Rigorous Projection for Image Stitching: a 3D-Informed Approach for Accurate Panoramic Photogrammetry

Riccardo Roncella [1], Luca Perfetti [2]

[1] Università degli Studi di Parma, Dept. of Engineering and Architecture (DIA), Italy – riccardo.roncella@unipr.it
[2] Università degli Studi di Brescia, Dept. of Civil Engineering, Architecture, Territory, Environment and Mathematics (DICATAM), Italy – luca.perfetti@unibs.it

Keywords: Multi-camera systems, Panoramic imaging, Parallax correction, Structure from Motion, 3D reconstruction.

Abstract

The paper presents a 3D-informed method for generating stitched panoramic images from multi-camera rigs through rigorous reprojection onto a scene model. Unlike conventional and parallax-tolerant stitching, the proposed approach explicitly accounts for camera calibration, relative orientation, and scene geometry, with the aim of reducing parallax effects while preserving metric consistency. The method is tested on confined environments, where non-coincident projection centres make stitching especially critical, and is evaluated with different rig configurations, including systems with both small and large sensor baselines. Two experiments are performed: a stitching-accuracy test against synthetic reference panoramas, and a Structure from Motion (SfM) test comparing rigorous panoramas with raw fisheye processing. Results show that the proposed approach yields geometrically consistent panoramas and substantial gains in processing efficiency, although raw fisheye images still provide the best overall metric performance in the most demanding reconstruction scenarios.

1. Introduction

In recent years, there has been a spread of acquisition systems composed of camera rigs: multi-camera systems enable more efficient and comprehensive image data acquisition, allowing immersive environment navigation experience in the style of Google Street View (Anguelov et al., 2010). Examples include so-called panoramic cameras, such as INSTA360, Leica Pegasus TRK, Mosaic (51, X, XPLOR), Teledyne Ladybug, etc. Such devices, consisting of a rigid body combining two or more fisheye cameras, are generally designed to minimize the distance between the optical centres of the different cameras to minimize parallax effects in panoramic images generation. In other cases, systems are designed with larger inter-sensors baselines (e.g., stereo-cameras, Ant3D (Perfetti et al., 2024), etc.). These systems, make it possible to reconstruct the scene in 3D from a single acquisition, something that would not be possible if all projection centres were coincident.

Classical stitching methods for panoramic images (Szeliski and Shum, 1997) rely on the assumption that all images involved in the stitching share a common projection centre (e.g., GigaPan - Lato et al., 2012). With the advent of panoramic multi-camera systems, the assumption of coincident projection centres is no longer valid. This makes (sometimes) the ratio between the baseline and the camera-object distance non-negligible, especially in narrow or confined environments. As a result, parallax issues become more pronounced especially for rigs with larger inter-camera distances and for scenes captured at short range. So-called “parallax-tolerant” approaches (Zhang and Liu, 2014) have been developed, applying local transformations, especially in the overlap regions between adjacent images, to mitigate parallax artifacts during stitching (Gao et al., 2011; Liao et al., 2025). Although these methods offer a simple, automated, and computationally efficient workflow, they do not perform a rigorous reprojection of image points based on the true camera’s relative orientation and acquisition and scene geometry. Consequently, important geometric information is lost in the final panoramic product. Most commercial (e.g., Agisoft Metashape) and open-source (e.g., COLMAP) photogrammetric software, have recently incorporated the possibility of using spherical or panoramic images (cylindrical, equirectangular, cube-map projections, etc.) in their processing pipelines. Nonetheless, scientific literature has shown that, to achieve the best metric accuracy, it is preferable to work directly with the raw multi-camera images (usually fisheye), while explicitly considering the geometric constraints imposed by the camera rig (Barazzetti et al., 2017; Perfetti et al., 2024).

The present work proposes a novel method for generating panoramic images from a three-dimensional model. Unlike conventional and parallax-tolerant stitching approaches, the proposed method explicitly considers scene geometry together with the known calibration and relative orientation of the cameras to model image formation rigorously, through a projection process conceptually analogous to true-orthophoto generation. Its main objective is not only to reduce visible stitching artifacts, but also to produce panoramas that preserve geometric consistency. Although direct processing of the original raw fisheye images generally remains the preferable solution when the highest metric accuracy is required, rigorously generated panoramas may still offer important practical advantages, including a reduced number of images, simpler image matching and point collimation, and more direct use in visualization-oriented applications. Particular attention is given in the experiments to large-baseline multi-camera systems, such as ANT3D, for which parallax-tolerant stitching is not applicable, and to confined environments, where parallax effects are expected to be most critical.

2. Materials and Methods

2.1 A rigorous 3D-informed approach for panoramic image generation

The method assumes that a three-dimensional representation of the scene is available, even if only approximate. Such a model may be obtained in different ways: during post-processing through photogrammetric block adjustment, either by using the automatically stitched panoramic images affected by distortion, or by processing the original fisheye images. Alternatively, the model may come from other sensors: for example, in mobile mapping applications, a spherical camera may be integrated with a LiDAR unit producing a point cloud in real time. In both scenarios, it is assumed that the Interior Orientations (IO) and Relative Orientations (RO) of all cameras in the rig are known, together with their absolute orientation in the same reference system as the 3D model.

After defining the position from which the panoramic image is to be generated with respect to the multi-camera rig, the projection is performed through a process conceptually similar to true-orthophoto production. For each pixel in the final panorama, the collinearity equations are used to trace the corresponding projection ray through object space until the related point is found on the original multi-camera imagery. The selected centre of the panoramic projection may coincide with the optical centre of one of the cameras, which can act as a master reference, but it may also be placed at the centroid of all camera centres (or at any other user-defined position) to reduce and more evenly distribute parallax and occlusion effects. In more detail, the workflow consists of the following phases:

1. **Definition of the projection geometry and raytracing.** Once the position and orientation of the panoramic image have been chosen (e.g., keeping the panorama fixed in the object reference system, or instead aligning it with the orientation of the camera rig) the projection ray associated with each panorama pixel is extended through object space until it intersects the 3D model.

2. **Back-projection onto the original images.** Starting from the point of intersection on the 3D model, the procedure projects this point back onto each image acquired by the multi-camera system, using the true internal and external orientation parameters of every camera.

3. **(Optional) Occlusion analysis.** Potential occlusions may be detected by checking whether each projection ray intersects the 3D model before reaching the intended object point. When this occurs, the ray is marked as occluded and is not used in the image synthesis.

4. **Selection of the visible image.** If the same object point is observable in more than one image, the algorithm identifies the frame that is expected to provide the most suitable representation. In systems with limited eccentricity this usually corresponds to the image in which the point lies nearer to the image centre, meaning that it is viewed with a smaller angle relative to the optical axis.

5. **(Optional) Radiometric blending.** To obtain smoother transitions in areas where adjacent cameras overlap, the user can define a blending zone within which the radiometric values are combined linearly.

As already mentioned, the complete workflow is strongly analogous to that adopted in true-orthophoto generation. The essential difference is that, in this case, the final product is not mapped onto a plane. Instead, the user can choose a spherical surface (e.g., such as in equirectangular panorama) or a cube-map representation, and the projection onto these surfaces is central rather than parallel.

2.2 Case studies

Multiple tests were carried out applying the proposed approach on two different case studies, deploying various camera systems. The test sites were selected considering narrow environments, where parallax is expected to have a stronger impact on the panoramic stitching, in order to highlight more clearly limitations and strengths of the proposed method.

The first test site consists of a small room measuring 1.8 × 0.8 × 2 m, previously used as a calibration field for panoramic and multi-camera rigs in earlier experiments (Perfetti et al., 2024 – Figure 1, top). The room is built from pink marble blocks, which provide a distinctive and clearly recognizable texture. In addition, circular coded photogrammetric targets were placed on all walls, as well as on the floor and ceiling, to define Ground Control Points (GCPs) and Check Points (CPs). For some of these targets (i.e., those visible from outside the room through the doorway) it was possible to determine their coordinates accurately using a Leica TCR1203 total station. For the remaining targets, a redundant photogrammetric survey was carried out with a Nikon D810 DSLR equipped with a 24 mm lens. In total, 94 coded targets were installed and measured. The purpose of this specific test site is to highlight any geometric discrepancies in the proposed rigorous reprojection approach by relying on an extremely redundant and highly accurate metric reference provided by the coded targets (i.e., with an estimated positioning precision of 0.88 mm). In this way, it becomes possible to reveal potential systematic effects or artifacts that may arise, for instance, from an imperfect estimation of the IO parameters, the EO (External Orientation) parameters, or the relative pose RO between the sensors of the camera rig.

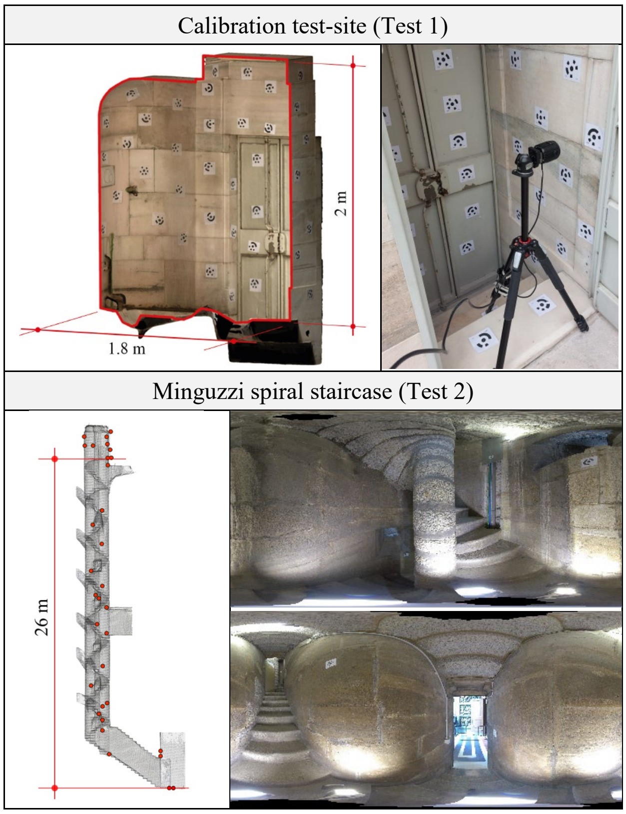

Figure 1. Schemes and images of the case of study environments used within the tests: calibration test-site (top); Minguzzi spiral staircase (bottom).

The second test site is a challenging architectural setting: the Minguzzi spiral staircase inside Milan Cathedral, Italy (Figure 1, bottom). This staircase has already been adopted in previous studies as an experimental site (Bruno et al., 2024), owing to the considerable difficulties it presents for any surveying approach. Its helical geometry is especially problematic, since the rotational development of the staircase prevents the upper portions from being imaged from lower viewpoints, thereby severely restricting visibility from any position along the path.

The staircase, made of marble, is located inside the right pylon of the cathedral’s main façade and extends for roughly 26 m in height. It is also characterized by particularly unfavourable lighting conditions: the environment is extremely dark, the artificial lighting is poor, and only a small number of openings connect the space to the outside. At the centre of the structure there is a marble column, approximately 40 cm in diameter, around which the stairway winds upward. The usable passage is extremely confined, with a width of only about 70 cm.

As in the previous case, the establishment of accurate GCPs along the staircase by means of a total station would be impractical, because of the narrowness of the space, the constrained survey geometry, and the very limited intervisibility between consecutive instrument stations. For this reason, the survey was performed using two cameras: a Canon 5D Mark III equipped with a Sigma 8 mm circular fisheye lens, and a Nikon D810 with a Sigma 12–24 mm rectilinear lens. The overall estimated accuracy of the survey was about 1 cm. This result allowed to identify and measure additional natural points on the structure to be used as CPs in subsequent experiments.

2.3 Equipment

Three distinct imaging systems (Figure 2) – MG1, ANT3D and INSTA360 Pro2 (hereby referred to as INSTA360) – were employed in the experimental campaign to assess how the spherical images generated with the proposed approach behave under different camera-rig configurations.

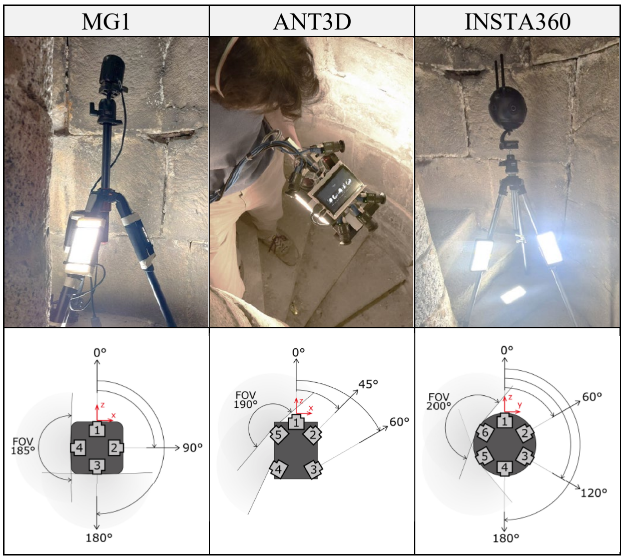

Figure 2. Camera systems confronted within the investigation: MG1 by Gexcel srl (left), ANT3D multi-camera (centre), INSTA360 (right). Images acquired during the survey activities for Test 2 at top row.

The main technical specifications of each device are summarized in Table 1. MG1 and INSTA360 are 360° panoramic rigs with limited (MG1) to moderate (INSTA360) inter-sensor spacing, able to generate parallax-tollerant panoramas. ANT3D, conversely, is a multi-camera system conceived for kinematic/handheld capture, prioritizing photogrammetric 3D reconstruction and thus designed to maximize inter-sensor baselines. Notably, the rear hemisphere is excluded from capture to avoid framing the operator.

It should be pointed out that, for the first test site, the INSTA360 dataset at the time of writing was not available.

Camera system:	MG1	ANT3D	INSTA360
Camera type	360° camera	Multi-cam.	360° camera
Max baseline	6 cm	20 cm	12 cm
N° sensors	4	5	6
Image res.	2530 x 4048	2448 x 2048	4000 x 3000
Pixel size	1.55 µm	3.45 µm	1.88 µm
Focal length	2.2 mm	2.7 mm	1.88 mm
FOV	185°	185°	200°
Pano res.	8192 x 4096	-	7680×3840

Table 1. Main specs of the employed camera systems.

2.4 Test methodology

Two tests were conducted to evaluate the proposed approach and to answer three main questions: *i)* how far Rigorous panoramas are geometrically more correct than conventional stitched panoramas, *ii)* whether they are accurate enough for use in Structure-from-Motion (SfM), and *iii)* whether the gain in processing efficiency can justify any loss in metric accuracy with respect to raw-image processing.

1. **Stitching accuracy (method validation):** comparison of Rigorous Panoramas against reference Synthetic Panoramas by measuring target displacements in image space to reveal residual parallax and other distortions.
2. **SfM performance (application validation):** evaluation of Rigorous panoramas’ SfM outcomes in terms of *i)* block orientation accuracy (via Check Points, CPs) and *ii)* processing time. This tests also aim at investigating the impact of the 3D model resolution used for panorama generation by comparing datasets produced from both high-resolution and low-resolution models. Results are contrasted with those obtained by directly processing the multi-camera Raw fisheye images.

2.4.1 Test 1 – Stitching accuracy (method validation)

This test was carried out employing two camera systems: ANT3D and MG1, offering a diverse dataset to assess the effectiveness of the proposed method on both short- and long-baseline’s multi-camera. For the MG1 system, we additionally compared Parallax-Tolerant panoramas with the Synthetic panoramas to provide a direct benchmark against a conventional stitching strategy. A comprehensive photogrammetric acquisition was carried out within the calibration test-site. To ensure calibration parameters decoupling and to avoid systematic effects, the multi-cameras have been moved and rotated around all axes. In total, the acquisition consists of respectively 83 poses for MG1 (83 x 4 = 332 images) and 250 poses for ANT3D (373 x 5 = 1865 images), forming the original **RF** (raw fisheye) image datasets. From these, 2 sets of overlapping panoramas, with identical EO, were generated per camera system: *i)* **RP** (rigorous panoramas): obtained by applying the proposed approach using the source 3D model of the test-site, *ii)* **SP** (synthetic panoramas) – reference: generated with the same EO as the Rigorous panoramas by directly rendering equirectangular views of the reference textured 3D model at matching resolution.

The test workflow comprises the following steps (Figure 3):

Generation of the reference textured 3D model of the test-site by carrying out image-base modelling using the redundant photogrammetric block acquired with the Nikon D810 DSLR and topographic control measurements. This yielded both a precisely oriented image block of the DSLR dataset, and a high-resolution 3D mesh with a consistent texture.

Orientation of raw multi-camera fisheye images. This step was performed separately for each tested camera system. IO and EO of each of the Raw fisheye datasets were estimated via incremental SfM, connecting the new image block to the previously oriented DSLR block kept fixed. This constrains the multi-camera network to the more stable DSLR solution, minimising misalignment with respect to the reference 3D model.

Rigorous panorama generation. On the basis of the estimated orientations and the reference 3D model, Rigorous panoramas were generated at the barycentre of the rig at each pose, aligning the spherical projection at the rig orientation.

Synthetic panorama generation. The reference data was generated by directly mapping an equirectangular view of the reference textured 3D model using the same EO and image resolution as for the corresponding Rigorous panoramas.

Mismatch measurement. For each system considered in the test, the image coordinates of the targets were automatically extracted using Metashape. For each target in each corresponding image pair (RP vs SP) the distance in image space was computed. Comparing the target image coordinates allows the geometric deviation of each panorama from the optimal one to be quantified, assuming the synthetic images as a gold-standard reference within the accuracy limits of the 3D model used to generate them.

When computing the residuals, the characteristics of the equirectangular projection must be considered: longitudinal distortion is zero at the equator and increases progressively toward the poles. Rather than measuring the distance directly on the equirectangular image plane, the angle between the two corresponding projection rays was computed. This angular difference was then converted into the equivalent pixel distance at the equator, to provide a normalized pixel unit quantity.

Datapoints of all target pairs in all images have been aggregated within a single equirectangular heatmap representing average mismatch (interpolating residuals between sparse points).

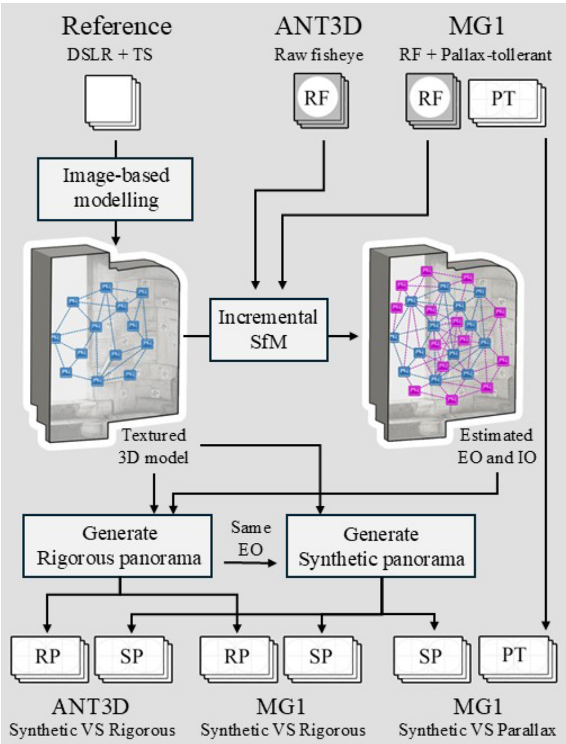

Figure 3. Processing workflow of Test 1

2.4.2 Test 2 – SfM performance (application validation)

Within the environment of the Minguzzi spiral staircase, three survey acquisition have been carried out with the tested camera systems. The panoramic multi-cameras: MG1 and INSTA360 were kept on tripod with the rig Z axis kept close to the vertical, images were acquired at each staircase step, plus few views connecting at the bottom floor and at the top terrace for a total of respectively 174 poses for MG1 and 173 poses for the INSTA360. With ANT3D, the acquisition was instead handheld and carried out dynamically acquiring the staircase both ways, top to bottom and then bottom to top, for a total of 1205 poses. Illumination in the dark environment is provided by continuous LED lights, incorporated within the handheld multi-camera for ANT3D, and attached to the tripod for the panoramic cameras. These three acquisitions formed the multi-camera Raw Fisheye datasets (**RF**), already subject of a previous investigation regarding optimal processing strategies for multi-camera rigs (Perfetti et al., 2024). The original datasets were completely reprocessed both using the optimal strategy identified in the previous experimentation applied on RF, and by applying the proposed approach to generate Rigorous panoramas before executing a new SfM process.

The metrics used to evaluate SfM performance are CPs RMSE, CPs maximum error, and the processing time recorded running all tests on the same workstation (CPU Intel Xeon 3104 @1.70GHz, RAM 32GB, GPU NVIDIA Quadro P1000). The accuracy evaluation has been conducted in two ways, applying two distinct ground control solutions:

GCPs solution A – ground control points are placed solely at the base of the staircases allowing for drift error to cumulate along the staircase extension. This solution allows evaluating the approaches in drift-prone scenarios and unideal conditions.

GCPs solution B – ground control points are placed both at the base and at the top of the staircase, at entrance and exit. This solution replicates a common survey scenario for confined spaces and multi-camera rigs with good ground control.

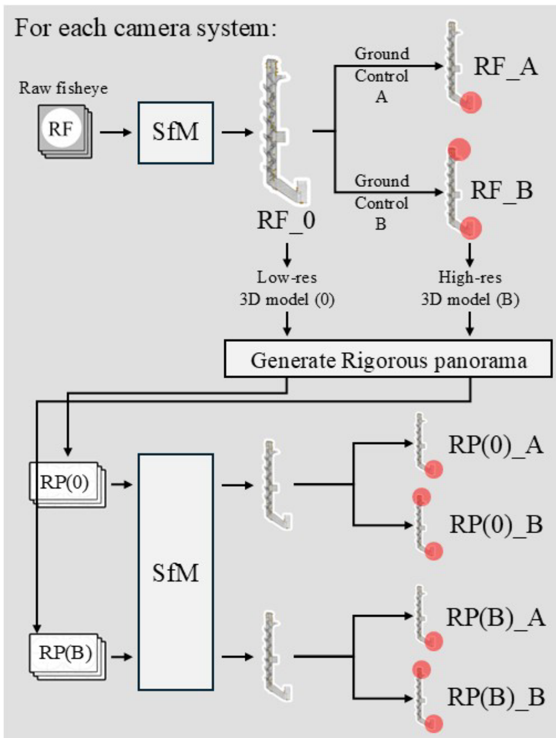

Figure 4. Processing outline of Test 2.

The test outline is as follows (Figure 4):

Computing SfM solution of the raw fisheye images: the three raw datasets have been processed in Agisoft Metashape (v.2.2.1) estimating rig RO and IO in the bundle adjustment. After orientation, each test has been differentiated into the different constraint solutions obtaining 3 results for each camera system: RF_0 (raw results with no constraints – required in the next step), and, RF_A and RF_B, applying the two GCPs solutions.

Generating the 3D models, for each camera system two 3D models have been constructed: *i*) a low-resolution simplified 3D model based on the results of RF_0, obtained by directly meshing the tie-points cloud, and *ii*) a high-resolution precise dense 3D model obtained from the result of RF_B (optimal solution), free from significant meshing artifacts.

Generating rigorous panoramas, for each camera system, the proposed approach has been applied on RF_0 and RF_B using the relative 3D models and the estimated rig's IO and EO. Rigorous Panoramas (**RP**) have been generated at the rig's barycentre aligned with the rig's orientation. This generated two new image datasets for each camera system: RP(0) and RP(B).

Computing SfM of the Rigorous panorama, the new 6 dataset have been processed applying the spherical camera model, later differentiating each RP dataset into the two GCPs solutions: RP(0)_A, RP(0)_B and RP(B)_A and RP(B)_B (see Figure 4).

3. Results

3.1 Results of Test 1

The purpose of the test is to highlight geometric distortions in the projection of image features introduced during panorama generation. MG1 offered the opportunity of comparing the parallax tolerant panoramas (PT) generated automatically, with rigorous panoramas (RP) generated via the proposed method. Two matching sets of synthetic panoramas (SP, see Section 2.4.1) were generated with the same resolution (8192 x 4096) and EO as the PT and RP datasets (see Figure 3).

For all image sets (i.e., RP, PT, and the two matching SPs) the target centres were automatically extracted, yielding around 5284 image points for the synthetic image sets, 4883 for RPs, and 4735 for the PT set, distributed over 83 frames and corresponding to 88 unique targets. To verify the geometric consistency of the SPs sets, the two synthetic image blocks were oriented by means of SfM to check the mean target reprojection residuals. This was found to be slightly below 0.2 pixels, a value considered representative of the order of magnitude of the uncertainty associated with automatic target localization.

Applying the same procedure to the tested RPs, led to higher reprojection residuals, approximately 1.2 pixels. This confirms, as expected, that when real data are used, the panoramic projection is influenced by residual factors that are not modelled perfectly (e.g., such as lens distortion), whereas these sources of error are absent in the synthetic images. Figure 5 shows an example of a synthetic equirectangular image use for the test and a close-up visual overlapping of the synthetic data with the rigorous and parallax-tolerant ones.

For the ANT3D dataset, the output of the proposed approach consists of panoramas with a resolution of 5760 × 2880 pixels. The different resolutions adopted for MG1 and ANT3D reflect the characteristics of the respective acquisition systems: ANT3D's cameras have a lower resolution than MG1's. The selected panorama resolutions were chosen to ensure an approximately comparable GSD (Ground Sampling Distance) between the equirectangular images and the raw original images. Given the large distance between the projection centres of the different sensors, generating panoramas with parallax-tolerant techniques was considered unnecessary.

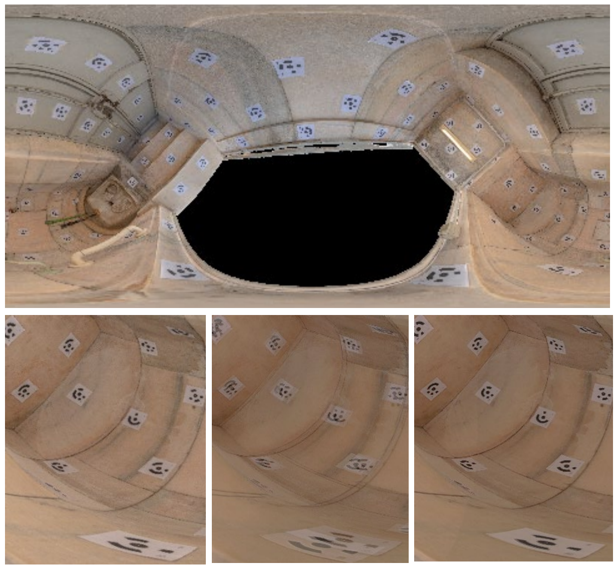

Figure 5. Stitching comparison: (*Top*) A reference Synthetic panorama. (*Bottom*) A zoomed-in comparison: Synthetic (*left*), Parallax-tolerant overlapped with Synthetic (*centre*), and Rigorous overlapped with Synthetic (*right*).

The two panoramic image blocks (RP and SP, 250 frames each) allowed the extraction of a total of 17,054 and 10,534 image points for the synthetic and rigorous sets, respectively. The reprojection residuals obtained from the orientation of the two blocks are fully consistent with the previous case, with a mean value of 0.15 pixels for the synthetic images and 1.2 pixels for the real ones. This shows that the rigorous panorama generation method provides results of comparable quality regardless of the sensor employed, even when the distance between the projection centres of the different systems varies substantially.

A more direct assessment of the metric correctness of the different panorama-generation methods can be obtained by comparing the image coordinates of the targets extracted from the real panoramas (RP and PT) with the reference coordinates measured on the synthetic ones. For the MG1 system, the comparison of 4,832 corresponding points (4642 for the parallax-tolerant) yielded a mean positional discrepancy, normalized at the equator as described in Section 2.4, of 0.92 pixels (sigma = 0.64 pixels) for the RPs, and 21.2 pixels (sigma = 10.9 pixels) for the PT panoramas. The improvement in geometric accuracy is evident and, as expected, the results confirm that PT panoramas are not advisable for photogrammetric purposes, especially in confined environments. Figure 6 shows the magnitude of these discrepancies in false colours.

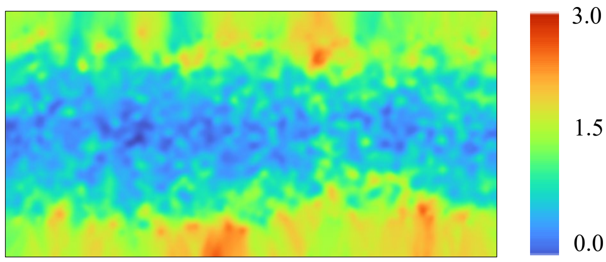

Figure 6. MG1 heatmap representation of image coordinate residuals (pixel units) after rigorous panoramic projection.

It is clearly visible that the average projection accuracy is mostly uniform, in the equatorial region at the different longitudes, with values ranging from 0 up to approximately 1 pixel. On the contrary, at higher latitudes, where the equirectangular projection produces significant deformation effects, the quality of the identification of the image points (i.e., the targets) is worse and can lead to some outlier/gross-error.

For ANT3D, the same comparison between rigorous projection and synthetic panoramas (10,439 correspondences) produced a mean positional discrepancy of 0.56 pixels (sigma = 0.43 pixels), which is significantly better than the corresponding MG1 result. It is important to note that in this case the panoramic image resolution is lower by about a factor of 1.4 than for MG1. As a result, in angular terms, the two systems do not yield critically different results, although ANT3D still retains a slight advantage (rescaling the residuals gives a mean value of 0.8 pixels). The better sensors quality of ANT3D, combined with greater overlap between sensors and a more robust rig geometry likely allows a more accurate geometric calibration and a better estimation of distortion characteristics, which in turn directly improves the geometric quality of the resulting panoramas.

3.2 Results of Test 2

Across both ground-control strategies, reconstructions from the raw multi-camera fisheye imagery (RF) systematically show lower CPs RMSE than those from Rigorous panoramas (RP), suggesting that the panorama-based workflow is more fragile in this challenging environment. With solution A (GCPs only at the base – Table 2), RF delivers RMSEs of 0.048 m (MG1), 0.067 m (ANT3D), and 0.066 m (INSTA360), while RP ranges from 0.055-0.134 m (MG1/ANT3D) up to the outlier result of 0.400 m for INSTA360. With solution B (GCPs at base and top – Table 3), the gap narrows: RF achieves 0.008-0.015 m RMSE, while RP shows only slight worse result in the 0.025-0.034 m range, a level that is acceptable for many architectural applications in confined spaces.

The most significant differences appear in the Z component, main direction of development of the staircase: RP tends to accumulate significantly larger vertical errors than RF in solution A. In solution B, errors tighten, but RP still shows higher Z RMSEs.

This is further illustrated by the qualitative drift plots in Figure 7, which, for each of the tests, show the drift trajectory as a spline interpolation of CPs errors. CPs reference location has been simplified projecting them on the vertical line (black segment) by taking only their Z coordinate. Starting from the simplified reference CPs locations, deviated points (grey points in the plots) have been found by adding to each CP the corresponding error vector. Scale factors have been applied for visual enhancing of deviations: for solution A, errors were scaled x50 in XY and x10 in Z; for solution B, errors were scaled homogeneously x100 in XYZ. Looking at MG1 and ANT3D in solution A, we observe similar XY deviation between RFs and RPs, this constitutes a translational drift that cumulates towards the top which affect RF and RPs similarly. RPs, however, also show higher Z deviations, especially visible in the ANT3D plots, and noticeably higher error spread increasing from the bottom to the top, especially visible in the CPs group at the top. This behaviour suggests that RPs are more affected by an additional scale drift that, as translational drift, cumulates toward the top and mainly shows as Z error since the Z axis is the main direction of development of the test environment.

RMSEs of the INSTA360 RPs, under solution A, are clear outliers, especially considering that the INSTA360 RF results are consistent with those of the other camera systems. On closer inspection, these outcomes can be attributed to differences in sensor orientation within the multi-camera rigs: in MG1 and ANT3D, the sensors' up-axes (-v) are roughly aligned with the multi-camera up-axis (z), whereas in INSTA360 the sensors' up-axes lie on the horizontal plane of the multi-camera system, and it is the sensors' u-axes that are roughly aligned with the multi-camera z-axis. This difference is significant. When the proposed method is applied with the option to align panoramas to the camera system, the generated equirectangular projection is based on the orientation of the first camera in the rig. For MG1 and ANT3D, this aligns the projection zenith with the multi-camera up-axis; for INSTA360, by contrast, the projection is rotated by 90°, effectively aligning the projection zenith with a direction in the multi-camera horizontal plane.

As a result, while in MG1 and ANT3D the higher-resolution portions of the Raw fisheye images are mapped to the least-distorted region of the equirectangular projection, in INSTA360 they are mostly mapped to the top and bottom borders, where distortion is greatest.

The outlier results of INSTA360 RPs suggest that the orientation of the panorama with respect to the multi-camera system plays a significant role in SfM accuracy. We can expect that optimal results should be reached by generating panoramas oriented to the multi-camera reference system with panoramas' zenith corresponding to the multi-camera up-axis (as in MG1 and ANT3D RPs), however, further investigations are required to examine the relationship between RP projection orientation and SfM outcomes.

Regarding the resolution of the 3D model used for panorama generation, switching from an approximate low-resolution to a high-resolution model does not lead to consistent, significant accuracy gains across systems. Under solution A, Z-RMSE improves for ANT3D (0.124 vs 0.088 m) but degrades for MG1 (0.026 vs 0.075 m). On average, reprojection residuals in pixels are slightly lower with the high-res model than with the low-res model (e.g., MG1: 2.0 vs 1.7 pixels; ANT3D: 2.0 vs 1.3 pixels), in line with the expectation that a cleaner surface model reduces small local misprojections. However, the magnitude of these differences is modest and varies by rig. Collectively, the data show no significant advantage in final metric accuracy in switching from an approximate 3D model to a refined one, this is clearly illustrated in the qualitative drift plots (Figure 7) where respective outcomes of RP(0)s and RP(B)s closely match. This suggests that a rough scene awareness is sufficient to effectively negate parallax effects and that only marginal gains can be excepted by further improving 3D model fidelity.

For what concerns processing times (see Table 4), the panorama-based pipeline is consistently faster, due to a drastic reduction in the number of images to be processed (one spherical frame versus multiple fisheyes) and a corresponding drop in tie-point counts. SfM times decrease by ~4x for MG1, ~3x for ANT3D, and ~6.5x for INSTA360. Considering the rate of reduction of SfM processing times between RFs and RPs, the Rigorous panorama generation and additional processing steps required after initial SfM of the Raw fisheye images, can already be justified by the expected reduction in processing times for computing further deliverables such as point clouds, 3D models and orthophotos, for which a similar speed-up rate is expected. Moreover, in scenarios where scene geometry can be estimated in real-time (e.g., camera coupled with LiDAR), and thus the approximate scene awareness is directly available, the proposed approach enables rapid image-based modelling able to achieve accuracies comparable to those of raw fisheye's outcomes (especially where robust constraints are available) at a fraction of the processing time (1/3rd to 1/6th).

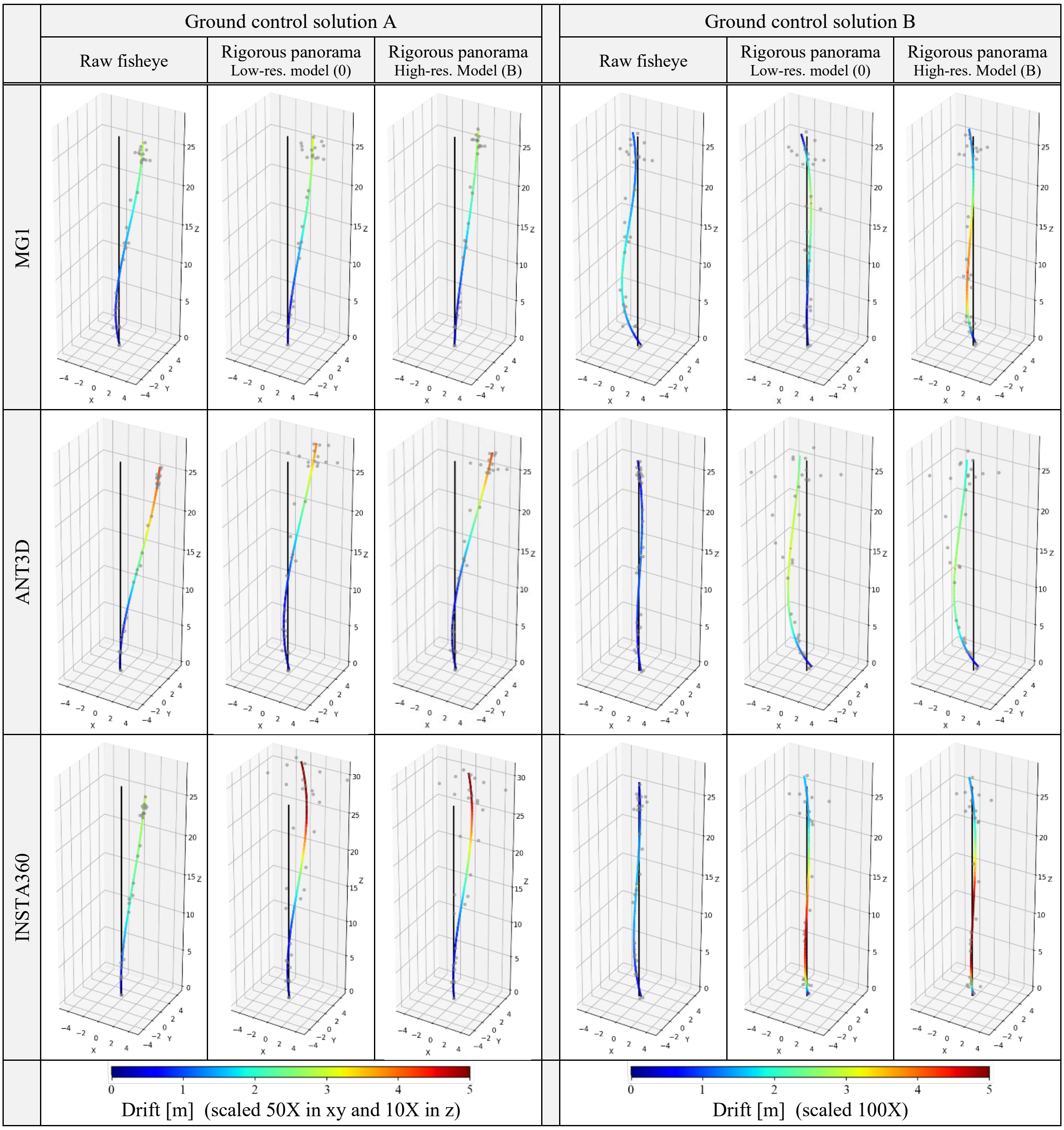

Figure 7. Qualitative drift plots.

Ground control – solution A								
Dataset		Reproj. Error [pixel]	CPs RMSE [m]				
			X	Y	Z	XYZ	max XYZ
Raw Fisheye **RF**	MG1	1.2	0.041	0.013	0.020	0.048	0.075
	ANT3D	1.2	0.063	0.018	0.011	0.067	0.090
	INSTA360	1.1	0.044	0.013	0.048	0.066	0.090
Rigorous Panorama **RP**	MG1 (0)	2.0	0.046	0.014	0.026	0.055	0.099
	MG1 (B)	1.8	0.038	0.016	0.075	0.086	0.129
	ANT3D (0)	2.0	0.034	0.039	0.124	0.134	0.193
	ANT3D (B)	1.3	0.052	0.030	0.088	0.106	0.153
	INSTA360 (0)	1.2	0.048	0.025	0.397	0.400	0.625
	INSTA360 (B)	1.0	0.046	0.022	0.316	0.321	0.498

Table 2. Accuracy evaluation results of all tests with ground control solution A.

Ground control – solution B							
Dataset		Reproj. Error [pixel]	CPs RMSE [m]				
			X	Y	Z	XYZ	max XYZ
Raw Fisheye **RF**	MG1	1.2	0.009	0.011	0.005	0.015	0.024
	ANT3D	1.3	0.004	0.005	0.005	0.008	0.016
	INSTA360	1.1	0.004	0.008	0.005	0.011	0.020
Rigorous Panorama **RP**	MG1 (0)	1.9	0.007	0.010	0.022	0.025	0.111
	MG1 (B)	1.7	0.006	0.017	0.019	0.026	0.043
	ANT3D (0)	2.0	0.017	0.023	0.008	0.030	0.060
	ANT3D (B)	1.3	0.014	0.020	0.004	0.025	0.049
	INSTA360 (0)	1.1	0.012	0.017	0.022	0.030	0.048
	INSTA360 (B)	1.0	0.011	0.015	0.029	0.034	0.058

Table 3. Accuracy evaluation results of all tests with ground control solution B.

Dataset		Processing time	N° of poses	N° of images	Image res.	Tie-points
Raw Fisheye	MG1	01:36	174	696	2530 x 4048	1.0 M
	Ant3D	14:21	1205	6,025	2448 x 2048	5.4 M
	INSTA360	02:48	173	1,038	4000 x 3000	2.0 M
Rigorous panorama	MG1	00:26	174	174	8192 x 4096	0.2 M
	Ant3D	05:10	1205	1205	5760 x 2880	1.6 M
	INSTA360	00:26	173	173	7680 x 3840	0.2 M

Table 4. Processing times and info of all tests.

4. Conclusion and future works

This work presented a scene-informed projection method for generating rigorous panoramas from the raw fisheye images of multi-camera rigs, with the aim of overcoming parallax effects typical of state-of-the-art parallax-tolerant stitching methods that hinder accurate scene geometry-based modelling. Two tests were conducted: *i*) a stitching-accuracy assessment (Test 1); and *ii*) a SfM performances assessment (Test 2).

Results indicate that the proposed method effectively resolve parallax-related distortions enabling significant efficiency gains in the SfM processing compared to processing the raw images. From a metric standpoint, raw fisheye processing remains the most accurate solution in the most demanding conditions: raw-image reconstructions mainly show translational drift, whereas rigorous-panorama reconstructions exhibit an additional scale-drift component. With well-distributed constraints, the panorama-based workflow reaches accuracy levels that are acceptable for many architectural applications. Contrary to what was expected, finds demonstrated how the resolution of the 3D model underlying the rigorous projection does not play a significant role in influencing SfM outcomes – thus, in prospective, scene geometry could be provided in real-time by active sensors. Instead, results highlighted two aspects that effectively influence SfM performance, i.e., the orientation of the panorama relative to the multi-camera reference frame (as shown by the degraded outcomes in sub-optimal configurations), and the observed additional scale drift, which may be tackled by introducing external scale information (e.g., from LiDAR).

To conclude, panoramic cameras and panoramic imagery promise more effective image networks (less images and faster processing), especially in narrow environments. However, readily available parallax-tolerant panoramas are known to be unreliable and unsuitable for image-based modelling due to the unavoidable parallax effects. The present investigation and the proposed method proved that effective parallax-corrected projection can be obtained even from approximate scene awareness and estimated camera orientations. Rigorously generated panoramas proved to be both, effective in image network simplification and processing times, and, suited for image-based modelling achieving adequate accuracies. As for now, rigorous panoramas generation is an added step after initial SfM processing, future works will investigate the coupling of panoramic camera with LiDAR sensors to skip the initial processing and to counteract the additional scale drift observed.

References

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., Weaver, J., 2010. Google street view: Capturing the world at street level. *Computer,* vol. 43, pp. 32–38.

Barazzetti, L., Previtali, M., Roncoroni, F., 2017. 3D Modelling with the Samsung Gear 360. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* vol. XLII-2-W3, pp. 85–90.

Bruno, N., Perfetti, L., Fassi, F., Roncella, R., 2024. Photogrammetric survey of narrow spaces in cultural heritage: Comparison of two multi-camera approaches. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-W4, 87-94.

Gao, J., Kim, S.J., Brown, M.S., 2011. Constructing image panoramas using dual-homography warping. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* pp. 49–56.

Lato, M.J., Bevan, G., Fergusson, M., 2012. Gigapixel Imaging and Photogrammetry: Development of a New Long Range Remote Imaging Technique. *Remote Sensing*, vol. 4, pp. 3006–3021.

Liao, T., Wang, C., Li, L., Liu, G., Li, N., 2025. Parallax-tolerant image stitching via segmentation-guided multi-homography warping. *Signal Processing,* 230, 109860.

Perfetti, L., Bruno, N., Roncella, R., 2024. Multi-Camera Rig and Spherical Camera Assessment for Indoor Surveys in Complex Spaces. *Remote Sensing* 2024, vol. 16, pp. 4505 16.

Szeliski, R., Shum, H.Y., 1997. Creating full view panoramic image mosaics and environment maps. *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH*, pp. 251–258.

Zhang, F., Liu, F., 2014. Parallax-tolerant Image Stitching, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 3262–3269.