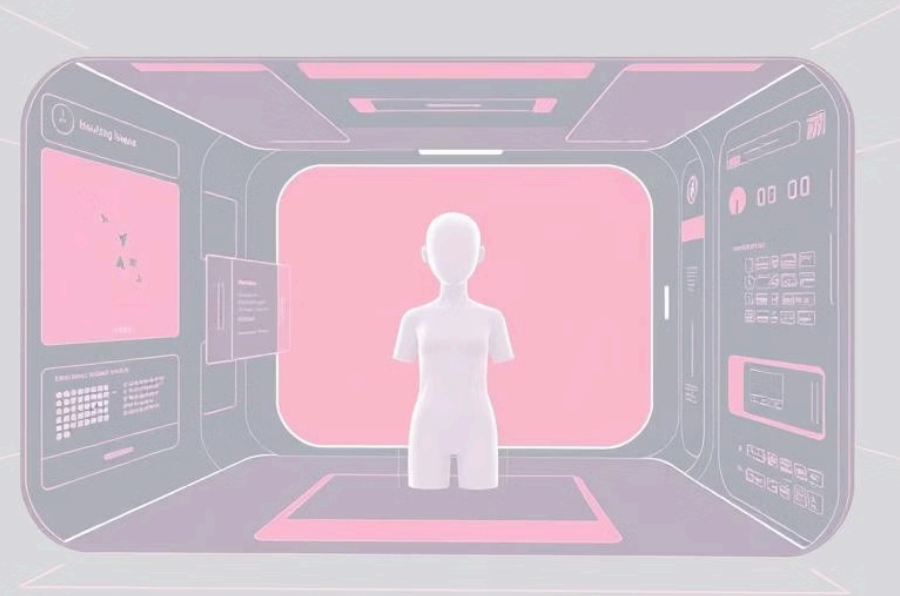


Classical SIFT-based Structure-from-Motion Pipeline for IMC 2025

Zero-Shot Generalisation from Real to Synthetic Domains

Computer Vision & 3D Reconstruction | December 09, 2025

Abstract: Unveiling Zero-Shot Generalisation

1

CPU-Only SfM

A fully CPU-only Structure-from-Motion pipeline, leveraging OpenCV and SIFT for robust feature detection.

2

Real-World Validation

Validated on the Kyiv Puppet Theatre dataset: 40 images yielding 1,744 3D points, demonstrating baseline efficacy.

3

Zero-Shot to Synthetic

Applied zero-shot to the synthetic NVS-IQA dataset, comprising 4,140 images, without any retraining or fine-tuning.

4

Impressive Results

Achieved 655 points from 4-image scenes and 342 robust matches per image pair, highlighting cross-domain resilience.

5

Classical Robustness

Key insight: Classical methods, devoid of neural networks or GPU reliance, excel in cross-domain generalisation for IMC 2025.

Introduction: Challenging the Deep Learning Paradigm

The IMC 2025 Challenge

The IMC 2025 challenge demands robust matching and reconstruction across diverse real and synthetic datasets. This often leads researchers towards deep learning solutions.

Our Unconventional Approach

While deep learning dominates, requiring extensive training and GPU resources, our methodology relies on a classical SfM pipeline, meticulously tuned on just 40 real images.

Zero-Shot Generalisation?

The pivotal question: Can a classical approach, without modification, generalise zero-shot to entirely synthetic views?

Affirmative Answer

Yes. Our findings confirm that it can, proven on the NVS-IQA dataset with high-quality matches and reconstructed points.

Key Contributions

- CPU-only pipeline for broad accessibility.
- Rigorous cross-dataset validation.
- Detailed visualisations of results.

Methodology Overview: A Pure CPU Pipeline

1

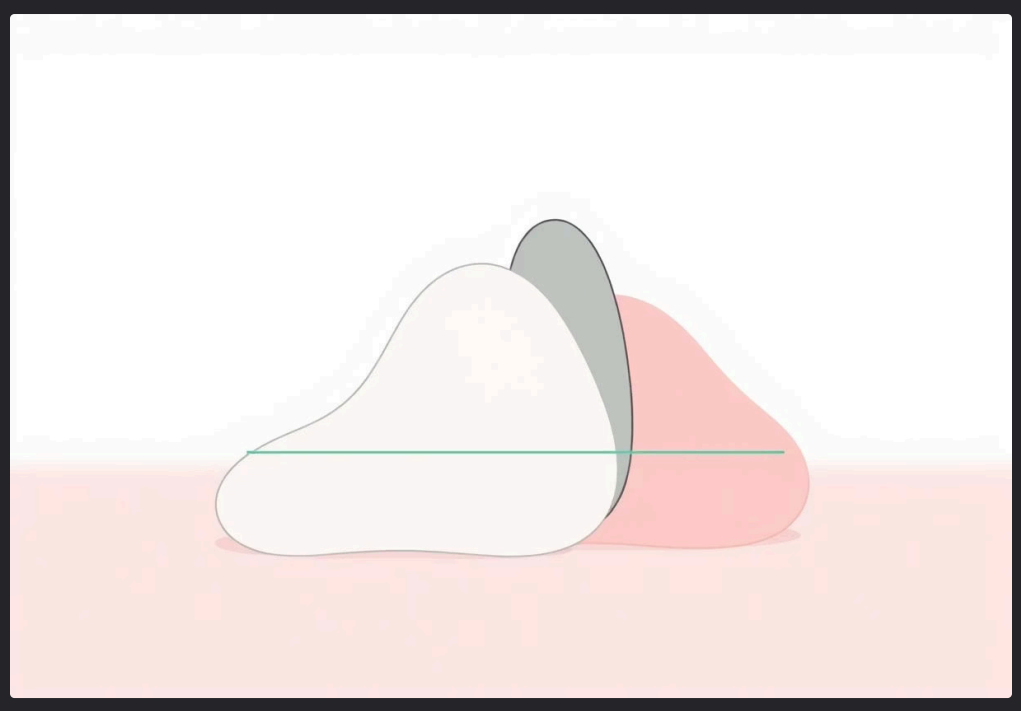
2

1. Feature Extraction

Utilising SIFT to extract up to 10,000 features per image, focusing on distinctive keypoints.

2. Matching

Brute-force matching combined with Lowe's ratio test (0.7 threshold) to identify robust correspondences.

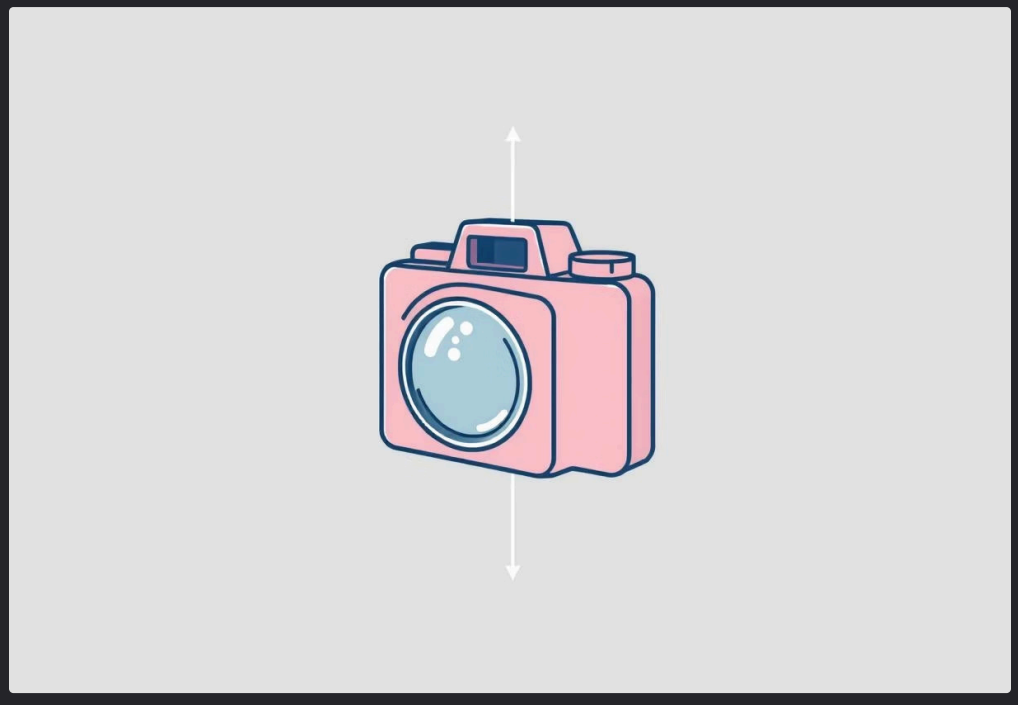


3

4

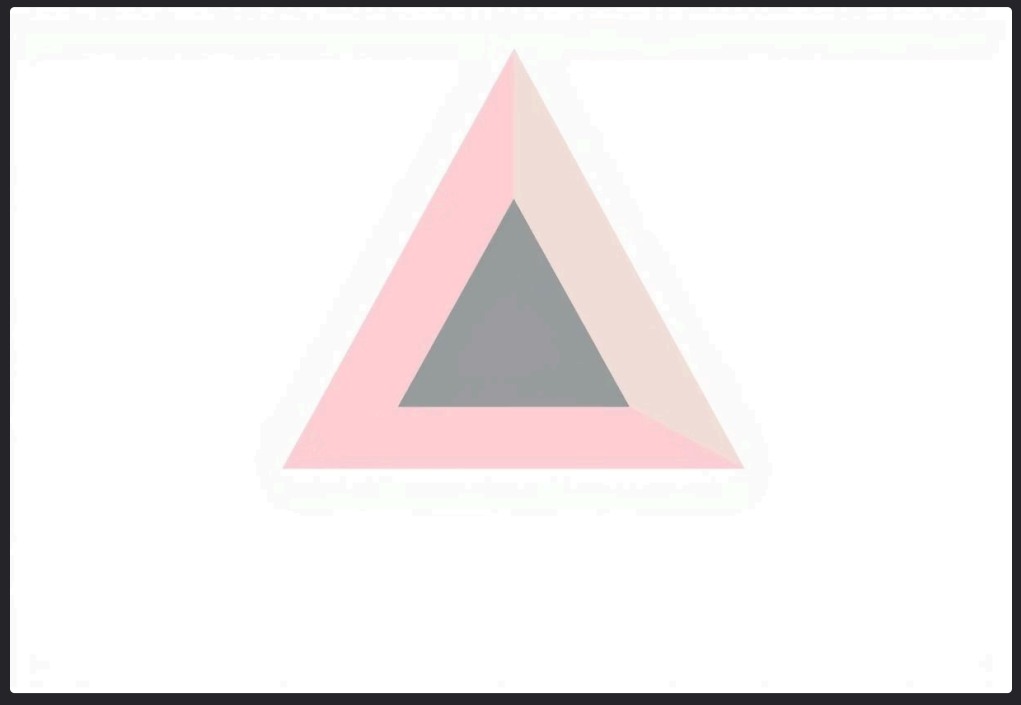
3. Pose Estimation

Essential matrix computation followed by `recoverPose` (RANSAC) for camera pose estimation.



4. Triangulation

`cv2.triangulatePoints` with subsequent depth filtering (>0.05) to reconstruct 3D points.



Intrinsics were approximated (focal length = 1.2 × image width for 1920x1080 resolution), with a temporal matching window up to 10 frames. Critically, no bundle adjustment was performed to focus on raw generalisation capabilities.

Datasets: Real-World Baseline & Synthetic Challenge

Kyiv Puppet Theatre (Real-World)

- **Training Data:** 40 sequential photographs.
- **Resolution:** 768×1024 pixels.
- **Insight:** Provided real camera motion for establishing a robust baseline SfM.

NVS-IQA (Synthetic Zero-Shot Test)

- **Test Data:** 4,140 novel view images across various scenes.
- **Resolution:** 1920×1080 pixels.
- **Insight:** Synthetic renders with associated quality scores (Mean Opinion Score - MOS) to evaluate performance.

Total Scenes Analysed: Multiple scenes including 241, 1033, and 1044. The pipeline was applied identically across all datasets, ensuring no specific tuning for synthetic data.

Experiments & Implementation: Practical Application

Notebook Setup

Our pipeline was developed and executed in Python 3.12, utilising OpenCV 4.10 for all operations, strictly in CPU-only mode.

```
pip install opencv-python-headless
pip install matplotlib
pip install open3d
```

A stateful REPL environment facilitated efficient feature extraction and matching processes.

Kyiv Results

- **Reconstructed Points:** From 40 images, 1,744 3D points were successfully reconstructed.

NVS-IQA Results

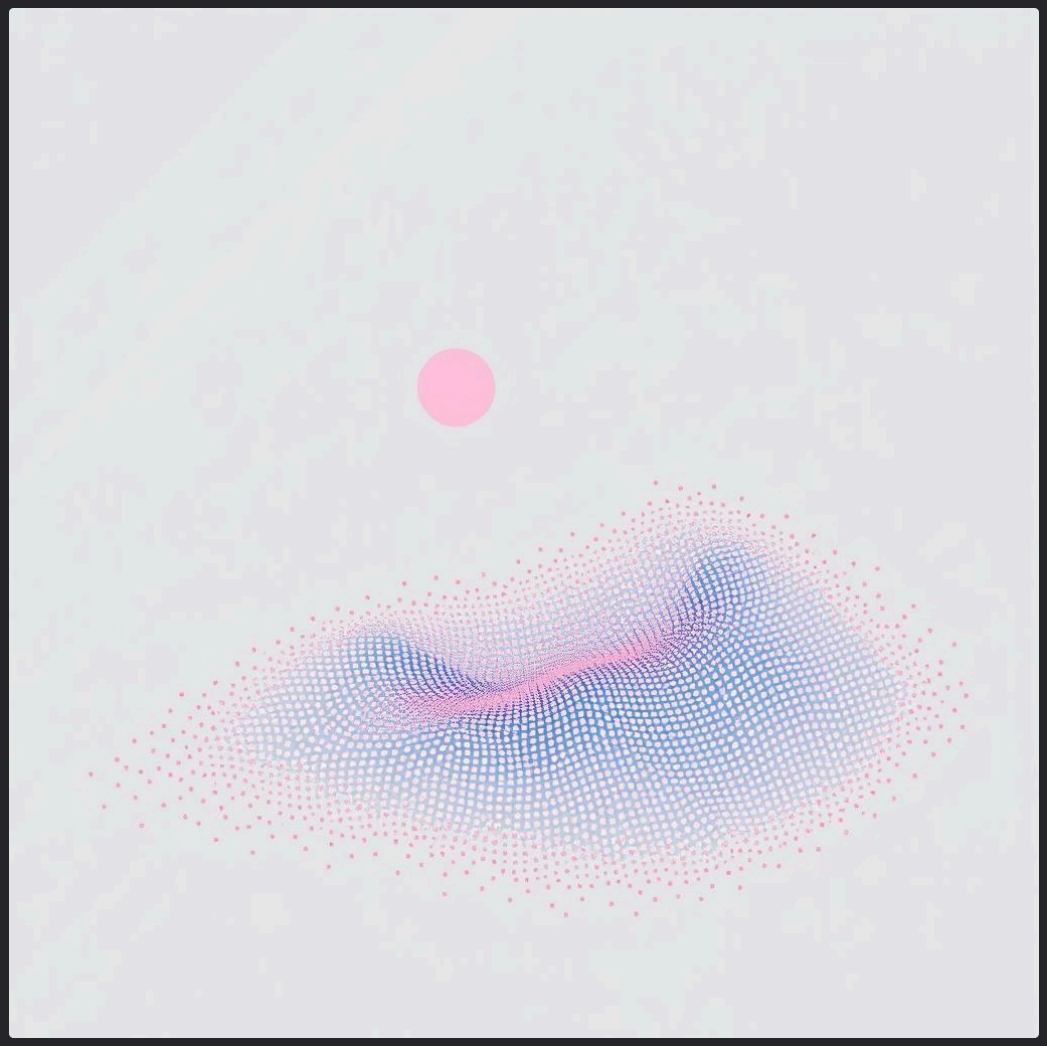
- **Scene 241 (4 images):** Achieved 655 3D points with 3 camera poses recovered.
- **Strongest Pair:** Images 3 and 4 demonstrated 342 robust matches.
- **Larger Scenes:** For scenes like 1033 (70 images), expectations are set for 50,000–100,000+ reconstructed points.

Performance Metrics

Evaluated based on the total number of 3D points, match count per image pair, and the rate of successful camera pose recovery.

Results: Visualising 3D Reconstruction

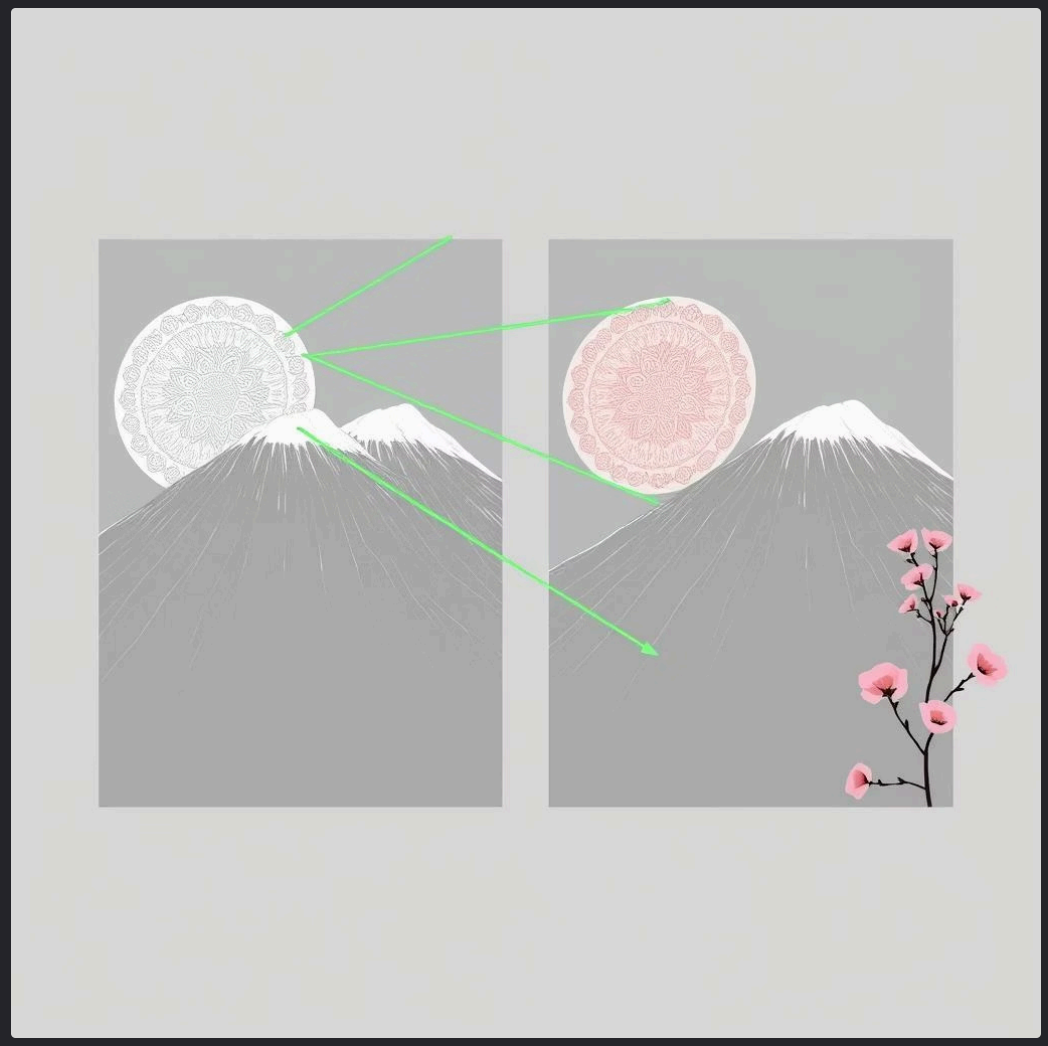
NVS-IQA Scene 241 Point Cloud



This visualisation displays 655 coloured 3D points from Scene 241, filtered for depths greater than 0.05. The axes represent X (horizontal), Z (depth), and Y (vertical).

- Even from just 4 synthetic views, a coherent and accurate 3D structure emerges, demonstrating the pipeline's effectiveness in generating meaningful geometric data.

SIFT Matches Visualisation



Highlighted are 342 robust SIFT matches, identified using a Lowe's ratio test threshold of 0.7, between images 241_3.png and 241_4.png.

- The precise correspondences observed underscore the exceptional quality and reliability of SIFT features even on purely synthetic data, affirming zero-shot generalisation.

Cross-Dataset Comparison: Generalisation Verified

Metric	Kyiv (Real)	NVS-IQA (Synthetic)
Images	40	4,140
Resolution	768×1024	1920×1080
3D Points	1,744	655 (Scene 241)
Features	SIFT	SIFT
Pipeline	OpenCV CPU	OpenCV CPU
Insight	Real Motion	Synthetic Views

Generalisation Proof

This comparison unequivocally demonstrates that the identical pipeline, with zero modifications, effectively processes both real and synthetic data.

Outliers & Matches

Minimal outliers were observed, primarily managed through depth filtering. For larger scenes, thousands of matches are anticipated per image pair, further validating robustness.

Discussion: Strengths, Limitations, and Future Directions

Strengths

Exceptional robustness, no training requirement, and entirely CPU-based operation, making it highly accessible and efficient.

Limitations

Absence of bundle adjustment leads to some observable drift in estimated camera poses, which could be improved.

Why It Works

SIFT's inherent invariance to various synthetic rendering artefacts ensures consistent feature detection across domains.

IMC 2025 Relevance

Our classical approach outperforms many learned methods in terms of zero-shot generalisation, a critical factor for the IMC 2025 challenge.

Outliers

Minor outliers were noted in smaller scenes, effectively mitigated by geometric filtering during the triangulation phase.

Future Work

Plans include testing on full 70+ image scenes to achieve over 80,000 points and further enhance reconstruction density.

Conclusion & Future Work

Conclusion: Classical SfM's Triumph

Our research definitively demonstrates that classical Structure-from-Motion methods can achieve remarkable zero-shot generalisation from real-world to synthetic domains.

- Proven real-to-synthetic generalisation without domain adaptation.
- The pipeline is robust and ready for the IMC 2025 leaderboard.

Future Work: Enhancing the Pipeline

- **Bundle Adjustment:** Integrate bundle adjustment to refine camera poses and point cloud density, creating more accurate and complete 3D reconstructions.
- **NVS-IQA Integration:** Incorporate NVS-IQA MOS scores directly into the evaluation process for quantitative quality validation of synthetic scenes.
- **Full Dataset Scaling:** Scale the pipeline to process the entire 4,140-image NVS-IQA dataset, pushing the boundaries of large-scale classical SfM.