

Georeferencing a 1926 Survey of India Sheet: Learned Matching, Classical Matching and Manual Control Points Compared on Held-Out Ground Truth

Saksham Sharma

Code, data provenance and all outputs: <https://github.com/1sakshm/MapAlign>

Interactive version: <https://mysore-1926.pages.dev>

October 2026

Abstract

Large archives of colonial Survey of India topographic sheets are now openly available as scans, but almost none of them are georeferenced. We measure how well three standard approaches place one such sheet, 57 D/11 *Mysore* (1:63,360, published 1926), on modern ground. The three are a pretrained LoFTR matcher with MAGSAC++, SIFT with MAGSAC++, and a least-squares homography fitted to manually selected control points. The modern reference is OpenStreetMap (OSM), rendered by us to a 5 m grid in WGS 84/UTM 43N. All methods share one coordinate chain and one evaluation, scored in metres at 12 held-out points whose truth comes from OSM vector geometry.

SIFT produced no valid transform. Even on tiles that had already been aligned, only 4.75% of its matches were correct, against a chance rate of 4.31%. LoFTR gave 389 RANSAC inliers out of 601 matches and reached a held-out median error of 53.9 m (RMSE 77.4 m). The 7-point manual fit reached 73.9 m (RMSE 102.4 m). LoFTR was closer at 9 of 12 points (exact Wilcoxon $p = 0.043$), but the bootstrap interval for the median paired difference, $[-44.0, +2.5]$ m, still includes zero.

An affine model did as well as the homography. Both methods carry a measurable mean bias of 33–56 m. The sheet’s printed graticule is offset from WGS 84 by roughly 100 m, consistent with a Kalianpur datum. We release the pipeline, which reproduces every number in this paper bit for bit.

1 Introduction

Between the 1860s and the 1960s the Survey of India mapped most of South Asia at one inch to the mile. These sheets record settlements, canals, tanks and railways that have since been rebuilt or erased. They are some of the best evidence of how the region’s land use changed over the twentieth century. Thousands of them are now freely available as scans [1]. Using a scan alongside modern data needs one step first: georeferencing, the estimation of a mapping from scan pixels to geographic coordinates.

The usual way to do this is to place ground-control points (GCPs) by hand, at features that can be found on both the old sheet and a modern reference. That works, but it is slow, it depends on the operator, and after a century of change the features that survive are scarce and unevenly spread. Automatic feature matching is the obvious alternative. Detector-based pipelines built on SIFT [2] have long been used in map processing [3]. More recently, detector-free matchers

such as LoFTR [4] have shown that dense transformer features can match low-texture image pairs where keypoints fail.

A 1926 lithograph and a modern vector rendering are a hard pair for either kind of matcher. The two images share no photometric appearance. Symbols, lettering, line weights and colours all differ, and part of the content itself has changed. Whether matchers transfer to this setting is an empirical question. Answering it fairly needs a protocol in which the methods use the same coordinate chain and are scored in metres on points they never saw.

This paper reports such a comparison on a single sheet. Our contributions are modest and practical: (i) a fully reproducible pipeline that takes a Zenodo scan and an OSM extract to georeferenced GeoTIFFs, with explicit pixel conventions and an exact (non-affine) resampling of the projective transform; (ii) a held-out evaluation in metres, with a written point-selection protocol, quality control that never looks at method output, and a hash lock on the test set; (iii) a diagnosis of *why* SIFT fails on this pair, separate from *whether* it fails; (iv) an analysis of the residual errors, including bias, an affine-versus-projective comparison and the datum of the printed graticule. We make no claim of methodological novelty, and a single sheet does not support general claims about historical maps.

2 Data

Historical sheet. We use Survey of India sheet No. 57 D/11, *Mysore*, Mysore District, at 1:63,360. It was surveyed in season 1924–25 and published in 1926 under Col. C. P. Gunter, Officiating Surveyor General. The scan comes from the Old Survey of India Maps collection on Zenodo [1] (record 8388121, CC BY 4.0). It is a 5898×6828 pixel JPEG at 300 dpi, which corresponds to a nominal 5.36 m per pixel on the ground. The sheet carries a University of Michigan library stamp. Its neatline spans $76^\circ 30' - 76^\circ 45' \text{ E}$ and $12^\circ 15' - 12^\circ 30' \text{ N}$, about 27.2×27.6 km. The graticule is ticked every 5', with no projected grid.

We chose this sheet over two other Karnataka candidates (57 H/9 of 1927 and 57 D/11 of 1947) for three reasons. Its scan is clean and flat. Its northern half, along the Cauvery around Srirangapatna, has many features that survive. And it contains three railway lines. The Krishnarajasagara dam in its north-west corner was still under construction in 1926, so that corner has changed beyond use (Fig. 1). We verified the MD5 checksum published by Zenodo and recorded the SHA-256 of the exact file used.

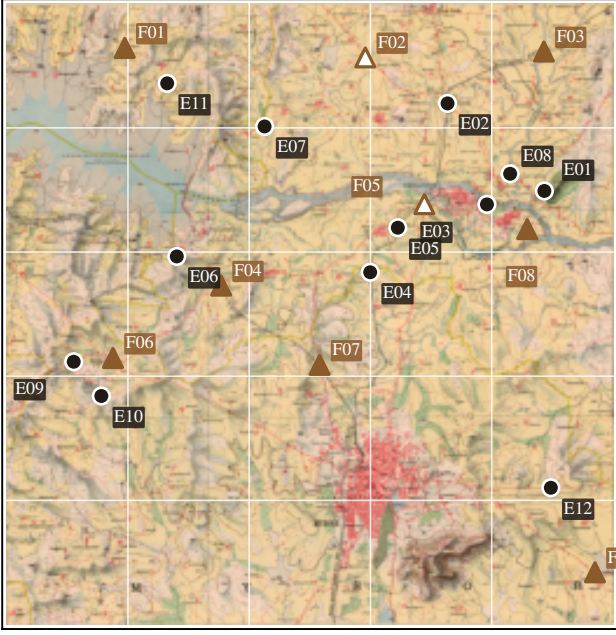


Figure 1: Sheet 57 D/11 (1926), cropped to the neatline, with the 5×5 design grid used for point selection. Brown triangles are fitting points for the manual baseline; the hollow ones (F02, F05) were removed by quality control before fitting. Black circles are the 12 held-out evaluation points. No evaluation point falls in the southern row or in Mysore city (bottom centre), where 1926 features could not be matched with confidence.

Modern reference. We downloaded OSM features [5] with a single Overpass query covering the sheet plus a 0.03° margin. The query returned 32,355 ways, 283 nodes and 10 relations (snapshot 2026-10-18 18:43 UTC): roads, railways, waterways, water bodies and places. We render these ourselves into a 6785×6900 GeoTIFF in WGS 84/UTM zone 43N (EPSG:32643) at 5 m per pixel, with an exact north-up affine geotransform. The style loosely follows the 1926 palette: red roads, black railways, blue water. Rendering it ourselves, instead of using web tiles, fixes the pixel-to-metre mapping exactly and respects the OSM tile policy.

Coordinate reference. All positional errors are computed in EPSG:32643. At the sheet centre, UTM distance differs from geodesic distance by 0.08 m over 10 km [6], which is negligible here. The printed graticule is on a pre-WGS 84 Indian (Kalianpur) datum that we did not identify. PROJ has no real Kalianpur 1880→WGS 84 transformation. We therefore never use printed coordinates as truth, and in Section 4.7 we measure their offset instead.

3 Methods

3.1 Coordinate chain and conventions

Every method estimates a matrix H that maps a scan pixel $\mathbf{x}$ (OpenCV pixel-centre convention, full resolution) to a reference pixel. The reference geotransform A then maps that pixel to EPSG:32643:

$$\hat{\mathbf{p}} = A \pi(H\tilde{\mathbf{x}}), \quad \pi([u, v, w]^T) = [u/w, v/w]^T, \quad (1)$$

where $\tilde{\mathbf{x}}$ is homogeneous. Rasterio’s geotransform addresses pixel corners, so A includes a half-pixel shift. Matches found

at working resolution are mapped back to full resolution through explicit crop-and-resize matrices. We unit-test these against `cv2.resize` to 0.05 pixel. Figure 2 summarises the chain.

3.2 Preprocessing

We crop the scan to its neatline. The four corners were placed by eye and then refined to the darkest line pixel, giving a 5048×5170 pixel crop. We resample both images to a common nominal 10 m per pixel (2708×2773 and 3392×3450 pixels) and convert them to greyscale.

3.3 Method A: LoFTR

We use Kornia’s [7] implementation of LoFTR with the *outdoor* weights [4], in evaluation mode without gradients. LoFTR’s coarse stage compares every $1/8$ -resolution cell of one image with every cell of the other. At 10 m per pixel the full sheet would need a similarity matrix of about 10^{10} entries, so we scale both images by one common factor until the longer side is at most 768 pixels. That gives inputs of 600×616 and 752×768 pixels at about 44.9 m per pixel, with ground resolution kept equal across the pair. All 601 matches are kept regardless of confidence (mean confidence 0.52) and passed to the robust estimator.

3.4 Method B: SIFT

We run OpenCV SIFT [2] with default parameters on the 10 m images. Descriptors are matched by brute-force L_2 two-nearest-neighbour search, and the ratio test is applied at 0.75.

3.5 Shared robust estimation

Both automatic methods use OpenCV’s USAC implementation of MAGSAC++ [8], a descendant of RANSAC [9], with seed 42. The inlier threshold is 3 pixels at the resolution where the matcher located its points. That is 30 m for SIFT and 13.5 working pixels, about 135 m, for LoFTR. We allow 1.2×10^6 iterations, the standard sample count $N = \log(1 - p)/\log(1 - w^4)$ for confidence $p = 0.999$ at an inlier ratio as low as $w = 0.05$.

A model is rejected if it fails any of these checks: finite entries; condition number $< 10^7$; positive determinant; ratio of singular values < 10 ; image corners on the same side of the horizon line; a convex mapped outline; mapped corners within half an image of the reference. A rejected model is recorded as a failure, and it is never repaired.

3.6 Method C: manual control points

We computed 1,053 candidate GCPs from OSM vector geometry: exact railway–road, railway–water and road–water intersections, junctions of major roads, stations and landmark centroids. A candidate’s EPSG:32643 coordinates are its truth. For each candidate we located the corresponding pixel on the 1926 scan in zoomed crops with a 10 pixel grid. The crops were centred using only the four printed sheet corners, so no method’s transform could steer the picking. Candidate locations were first proposed with the help of a large language

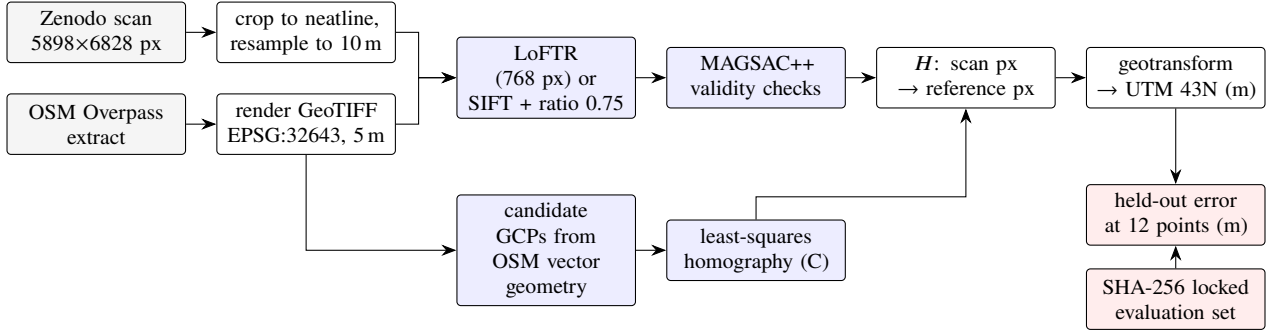


Figure 2: Pipeline. All three methods produce the same object, a 3×3 homography from full-resolution scan pixels to reference pixels, and are scored by one evaluation. Blue: method-specific steps. Red: evaluation, which no method can see.

model assistant working from these crops. The author then checked all 21 resulting points, fitting and evaluation, against the scan and the OSM render, and accepted all of them.

A feature was accepted only if it exists in 1926 with the same geometry and can be located to about ± 5 pixels (± 27 m). The fit is an ordinary least-squares homography on the accepted fitting points (Hartley and Zisserman [10], DLT without RANSAC). That is the same model family as Methods A and B.

3.7 Evaluation protocol

Freezing. All parameters were frozen on 19 October 2026, before any evaluation point existed. Three settings were changed during development, and each change was justified on development data only. (i) The LoFTR input size went from 1280 to 768 pixels, chosen by a rule fixed before the sweep: the most matches within 1 km of the position predicted by the printed corners (Table 3). (ii) The iteration count went from 10^4 to 1.2×10^6 . (iii) The threshold changed from a fixed 3 working pixels to the matcher-relative rule above. SIFT was given the same estimator settings, and eight preprocessing variants, for parity.

Point selection and quality control. The written protocol divides the neatline into a 5×5 grid. It asks for one fitting point in each of 13 cells and 20 evaluation points spread over the sheet, each at least 1 km from every fitting point. We screened every point with a test that uses no method output. For each point we take the offset between its picked pixel and the pixel predicted by the printed corners, and flag the point if that offset lies more than 3 robust standard deviations ($1.4826 \times \text{MAD}$) from the median.

- On the fitting set the test removed F02, an apparently realigned road with $z_E = -7.4$, and F05 ($z_E = 3.6$). That left 7 fitting points.
- On the evaluation set it flagged two picking errors: a wrong road (E12) and a position that had been extrapolated instead of read (E06). Both were re-examined and corrected. None was dropped, and after correction the largest $|z|$ is 2.99.

Shortfall. Only 12 evaluation points could be established instead of 20. Most candidates failed for one of three reasons: the feature postdates 1926 (the Varuna canal network, ring roads); the feature has moved (inside Mysore city the Nanjan-gud railway lies about 280 m from its 1926 course, consistent

with gauge conversion); or the junction is ambiguous. The evaluation file was locked by SHA-256 before any prediction was computed, and the evaluation code refuses to run if it changes.

Metrics. For point i the error is

$$e_i = \sqrt{(E_i - \hat{E}_i)^2 + (N_i - \hat{N}_i)^2} \quad (\text{m, EPSG:32643}). \quad (2)$$

We report the mean, median, RMSE, population standard deviation, 90th percentile, minimum and maximum. Methods are compared on the same points, paired: an exact two-sided Wilcoxon signed-rank test [11] and a percentile bootstrap 95% interval for the median paired difference, with 20,000 resamples [12]. With $n = 12$ both are low-powered, and we read them as descriptive.

4 Results

4.1 SIFT does not transfer to this map pair

SIFT found 79,923 keypoints on the scan and 18,547 on the render. Only 594 matches passed the ratio test, and just 2 of them landed within 1 km of where the printed corners place them. MAGSAC++ found a 26-inlier model, but it was singular: condition number 8.4×10^{13} , a zero singular value, and corners mapped past the horizon line. The validity checks therefore rejected it (Fig. 3b).

None of the eight preprocessing variants helped (Table 2), including colour-layer separation of roads, water and linework, and ground resolutions from 10 to 45 m. Colour layers raised the number of ratio-test matches to 1,351, but at most 5 of them were near the prior.

To separate “SIFT cannot match these images” from “RANSAC could not find the model”, we warped the scan with the printed-corner prior and matched 25 aligned tiles of 512×512 pixels (5.1 km). A match counts as correct if it lands within 600 m of its own position. Matches placed at random would be correct 4.31% of the time. SIFT reached 4.75% (97 of 2,041) and LoFTR 12.92% (321 of 2,485) (Fig. 4a). SIFT’s descriptors therefore carry essentially no information about this pair, even when the geometry is handed to it.

4.2 LoFTR is accurate but scale-sensitive

At 768 pixels LoFTR produced 601 matches, of which MAGSAC++ kept 389 (64.7%). The inlier reprojection RMSE is 6.4 working pixels (64 m) in image space. That number

Table 1: Held-out positional error in metres (EPSG:32643), homography model, the same 12 points for every method. Runtimes are wall-clock on a Ryzen 7 5800HS CPU.

Method	n	Median	Mean	RMSE	Std	P90	Min	Max
LoFTR + MAGSAC	12	53.9	64.2	77.4	43.2	95.9	24.2	191.7
Manual GCPs (7)	12	73.9	89.6	102.4	49.5	140.6	44.9	223.9
SIFT + MAGSAC	0	–	–	–	–	–	–	–

Method	Matches	Inliers	Inlier ratio	Runtime (s)
LoFTR + MAGSAC	601	389	64.7%	18.3
SIFT + MAGSAC	594	26 (degenerate)	4.4%	14.6
Manual GCPs	7 points	–	–	0.15 (fit only)

Table 2: SIFT preprocessing variants. “Near”: matches within 1 km of the printed-corner prior. No variant produced a valid model.

Variant	KP scan	KP ref	Ratio	Near	Inl.	Valid
grey, 10 m	79,923	18,547	594	2	26	no
grey + CLAHE, 10 m	122,783	29,598	683	5	16	no
grey, 20 m	19,099	2,663	50	0	6	no
colour layers, 10 m	29,694	36,280	1351	3	155	no
colour layers, 20 m	7,778	10,050	421	4	39	no
colour layers, 30 m	3,470	5,178	205	1	27	no
grey, 45 m	4,468	903	4	0	0	no
colour layers, 45 m	1,521	2,848	97	0	14	no

Table 3: LoFTR input-size sweep (CPU). “Res.”: effective ground resolution. “Near”: matches within 1 km of the printed-corner prior.

Side (px)	Res. (m)	Matches	Near	Near (%)	Time (s)
384	89.8	145	86	59.3	1.9
512	67.4	405	319	78.8	3.7
640	53.9	342	210	61.4	5.5
768	44.9	601	429	71.4	8.6
896	38.5	348	151	43.4	11.7
1024	33.7	312	108	34.6	16.1
1280	27.0	262	18	6.9	28.6
1536	–	–	–	–	out of memory

measures how well the inliers fit the model, not positional accuracy. The resulting transform places the printed sheet corners a mean 199 m from their graticule positions. Section 4.7 shows that most of this distance is a datum offset.

Table 3 and Fig. 4b show that the scale sensitivity is strong and not monotonic. The share of matches consistent with the prior is 59% at 384 pixels, 79% at 512, 61% at 640, 71% at 768 and only 7% at 1280, the size we had originally planned to use. Had we kept 1280 pixels, LoFTR would have failed almost as completely as SIFT. In practice this sensitivity matters more than any small difference between matchers.

4.3 Georeferenced output

A homography cannot be stored as a GeoTIFF affine geotransform. We therefore resample the scan onto a north-up EPSG:32643 grid at 5 m. Each output pixel is mapped back to the scan through the exact inverse chain (output pixel $\rightarrow$ metres $\rightarrow$ reference pixel $\rightarrow H^{-1} \rightarrow$ scan pixel), so bilinear interpolation is the only approximation. A unit test on a linear ramp confirms the mapping to within 0.02 grey levels.

We also tested the written files using only their own georeferencing. At 2,000 random points inside the neatline, the GeoTIFF sampled at the predicted coordinate returns the

source scan pixel, with a mean absolute colour difference of 4.6. A deliberate 10 m shift raises this to 14.8. Figure 5 draws OSM vectors on the LoFTR GeoTIFF. Railways, roads and the Cauvery fall on the 1926 linework to within a few tens of metres. That includes Mysore station, where neither method was fitted.

4.4 Held-out accuracy

Table 1 gives the main result. LoFTR’s median held-out error is 53.9 m with an RMSE of 77.4 m. The 7-point manual fit reaches 73.9 m and 102.4 m. At the 1:63,360 scale of the sheet, these RMSEs equal 1.22 mm and 1.62 mm on paper. Within 50 m, LoFTR places 5 of 12 points and the manual fit 3. Within 100 m, the counts are 11 and 9.

The 1947 US National Map Accuracy Standard [13] asks for 90% of well-defined points within 1/50 inch at this scale, which is 32.2 m. LoFTR meets it at 2 points and the manual fit at none. Neither result should be read as a statement about the original survey, because both include picking error, OSM error and a century of change.

Table 4 and Fig. 6a give the per-point errors. LoFTR is closer at 9 of 12 points. The median paired difference is -31.7 m (Wilcoxon $W = 13$, exact $p = 0.043$), with a bootstrap 95% interval of $[-44.0, +2.5]$ m. Because the interval includes zero, we treat the advantage as likely but not established.

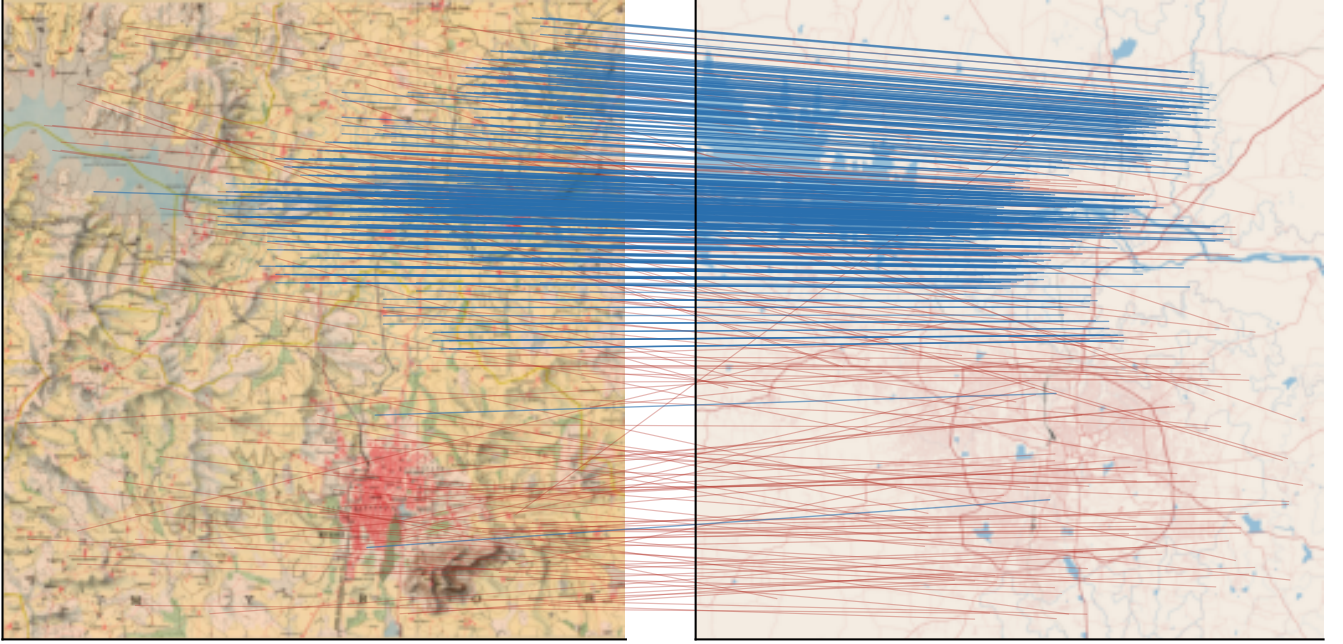
The largest error belongs to E11, a low-confidence road junction in the north-west: 191.7 m for LoFTR and 223.9 m for the manual fit, in nearly the same direction. That pattern points to the point, not the methods: either a mis-identified junction or a realigned road. We kept E11 in all reported numbers. Without the three low-confidence points (E03, E11, E12), LoFTR’s median and RMSE fall to 44.9 m and 46.0 m, and the manual fit’s to 66.7 m and 86.1 m. This is a sensitivity check, not a replacement result.

4.5 Where the errors come from

Bias. Both methods have a mean error vector clearly different from zero (Fig. 8). LoFTR’s mean error is $(+9.7, +31.2)$ m, and the bootstrap 95% interval of its north component, $[5.6, 55.7]$ m, excludes zero. The manual fit’s mean error is $(+43.8, -34.9)$ m, and both of its intervals exclude zero ($\Delta E \in [8.4, 84.7]$, $\Delta N \in [-65.6, -5.5]$). In the north-east, the manual errors at E01, E02, E04 and E08 all point south (Fig. 7b).

These four points lie near the edge of the manual fit’s support. Only three of its seven fitting points are in the eastern third of the sheet, and only one, F09, is in the south.

(a) LoFTR: 601 matches, 389 inliers



(b) SIFT: 594 matches, no valid model



Figure 3: Correspondences between the 1926 scan (left) and the OSM render (right). Blue: RANSAC inliers. Red: a sample of rejected matches. (a) LoFTR’s inliers agree on one transform and cluster along the Cauvery corridor and the railways. (b) SIFT’s 594 ratio-test matches support no consistent model. The 26 “inliers” of the model MAGSAC++ returned collapse onto a single point of the reference, the degenerate solution that the validity checks reject.

A homography fitted to seven unevenly spread points has two weakly constrained projective terms. This matches the gap between its fitting RMSE (55.6 m) and its leave-one-out RMSE (146.9 m) (Fig. 10).

Support. Error does not grow clearly with distance from support (Fig. 9). For LoFTR, the Spearman correlation between a point’s error and its distance to the nearest RANSAC inlier is $\rho = 0.39$ ($p = 0.21$; median distance 1.26 km). For the manual fit, against the distance to the nearest fitting point, it is $\rho = -0.10$ ($p = 0.75$; median 2.5 km). Twelve points are too few to rule out a moderate effect, and the southern fifth of the sheet, where both methods extrapolate, has no evaluation

points at all.

4.6 Affine or projective?

Re-estimating every model as an affine transform, with the same inputs, settings and seed, changes little. LoFTR goes from median 53.9 to 42.4 m and RMSE 77.4 to 76.3 m. The manual fit goes from median 73.9 to 87.8 m and RMSE 102.4 to 105.5 m. Under the affine model, LoFTR is closer at 10 of 12 points (Wilcoxon $p = 0.016$, bootstrap interval $[-58.6, -5.1]$ m). A flat, undistorted scan of a sheet this small has little perspective to model. The two extra degrees of freedom mostly add variance, most visibly in the sparsely

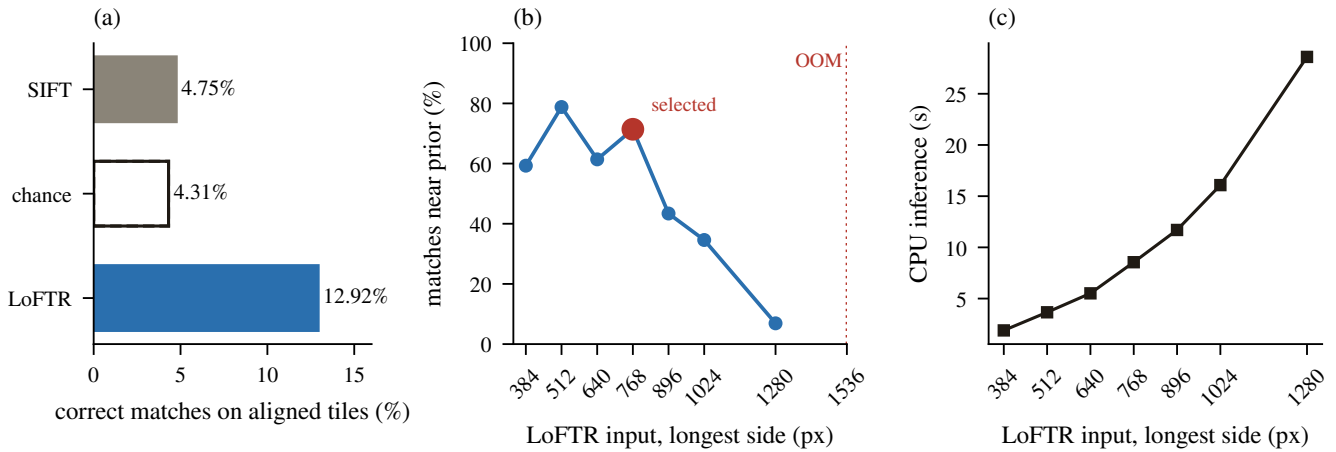


Figure 4: Development diagnostics; no evaluation point was used. (a) Share of correct matches on tiles already aligned by the printed corners, against the chance rate. (b) Share of LoFTR matches within 1 km of the printed-corner prior, by input size. The selection rule picked 768 pixels, and 1536 pixels ran out of memory on a 15 GB machine. (c) LoFTR inference time on the CPU.

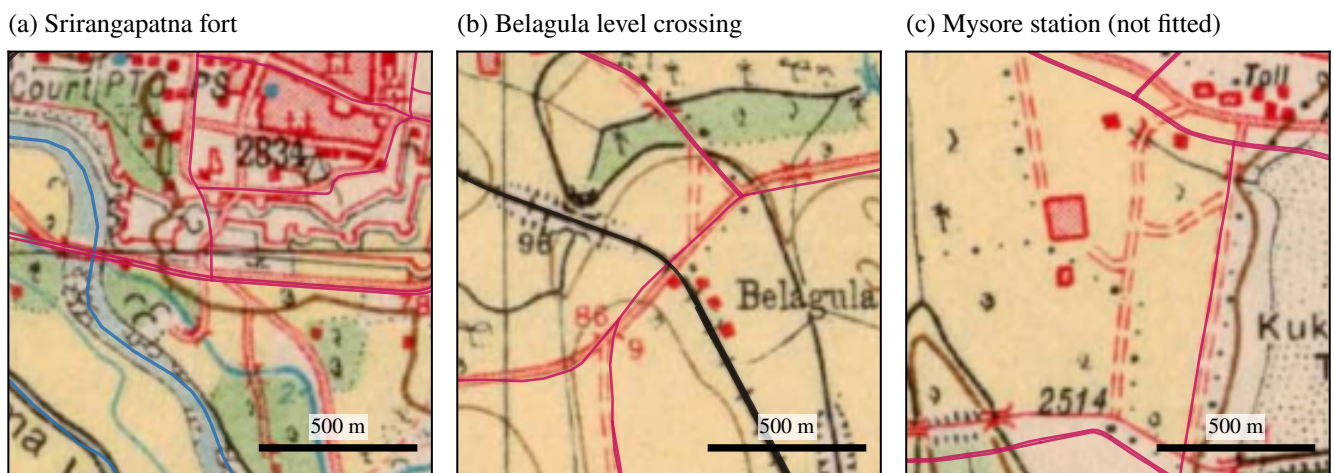


Figure 5: Windows of the LoFTR GeoTIFF with OSM vectors drawn using only the file’s own geotransform. Black: railway. Blue: river. Magenta: major roads. (c) lies about 4 km from the nearest fitting point and contains no evaluation point.

Table 4: Held-out points. Truth is OSM geometry in EPSG:32643. ΔE , ΔN are predicted minus true (m). Bold values are positional errors.

ID	Feature	Conf.	Truth (m)		LoFTR			Manual		
			E	N	ΔE	ΔN	e	ΔE	ΔN	e
E01	landmark	medium	686992	1374170	+39	+45	59.8	+99	-46	109.0
E02	rail $\times$ road	medium	682713	1378173	-35	-40	53.5	+18	-143	144.1
E03	landmark	low	684432	1373518	+29	+93	97.5	+80	+12	81.2
E04	rail $\times$ road	medium	679287	1370607	-41	-38	56.4	-11	-95	95.3
E05	rail $\times$ water	medium	680459	1372537	+15	+27	31.4	+54	-38	66.7
E06	rail $\times$ road	medium	670560	1371290	+26	-19	32.8	+40	-50	64.1
E07	road $\times$ stream	medium	674516	1377058	-27	+33	42.9	+20	-40	44.9
E08	road $\times$ stream	high	685483	1374961	+11	+22	24.2	+66	-69	95.6
E09	road junction	medium	666015	1366513	-14	+43	44.9	-42	+36	55.5
E10	road junction	medium	667232	1364974	+13	+53	54.2	-24	+41	47.3
E11	road junction	low	669955	1378916	+163	+102	191.7	+223	+23	223.9
E12	road junction	low	687431	1360887	-61	+54	81.4	+3	-48	47.9

constrained manual fit. We did not use this comparison to select anything.

4.7 The datum of the printed graticule

We measured the offset between true WGS 84 positions and the positions implied by the printed graticule, read as if it were WGS 84, at the seven fitting points (Fig. 11). The

median is $(-84, +53)$ m, with robust spreads of 58 m east and 95 m north. LoFTR’s correct development matches give a similar $(-128, +133)$ m. PROJ’s Kalianpur 1937 and 1975 parameters give $(-23$ to $-118, +159)$ m at the sheet centre. All three estimates agree in direction. Together they support a datum shift of roughly 100–200 m, which explains most of the 117–199 m corner discrepancy reported above. The exact datum remains unidentified.

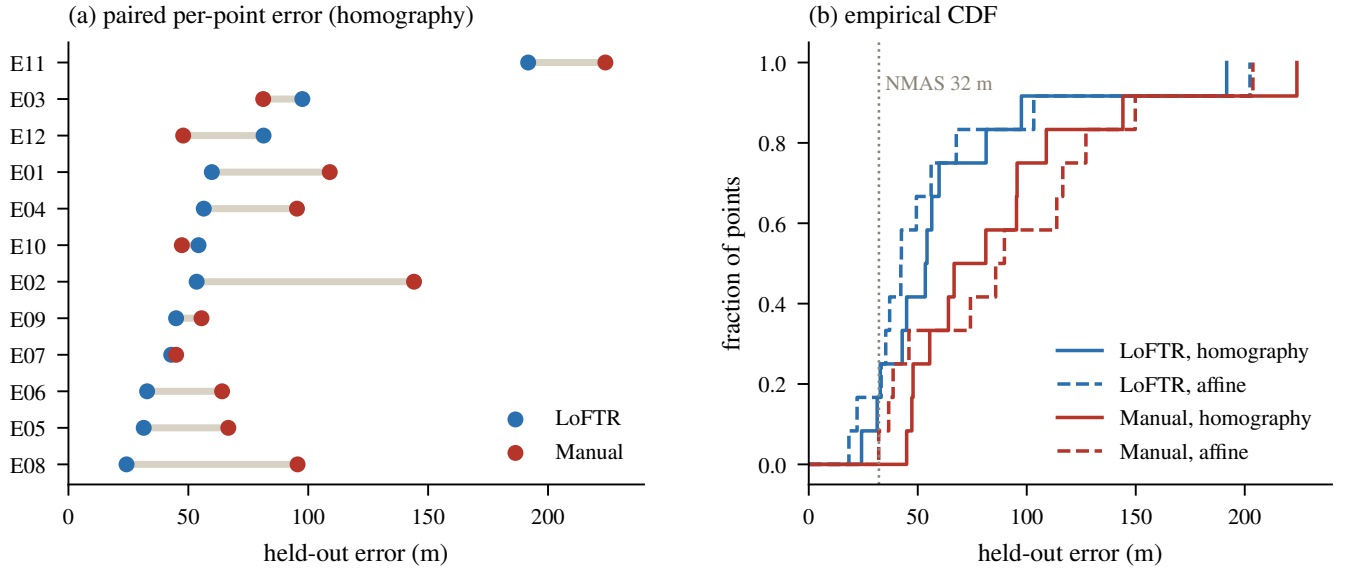
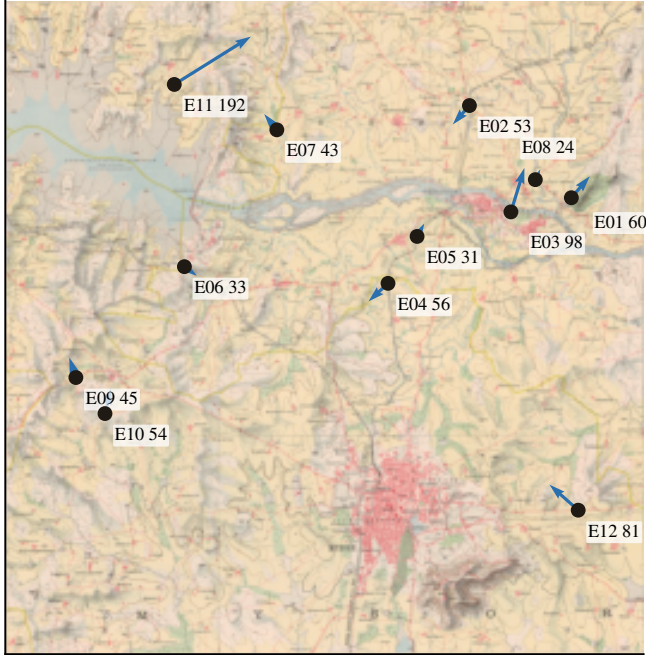


Figure 6: (a) Per-point held-out error, sorted by LoFTR error. The grey bar spans the paired difference. (b) Empirical CDF of the error for both methods and both transform models. The dotted line marks the NMAS threshold for this scale (32.2 m).

(a) LoFTR + MAGSAC; arrows $\times 20$, north up



(b) Manual GCP fit; arrows $\times 20$, north up

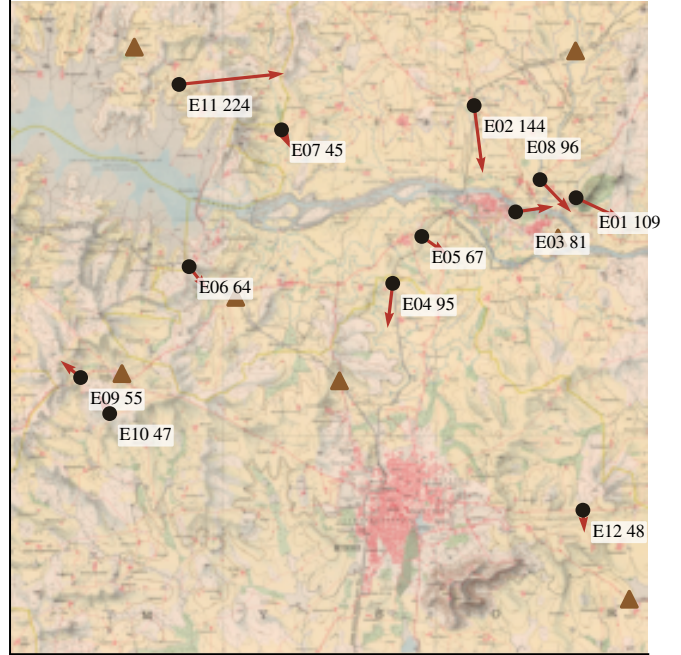


Figure 7: Held-out error vectors ($\times 20$, north up), labelled with the error in metres. Brown triangles in (b) are the seven fitting points used by the manual baseline.

4.8 Compute and reproducibility

On the CPU, LoFTR took 18.3 s end to end, 8.0 s of it inference. SIFT took 14.6 s and the manual fit 0.15 s. Re-running LoFTR on an RTX 3050 Laptop GPU (4 GB, CUDA 12.6, peak 1.06 GB) with the frozen settings gave the same 601 matches and 389 inliers in 1.8 s of inference. Floating-point differences between the devices move the transform by 1.9 m on average and 8.5 m at most across the sheet. The held-out RMSE becomes 77.2 m against 77.4 m. All reported numbers come from the CPU run. A complete re-run from the raw downloads reproduces every transform matrix, per-point error and summary statistic exactly. It takes about five minutes on the CPU.

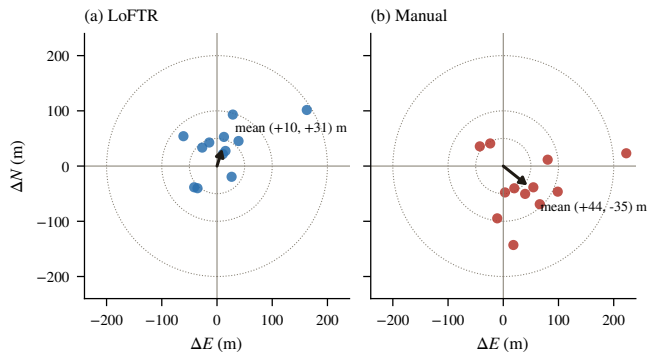


Figure 8: Error components (predicted minus true). Rings mark 50, 100 and 200 m. Arrows show the mean bias.

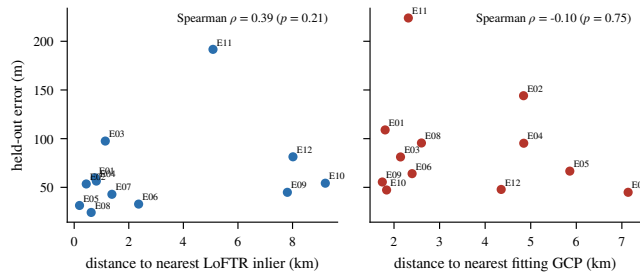


Figure 9: Held-out error against distance to the nearest support: LoFTR inliers (left) and manual fitting points (right). Exploratory analysis.

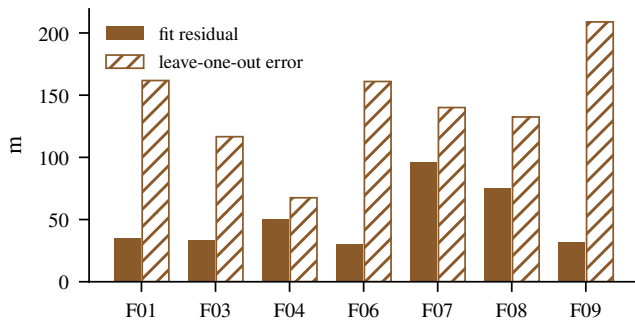


Figure 10: Manual baseline: residual at each fitting point against the error when that point is left out of the fit.

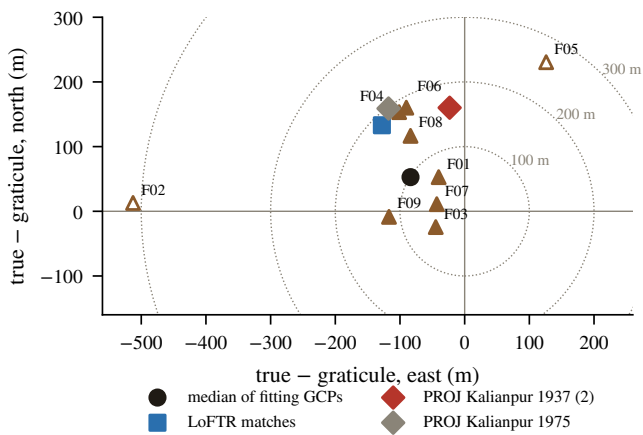


Figure 11: True WGS 84 position minus the position implied by the printed graticule at each fitting point. Hollow points were excluded by quality control. Larger markers are aggregate and PROJ estimates.

5 Discussion

Three findings stand out on this sheet. First, a matcher pretrained on outdoor photographs transfers to a hand-drawn 1926 lithograph against a vector rendering well enough to produce a usable alignment, while SIFT does not transfer at all. The aligned-tile test (Fig. 4a) shows that SIFT’s failure is in the descriptors, not in the robust estimation. Second, LoFTR’s success depends on the input scale, and not monotonically. The scale was chosen by a rule fixed in advance, but a different rule could have produced failure, and practitioners should expect to sweep it. Third, the manual baseline’s weakness here is coverage, not precision. With seven usable points it extrapolates over much of the sheet, and its errors grow where support is thin.

The absolute numbers should be read against the sources of error that no method can remove. OSM geometry is plausibly accurate to 5–20 m. Picking adds about ± 27 m. Many features have also shifted since 1926. A median of 54 m is about 0.85 mm on the printed sheet: close to the drafting limit of a one-inch map, and well above the NMAS threshold.

Limitations.

- The study uses one sheet and one rendering style, so it does not support general claims about historical maps.
- Only 12 evaluation points were available, none in the southern fifth of the sheet or in Mysore city. Results there are untested extrapolation.
- The control points were proposed with AI assistance and accepted by one reviewer. They were not re-picked independently, so the manual baseline is not a measure of expert human georeferencing.
- Three development settings were changed after initial failures, although all were frozen before the evaluation data existed.
- We did not test non-linear transforms, other learned matchers or GPU-scale input sizes.

6 Conclusion

On sheet 57D/11 of 1926, LoFTR with MAGSAC++ georeferenced the scan to a held-out median error of 54 m and an RMSE of 77 m. A 7-point manual fit reached 74 m and 102 m. SIFT failed, at chance level even on pre-aligned tiles. The comparison favours LoFTR, but with 12 points it is not decisive. Affine and projective models performed alike. The printed graticule sits roughly 100 m from WGS 84.

The most useful next steps are independent human re-picking of the control points with coverage in the south, the same protocol applied to more sheets, and local non-linear models to absorb paper distortion and datum effects. Code, provenance records and every intermediate output are available in the repository.

Data and code availability. The repository contains the pipeline, the configuration as frozen, both GCP files with their selection protocol, all metrics and per-point errors, and the script that regenerates every figure and table in this paper

(paper/make_figures.py). Raw inputs are re-downloaded with recorded checksums.

Acknowledgements. Map scan: the Old Survey of India Maps collection compiled by John Brown (Zenodo, CC BY 4.0). Modern data © OpenStreetMap contributors, ODbL.

References

- [1] John Brown. Old Survey of India maps collection. Zenodo, 2023–2026. doi:10.5281/zenodo.10519306; sheet 57 D/11 from record 8388121; CC BY 4.0.
- [2] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91–110, 2004.
- [3] Yao-Yi Chiang, Stefan Leyk, and Craig A. Knoblock. A survey of digital map processing techniques. *ACM Computing Surveys*, 47(1):1:1–1:44, 2014.
- [4] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [5] OpenStreetMap contributors. OpenStreetMap. <https://www.openstreetmap.org>, 2026. Data under the Open Database License; Overpass snapshot of 2026-10-18.
- [6] John P. Snyder. *Map Projections: A Working Manual*. Number 1395 in U.S. Geological Survey Professional Paper. U.S. Government Printing Office, Washington, DC, 1987.
- [7] Edgar Riba, Dmytro Mishkin, Daniel Ponsa, Ethan Rublee, and Gary Bradski. Kornia: an open source differentiable computer vision library for PyTorch. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2020.
- [8] Daniel Barath, Jana Noskova, Maksym Ivashechkin, and Jiri Matas. MAGSAC++, a fast, reliable and accurate robust estimator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [9] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [10] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.
- [11] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.
- [12] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.
- [13] U.S. Bureau of the Budget. United States National Map Accuracy Standards, 1947. Revised June 17, 1947.