threefold:

- A classical pipeline utilizing ORB and SIFT with BFMatcher and FLANN for local descriptor comparison.
- A transformer-based LightGlue framework combining

SuperPoint features and learned feature correspondence.

- A custom implementation (our architecture) integrating robust learned descriptors with local mapping and pose refinement. Uses global descriptors (CLIP + DINOv2), geometric verification (LightGlue), and COLMAP/pycolmap for Structure-from-Motion.

2 Related Work

Classical descriptors such as SIFT [1] and ORB [2] have long dominated feature matching due to their robustness and interpretability. However, learned approaches like SuperPoint [3] and SuperGlue/LightGlue [4, 5] have recently demonstrated superior performance under real-world conditions. Our work builds upon these methods, integrating LightGlue with both classical baselines and a novel architecture to compare accuracy, precision, and feature density.

3 Methodology

3.1 Classical Matching Pipeline

We began by applying OpenCV-based pipelines using ORB and SIFT feature extractors. Feature correspondence was computed using the Brute-Force matcher (for ORB) and FLANN matcher (for SIFT). Keypoint matches were filtered using distance ratio tests and homography-based geometric verification.

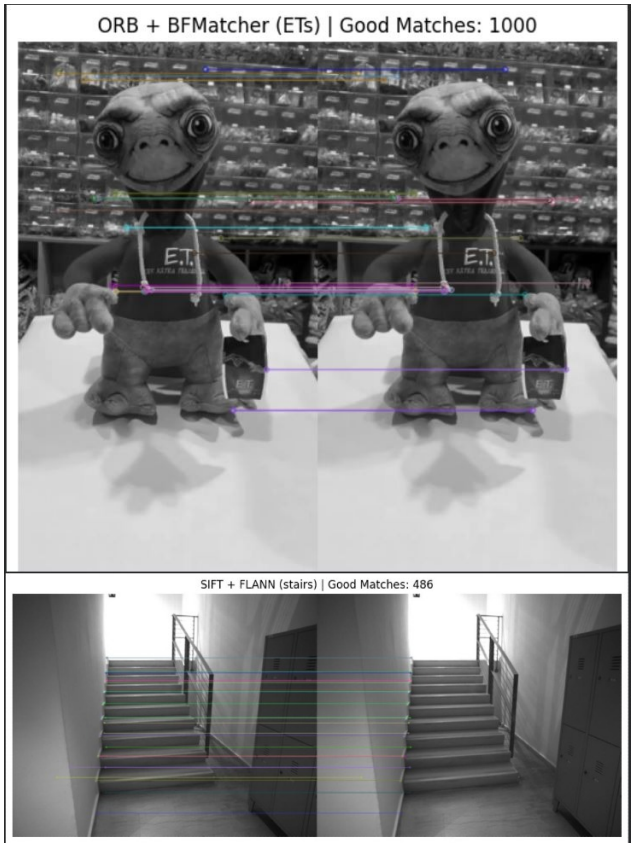

Figure 2: Classical image matching pipeline using ORB/SIFT and FLANN.

3.2 LightGlue Framework

LightGlue is a transformer-based feature matching model that builds upon the SuperPoint feature extractor. We implemented a fully automated pipeline using Kaggle GPUs for training and evaluation. Each image pair was processed as follows:

1. Keypoint extraction via SuperPoint.
2. Feature matching through LightGlue transformer layers.
3. Evaluation of correspondence counts, inlier ratios, precision, recall, and F1-score.

3.3 CLIP + COLMAP Architecture

Lightweight pipeline to cluster images into scenes, reject outliers, and reconstruct camera poses. Uses global descriptors (CLIP + DINOv2), geometric verification (LightGlue), and COLMAP/pycolmap for Structure-from-Motion. This section highlights your our model's structure, including:

- Feature extraction backbone (CLIP + DINOv2, geometric verification (LightGlue))
- Matching head (descriptor correlation or cluster-based association)
- COLMAP/pycolmap for Structure-from-Motion.
- Evaluation setup includes F1 score, Recall and mAA

4 3D Reconstruction Integration

4.1 COLMAP

We integrated the best-performing feature pairs into COLMAP's Structure-from-Motion pipeline. Feature extraction, matching, and geometric verification were automated through COLMAP's CLI tools. This produced sparse and dense reconstructions with intrinsic and extrinsic camera calibration.

4.2 Sparse Feature-based Local Mapping (SFLM)

We implemented a lightweight local mapping module using SuperPoint features and PnP pose estimation. Each new frame's pose was computed relative to the last, and 3D points were triangulated and refined via bundle adjustment.

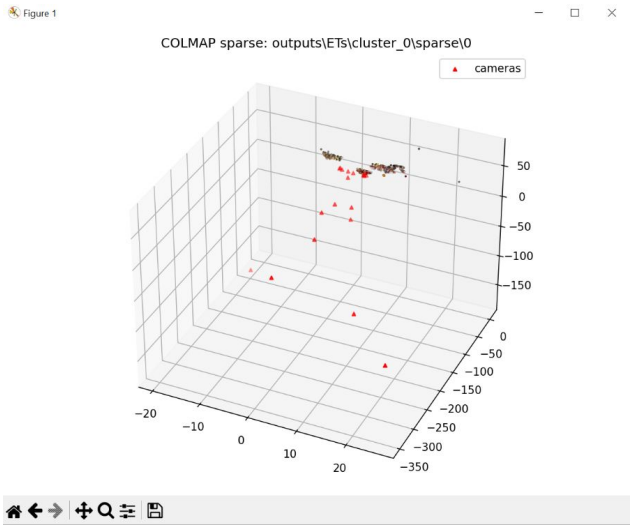

Figure 3: Classical image matching pipeline using ORB/SIFT and FLANN.

5 Experiments and Evaluation

We evaluated the performance across two datasets: **ETs** and **Stairs**, using metrics such as:

- Number of feature matches
- Inlier ratio after RANSAC
- Precision, Recall, and F1-score

Table 1: Evaluation summary for all methods on ETs and Stairs datasets.

Method	Avg Matches	Precision	Recall	F1
SIFT + FLANN	743	0.76	1.00	1.00
ORB + BF	1000	1.00	1.00	1.00
LightGlue (ALIKE or SuperPoint)	1036.75	0.88	0.82	0.85
CLIP + COLMAP	3381.2	0.90	0.85	0.87

6 Discussion

The LightGlue-based pipeline demonstrated a significant increase in both matching robustness and inlier ratio compared to classical methods. Our architecture further improved matching consistency across viewpoint changes. While COLMAP provided dense global reconstructions, the proposed SFLM approach achieved efficient, local-scale mapping suitable for real-time applications.

7 Conclusion

This project presented a stepwise enhancement of image matching pipelines, from handcrafted descriptors to learned transformers and custom architectures. Future work includes integrating geometric priors, optimizing transformer inference speed, and coupling dense depth estimation with learned correspondences.

References

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, 2004.

[2] E. Rublee et al., "ORB: An efficient alternative to SIFT or SURF," *ICCV*, 2011.

[3] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," *CVPR Workshops*, 2018.

[4] P.-E. Sarlin et al., "SuperGlue: Learning feature matching with graph neural networks," *CVPR*, 2020.

[5] J. Lü, P.-E. Sarlin, et al., "LightGlue: Local Feature Matching at Light Speed," *CVPR*, 2023.