The FINE Benchmark (Fisheye Indoor Narrow spaces Evaluation)

Many researchers are working with fisheye photogrammetry, low-cost image-based sensors and SLAM technologies to define reliable approaches to survey indoor and complex spaces. The different approaches become challenging in narrow environments, such as underground tunnels, mining areas, stairways, etc.
The FINE Benchmark wants to provide a set of data to evaluate the performances of different image-based processing methods when surveying complex spaces (Fig. 1). It offers the opportunity to (i) tackle one or more topics that influence the final accuracy of the results and (ii) share and challenge your approach with the community.
Participants from academia, research institutes and companies are invited to participate to use the benchmark data and demonstrate their tools, programmes, processing methods and developments in elaborating image sets of different types and range-based clouds for the 3D reconstruction of narrow spaces. Presentations dealing with the benchmark data will be included in a special session of the 3D-ARCH 2019 event - no paper required. All participants will jointly prepare a journal paper after the event.
If you are willing to participate to the FINE Benchmark, please contact Fabio Remondino (remondino@fbk.eu) or Francesco Fassi (francesco.fassi@polimi.it) to receive the data.

1. The case study

The benchmark data are acquired in the historical rooms and tunnels of the San Vigilio Castle, located at the very top of Città Alta (Bergamo, Italy) – Figure 1.

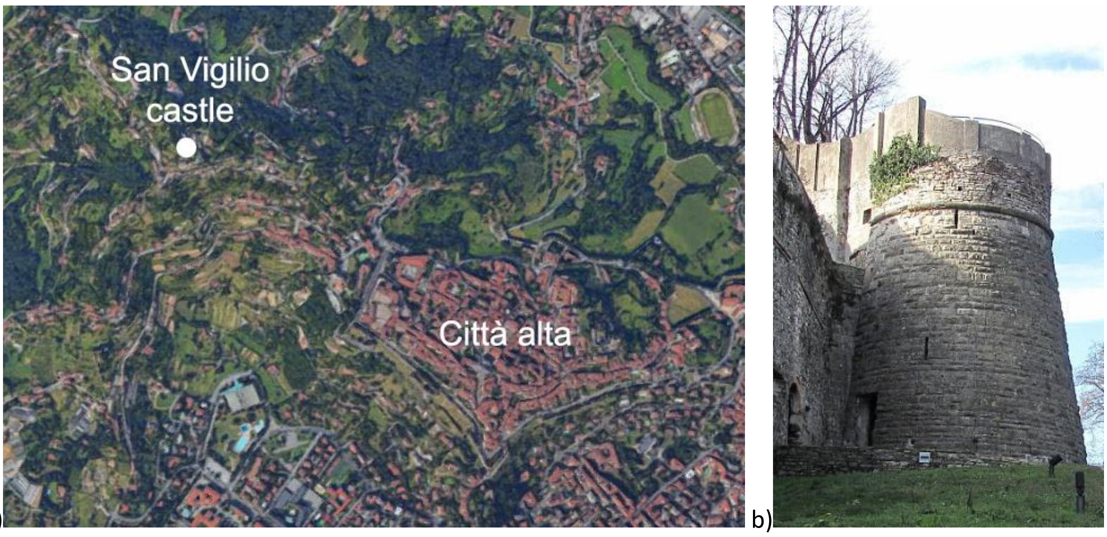
a)  b)

Figure 1: a) Position of the San Vigilio Castle in the upper part ("Città Alta") of Bergamo (Italy) and b) the entry of the tower leading to the interior passage and underground tunnels.

The benchmark is composed of data acquired in:

- an underground dark tunnel (circa 80 meters long) excavated in the rock, with a muddy floor, humid walls with some areas with a ceiling lower than 1.5 metres (Fig. 2).

Figure 2: Some view of the underground rock tunnels starting from the tower room.

- a tower room with a circular / semi-spherical shape.

Figure 3: Some views of the tower room which connect the underground rock tunnel and the upper passage.

- an interior passage (Fig. 4), starting from the tower's ground floor and leading to the castle's upper part, constituted of staircases, planar surfaces, sharp edges, walls with squared rock blocks and quite uniform texture.

Figure 4: Some views of the passage which goes from the tower room to the upper part of the castle.

2. The acquired data

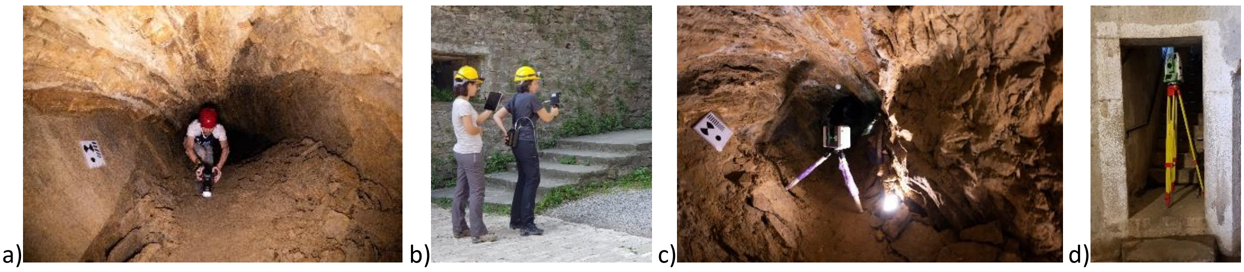


Figure 5: Some views of the data acquisition campaigns: fish-eye photogrammetry (a), hand-held (b) and static (c) laser scanning, topography (d).

5 survey acquisitions were performed on site:

1. a photogrammetric survey using a DSLR camera with Fisheye-lens:
2. a photogrammetric survey using a multi-camera Rig with 6 Gopro;
3. a hand-held mobile laser scanning survey using a ZEB-Revo Geoslam;
4. a static laser scanning survey using a Leica RTC360;
5. a topographic survey with a Leica TCRA1203 total stations and 9 stations.

2.1 Photogrammetry: Full frame DSLR – 8mm equisolid fisheye

The benchmark releases three datasets acquired with the full frame camera Canon 5D MKIII coupled with the fisheye lens Sigma 8mm f/3.5 EX DG:

1. One datasets for the underground tunnel ("Tunnel")
2. One dataset for tower interiors ("Tower")
3. One dataset for camera calibration ("Calibration").

All the images were captured hand-holding the camera and using a speedlight to illuminate the scene.
The capturing geometry used for the tunnel acquisition consists of 5 pictures taken trying to respect a 1:1 base/distance ratio (Fig. 6):

a) 2 photos taken from the very closest position to the two sides of the tunnel by pointing toward the opposite side,
b) 1 photo taken roughly 20cm from the ground upwards,
c) 2 photos from a lower position by pointing up with the camera tilted roughly 45° degrees both ahead and backwards.

For practical reasons the capturing geometry was not repeated sequentially but instead the pictures were acquired type-wise. Occasionally, extra pictures were added for more articulated areas like sharp turns, corners and narrow connections.

Figure 6: Fisheye acquisitions in the underground tunnel, with the used 5 camera position geometry.

For the tower acquisitions, the capturing geometry was kept rather similar, extra pictures were taken for the staircases to take into consideration the steps. In the two larger rooms we followed a classical indoor acquisition.
The third dataset provides pictures of a 3D calibration testfield created onsite and acquired immediately after the survey (Fig. 7). The dataset was acquired in a classical way by moving and rotating the camera around the testfield.

Figure 7: A fisheye image of the onsite calibration testfield (left) and the multi-camera rig acquisition of the testfield.

2.2 Multi-camera rig – six Gopros

An array of action cameras was used to perform a fast video acquisition of both tunnel and tower areas. The rig consists of six GoPro cameras, five Hero3 and one Hero4 (G5) mounted rigidly on a rectangular aluminium structure (Fig. 8). Continuous light is provided by two LED illuminators mounted on the back.
The rig was designed in order to have a sufficient base-distance between the six cameras in relation with the width of the narrow passages to be able to reconstruct the object geometry at every single position of the rig. Two cameras are mounted on the top (G6) and on the bottom (G5) of the structure, tilted roughly 45° degrees downwards and upwards, respectively. Four cameras were mounted on the sides of the rig, two of them (G1, G2) in a convergent manner oriented horizontally, and two in a divergent manner (G3, G4) oriented vertically.

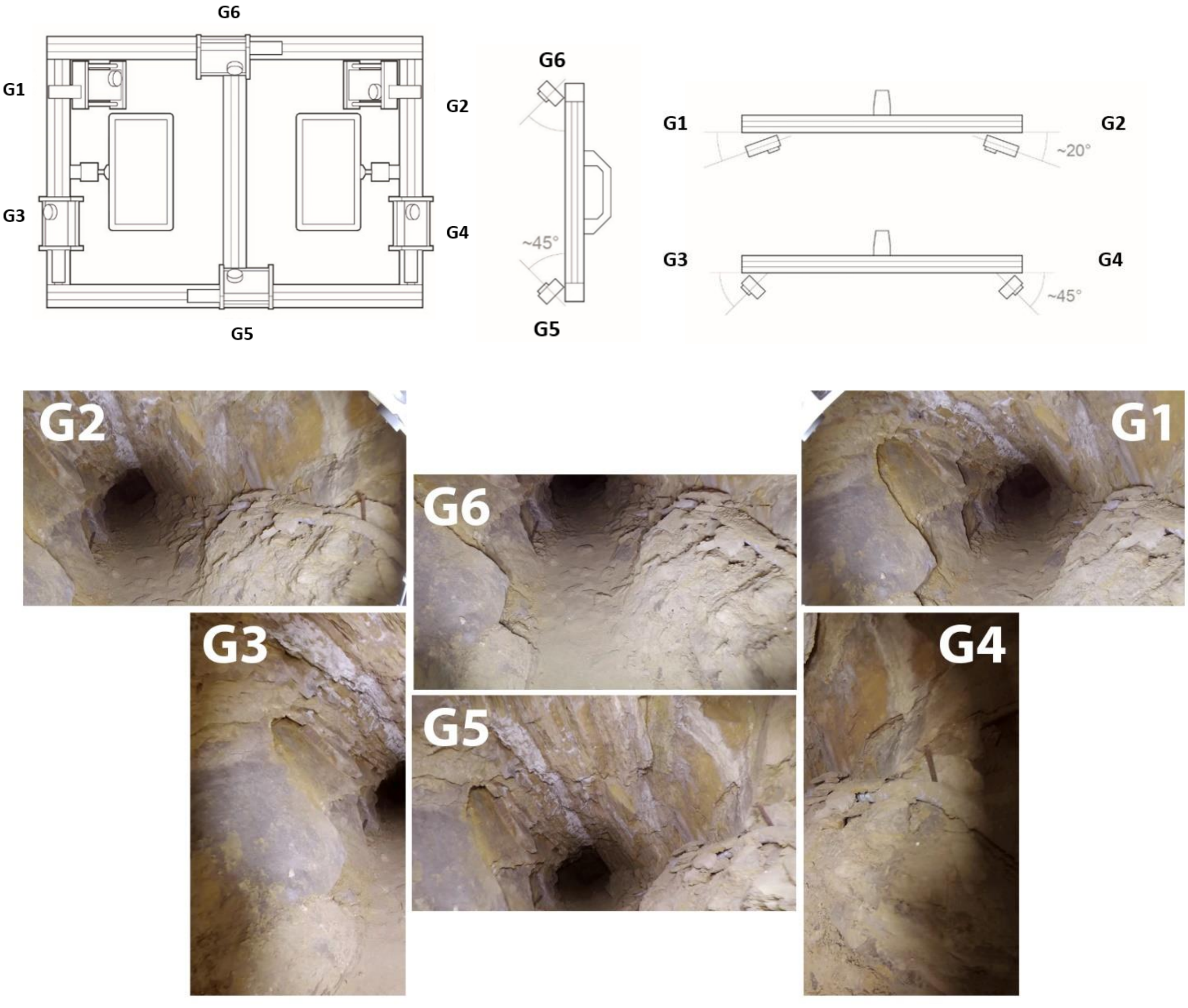


Figure 8: The Gopro array schema (top) and an example of the 6 views inside the underground tunnel (bottom).

The benchmark provides three videos per camera:

1. one video for the tunnel acquisition: the videos start inside the tunnel, they proceed to the end of the passage and back to the starting point. Then the acquisition continues outside the tunnel providing a connection with the entrance of the tower through a very narrow manhole. The camera goes out of the manhole, acquire the entrance room and goes down into the manhole again. The videos end with the acquisition of the calibration testfield used also for Fisheye photogrammetry (Section 2.1).

2. one video for the tower acquisition: the tower acquisition starts at the entrance room of the tower, it goes one way up till the upper exit and then down again, it continues outside of the tower and finishes with the acquisition of the calibration test-field used for Fisheye photogrammetry (Section 2.1).
3. one video for single camera calibration purposes: it was performed in the lab a few days after the survey and using a dedicated testfield (Fig 9).

Figure 9: The testfield in the lab to calibrate the Gopro cameras.

The acquisition of the onsite testfield may be used to calibrate the relative distances and orientations between the cameras. <u>The videos are not synchronized</u>.

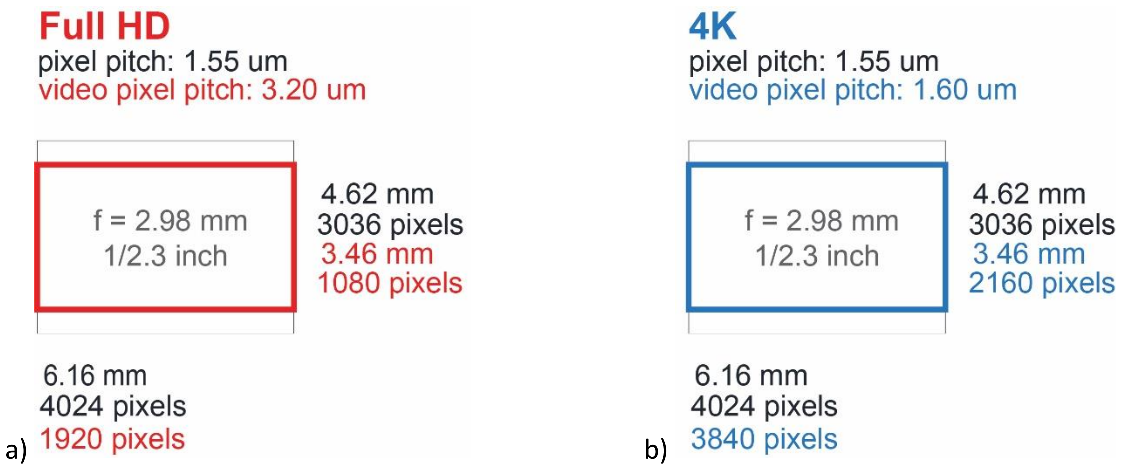


Figure 10: The sensor information for the Hero3 Gopro cameras in G1, G2, G3, G4 and G6 positions which acquired full HD resolution video sequences (a). The sensor information for the Hero 4 GoPro in position G5 which acquired 4K resolution video sequences (b).

<u>2.3 ZEB Revo Geoslam</u>

The **ZEB Revo Geoslam** range instrument was chosen as good and low-cost <u>mobile</u> solution to survey the very narrow spaces of the FINE benchmark. The offered range-based data might be used for alignment purposes or as ground truth comparison or for integration with the image-based data.
Two scans ("Tower" and "Tunnel") have been acquired performing roundtrips and closed loops in order to reduce possible drift error propagation in "swing" effects or linear deviation. The first scan involved the outside and the inside of the tower, while the second scan surveyed the underground tunnel and part of the tower. The room in the ground level of the tower was the common area between the two scans in order to mutually register them.
In the Preview folder, the reduced point clouds are given, for a quick visualization of the two scans.
In the Raw Data folder, the initial Geoslam data of both scans are available. These are the files stored by the instrument after each scanning session and include the point clouds, the trajectories and time information.

In the Final Cloud folder, the final corrected and merged point clouds (using the Geoslam proprietary software) for both scan sessions are given.

2.4 Leica RTC360 laser scanner

The **Leica RTC360** range-based static survey was performed to collect ground truth data to verify elaborations and results coming from the other techniques. The new RTC360 scanner provides full control of the acquisitions directly using a tablet and assures an automatic cloud to cloud pre-registration during the field acquisition using a new VIS (Visual Inertial System) system embedded into the instrument and based on SLAM technology. This is useful to speed up the subsequent fine registration among the scans. Moreover, the limited dimensions of the instrument, its light weight and the fast acquisition time promise fast scanning survey of very complex, narrow and hostile environments which are normally not possible or very critical in terms of time and feasibility with classical laser scanner instruments. For all these reasons we choose this scanner for the activity.
To complete the whole survey, a total of 116 scans were performed. Inside the tunnel and in the very narrow spaces of the stairs, a distance of circa 1 m between two consequent scans was kept to assure a very good overlap necessary for the cloud-to-cloud registration. A led illumination was used to allow the use of VIS.
Some paper black and white markers (mounted on a rigid plastic support) were distributed into the area together with 30 spheres (20 for the tunnel and 10 for the interiors of the tower). All the external black and white markers and the first 4 markers inside the tunnel were measured with the total station. The 12 cm diameter spheres are white and made by Styrofoam, they can be used to register or to check the point clouds. Only the black and white targets inside the tunnel are also visible in the photogrammetric acquisitions.
The Leica scan data were registered using the Leica Register360 tool and a cloud-to-cloud registration method. The process is validated using checkpoint measured with the total station and with a visual inspection sectioning the whole cloud in 10 different areas. The final error on the marker is 4 mm. The visual inspection does not highlight misalignments, double surfaces or holes.

2.5 Topographic acquisitions

Using a Leica TCRA 1203 total station, a closed topographic network was created to georeferenced the laser scanner data and support for the other photogrammetric elaborations. A total of 85 points were collimated from 9 different stations. The overall RMS of the 3D coordinates of the network is less 1 mm.

3. Research objectives of the benchmark

The benchmark data are suitable for different kind of analysis and studies. The **photogrammetric acquisitions** are the real core of the benchmark whereas the range-based point clouds and the topographic points are provided as ground truth and for scaling/control purposes. The main questions that the benchmark poses regard the potential of image-based fisheye techniques for 3D reconstruction of indoor narrow spaces and whether those can be considered valid low-cost alternatives to static or mobile/hand-held laser scanning instruments. We have collected some research questions to be investigated and discussed based on various topics and steps of the photogrammetric pipeline:

Image capturing geometry:
The provided DSLR fisheye acquisitions include a 5-images capturing geometry whereas the multi-camera acquisition provides a 6-viewpoints capturing geometry:

- are all these viewpoints necessary to achieve accurate and complete results?
- are they insufficient or redundant?
- what is the minimum requested number of image couples or triplets?
- what is the minimum base-to-distance ratio?

- is it possible to find a fixed rule about the capturing geometry for this kind of techniques in this type of environment?
- what is the best image capturing geometry in complex spaces?
- how is an irregular and complex scene (narrow spaces, sharp edges, low light conditions, etc.) affecting the capturing procedure and network geometry?

Synchronization and frame extraction:
The GoPro videos are not synchronized but they are provided with a clear audio track recorded during the survey by each camera as well as with time marks in the form of flashlight lamps at the beginning of the videos.
- how can we overcome a precise internal triggering in case of low-cost / action cameras?
- is there any way to synchronize videos during the processing?
- what is the best / most efficient way to extract frames from the available video sequences?

Image pre-processing:
- is it useful to run some pre-processing algorithm to balance tones across the frame, recover shadow details or enhance textureless areas?
- is any pre-processing more useful for tie point extraction or dense image matching?

Camera calibration and relative orientation:
The benchmark provides two calibration testfields for the DSLR and multi-camera acquisitions, acquired onsite or after the survey in the lab. In this way it is possible to test different strategies to derive internal distortion parameters of the single cameras and relative orientation constraints between the cameras that compose the multi-camera rig. More specifically:
- which method is delivering more precise results among self-calibration, calibration onsite and calibration in the lab?
- is self-calibration suitable with the used capturing geometry and surveyed complex environments?
- once the relative orientation of the multi-camera rig is computed, what is the best way to exploit it during the bundle adjustment phase?
- is an accurate relative orientation enough to achieve reliable results also without using external constraints?
- are GCPs or marker always necessary?

Tie point extraction and bundle adjustment
- what is the best method (detector/descriptor) to find correspondences in fisheye images?
- can fisheye photogrammetry guarantee reliable and accurate results in difficult scenarios like the one presented in the benchmark?
- is it possible to achieve, provided the right processing and external constraints, errors within a few centimetres from the reference ground truth?
- how can we deal with drift errors in such small and elongated environments?
- what are the main sources of drift and how are they linked to images resolution, capturing geometry, processing or external constraints?

Sensor and data integration:
The fusion of data coming from different devices is normally offering more chances for a better surveying result. Given all the available data,
- how can we precisely and reliably co-register and merge image- and range-based data in order to close possible gaps in the surveyed areas?
- how can we map RGB information on the range-based point clouds?
- could we integrate SLAM techniques and image-based approaches to provide better, more reliable or faster results?
- can only one technique be successful and sufficient for surveying complex environments?

4. List of distributed data:

Scanning:

Leica RTC360 – ground truth.

- 1 point cloud containing the tunnel and the tower together (ca 17 GB)
- 1 point cloud of the tunnel
- 1 point cloud of the tower

Geoslam ZEB Revo

- clouds preview
- raw scans for Tower and Tunnel areas
- final clouds for Tower and Tunnel areas

Topography:

- Cross markers coordinates
- Circular markers coordinates
- Network points coordinates
- Tunnel sketch plan
- Tower sketch plan
- Self-calibration testfield images

Photogrammetry:

Full frame DSLR – 8mm equisolid fisheye

- Tunnel acquisition, 872 .dng image files
- Tunnel acquisition, 872 .jpg processed image files
- Tower acquisition, 778 .dng image files
- Toer acquisition, 778 .jpg processed image files
- Calibration acquisition, 36 .dng image files
- Calibration acquisition, 36 .jpg processed image files

Multi-camera rig – six Gopros

- G1, tunnel acquisition, one video Full HD
- G1, tower acquisition, one video Full HD
- G1, calibration, one video Full HD
- G2, tunnel acquisition, one video Full HD
- G2, tower acquisition, one video Full HD
- G2, calibration, one video Full HD
- G3, tunnel acquisition, one video Full HD
- G3, tower acquisition, one video Full HD
- G3, calibration, one video Full HD
- G4, tunnel acquisition, one video Full HD
- G4, tower acquisition, one video Full HD
- G4, calibration, one video Full HD
- G5, tunnel acquisition, one video 4K
- G5, tower acquisition, one video 4K
- G5, calibration, one video 4K
- G6, tunnel acquisition, one video Full HD
- G6, tower acquisition, one video Full HD
- G6, calibration, one video Full HD