

# ACCURACY ASSESSMENT OF 3D MODELS GENERATED FROM GOOGLE STREET VIEW IMAGERY

N. Bruno\*, R. Roncella

Università degli Studi di Parma, Dipartimento di Ingegneria e Architettura, Parco Area delle Scienze 181A, 43124, Parma, Italy –  
nazarena.bruno@studenti.unipr.it, riccardo.roncella@unipr.it

## Commission II

**KEY WORDS:** Spherical Photogrammetry, Equi-rectangular Panoramas, Accuracy, 3D reconstruction, Google Street View

### ABSTRACT:

Google Street View is a technology implemented in several Google services/applications (e.g. Google Maps, Google Earth) which provides the user, interested in viewing a particular location on the map, with panoramic images (represented in equi-rectangular projection) at street level. Generally, consecutive panoramas are acquired with an average distance of 5-10 m and can be compared to a traditional photogrammetric strip and, thus, processed to reconstruct portion of city at nearly zero cost. Most of the photogrammetric software packages available today implement spherical camera models and can directly process images in equi-rectangular projection. Although many authors provided in the past relevant works that involved the use of Google Street View imagery, mainly for 3D city model reconstruction, very few references can be found about the actual accuracy that can be obtained with such data. The goal of the present work is to present preliminary tests (at time of writing just three case studies has been analysed) about the accuracy and reliability of the 3D models obtained from Google Street View panoramas.

## 1. INTRODUCTION

Nowadays images and open data available on-line are constantly increasing. Literature provides many examples of public domain and free available images used for 3D reconstruction. For instance, 3D reconstruction can be obtained from generic touristic photos gathered from the web (Wahbeh et al., 2016), frames extracted from videos (Condorelli and Rinaudo, 2018), crowdsourcing images (Somogyi et al., 2016; Frahm et al., 2013) and so on.

Among all the web images collections, service such as Google Street View represents a wide database of available images at street level (Anguelov et al., 2010), freely available and continuously updated. It was launched in several cities in the United States in 2007 and today supplies a great amount of geo-referenced panoramic images that cover many areas of the world, including cities and rural areas. Such a database can represent a great opportunity for several applications, spreading from quick 3D city models reconstruction (Cavallo, 2015; Torii et al., 2009; Micusik and Kosecka, 2009), to historical documentation of urban areas, monitoring, reconstruction of lost building, forensic analyses (Abate et al., 2018) and so on.

In this context, several works addressed the possibilities given by Google Street View imagery (Cavallo, 2015) but few have investigated metric applications and have evaluated accuracy of 3D models obtained by processing multiple Google Street View images referred to the same area.

To record the actual dimensions of the space being photographed, Google vehicles are equipped also with laser scanners that measure up to 50 meters 180° in the front of the vehicle (Cumminis, 2012). Generally, consecutive panoramas are acquired with an average distance of 5-10 m (Agarwal et al., 2015) and can be compared to a traditional photogrammetric strip and, thus, processed to reconstruct portion of city at nearly zero cost.

The latest updates in many software solutions (e.g. PhotoScan by Agisoft, Pix4Dmapper by Pix4D, MicMac, etc.) have

implemented spherical camera model, making possible to process directly spherical (equirectangular) images. Today, in fact, the rapid development of low cost sensors for the acquisition of spherical images (Barazzetti et al., 2018), without the need for complex processes of stitching of images acquired with traditional pinhole cameras (Fangi, 2007; Fangi, 2009), has rekindled the interest in spherical photogrammetry.

In this context, the goal of the present work is to test the accuracy and reliability of the 3D models obtained from Google Street View panoramas. In addition, a workflow was implemented to automatic download and processing the images, in order to speed and simplify images elaboration, also over wide areas.

## 2. GOOGLE STREET VIEW IMAGES

Google Street View is a technology implemented in several Google services/applications (e.g. Google Maps, Google Earth) which provides the user, interested in viewing a particular location on the map, with panoramic images at street level. From its introduction in 2007, data quality, location coverage and connected services has grown constantly with an ever increasing base of active (from 2012 users can provide panoramas acquired from a single spot and from 2017 can create their own street view route) and passive users. In recent years the location considered by the service has grown as well, including interior location of relevant monument/landmarks (e.g. the White House), museums, natural/recreational parks, and so on. From 2014 the platform allows to show (where available) panoramas acquired at different epochs, making the tool particularly useful to assess previous state of relevant features. Finally, in recent years (2017), the equipment used for panorama acquisition has been improved with higher resolution cameras, the introduction of a laser scanner and with better GNSS (Global Navigation Satellite System) and Inertial Navigation Systems (INS) to evaluate the location and pose of every panorama. The equipment is usually mounted on vehicles (vans, cars, etc.), but can also be used with tricycles, boats, snowmobiles, or inserted in backpacks and

\* Corresponding author

trolleys to allow the acquisition of the data in otherwise unreachable locations.

Depending on the devices used to record panorama and navigation data, Google identifies two different products quality level: Street View ready and Street View ready Pro Grade products, with the latter considered providing a high degree of accuracy and image quality.

As far as data access is concerned, several options can be considered, the simplest, but less customizable, being the use of web or desktop applications such as Google Maps or Google Earth. At the same time, Google provides public APIs (Application Programming Interface) for requesting panoramas and associated data/metadata of a given area. REST (Representational State Transfer) Street View Static API allows the transfer of data with standard HTTP (HyperText Transmission Protocol) requests, but the same operations can be performed with JavaScript commands using the Javascript google Map API. Several free and open source tools, based on both APIs, are available on the web for downloading and processing Streetview Panoramas, with different level of customization. In the present work an in-house developed code based on the REST API has been used.

Panoramas are distributed in equi-rectangular projection with different level of resolutions ("zoom") ranging from 416x208 pixels (zoom = 0) to 13312x6656 (zoom = 5).

Most of the photogrammetric software packages, nowadays, implement spherical camera models and can directly process images in equi-rectangular projection. However, especially for commercial software, it's usually hard to infer how exactly the Structure from Motion (SfM), Bundle Adjustment (BB) and dense image matching procedures are implemented. Even if image formation follows the same rules (central projection of object points) regardless of the projection used to represent image data, with equi-rectangular projection, beside actual perspective transformation, the images have an additional distortion effect. In particular, especially outside the central (equatorial) part of the image, image features are stretched along horizontal (latitude) direction. How this issue is tackled by the software is often omitted in user reference manual/user forum and is therefore unknown. Some authors (see for instance Barazzetti et al., 2014) showed that using a pin-hole based dataset might lead to better results than others where different projection, with stronger scale (usually anisotropic) variation, are used.

Accessing Google Maps REST API, further information can be extracted and injected in the EXIF header of the images (GPS location, panorama orientation, associated depth map, etc.). The methodology used for the automatic download of the Google panoramas is summarized in Figure 1.

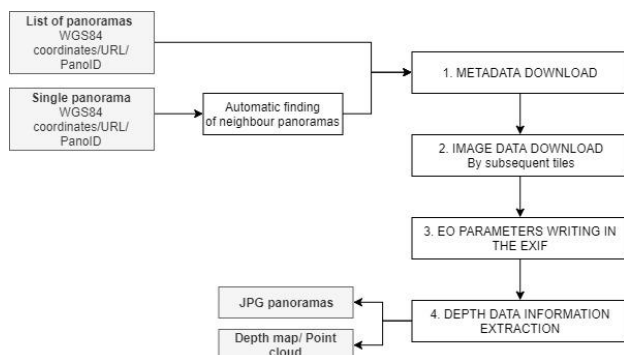


Figure 1.\_The developed pipeline.

The developed application allows the user to specify a specific location expressed in WGS84 coordinates or with its corresponding PanoID (a unique alphanumeric string of character

that identifies the Panorama image, that can be found inside the URL (Uniform Resource Locator) of the selected panorama) and automatically find all the neighbour panoramas. Otherwise, the user can specify a list of (WGS84) locations or PanoIDs (or panoramas URL) and the application downloads every single image.

Subsequently, for each image, the application (1) sends an http request to download the information (metadata) associated with the selected panorama image. Such data is transferred with a JSON (JavaScript Object Notation) file which contains several information: time of acquisition, panorama location and orientation in WGS84 reference system, the maximum zoom (i.e. resolution) available for the panorama, and the depth map associated with the panorama image.

At the same time image data is downloaded (2): the API requires that each single tile (the panorama is divided in tiles of 512x512 pixels) is requested separately; the total image is then composed by subsequent requests.

Exterior orientation parameters, stored in the JSON metadata file, are written in the EXIF of each single panorama (3).

Metadata are further processed (4) to extract depth data information: the data is optimized (compressed) for web transmission, and additional operations are required to obtain depth map representation of the scene acquired in the panorama (in-depth description of computation algorithm can be found in Cavallo, 2015). The final depth map (512x256 values) can then be used to obtain a point cloud with associated RGB information coming from the panoramic image.

It's worth noting that to limit the size of depth data stored in the JSON file, the scene is simplified in a set of planar region (the maximum allowed number of planes is 255), which approximate the spatial distribution of the points acquired by the laser scanning equipment. The available data should not be considered useful for accurate 3D reconstruction, but can provide some information about the approximate depth of the scene.

At the end of the image download procedure, the equi-rectangular panoramas can be converted in pin-hole projection: the user specifies the area of the panorama that he wants to extract: since a spherical image span a 360°x180° Field of View (FOV), it cannot be entirely converted in a single planar pin-hole. The conversion is pretty straightforward: image data is resampled with the following procedure:

1. The object space reference system (XYZ) is rotated (X'Y'Z') so that its Z' axis passes through the (planar) image space ( $\xi\eta\zeta$ ) of pin-hole model, and its X' and Y' axes are parallel respectively to the  $\xi$  and  $\eta$  axes;
2. The equi-rectangular image data (whose resolution is  $w \times h$  pixel) is considered lying on a spherical surface (unit sphere): image coordinates (xy), are converted in longitude ( $\varphi$ ) and latitude ( $\lambda$ ) of the projected point on unit sphere:

$$\varphi = \frac{(x-x_0) \cdot 2\pi}{w} \quad (1)$$

$$\lambda = \frac{(y-y_0) \cdot \pi}{h} \quad (2)$$

Where  $x_0 = w/2$  and  $y_0 = h/2$  represent the center of the equi-rectangular image.

3. The generic point ( $\varphi, \lambda$ ) casts a ray in object space which passes through the origin with the following parametric equation:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} t \cdot \cos \varphi \sin \lambda \\ t \cdot \sin \varphi \sin \lambda \\ t \cdot \cos \lambda \end{bmatrix} \quad (3)$$

4. The rotation computed at point (1.) is applied to the parametric vector of eq. 3;
5. Considering a valid value for the parameter  $t$  (e.g.  $t = 1$ ) the resulted array represents the homogenous

coordinates of the corresponding point in image space equivalent to the point  $[\xi \ \eta \ 1]$ .

If inverse transformation (from pin-hole to equi-rectangular image space) is required the same approach can be used:

6. The homogeneous coordinates of a generic pin-hole image points are considered  $([\xi \ \eta \ 1])$ ;
7. The homogeneous vector is normalized and represents the corresponding ray in (pin-hole) image space;
8. The array is transformed with the inverse of the rotation described in (1.). The resulting array  $([a \ b \ c])$  represents the intersection between the casted ray and the unit sphere;
9. It's corresponding longitude and latitude on unit sphere can be easily computed:

$$\varphi = \sin^{-1} b \quad (4)$$

$$\lambda = \tan^{-1}(a/c) \quad (5)$$

10. Inverting eqs. (1) and (2) it's possible to derive image coordinate of the corresponding point in equi-rectangular image space.

If exterior orientation (EO) parameters of the original panorama are known, it's quite simple to derive the corresponding EO parameters of the newly generated pin-hole (planar) image, considering that the centre of perspective is the same for the two projective models (i.e. camera centre is the same) and the rotation of the planar image space derives from the composition of the equi-rectangular pose and the rotation computed in the previous step (1.).

At the same time, as far as interior orientation (IO) parameters are concerned, once the area to be extracted (selected by the user) is known, considering that, in the previous approach the pin-hole image plane is considered tangent to the unit sphere, principal distance (1 unit) and image plane size (depending on the choice of the user) can be easily computed. Usually spherical images are considered already corrected w.r.t. lens distortion.

The whole procedure was automated with a Matlab code and an example of output is shown in Figure 2.

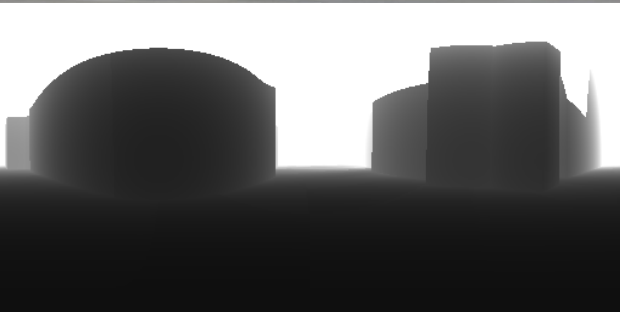


Figure 2\_Equirectangular projection and associated depth map downloaded from Google.

### 3. ACCURACY EVALUATION

To evaluate the accuracy of the photogrammetric models obtained processing images provided by Google Street View,

different approaches have been tested. To grant result repeatability, the methodology described has been tested in three different case studies with different geometric characteristics:

1. a square (Piazza Duomo in Parma, Italy);
2. a street (via Cardinal Ferrari, on the south side of Parma Cathedral, Italy);
3. the internal courtyard of a palace (Costabili palace in Ferrara, Italy).

For each of these case study, a reference dataset was available and was used to evaluate the accuracy of Google images models. Piazza Duomo and via Cardinal Ferrari were surveyed in 2017, integrating total station, GPS, terrestrial laser scanning (TLS) (Bruno and Roncella, 2018). Palazzo Costabili, instead, was subject in 2015 to a topographic and high-resolution close range photogrammetry survey.

To evaluate the accuracy, a first aspect considered was the effect that different ground control solutions have on images orientation and dense matching. Three different control solutions have been applied to each dataset analysed.

1. "NO GPS" solution: in this case, images were oriented using only Ground Control Point (GCP) coordinates. The GPS coordinates of the camera centre position provided by Google were not used as observation in Bundle Block Adjustment (BBA). In this way the accuracy was evaluated without considering possible errors introduced by inaccurate GPS measurements.
2. "GPS LW (Low Weight)" solution: the ground control was based only on the GPS coordinated of camera centre position provided by Google. No GCP were used. However, the camera centre positions were used in the stochastic model of BBA with a low weight, so as not to constrain too much the relative orientation solution between panoramas and act only as georeferencing. This was important to evaluate the internal coherence of the panoramas without inserting particular constraints to the block.
3. "GPS HW (High Weight)" solution: also in this case the ground control was based only on GPS camera centre coordinates, without any GCP. Differently from the previous solution, such coordinates were used with a high weight in BBA. The aim was to evaluate the accuracy of the GPS measurements provided by Google and their consistency with the topographic reference dataset.

In addition, the evaluation of how much the conversion from equi-rectangular to planar (pin-hole) projection model could improve image orientation and dense matching has been done. In particular, from each dataset the planar projections of some limited areas have been extracted and then processed applying the three different ground control solutions described before.

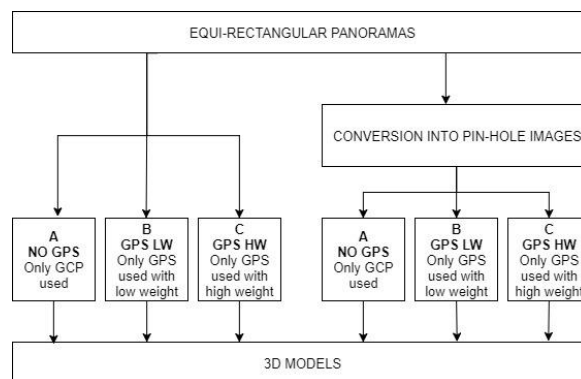


Figure 3. The different options for images elaboration.



Summarising, for each case study, six different processing tests have been done: using equi-rectangular or planar images and applying different ground control solutions (Figure 3). All data were processed with Agisoft Photoscan using the same processing pipeline.

As far as image orientation is concerned, equi-rectangular images have been processed using spherical camera model, while planar images were oriented using the IO parameters estimated during the conversion process. The tie points extraction was performed working with images at the original size ("high quality accuracy" in Photoscan terminology) and using the following set of parameters to define the different control solution:

| Control solution | CC* enabled | GCP enabled | CC* precision | GCP Precision** |
|------------------|-------------|-------------|---------------|-----------------|
| NO GPS           | no          | yes         | /             | 0.02 m          |
| GPS LW           | yes         | no          | 10 m          | /               |
| GPS HW           | yes         | no          | 0.01 m        | /               |

\* Camera centre position

\*\* Set equal to 2 cm considering the accuracy of the reference surveys

Table 1. Parameters used for orientation process.

Dense matching was performed using down-sampled images by factor of 2 ("high quality" in Photoscan terminology) to not overload the processing times. The Digital Surface Model (DSM) was generated setting the maximum number of polygons in the final mesh as 1/5 of the number of points in the previously generated dense point cloud ("high quality" in Photoscan terminology).

The results were checked analysing the residuals on:

- re-projections (tie point BBA residuals of collinearity equations), to evaluate the internal consistency of the photogrammetric block;
- camera centres position residuals: in the solutions where (independent) GCP were used, camera centre residuals allowed to evaluate the accuracy of the GPS measurements; in the solutions where the ground control was based only on GPS camera centre coordinates, an evaluation of the internal consistency of the block is provided;
- check points residuals, to estimate the accuracy of the final object reconstruction.

Finally, the DSMs obtained from each dataset and from each processing pipeline were compared with the mesh surface model obtained by the high resolution survey just mentioned. The two dataset have been referred to the same reference system and a further alignment with ICP algorithm was performed to reduce possible residual systematic effects.

## 4. RESULTS

### 4.1 Piazza Duomo case study



Figure 4. Piazza Duomo

Piazza Duomo (Cathedral square) is one of the main squares located in the city centre of Parma (Italy). It is a rather wide square of 45x55 m. Google panoramas have been acquired along the four sides of the perimeter, with an average base-length equal to 10 m.

The 2015 and 2017 datasets have been processed independently and together (2015-17 in the table below), using both equi-rectangular and planar images.

All equi-rectangular images have been processed to reconstruct the entire square, while only the cathedral façade was analysed with planar images. In particular, only panoramas located in front of the façade were converted in pin-hole projections.

During the processing of equi-rectangular images, it has been observed that in some cases the automatic extraction of tie points is not optimal. In open spaces such as Piazza Duomo, a large part of the image frame shows sky or ground areas. If, on one hand matched ground features might be erroneous due to element repetition, on the other hand points matched in the sky represents elements at (nearly) infinite distance and can make the orientation solution unstable. A typical example of matched point in an outdoor spherical image is shown in Figure 5, where relevant object features are not considered/matched or considered invalid (grey dots), while the vast majority of valid corresponding points are recognized in the sky.

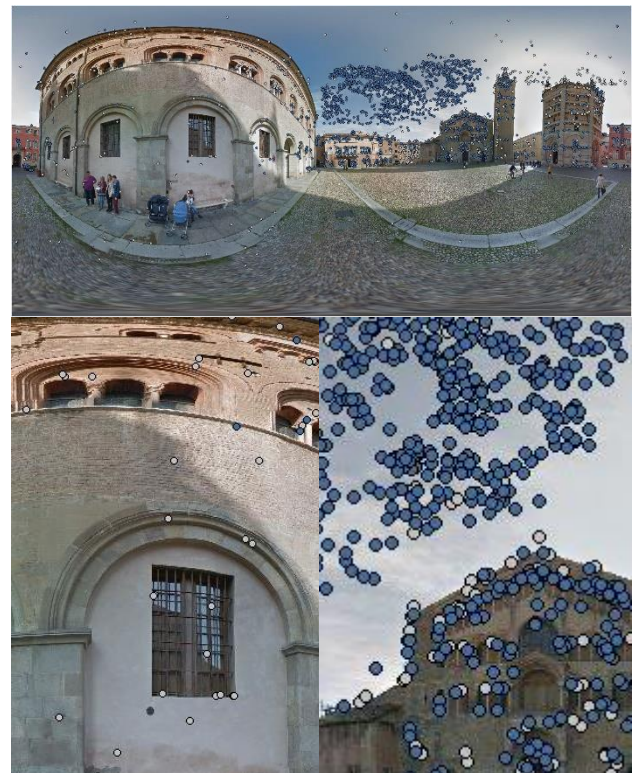


Figure 5. Example of typical matched point in an outdoor spherical image

In this regard, masking the images excluding the areas that shows the sky, and identifying manually some tie points in the areas where object features are not matched, improved significantly panoramas orientation.

Another problem observed during dense matching is that some parts of the image block are not registered correctly with other (maybe due to the limited number of corresponding tie points) producing relevant shifts between subset of dense points in the final cloud. Figure 6 shows this effect over three sides of the square.



Figure 6. Example of shifts between subset of dense points.

Again, to reduce this relative misalignment, manual tie points identification have proven to be effective: tie points creates additional constraints between portions of the model, reducing shift and rotation effects that, otherwise, occur.

On the basis of these considerations, in the equi-rectangular images presented in this paper the areas occupied by sky have been masked during feature extraction and manual tie points (used as GCP or CP) have been collimated.

On the contrary, planar images are not affected by the above mentioned effects, probably because the most part of the image frames the object to reconstruct. For this reason, in the following case studies, planar images were not masked, but manual tie points have been collimated anyway and used as GCP or CP to check the orientation solution.

Below, stats on collinearity equation residuals, check points error and DSM comparisons are provided.

| Dataset (equi-rectang.) | tie points n. | Projections n. | Reproj. residuals [pix] | Camera center res.[m] |
|-------------------------|---------------|----------------|-------------------------|-----------------------|
| 2015 NO GPS             | 7534          | 18941          | 14.7                    | 4.87                  |
| 2015 GPS LW             | 7987          | 20068          | 8.96                    | 3.219                 |
| 2015 GPS HW             | 8438          | 21340          | 11.6                    | 2.991                 |
| 2017 NO GPS             | 2812          | 10825          | 13.3                    | 5.99                  |
| 2017 GPS LW             | 2610          | 9986           | 13.4                    | 4.766                 |
| 2017 GPS HW             | 2645          | 9927           | 21.6                    | 3.541                 |
| 2015-17 NO GPS          | 20316         | 58257          | 13.9                    | 6.4                   |
| 2015-17 GPS LW          | 19942         | 56637          | 18.5                    | 5.675                 |
| 2015-17 GPS HW          | 13458         | 37470          | 67.2                    | 8.976                 |

Table 2. Equirectangular dataset: stats on collinearity residuals.

| Dataset (equi-rectang.) | GCP | CP | RMS CP | St. dev. CP |       |       |
|-------------------------|-----|----|--------|-------------|-------|-------|
|                         |     |    |        | X           | Y     | Z     |
| 2015 NO GPS             | 6   | 16 | 0.327  | 0.208       | 0.159 | 0.141 |
| 2015 GPS LW             | 0   | 22 | 3.307  | 0.446       | 0.559 | 0.349 |
| 2015 GPS HW             | 0   | 22 | 3.493  | 0.694       | 0.559 | 0.327 |
| 2017 NO GPS             | 6   | 16 | 0.134  | 0.091       | 0.036 | 0.077 |
| 2017 GPS LW             | 0   | 22 | 5.891  | 0.595       | 1.536 | 0.726 |
| 2017 GPS HW             | 0   | 22 | 5.148  | 0.447       | 1.423 | 0.623 |
| 2015-17 NO GPS          | 6   | 16 | 0.211  | 0.159       | 0.176 | 0.093 |
| 2015-17 GPS LW          | 0   | 22 | 6.223  | 0.744       | 1.377 | 0.810 |
| 2015-17 GPS HW          | 0   | 22 | 10.94  | 2.502       | 1.908 | 1.544 |

Table 3. Equirectangular dataset: stats on residuals [m].

| Dataset (planar)   | tie points n. | Projections n. | Reproj. res. [pix] | Camera center res.[m] |
|--------------------|---------------|----------------|--------------------|-----------------------|
| Façade 2015 NO GPS | 2528          | 7448           | 2.14               | 5.382                 |
| Façade 2015 GPS LW | 2510          | 7395           | 2.15               | 2.824                 |
| Façade 2015 GPS HW | 2601          | 7673           | 4                  | 1.566                 |
| Façade 2017 NO GPS | 6680          | 23849          | 2.66               | 6.227                 |
| Façade 2017 GPS LW | 6680          | 23849          | 2.67               | 4.225                 |
| Façade 2017 GPS HW | 6680          | 23849          | 4.73               | 1.686                 |

|                       |      |       |      |       |
|-----------------------|------|-------|------|-------|
| Façade 2015-17 NO GPS | 2622 | 17662 | 2.88 | 6.294 |
| Façade 2015-17 GPS LW | 2615 | 17630 | 2.81 | 4.542 |
| Façade 2015-17 GPS HW | 1720 | 9422  | 4.04 | 2.637 |

Table 4. Planar dataset: stats on collinearity residuals.

| Dataset (planar)      | GCP | CP | RMS CP | St. dev. CP |       |       |
|-----------------------|-----|----|--------|-------------|-------|-------|
|                       |     |    |        | X           | Y     | Z     |
| Façade 2015 NO GPS    | 6   | 11 | 0.15   | 0.088       | 0.046 | 0.068 |
| Façade 2015 GPS LW    | 0   | 17 | 7.218  | 1.056       | 0.534 | 0.694 |
| Façade 2015 GPS HW    | 0   | 17 | 7.006  | 1.453       | 0.402 | 0.735 |
| Façade 2017 NO GPS    | 6   | 11 | 0.261  | 0.169       | 0.067 | 0.157 |
| Façade 2017 GPS LW    | 0   | 17 | 9.284  | 0.704       | 1.844 | 1.170 |
| Façade 2017 GPS HW    | 0   | 17 | 9.784  | 0.906       | 1.980 | 1.366 |
| Façade 2015-17 NO GPS | 6   | 11 | 0.296  | 0.195       | 0.100 | 0.152 |
| Façade 2015-17 GPS LW | 0   | 17 | 7.839  | 0.833       | 1.390 | 1.068 |
| Façade 2015-17 GPS HW | 0   | 17 | 6.906  | 0.748       | 1.250 | 0.891 |

Table 5. Planar dataset: stats on residuals [m].

The analysis of camera centre and CP residuals shows the inconsistency between GPS and CP observations: the solutions where GPS control was used in BBA (regardless of the weight imposed) provided high CP RMS.

No information is provided by Google on the GPS equipment used, nor on precisions and expected accuracy. Nevertheless, considering the acquisition environment (in most cases city centre with frequent satellite occlusions, multipath, etc.), remarkable variability in GPS accuracy is to be reasonably expected. Using the GCP, although the not optimal orientation solution introduces quite certainly some errors on the camera centre estimation, residuals on camera centre position can be used to evaluate the level of accuracy of GPS data (i.e. ca. 4÷6 meters). To check the effect of the different orientation solution on object reconstruction, some relevant distance/length on the object (e.g. the width and height of the main façade) has been measured manually on the image block and compared with known values. While with ground control points the photogrammetric measurements are consistent with the real lengths (i.e. few centimetres discrepancies have been observed), with GPS control the façade is significant oversized, with errors up to some meters. Re-projection residuals, as should be expected, are much lower in planar image datasets.

Nevertheless, the analysis of CP residuals shows that, in this case study, the use of planar images instead of panoramas does not improve considerably the orientation solution, making the use of planar or equi-rectangular datasets comparable.

As far as the dense matching is concerned, the comparison between the different approaches has been done considering only the main façade of the cathedral in order to have a common test field both for equi-rectangular and planar datasets. Nevertheless, only the DSM obtained using GCP control have been used for comparisons, because of the inconsistency of the model obtained using GPS control with real dimensions, as described above.

In general, the DSMs obtained are very noisy, probably because of the low resolution of the images processed. The comparisons results show standard deviation values of about 20 cm for equi-rectangular dataset and 10 cm for planar (Figure 7 and Table 6). In this context, thus, the use of planar images improves dense matching process and the quality of the model, halving the RSM values.

| Equi-rect/planar | Dataset               | Standard deviation [m] |
|------------------|-----------------------|------------------------|
| Equi-rectangular | 2015 NO GPS           | 0.244                  |
| Equi-rectangular | 2017 NO GPS           | 0.211                  |
| Equi-rectangular | 2015-17 NO GPS        | 0.208                  |
| Planar           | Façade 2015 NO GPS    | 0.119                  |
| Planar           | Façade 2017 NO GPS    | 0.113                  |
| Planar           | Façade 2015-17 NO GPS | 0.11                   |

Table 6. DSM comparison.



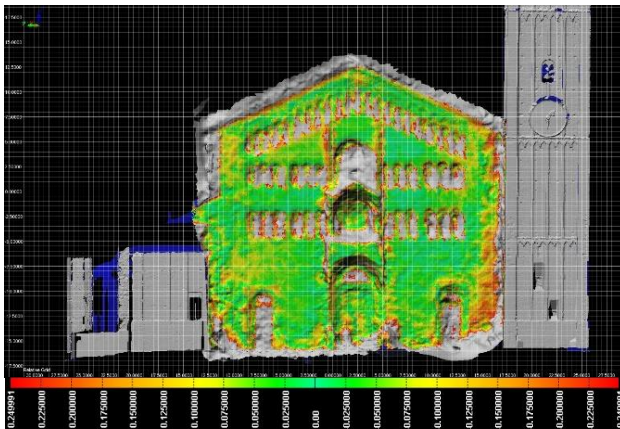


Figure 7. “Façade 2017 NO GPS” DSM comparison. False colour representation.

#### 4.2 Cardinal Ferrari street case study

Cardinal Ferrari street is a quite narrow street (10x100 m) located on the south side of the Cathedral of Parma. Also in this case the available datasets are three (2014, 2015 and 2016) composed of 10, 11, 11, images respectively, with an average base-length of 11 m. This case study represents the typical geometry of Google Street view image acquisition, where images are acquired along an approximate straight trajectory.

This block geometry is particularly disadvantageous for image orientation, first of all because the camera centres are aligned, leaving one degree of freedom (rotation of image block around the direction of vehicle trajectory) hardly evaluable if only GPS camera station positions are used. In addition, due to the conformation of a road (narrow and long) the distance from the camera centre of the elements framed in the image is very variable. In particular, the elements on the street sides located in front of the camera are very close to the shooting point (so deformed in panoramas), the objects along the longitudinal development of the road have a high perspective distortion and, finally, the elements at the end of the street represent points almost at infinite distance. Therefore, between consecutive panoramas there are remarkable depth changes and the perspective distortion that affect the same element changes considerably. This jeopardizes image orientation and dense matching due to difficulties in finding features correspondences.



Figure 8. Cardinal Ferrari street.

In this case study, processing the three datasets (2014, 2015 and 2016) independently resulted in invalid image orientation or incomplete reconstruction of most of the elements during dense matching. Therefore, to entirely reconstruct the model, it was necessary to process the three datasets together, both in the case of equi-rectangular and planar images. In this case the base-lengths (and perspective variations) between consecutive images

have been reduced, improving the matching. Nevertheless, the dataset (“All GPS LW” in the table below), where no GCP have been used and the GPS control was used with a low weight, did not oriented regardless of tie point manual identification.

| Dataset (equi-rectang.) | tie points n. | Projections n. | Reproj. res. [pix] | Camera center res.[m] |
|-------------------------|---------------|----------------|--------------------|-----------------------|
| All NO GPS              | 18487         | 43429          | 15                 | 2.677                 |
| All GPS LW              | Not oriented  |                |                    |                       |
| All GPS HW              | 18527         | 43852          | 14.3               | 1.42                  |

Table 7. Equirectangular dataset: stats on collinearity residuals.

| Dataset (equi-rectang.) | GCP | CP | RMS CP | St. dev. CP |       |       |
|-------------------------|-----|----|--------|-------------|-------|-------|
|                         |     |    |        | X           | Y     | Z     |
| All NO GPS              | 7   | 16 | 0.181  | 0.086       | 0.094 | 0.121 |
| All GPS HW              | 0   | 23 | 2.288  | 0.068       | 0.228 | 0.149 |

Table 8. Equirectangular dataset: stats on residuals [m].

The test using planar images has been made on two areas of the south side of the cathedral, using only the panoramas taken in front of these areas. Also, in this case, it was necessary to process together all the images belonging to different years. The use of planar projection, indeed, does not overcome the previous mentioned problems related to disadvantageous block geometry, since re-projection simply removes the deformations produced by spherical mapping.

On the contrary, pin-hole images allowed the orientation of all the datasets, regardless the different control solution adopted (Table 9).

Also in this case, the analysis of camera centre and CP residuals demonstrates the low consistency between GPS and GCP observations, even if it is higher than in Piazza Duomo case study.

Despite the high residuals observed, all the models obtained have correct dimension (i.e. scale factor). Sampling relevant distances on the object but no significant discrepancies have been observed (differences with real lengths up to few centimetres).

| Dataset (planar)     | tie points n. | Projections n. | Reproj. res. [pix] | Camera center res.[m] |
|----------------------|---------------|----------------|--------------------|-----------------------|
| sacristy. All NO GPS | 11886         | 26425          | 7.24               | 2.722                 |
| sacristy. All GPS LW | 12207         | 27520          | 7.6                | 2.362                 |
| sacristy. All GPS HW | 11886         | 26425          | 7.2                | 1.53                  |
| 3rd bay All NO GPS   | 12561         | 28506          | 3.17               | 2.764                 |
| 3rd bay All GPS LW   | 12830         | 29178          | 3.14               | 1.142                 |
| 3rd bay All GPS HW   | 12561         | 28506          | 5.92               | 1.064                 |

Table 9. Planar dataset: stats on collinearity residuals.

| Dataset (planar)     | GCP | CP | RMS CP | St. dev. CP |       |       |
|----------------------|-----|----|--------|-------------|-------|-------|
|                      |     |    |        | X           | Y     | Z     |
| sacristy. All NO GPS | 4   | 5  | 0.228  | 0.166       | 0.102 | 0.074 |
| sacristy. All GPS LW | 0   | 9  | 1.542  | 0.234       | 0.642 | 0.090 |
| sacristy. All GPS HW | 0   | 9  | 2.358  | 0.237       | 0.642 | 0.091 |
| 3rd bay All NO GPS   | 4   | 5  | 0.254  | 0.046       | 0.127 | 0.077 |
| 3rd bay All GPS LW   | 0   | 9  | 2.105  | 0.344       | 0.184 | 0.055 |
| 3rd bay All GPS HW   | 0   | 9  | 2.062  | 0.262       | 0.249 | 0.061 |

Table 10. Planar dataset: stats on residuals [m].

As far as the analysis of DSM is concerned, also in this case the reconstructed surface is rather noisy. All the models have been compared with the reference DSM obtained from laser scanning survey and no significant discrepancies have been noticed between the different processing approaches. All the control solutions (GCP, GPS LW, GPS HW) have been proven to be reliable, with residuals’ standard deviations of ca. 10 cm. In

particular, planar projection slightly improved DSM reconstruction (Table 11).

| Equi-rect/planar | Dataset            | Standard deviation [m] |
|------------------|--------------------|------------------------|
| Equi-rectangular | All NO GPS         | 0.101                  |
| Equi-rectangular | All GPS HW         | 0.102                  |
| Planar           | sacristy NO GPS    | 0.086                  |
| Planar           | sacristy GPS LW    | 0.085                  |
| Planar           | sacristy GPS HW    | 0.091                  |
| Planar           | 3rd bay All NO GPS | 0.096                  |
| Planar           | 3rd bay All GPS LW | 0.097                  |
| Planar           | 3rd bay All GPS HW | 0.100                  |

Table 11. DSM comparison.

### 4.3 Costabili palace case study

Costabili palace is located in the city centre of Ferrara (Italy) and houses the National Archaeological Museum of Ferrara. It is a particular case study, since, due to the increasing interest of Google in documenting public/cultural sites, panoramas have been acquired also inside the palace, enabling a virtual tour of the Museum.



Figure 9. Palazzo Costabili internal courtyard. Northern and southern façades.

The test presented here take into account only the images acquired in the internal courtyard (Figure 9). It is a pedestrian area, so the acquisitions have been done by walking operators and not by a moving car. This acquisition method, reasonably has improved the GPS measurements accuracy, as it will be shown by the results provided below.

Currently only one dataset is available: it dates to 2013 and is composed of 30 panoramas acquired with an average base-length of 3 m. Block geometry, conformation and site dimensions are advantageous: it is a small size (20x25 m) closed courtyard, avoiding the drawbacks related to acquisition along streets or in wide spaces, as seen in the previous examples. Images are arranged along the four sides of the courtyard perimeter, with a short base-length, which produce a high overlap between images. In addition, the shooting points are quite close to the building façades, thus buildings occupy the main part of the image frame, reducing the areas that show the sky. Nevertheless, for uniformity with the previous examples, also in this case the parts of the images that depicted the sky were masked.

On the basis of these considerations, and as will be confirmed by the stats below, the expected accuracy should be higher than in previous case studies.

| Dataset (equi-rectang.) | tie points n. | Projections n. | Reproj. res. [pix] | Camera center res.[m] |
|-------------------------|---------------|----------------|--------------------|-----------------------|
| NO GPS                  | 12393         | 32052          | 3.8                | 0.260                 |
| GPS LW                  | 12393         | 32052          | 3.8                | 0.213                 |
| GPS HW                  | 12444         | 31842          | 3.93               | 0.101                 |

Table 12. Equirectangular dataset: stats on collinearity residuals.

| Dataset (equi-rectang.) | GCP | CP | RMS CP | St. dev. CP |       |       |
|-------------------------|-----|----|--------|-------------|-------|-------|
|                         |     |    |        | X           | Y     | Z     |
| NO GPS                  | 9   | 11 | 0.068  | 0.025       | 0.051 | 0.025 |
| GPS LW                  | 0   | 20 | 0.086  | 0.035       | 0.063 | 0.037 |
| GPS HW                  | 0   | 20 | 0.141  | 0.059       | 0.106 | 0.043 |

Table 13. Equirectangular dataset: stats on residuals [m].

Also in this case, the re-projection to planar has been applied only to limited parts inside the scene, i.e. the southern and the eastern façades (Figure 9).

| Dataset (planar) | tie points n. | Projections n. | Reproj. res. [pix] | Camera center res.[m] |
|------------------|---------------|----------------|--------------------|-----------------------|
| South NO GPS     | 5436          | 23288          | 1.26               | 0.057                 |
| South GPS LW     | 5436          | 23288          | 1.26               | 0.066                 |
| South GPS HW     | 5436          | 23288          | 1.25               | 0.061                 |
| Est NO GPS       | 6277          | 31066          | 1.16               | 0.073                 |
| Est GPS LW       | 6277          | 31066          | 1.16               | 0.069                 |
| Est GPS HW       | 6277          | 31066          | 1.15               | 0.068                 |

Table 14. Planar dataset: stats on collinearity residuals.

| Dataset (planar) | GCP | CP | RMS CP | St. dev. CP |       |       |
|------------------|-----|----|--------|-------------|-------|-------|
|                  |     |    |        | X           | Y     | Z     |
| South NO GPS     | 5   | 5  | 0.038  | 0.023       | 0.020 | 0.019 |
| South GPS LW     | 0   | 10 | 0.069  | 0.039       | 0.035 | 0.044 |
| South GPS HW     | 0   | 10 | 0.066  | 0.038       | 0.032 | 0.043 |
| Est NO GPS       | 6   | 6  | 0.055  | 0.020       | 0.031 | 0.017 |
| Est GPS LW       | 0   | 11 | 0.057  | 0.026       | 0.042 | 0.026 |
| Est GPS HW       | 0   | 11 | 0.069  | 0.030       | 0.050 | 0.023 |

Table 15. Planar dataset: stats on residuals [m].

The analysis of residuals demonstrates a good accuracy of the results: block internal coherence is very high with re-projection residuals of only 1 pixel and the stats on camera centre positions and check points show residuals of few centimetres (mean Check points RMS equal to 7 cm).

In particular, GPS observations on camera centres position are consistent with GCP measurements, with residuals up to 14 cm (equi-rectangular) and up to 7 cm (planar). Using planar projections instead of equi-rectangular improves results in particular on orientation solution based on GPS control. In these cases the residuals can be halved. Using GCP control, instead, the use of planar or panoramic images is equivalent.

The DSM analysis shows results in accordance with the ones relating to orientation. The average value of standard deviation is 4 cm and only the model processed using equi-rectangular images and high weight of the GPS control (GPS HW) has a higher standard deviation.

In general, the model surface is less noisy than the previous case study, probably thanks to the good overlap between panoramas and the high resolution of the object on the images.

| Equi-rect/ planar | Dataset      | Standard deviation [m] |
|-------------------|--------------|------------------------|
| Equi-rectangular  | NO GPS       | 0.046                  |
| Equi-rectangular  | GPS LW       | 0.041                  |
| Equi-rectangular  | GPS HW       | 0.051                  |
| Planar            | South NO GPS | 0.042                  |
| Planar            | South GPS LW | 0.041                  |
| Planar            | South GPS HW | 0.042                  |
| Planar            | Est NO GPS   | 0.045                  |
| Planar            | Est GPS LW   | 0.044                  |
| Planar            | Est GPS HW   | 0.044                  |

Table 16. DSM comparison.

## 5. CONCLUSIONS

The research presented in this paper aimed to evaluate the accuracy of photogrammetric models obtained from Google

Street View images. The tests conducted on three case studies showed a great variability of results, connected to the geometric characteristics of the image block and to the accuracy of the GPS measurements.

As far as the geometry is concerned, wide places and narrow streets present different obstacles: in wide spaces (such as squares), a large part of the equi-rectangular image shows the sky and, if not masked, many tie points could be matched on it, making the orientation solution unstable. In addition, the objects far from the shooting points have low resolution, leading to noisy DSM reconstruction. In this context, converting equi-rectangular images into pin-hole ones, framing only a restrict part of the scene, improves DSM reconstruction.

In narrow streets, camera centres are generally aligned, introducing possible systematic rotations if only GPS camera station positions are used as ground control. Again, if the base-length of acquisition is high (ca. 10 m), consecutive panoramas are characterized by remarkable depth changes and perspective effects which make difficult to match corresponding feature. Using different datasets together (such as images acquired in different epochs) improves the solution: having more images, in many occasions more favourable base-lengths can be found.

The quality of the GPS observations plays a very important role in the accuracy of the final 3d models. No information is provided by Google about precisions and accuracy, but in two of the three cases analysed it has proven to be quite inaccurate (up to few meters). In the Palazzo Costabili case study, instead, the residuals on camera centres GPS coordinates are ca. 25 cm (equi-rectangular) and ca. 6 cm (pin-hole). Anyway, the use of GPS observations as ground control is useful to improve the orientation solution convergence, providing initial camera centre positions.

The use of Google Street View images has proven to be generally suitable for 3D reconstruction if high accurate model are not required. To improve accuracy and verify the quality of the solution obtained, the use of some GCP and CP is desirable.

## REFERENCES

Abate, D., Toschi, I., Sturdy-Colls, C., Remondino, F., 2018. Panoramic images, 2d feature-based and change detection methods for the documentation of contaminated crime scenes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-2, 1-8.

Agarwal, P., Burgard, W., Spinello, L., 2015. Metric Localization using Google Street View. *IEEE Conference on Computer Vision and Pattern Recognition*.

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. 42.

Barazzetti, L., Previtali, M., Roncoroni, F., 2018. Can we use low-cost 360 degree cameras to create accurate 3D models?, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 69-75.

Barazzetti, L., Previtali, M., & Scaioni, M., 2014. Simultaneous registration of gnomonic projections and central perspectives. *The Photogrammetric Record*, 29(147), 278-296.

Bruno, N. and Roncella, R., 2018. A restoration oriented HBIM system for cultural heritage documentation: the case study of Parma cathedral. *The International Archives of the*

*Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 171-178.

Cavallo, M., 2015. 3D City Reconstruction From Google Street View, *Comput. Graph. J*

Condorelli, F. and Rinaudo, F., 2018. Cultural heritage reconstruction from historical photographs and videos, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 259-265.

Cummins, M., 2012. Google Street View – Soon in 3D?. *Educating Silicon*, Retrieved January 3, 2012.

Fangi, G., 2007. The multi-image spherical panoramas as a tool for architectural survey. In: *The Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36(5/C53), pp. 311–316.

Fangi, G., 2009. Further developments of the spherical photogrammetry for cultural heritage. In: *XXII International Committee for Cultural Heritage (CIPA)*, pp. 11-15.

Frahm, J.M., Heinly, J., Zheng, E., Dunn, E., Fite-Georgel, P. & Pollefeys, M., 2013. Geo-registered 3D models from crowdsourced image collections, *Geo-spatial Information Science*, 16:1, 55-60.

Micusik, B. and Kosecka, J., 2009. Piecewise planar city 3d modeling from street view panoramic sequences. In *CVPR09*, 2009.

Somogyi, A., Barsi, A., Molnar, B., Lovas, T., 2016. Crowdsourcing based 3D modeling, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5, 587-590.

Torii, A., Havlena, M., Pajdla, T., 2009. From Google Street View to 3D City models. *IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*.

Wahbeh, W., Nebiker, S., Fangi, G., 2016. Combining public domain and professional panoramic imagery for the accurate and dense 3D reconstruction of the destroyed bel temple in Palmyra. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-5, 81.88.