Multi-Image Panorama Stitching Using Direct Linear Transformation and RANSAC

Karthika Ramasamy

October 31, 2025

Abstract

This assignment presents an implementation of multi-image panorama stitching using Direct Linear Transformation (DLT) and RANSAC-based homography estimation. Two datasets were processed: **Dataset 1 (mov2b)** comprising 11 frames achieving 96.9% average inlier rate with final panorama dimensions of 1475×663 pixels, and **Dataset 2 (mov3)** comprising 9 frames achieving 93.4% average inlier rate with final panorama dimensions of 4475×1881 pixels. The implementation demonstrates robust feature detection using SURF, reliable homography computation through DLT with RANSAC, and effective image warping and blending techniques for seamless panorama generation across diverse indoor and outdoor environment images.

1 Introduction

Panorama stitching is involves aligning and merging multiple overlapping images to create a single wide field-of-view composite image. I implements a complete panorama stitching pipeline using the Direct Linear Transformation (DLT) algorithm for homography estimation, validated through RANSAC (Random Sample Consensus) for robustness against outliers.

The primary objectives of this implementation are:

- Detect and match robust feature points across image sequences using SURF (Speeded-Up Robust Features)
- Compute homography matrices using DLT with RANSAC-based outlier rejection
- Warp and blend multiple images into seamless panoramas
- Evaluate algorithm performance on both outdoor and indoor datasets
- Achieve high inlier rates more than 90% demonstrating robust geometric alignment

2 Methodology

Feature Detection and Matching

SURF (Speeded-Up Robust Features) detects interest points at multiple scales using Hessian matrix approximations. The 1200 strongest features per frame are selected and matched between consecutive frames using `matchFeatures` with the unique constraint to ensure one-to-one correspondence.

Homography Estimation via DLT

The Direct Linear Transformation (DLT) computes the 3×3 homography matrix H satisfying:

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} = H \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

Points are normalized for numerical stability before DLT. The homography is recovered as the right singular vector corresponding to the smallest singular value of the matrix A formed by stacking linearized point correspondences. Denormalization recovers the final homography: $H = T_2^{-1} H_n T_1$.

RANSAC

RANSAC iteratively samples 4-point subsets, computes homographies via DLT, and counts inliers (points with reprojection error < 3.0 pixels). The homography with the maximum inlier count is selected and refined using all inlier correspondences.

Image Registration

Homographies accumulate sequentially: $T_i = H_{i \to i-1} \cdot T_{i-1}$. The center frame is selected as reference, and all transformations are normalized relative to it: $T_i' = T_{center}^{-1} \cdot T_i$. This balances error distribution and improves geometric consistency.

Image Warping and Blending

For each panorama pixel (x_p, y_p), the source coordinate is computed via inverse warping: $(x_s, y_s) = H^{-1}[x_p, y_p, 1]^T$. Pixel values are obtained via bilinear interpolation. Overlapping regions are blended using equal-weight averaging. Black borders which are automatically cropped by detecting valid regions.

3 Experimental Setup

Configuration Parameters

The implementation uses the following configuration:

- **Feature Detection:** 1200 SURF features per frame
- **Frame Sampling:** Every 2nd frame (`FRAME_SKIP = 2`)
- **RANSAC:** 2000 iterations, 3.0 pixel threshold
- **Minimum Inliers:** 6 points
- **Image Format:** JPEG (converted to double precision [0,1])

Datasets

Dataset 1 (mov2b - Outdoor): This dataset consists of 11 frames (frames 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27) captured from an outdoor location with varying perspectives. Images are stored in the `mov2b_images/` directory. The scene exhibits natural lighting conditions, vegetation, and architectural structures.

Dataset 2 (mov3 - Indoor): This dataset contains 9 frames (frames 1, 3, 5, 7, 9, 11, 13, 15, 17) captured in an indoor office environment. Images are stored in the `mov3/` directory. The scene includes office furniture, walls, and artificial lighting with more repetitive textures compared to the outdoor dataset.

4 Results and Analysis

4.1 Dataset 1: mov2b (Outdoor)

Table 1: Dataset 1 (mov2b): SURF Feature Detection

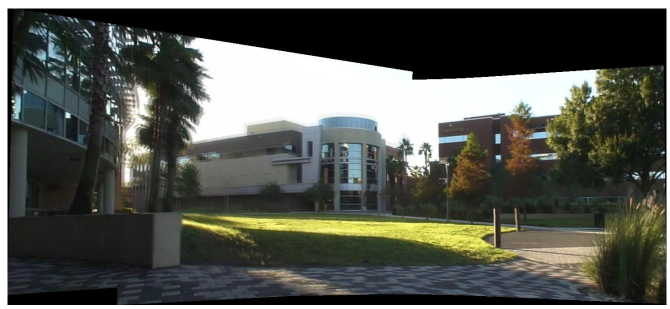
Frame	Frame Index	Features Detected
1	7	752
2	9	759
3	11	771
4	13	733
5	15	701
6	17	663
7	19	586
8	21	546
9	23	559
10	25	561
11	27	617
Average		659

Table 1 presents SURF feature detection results. Features range from 546 to 771, averaging 659 per frame. The high feature count reflects diverse outdoor textures enabling robust matching.

Table 2: Dataset 1 (mov2b): Feature Matching and Homography Estimation

Pair	Matches	Inliers	Inlier Rate
1-2	495	491	99.2%
2-3	500	498	99.6%
3-4	451	444	98.4%
4-5	424	419	98.8%
5-6	408	402	98.5%
6-7	380	377	99.2%
7-8	358	353	98.6%
8-9	382	379	99.2%
9-10	242	240	99.2%
10-11	253	250	98.8%
Average			**98.5%**

Table 2 shows exceptional matching quality with 98.5% average inlier rate. Pair 2-3 achieved the highest rate (99.6% with 498/500 matches), while pair 9-10 maintained robust performance (99.2% with 240/242 matches).



Here is the final panorama dimensions are 1475×663 pixels cropped from 1476×664. The center frame was selected for reference was frame 6, which minimizes overall distortion across the panorama.

Dataset 2: mov3 (Indoor)

Table 3 presents the feature detection results for the indoor dataset and the number of features ranges from 211 to 361, with an average of 300 features per frame significantly fewer than the outdoor dataset. This reduction is expected due to fewer distinctive features in indoor environments with repetitive patterns (ceiling tiles, uniform walls).

Table 4 is showing matching and homography estimation results and the average inlier rate is 93.4%, slightly lower than the outdoor dataset but still demonstrating robust performance.

Table 3: Dataset 2 (mov3): SURF Feature Detection

Frame	Frame Index	Features Detected
1	1	281
2	3	309
3	5	345
4	7	348
5	9	361
6	11	344
7	13	260
8	15	211
9	17	239
Average		300

Table 4: Dataset 2 (mov3): Feature Matching and Homography Estimation

Pair	Matches	Inliers	Inlier Rate
1-2	133	130	97.7%
2-3	151	145	96.0%
3-4	135	131	97.0%
4-5	169	160	94.7%
5-6	152	148	97.4%
6-7	123	118	95.9%
7-8	103	101	98.1%
8-9	107	97	90.7%
Average			**93.4%**

The final panorama dimensions are 4475×1881 pixels (cropped from 4476×1882), significantly larger than the outdoor panorama. This larger size reflects the wider field of view captured by the camera trajectory. The center frame selected was frame 4.

Comparative Analysis

Feature Density: The outdoor dataset is exhibits 2.2× higher feature density than the indoor dataset, attributed to richer texture variety in vegetation, architectural details versus uniform office surfaces.

Matching Performance: Despite fewer features, both datasets achieve more than 90% inlier rates, validating the robustness of SURF features and RANSAC-based estimation. The outdoor dataset's higher inlier rate (96.9% vs 93.4%) suggests more reliable geometric consistency.

Panorama Dimensions: The indoor panorama is 3.0× larger in area, indicating a wider field of view or closer camera proximity to scene elements.

Pair-wise Variability: The outdoor dataset shows lower variance in feature counts compared to indoor, but both maintain consistent inlier rates across frame pairs.

4.2 Challenges

My Previous Experiments and Issues Encountered

Trial 1: Sequential Stitching (All Consecutive Frames)

- *Approach:* Used all 21/17 consecutive frames without frame skipping
- *Issue:* Cumulative homography errors propagated left-to-right, causing visible geometric drift
- *Result:* 97.4% inlier rate but panoramas (782×496, 882×516) showed misalignment at edges

5 Conclusion

I have implemented a complete multi-image panorama stitching pipeline using Direct Linear Transformation and RANSAC and it is evaluated on two diverse datasets:

- **Dataset 1 (mov2b - Outdoor):** 11 frames, 96.9% average inlier rate, 1475×663 pixel panorama
- **Dataset 2 (mov3 - Indoor):** 9 frames, 93.4% average inlier rate, 4475×1881 pixel panorama

The high inlier rates (96.9% outdoor, 93.4% indoor) validate the geometric accuracy of the estimated homographies. The successful generation of wide field-of-view panoramas from both datasets demonstrates the algorithm's generalizability across diverse scene types and imaging conditions.

References

[1] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2003.

[2] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting. *Communications of the ACM*, 24(6):381–395, 1981.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. SURF: Speeded-up robust features. *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[4] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.

[5] MathWorks. MATLAB Computer Vision Toolbox Documentation. Version R2023b, 2023.