COMPUTER VISION FINAL PROJECT

MONOCULAR TWO-VIEW POSE ESTIMATION

Recovering relative camera motion through iterative feature detection, geometric constraints, and robust matching strategies.

FEATURE DETECTION
ORB vs. SIFT

GEOMETRY
Essential Matrix

ROBUSTNESS
RANSAC / Ratio Test

VALIDATION
ArUco Ground Truth

Design and Evaluation of a Monocular Pose Pipeline

Problem Statement & Scope

Defining the geometric objectives and system boundaries

Problem Statement

Given two images I_1 and I_2 of a static scene captured by the same camera, estimate the relative camera pose (R, t).

$$x_2 \sim K(Rx_1 + t)$$

INPUTS

- 2 RGB Images
- Intrinsics Matrix (K)

OUTPUTS

- $R \in SO(3)$
- $t \in \mathbb{R}^3$ (Up to Scale)

Project Scope

Core Focus

- Epipolar geometry estimation
- Feature detector & descriptor analysis
- Camera calibration & stabilization
- Quantitative evaluation of robustness

Deliberate Exclusions

- Dense Depth
- Multi-view VO
- Bundle Adjustment
- Sensor Fusion (IMU)

** Monocular constraints limit translation recovery to direction only (unit vector).*

Pipeline Overview

From raw image pairs to calibrated relative pose estimation

01

Input & Calibration

Capture two RGB images of a static scene. Integrate **calibrated intrinsics (K)** to account for focal length and principal point.

02

Extraction & Downscaling

Apply **Area-based Downscaling** for efficiency, then extract keypoints via **ORB** or **SIFT** descriptors.

03

Adaptive Matching

Filter correspondences using **Adaptive Ratio Testing** and **Adaptive Symmetric Selection** for high-fidelity pairs.

04

Robust Geometry

Employ **Robust Essential Matrix Handling** with refined outlier rejection to ensure stable pose recovery.

05

Pose Recovery

Decompose the Essential Matrix via **SVD** to recover the relative rotation $\mathbf{R} \in \mathbf{SO(3)}$ and translation $\mathbf{t}$.

FINAL RESULT

[R | t]

Relative Camera Motion

PIPELINE_STATUS: CALIBRATED_ACTIVE

FROM ORB TO SIFT

Balancing Computational Efficiency and Geometric Stability

ORB (ORIENTED FAST)

- Fast and lightweight; designed for real-time systems.
- Binary descriptors with Hamming Distance matching.
- **Highly sensitive to image resolution and scaling.**
- Unstable results: Small resolution shifts cause large rotation errors. errors.

SIFT (SCALE-INVARIANT)

- Explicit scale-space construction for high robustness.
- Floating-point descriptors with L2 Norm matching.
- **Stable rotation estimates across varying resolutions.**
- High cost: ~5x slower computation than ORB.

PERFORMANCE INSIGHTS

75% ERROR REDUCTION (AVG)

"Just by changing to SIFT, the error dropped significantly... from 8.4 to 2.87 degrees."

Camera Calibration & Intrinsics

From heuristic approximations to metric precision

Offline Calibration Pipeline

STEP 1: CAPTURE

Multiple chessboard images from various viewpoints.

STEP 2: DETECTION

Sub-pixel corner detection in 2D image space.

STEP 3: OPTIMIZATION

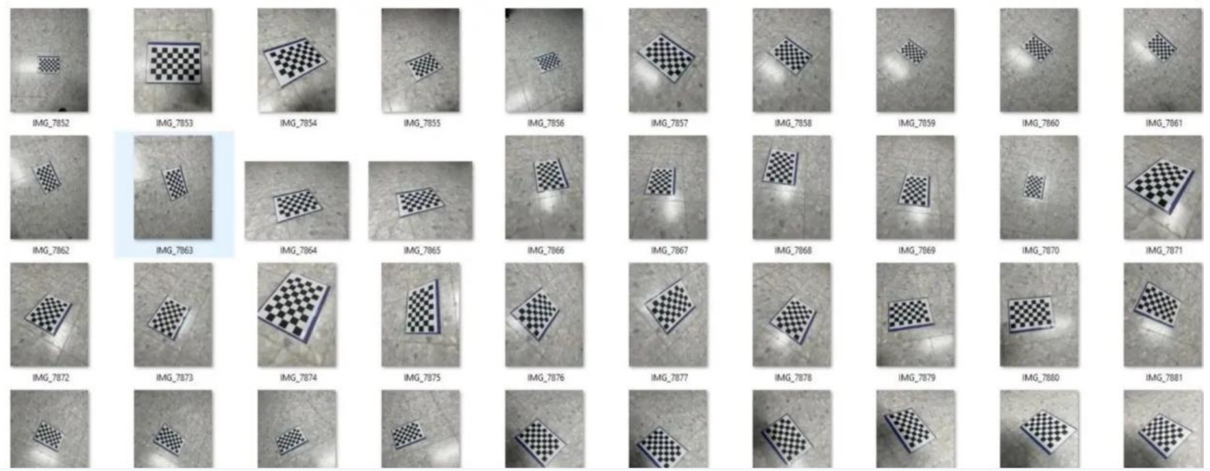
Estimating Intrinsic Matrix (K) and lens distortion.

Implementation Note

To maintain geometric accuracy across different scales, the intrinsic matrix ***K*** was dynamically scaled whenever image resizing occurred during preprocessing.

CALIBRATION DATASET (MULTI-VIEW)

OpenCV Pipeline



BEFORE: APPROXIMATION

K estimated from image resolution (w/2, h/2)

- High scale sensitivity
- Unstable rotation estimates

AFTER: CALIBRATION

Metric matrix via chessboard pattern

- Reduced rotation variance
- Consistent Essential Matrix

Key Enhancements: Pipeline Optimization

Advanced techniques for robust matching and performance efficiency

Adaptive Ratio Testing

Dynamically adjusts the ratio threshold based on local descriptor density and scene complexity.

- ✓ Goes beyond fixed Lowe's ratio (0.7-0.8) for better sensitivity.
- ✓ Retains more valid matches in challenging texture-less regions.

> **INNOVATION**
> *"Balances precision and recall by adapting to local feature distribution."*

Adaptive Symmetric Selection

Refines symmetric matching by selectively applying strict constraints only where ambiguity is high.

- Adaptively applies symmetric matching when match ambiguity is high
- Otherwise, falls back to one-way matching to preserve inliers.

+ Reduces compute overhead compared to global symmetric matching.

+ Maintains high-quality matches for Essential Matrix estimation.

Area-based Image Downscaling

Optimizing input resolution for robust feature detection.

Efficient
COMPUTE SAVINGS

Stable
FEATURE DENSITY

Geometry-Aware Essential Matrix Estimation from Refined Matches

Ensuring geometric integrity through advanced outlier rejection

Robust MAGSAC Estimation

The 5-point algorithm is combined with a MAGSAC-based robust estimator operating on a refined correspondence set produced by adaptive ratio testing and selective symmetry.

- Robust inlier selection using MAGSAC on high-confidence correspondences.
- Cheirality validation during pose recovery to select the correct solution.
- Conservative failure handling under low-parallax or ambiguous configurations.

Upstream-Guided Geometric Validation

Improved match quality upstream reduces degeneracy in Essential Matrix estimation, improving pose stability in low-parallax and near-planar scenes.

GEOMETRIC VERIFICATION

MORE STABLE

INLIER RELIABILITY

CONSISTENT

GEOMETRIC ERROR

Operating MAGSAC on refined correspondences significantly improves geometric consistency under noisy matches.

EPIPOLAR GEOMETRY

POSE STABILITY

ESTIMATION_MODE: ROBUST_GEOMETRY

Why ArUco Markers?

Why ArUco Markers?

To check how accurate our monocular pose estimation really is, we needed an external and reliable ground-truth reference.

- **Known Geometry:**
  The marker size and geometry are known in advance, which allows us to compute an accurate reference pos
- **Robust Pose Estimation:**
  ArUco provides a stable 6-DOF pose using a calibrated camera. camera.
- **Relative Comparison:**
  This lets us directly compare the relative rotation and translation between two images.

Note: ArUco was used exclusively for internal evaluation and was not part of the final pose estimation system.

EXPERIMENTAL SETUP

Error Metrics

Rotation Error	Degrees (°)
Euler Angle Error	R / P / Y
Translation Direction	Qualitative

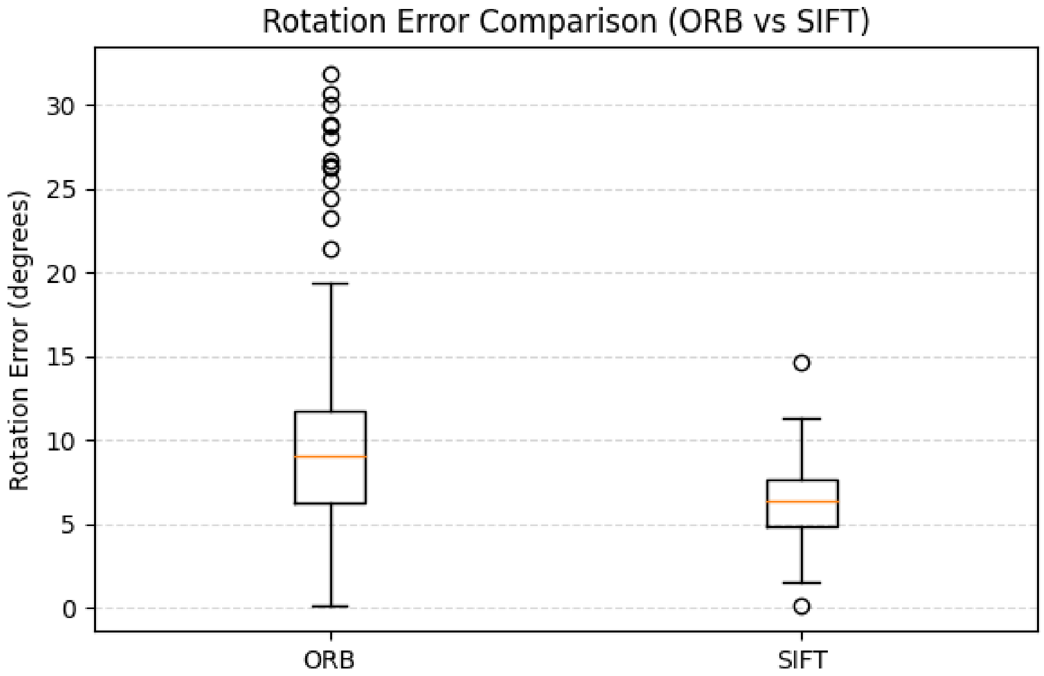
GROUND TRUTH WORKFLOW

IMAGE CAPTURE › MARKER DETECT › POSE CALC › ERROR ANALYSIS

Quantitative Results: ORB vs SIFT

PERFORMANCE BENCHMARK

Rotation Error Comparison (Degrees)



Key Observations

Stability

SIFT shows together error distribution.

Outliers

ORB produces larger error spikes

Tradeoff

SIFT is approximately **5x-20x slower** than ORB, but more reliable.

VERDICT

SIFT provides more stable pose estimates than ORB under varying image scales.

LARGE ORB OUTLIERS

ORB PEAK ERROR

Extreme sensitivity to resolution and different scenes.

~25%

LOWER MEDIAN ERROR

SIFT shows lower and more stable rotation error.

VISUALIZATIONS & DEBUGGING TOOLS

Visualization tools were implemented to validate results, run diagnostics, and identify algorithmic failures.

Keypoint Detection

Validating feature distribution across texture-rich and low-texture regions.

Feature Matching

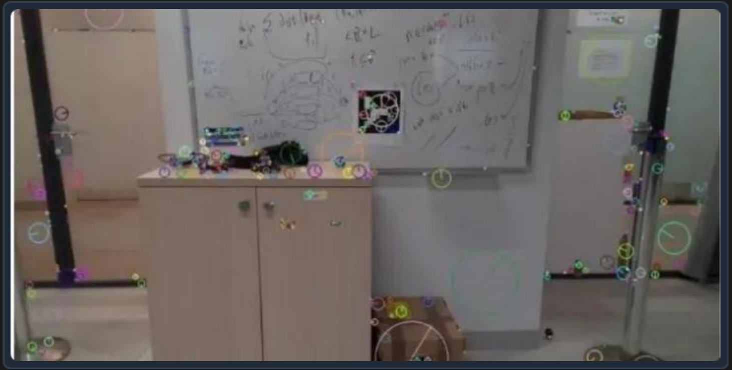
Visualizing raw matches vs. inliers after geometric filtering.

Epipolar Geometry

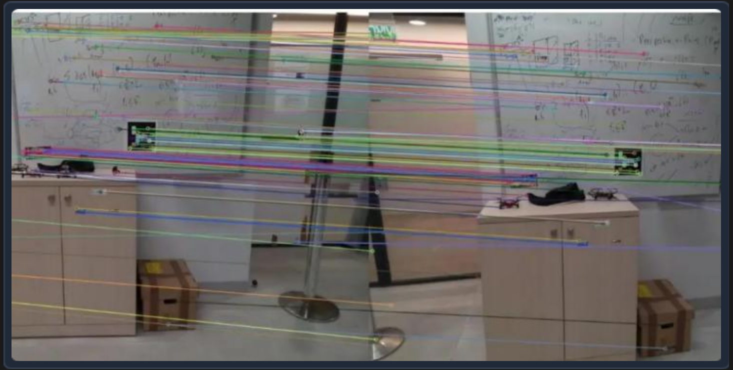
Verifying geometric consistency through epipolar line projections.

Pose Consistency

Compare the estimated rotation and translation to ArUco ground truth.



KEYPOINT DISTRIBUTION



MATCHED FEATURE PAIRS

MATCHED INLIERS AFTER OUTLIER FILTERING

DIAGNOSTIC_MODE

Validation using geometric constraints and ground truth

VIDEO EXPERIMENTS & FUTURE DIRECTIONS

Transitioning from static pairs to sequential frame estimation

IMPLEMENTED IMPROVEMENTS

- **Adaptive Ratio Testing**: Replaced fixed thresholds for dynamic scenes.
- **Robust E-Matrix Handling**: Reduced pose instability and degenerate cases.
- **Area-based Downscaling**: Optimized the throughput without losing losing features.

FUTURE DIRECTIONS

Moving towards **Full Monocular SLAM**: Multi-camera setups and more advanced optimization methods.

Frame-to-Frame Pose Behavior (Video) – Below is a video turned into a GIF