Use of SfM-MVS approach to nadir and oblique images generated throught aerial cameras to build 2.5D map and 3D models in urban areas

Massimiliano Pepe[a], Luigi Fregonese[b] and Nicola Crocetto[c]

[a]DICATECh, Politecnico di Bari, Italy; [b]Department of "Architecture, Built Environment and Construction Engineering", Politecnico di Milano, Mantova, Italy; [c]Dipartimento di Ingegneria, Università degli Studi della Campania "Luigi Vanvitelli", Aversa, Italy

ABSTRACT

The use of Structure-from-Motion (SfM) and Multi-View-Stereo (MVS) approaches to build 3D models of structures belonging to the Cultural Heritage environment is becoming more widespread. Due to the big dimensions of the digital aerial images generated by airborne sensors and, consequently, computation problems, these images were rarely used for the construction of 3D models through SfM-MVS approach. In addition, the values of overlap used in traditional aerial surveys would lead to a geometry of acquisition that is quite weak. For these reasons, the use of aerial images generated by airborne sensors in the SfM-MVS approach was limited. However, taking into account the possibility to acquire aerial nadir and oblique images according to multi-view, the increasing of High Performance Computation and the use of Direct Georeferencing, the quality of the city model obtained with the SfM/MVS approach was evaluated on a dataset of aerial images involving the Old Town of Bordeux.

KEYWORDS

SfM; oblique camera; 3D city model; aerial platforms; hybrid sensors

Introduction

The use of airborne camera and sensors plays an important role as survey methods in the representation of urban and cultural heritage sites. The passive sensors have been success-fully applied for a lot of years (Gruen and Spiess 1987; Jensen and Cowen 1999; Petrie and Walker 2007) by aerial film camera and digital sensors (frame camera and three-line scan-ner) (Dold and Flint 2007). In this context, the use of the stereoscopy approach by nadir images makes possible to obtain numerical cartography and orthophoto (Falchi 2017).

Recently, the introduction of a system constituted by multi and oblique cameras allows to represent the urban area according multi view. Some examples of these acquisition sys-tems are the pictometry (Wang et al. 2008), Leica RCD30 Oblique Penta System (Wagner et al. 2013; Pepe and Prezioso 2016) developed by Leica geosystems and Microsoft UltraCam Osprey (Höhle 2013). Leica Geosystems Company has introduced a new concept of the geo-data acquisition for cities developing 'Leica City mapper'. This system is able to

CONTACT Massimiliano Pepe massimiliano.pepe@poliba.it

acquire simultaneously data by active and passive sensors. So it is possible to obtain both the multi-view of an area and a dense point cloud useful for the 3D modelling.

In the last years the Structure- from-Motion (SfM) and Multi-View-Stereo (MVS) photogrammetric approach was applied with success in a great number of studies (Pavlidis et al. 2007; Costantino et al. 2017; Peña-Villasenín et al. 2017; Pepe 2018a) above all in the survey of Cultural Heritage environment.

For this reason, the applications which use images generated by airborne sensors are a few ones. This is due to two main factors: the first factor is linked to the computation issue in relation to the big memory size of the single image. Every aerial image has a dimension of beyond 300 Mb. The greater the number of pixel in an image is, the greater the time that the software uses for processing images. So a significant portion of this time is spent on matching of detected features across the photos. As a consequence, the man-agement and processing of large matrices involve some important computational prob-lems that can be overcome by using of more computers working simultaneously.

The second factor is related to the weak geometry of acquisition. The flight planning is generally realized according to aerial survey tradition, i.e., using a value of over-lap = 60%and side-lap = 15%÷20%. This configuration brings to weak geometry in software based on SfM approach. The SfM approach is based on the fundamental principle of the stereoscopic photogrammetry but in a different way from the traditional photogrammetric methods. In fact, this approach solves the scene geometry without any need to specify a priori targets or check points in the acquired images. To achieve this aim, SfM approach is based on algo-rithms which allow to detect and describe local features of the images, such as SIFT (Scale Invariant Feature Transform) and SURF (Speeded Up Robust Features). So the points of interest well known as keypoints are automatically identified in each image. In general, the number of keypoints found in a single image depends on the image texture and resolution. In order to reconstruct the 3D camera poses, subsequently, the keypoints are matched to other similar ones recognized in other images. The greater the number of images, the greater the likelihood of image matching success. Indeed, the need to acquire the full geometry of the object or interest scene with a high degree of image overlap, gives rise to the name: structure derived from a moving sensor (Westoby et al. 2012).

The acquisition by aerial multi-camera makes possible to improve geometry and increase the probability of the success of the image matching as a consequence.

SfM approach requires, for camera models based on collinearity equations, a minimum of three corresponding points per image acquired according to convergent angles.

A further contribution can be provided by knowledge of the External Orientation (EO) parameters obtained from the connection between the aerial sensor with GNSS (Global Navigation Satellite Systems) and INS (Inertial Navigation System), such as Applanix (developed by Trimble Company), IPAS (developed by Leica Geosystems Company) or Aerocontrol (developed by IGI Company). To achieve high accuracy of the EO, the GNSS/INS data (rover) must be combined with the GNSS data from a terrestrial station (master) whose data can be obtained by the use of the CORS infrastructure (Continuously Operating Reference Station) or positioning a GNSS instrumentation on a vertex (or several vertices) of known coordinates (Pepe 2018b). Performing an adequate post-processing of the airborne trajectory with dedicated software, it is possible to deter-mine the spatial coordinates (East, North, ellipsoidal height) and the attitude angles (*Roll, Pitch, Heading*) for each image, i.e., the EO of the image. The technique that allows to determine the parameters of external orientation is known as Direct Georeferencing (DG)(Cramer 2001; Pepe et al 2015); in fact, the benefits of DG in SfM approach was evaluated in order to build 2.5D map and 3D models in urban areas. In the case some parts of the structure cannot be observed using only aerial images it is possible to consider the integration by terrestrial images.

Therefore, taking into account the dataset of the images and lidar point acquired over the city of Bordeaux (France), three main aspects are addressed and discussed in the paper:

- quality of the aerial images alignment with and without the use of EO data;
- evaluation of the 2.5D map generated by aerial images and comparison with the 2.5D lidar map;
- possibility of building accurate 3D models: two cases of study taken into consideration. In the first case it has been used a dataset with only aerial images to build 3D model and, subsequently, 2D orthophoto of the façades. In order to evaluate the accuracy of 3D model in the best possible configuration and using images in the TIFF (*Tagged-Image File Format*) format, more tests were carried out to verify the impact about the quality of the 3D model through the use of compressed images in JPEG (*Joint Photographic Experts Group*) format and weakening the geometry of the acquisition scene. In the second case study, the 3D model has been obtained through the integration of aerial images and free terrestrial dataset.

Organization of the paper

The section called ' Aerial platforms and air space' describes the use of aerial platforms for photogrammetric and remote sensing purposes, as well as, the constraints imposed on the flight in some airspace.

The section called 'Data and Methods' describes the airborne dataset used in the experimentation and the workflow that it is necessary to develop in order to obtain the urban models, such as 2.5D map and 3D models. In particular, the aerial images used for the several tests were acquired with multi-cameras over the historic centre of Bordeaux, in France. The use of SfM-MVS approach was applied and evaluated in two specific cases of study in order to build 3D models. In addition, using this photogrammetric approach, a 2.5D map was produced and compared with one generated from Airborne Laser Scanner sensor (ALS).

'Discussion' and 'Conclusions' are summarized at the end of the paper.

Aerial platforms and air space

The photogrammetric survey can be realized using different aerial platforms and applied in a lot of fields, such as archaeology, 3D City Mapping, Digital Cultural Heritage and vegetation monitoring (Everaerts 2008). Historically, the aerial surveys were carried out using fixed-wing aircraft. This type of aircraft was used as an aerial platform using frame cameras, three line-scanner (push-broom), hyperspectral, RADAR and lidar sensors (Olsen and Olsen 2007; Pepe et al. 2018). In the survey of special mapping, the helicopters can be used with great advantage to obtain geo-data with high resolution (Morgan and Falkner 2001). Reducing the helicopter speed, it is possible an enhanced observation of the area of interest. For this reason, this aerial platform is widespread in the survey which use push-broom and Airborne Laser Scanner (ALS) sensors. In the last years, the use of UAVs (Unmanned Aerial Vehicles) well known also as 'drones' or 'UAS' (Unmanned Aerial System) is becom-ing increasingly widespread as aerial platform and technological solution to optimize photo-grammetric flight. This wide and rapid diffusion of UAS is due to the great flexibility in the execution of the survey and the high spatial resolution that can be obtained with lighter and more and more performing optical sensors. Nowadays, there are a lot of aerial platform solutions for the photogrammetric survey on the market, such as aircraft, helicopter and UAV rotary- or fixed-wing and single- or multi-rotor.

Anyway the UAS must adapt to the rules and directives of airspace too. So, before starting a flight, it is necessary to check the availability of the air space about the investi-gated area because there are several areas where it is not allowed to use drones. For example, one of these areas is the historic centre of Bordeaux (France), as shown in the web mapping service of the French government (https://www.geoportail.gouv.fr/donnees/restrictions-pour-drones-de-loisir).

It is possible to overcome this limitation in the data acquisition using a system of digital multi-cameras mounted on aerial platform, i.e., on aircraft with pilot and observer. Considering that the aircraft flight must be performed at a greater altitude than the UAV aerial platform, this means that the distance between object and sensor could be increased and, as a consequence, obtained an image with a lowest geometric resolution.

SfM-MVS approach

In general, a dataset of stereoscopic images can be processed by Structure-from-Motion (SfM) approach that is the process of reconstructing 3D structure from its projections into a series of images taken from different viewpoints. This very popular approach provides fully automated 3D modelling for arbitrary imagery without any pre-knowledge or on-site measurements (Luhmann et al. 2007). The MVS approach is used to increase the density of points starting from the point cloud generated in SfM process (Skarlatos and Kiparissi 2012). Therefore, the several steps that are necessary in order to obtain 3D model, can be summarized as follow:

- *Alignment of the images.* This task allows to determinate the camera position and orientation of each image. In addition, the SfM software produces a sparse point cloud. The two main steps that lead to the reconstruction of the model can be summarized as follows (Snavely 2008):

  Correspondence search:
  - *Feature extraction*: this task can be realized, for example, using the Scale-Invariant Feature Transform (SIFT) feature detector (Lowe 2004) which it is a feature detection algorithm in computer vision to detect and describe local features in images.
  - *Matching*: the system searches for feature correspondences by finding the most similar feature in image $\mathbf{I}_a$ for every feature in image $\mathbf{I}_b$, using a similarity metric comparing the appearance.
  - *Geometric verification*: this stage verifies the potentially overlapping image pairs using a robust estimation technique, such as RANSAC (RANdomSAmple Consensus) (Fischler and Bolles 1987).

  Reconstruction stage;
  - *Initialization*: the system begins with a carefully selected two-view reconstruction.
  - *Image registration*: in this task, using feature correspondences, new images can be registered to the current model by solving the Perspective-n-Point (PnP) problem (Fischler and Bolles 1987).
  - *Triangulation*: a newly registered image must observe existing scene points and, even in this step, the use of robust estimation technique method is crucial.

- *Bundle Adjustment (BA)*: this is a refinement method used to improve SfM solutions
- *Building dense* clouds. Based on the estimated camera positions, it is possible to calculate depth information for each image in order to be combined into a single dense point cloud. On the base of MVS algorithm, this task allows to increase the density of the point cloud generated in SfM process.
- Building mesh. The process of transforming from the point cloud into a single-surface triangulated mesh can be fulfilled with different algorithmic approaches in relation to morphology of the object or area under investigation. For example, Agisoft PhotoScan software offers, in relation to the type of the surface, three possible setting: *Exact, Smooth and Height-field*. These several approaches are based on pair-wise depth map computation.

Data and method

Historic Centre of Bordeaux

The aerial dataset was acquired over the historic centre of Bordeaux, in France. Bordeaux is a port city located on the banks of the Garonne River, in the south-west of France, whose distance is about a few hundred kilometers from the Atlantic Ocean. The *Port of the Moon* (port city of Bordeaux) is inscribed as an inhabited historic city, an outstanding urban and architectural ensemble, created in the age of the Enlightenment, whose values continued up to the first half of the 20th century. For this reason, since 2007 the historic centre of Bordeaux has been declared by UNESCO World Heritage. More than 350 buildings are classified as historical monuments, such as: *Cathédrale Saint-André*, *Le Grand Théâtre*, *Bordeaux Palaisde la Bourse*, *Basilique Saint-Seurin*, *Basilique Saint-Michel* and *PalaisGallien*.

In this paper, two historical buildings have been considered: the 'Grand Théâtre de Bordeaux' and the 'Porte de Bourgogne'. The basic criterion that led to the choice of these two buildings is connected to the visibility from the aerial images of the entire structure. Indeed, while the building 3D model of 'Grand Théâtre de Bordeaux' was realized using only aerial images, for the 'Porte de Bourgogne' the aerial (images) dataset was integrated with terrestrial images. It was not possible to represent the intrados of the arch of the 'Porte de Bourgogne' using aerial images only, the terrestrial images were obtained from internet resources in free way.

The *Grand Théâtre de Bordeaux* was designed by the architect Victor Louis and was conceived as a temple of the Arts and Light. The main façade, in neo-classical architecture, has a portico of 12 Corinthian style colossal columns that support an entablature on which stand 12 statues representing the nine Muses and three goddesses. The dimensions (taken from 2.5D map) in plan of the structure are 81 m in length and 45 m in width. The height of the structure is variable because of the shape of the dome roof; the highest part of the dome is about 15 m.

The *Porte de Bourgogne* (formerly known as *Salinières*) is a medieval gate located at the end of *Cours Victor Hugo*, one of the main roads in the historic centre of Bordeaux. This stone arch was erected in 1757 and marked the official entrance to the city on the old road from Paris. Until 1807, two smaller arches stood next to the main arch and were joined to the surrounding buildings.

The two buildings chosen for the experimentation is located in the historic centre of Bordeaux, as shown in the orthophoto realized from optical dataset (see Figure 7).

Hybrid (optical and lidar) sensors and software used in the experimentations

The geo-data were acquired by 'Leica CityMapper' hybrid sensor mounted on the Partenavia P68C airborne (a twin piston engine) over the city of Bordeaux (approximate geographic coordinates: $\varphi = 40°50'33''$; $\lambda = 0°34'25''$) at an altitude of 850 m above the ground, i.e., 850 m AGL (Average Ground Level). Since the altimetry of the study area is rather flat, the height of each Flight Line (FL) has been the same.

The optical part of Leica CityMapper sensor is constituted by one nadir frame multispectral camera that uses 83 mm lens and four oblique frame cameras using 156 mm lens according a tilt angle of 45° (Figure 1(a,b)). In particular, both the nadir and oblique camera allow an acquisition of the image by a sensor of 10,320 × 7,752 pixels (80 MP). These cameras get a better image quality due to the Mechanical Forward Motion Compensation (FMC) along two axes and the nadir camera. To obtain RGB colour is able to acquire in Near Infrared (NIR 780–880 nm). Anyway in this paper RGB images have been used only.

The flight planning was designed in order to obtain a Ground Sample Distance (GSD) of 5 cm for Nadir images, where the relation between the diverse elements can be calculated as (Neumann et al. 2016):

$$GSD = \frac{Z}{c} \cdot \text{CCD pixel size} \tag{1}$$

where Z is the flight height (also indicated with the letter H) and c is the focal length;

$CCD_{\ pixel\ size}$dimension of the sensor (5.2 µm).

Using oblique frame camera, the formula for GSD calculation can be calculated by following formula (Höhle 2013):

$$GSD = \text{pixel size} \cdot \frac{Z}{c} \cdot \frac{\cos(\beta - t)}{\cos\beta} \tag{2}$$

where β is the angle between the direct line from the lens to a target and the vertical; t is the tilt angle of the considered camera.

In Figure 2a,b is reported the nadir and oblique ($t = 45°$) acquisitions configurations in the hypothesis of flat terrain. From the observation of these figures, it is easy to note as, especially in the case of oblique cameras, the GSD varies considerably as one moves away from the principal point (PP').

Considering the features of the Leica CityMapper Airborne Hybrid Sensor (i.e., the focal length of 86 mm for nadir camera, 156 mm for oblique camera and the pixel size of 5.2 µm) and using the formula for nadir (1) and oblique (2) images, it was possible to build the following graph (Figure 4). Here it is possible to note as, for the same flight height, the GSD in the principal point is lower than that can be obtained in the normal

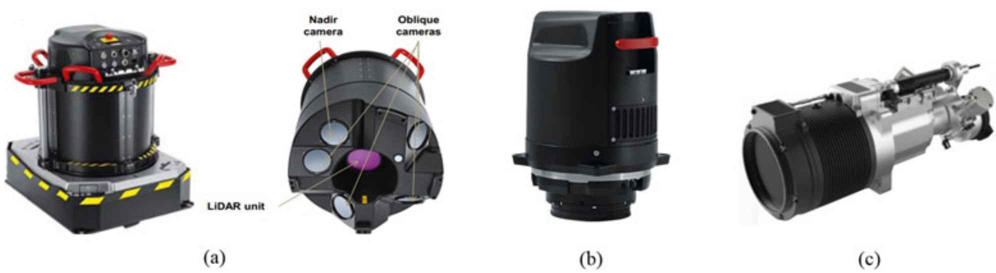


Figure 1. a) Leica CityMapper hybrid sensor: 1 Leica RCD30 CH82 multispectral camera in nadir, 4 Leica RCD30 CH81m oblique camera, viewing angle 45° and 1 Leica LiDAR unit (positioned in the central part of the sensor); b) Leica RCD30; c) Leica Hyperion LiDAR unit

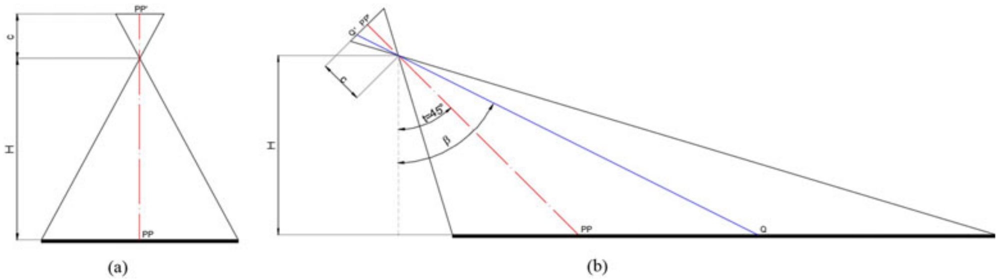


Figure 2. Acquisition geometry. a) nadir acquisition configuration, i.e. the so-called "normal case"; b) oblique acquisition configuration with tilt angle of the camera of 45°

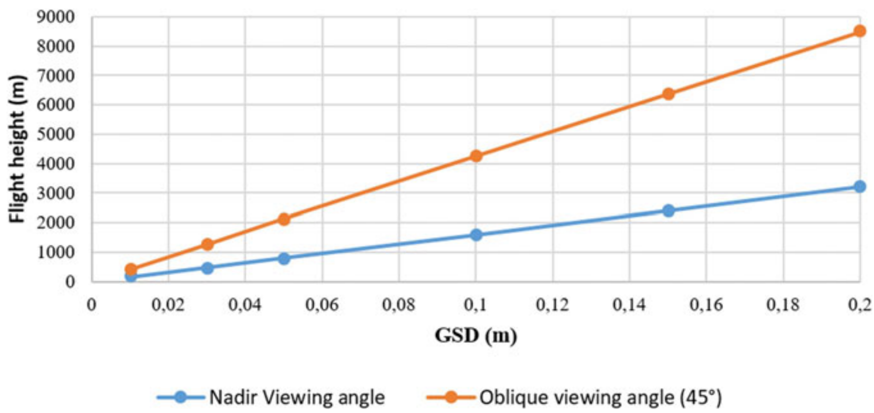


Figure 3. Relation between flight height and GSD for Leica CityMapper Airborne Hybrid Sensor in the principal point. a) nadir acquisition configuration; b) oblique acquisition configuration with tilt angle of the camera of 45°.

case. In addition, from the graph of Figure 3, it is possible to note as, for the same height, the GSD in the main point is lower than that obtained in the normal case.

Beyond optical sensors, the *CityMapper* sensor is also equipped with an ALS (Airborne Lidar Scanner), as shown in the Figure 1a,c. The ALS sensors provide the range measurements between the laser scanner and the Earth topography by the time-of-flight between the emitted and backscattered laser pulse (Tran et al. 2015). In addition, the last versions of ALS sensors are able to record the full waveform of the electromagnetic wave.

Using hybrid sensors it is possible to obtain not only optical data but also a georeferencing point cloud of the scene surveyed where for each point are known the spatial coordinates, the intensity and return echoes for each pulse sent to the target. The main features of the ALS sensor provided in CityMapper system are:

- Pulse repetition frequency up to 700 KHz;
- Return pulses Programmable up to 15 returns, including intensity, pulse width, area under curve and skewness waveform attributes; full waveform recording option at down-sampled rates;
- Oblique scanner, with various scan patterns;
- Real time LIDAR waveform analysis, including waveform attribute capture.

The Flight Planning was realized using Leica MissionPro in order to obtain a GSD of 5 cm for Nadir images. The optical dataset consists of 242 Nadir images and 1210 images according backward, forward, left and right geometric positions. The external orientation (EO) of every image captured is given in UTM Zone 30 North. As concerns the geo-data acquired by ALS sensors, the flight planning was designed in order to obtain a point density of 7-8pts/m^2. In particular, the lidar dataset consists of 3 point clouds file in LAS ASPRS format.

Agisoft PhotoScan Pro Software Version 1.4.4. was used for the post-processing of the images, i.e., to build 2D, 2.5D and 3D models based on SfM/MVS algorithms.

Method

In order to apply the SfM-MVS method to aerial images it is necessary to consider several factors, such as the size of the image, the contribution of EO data in the alignment step and the actual coverage of buildings with aerial images.

Considering the large dimension of the single image, it is necessary to transform the image *TIFF* in a format more manageable, such as the format *JPEG*. Indeed, the aerial dataset that covers a city is usually quite rich of images. As a consequence image transformation task depends on the number of images contained in the dataset and the potential of the computers by which it is possible to perform the calculation. Image format

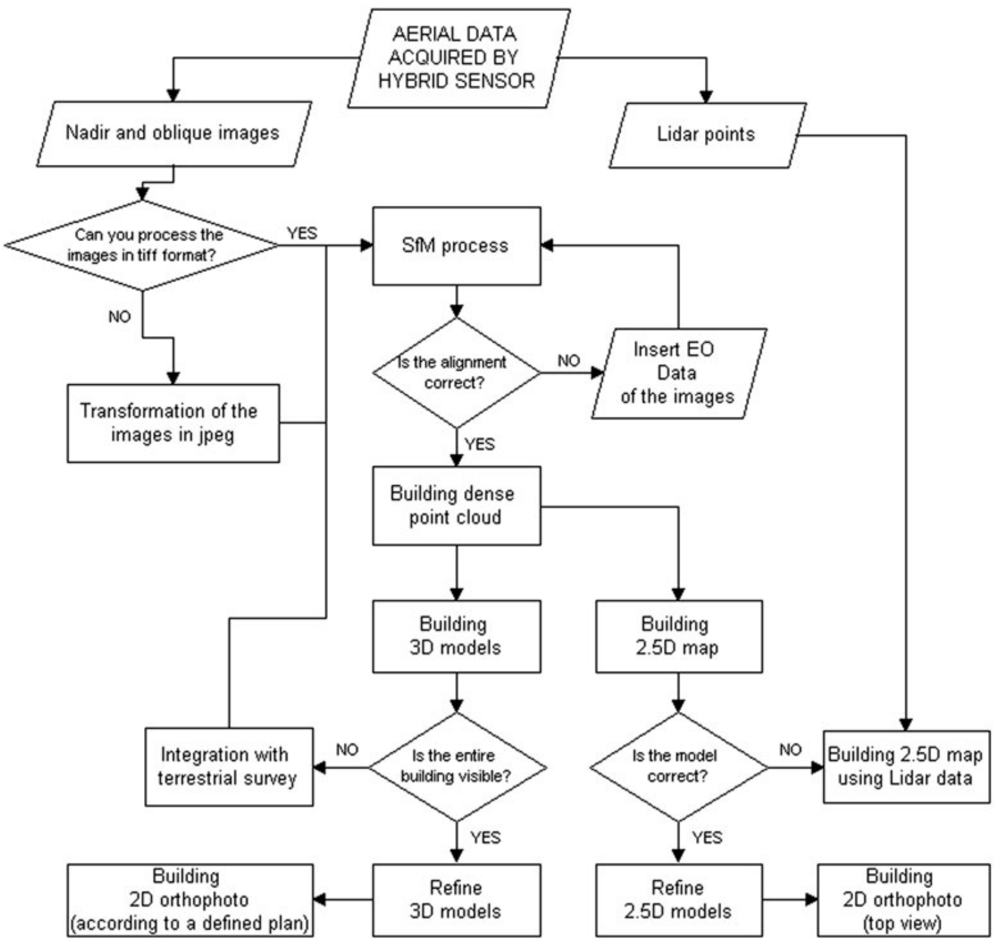


Figure 4. Flowchart of the developed method

transformation can be performed easily and quickly in image management and editing software, such as Adobe Photoshop. This software allows thirteen total levels of compres-sion from 0 to 12: choosing a lower number results a greater compression and a smaller file size. The compression value equal to 10 was chosen in this case study.

About the alignment of the images, if the external orientation of each image is known (calculated by the use of kinematics techniques) these values can be inserted into the SfM-MVS software. This possibility could speed up, improve the image alignment process and generate georeferenced point cloud in a specific coordinate system.

In addition, depending on the distance between the buildings, the height of the buildings and the overlap of the images, some parts of a building could not be visible. So it is necessary to integrate the aerial survey by a terrestrial one.

The method developed for the construction of the 2.5D map, 3D model and 2D orthophoto of the building facades or orthophoto of the city can be summarized in the following pipeline (Figure 4).

Processing of the areal images by SfM approach: alignment of the images with and without EO

All the images of the historic centre of Bordeaux were used in order to verify the ability of the software to align the images and to generate the 3D city model. Anyway consider-ing the large dimension of the single image and the large number of images (1210), the first step was the reduction of the image size transforming the extension from *TIFF* to *JPEG* format. This way the overall dataset was reduced from 383 GB to 64.4 GB. Since PhotoScan software can be configured to run on a computer cluster where processing is distributed on several connected computer nodes, all the images were processed in a single chunk. In particular, the processing was divided onto 4 computers set up in the minimum configurations suggested by the software developer. The main features of the PCs are reported in the following table 1:

In the PhotoScan setting the parameter ‘Higher accuracy’, that’s the full resolution photo use, it has been possible to obtain the best accurate camera position and orientation estimations by SfM approach. Despite the use of this setting, the image alignment proc-essing did not produce satisfactory results. Anyway many images of the dataset were incorrectly aligned. This is due to the weakness of the image geometry, i.e., small values of side-lap and over-lap. Of consequence, the sparse point cloud represents incorrectly state the places, as shown in the figure 5a.

For this reason a new project in SfM PhotoScan software was created using EO data obtained by post-processing of the GNSS/INS data. Since the EO data contain the altitude in ellipsoidal height, it was necessary to transform this height in orthometric using

Table 1. Main features of the PCs used for the experimentations.

	PC#1	PC#2	PC#3	PC#4
CPU (Central Processing Unit)	Intel(R) Xenon(R) CPU E5-2643@3.30ghZ	Quad-core Intel Core i7 CPU, Socket LGA 1150	Quad-core Intel Core i7 CPU, Socket LGA 1150	Quad-core Intel Core i7 CPU, Socket LGA 1150
RAM (Random Access Memory)	128GB	16 GB	16 GB	8 GB
GPU (Graphics Processing Unit)	Nvidia Quadro K5000	Nvidia GeForce GTX 980	Nvidia GeForce GTX 980	Nvidia GeForce GTX 980

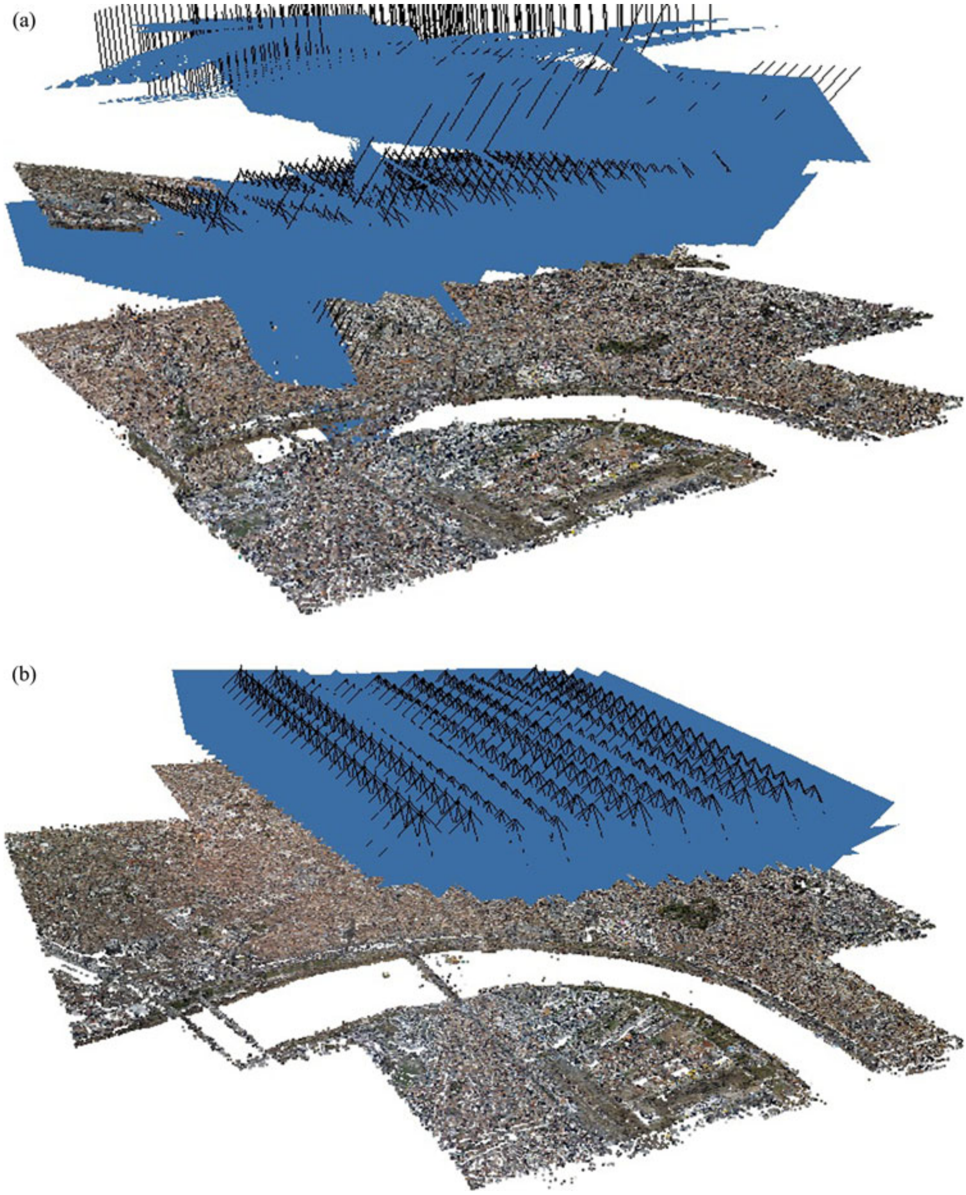


Figure 5. Representation of sparse point cloud and camera poses within Photoscan software; a) without the use of Direct Georeferencing; b) using Direct Georeferencing

adequate algorithms developed in Matlab software and specific geoid model relative to the area affected by the survey. In particular, the geoid model adopted was RAF09, which is a hybrid geoid model for France with a grid spacing of 1.5' in latitude and 2' in longitude. By performing a bilinear interpolation, the ellipsoidal height was transformed into an orthometric height. Of consequence, the EO file with the orthometric height was reconstructed. Using EO data, the images were correctly positioned in the space and the time of the alignment of the images were halved compared to the previous project where the EO parameters were unknowns. This way the sparse point cloud was positioned on a single and continuous plane (Figure 5b).

Use of SfM-MVS photogrammetry approach to build 2.5D map

Since the use of EO parameters allowed to align the images correctly, the use of MVS algorithm made possible to build a dense point cloud. PhotoScan software allows different level of point density in relation to the downscaled of the images. For example, the option 'Ultra High' means that the image does not scale, 'High' the images are downscaled two times by each side, 'Medium' the images are downscaled four times by each side, 'Low' the images are downscaled eight times by each side and 'Lowest' are downscaled sixteen times by each side.

In this project, the option 'Medium' was considered. This way, dense point cloud of the historic centre of Bordeaux was generated.

The processing of the (dense) point cloud has requested a time of about 18 hours. At the end of the process, about 100 million points were generated. Subsequently, the point cloud was exported in LAS ASPRS file because this type of extension can be managed in many GIS (Geographic Information System) software and in point cloud edit-ing software.

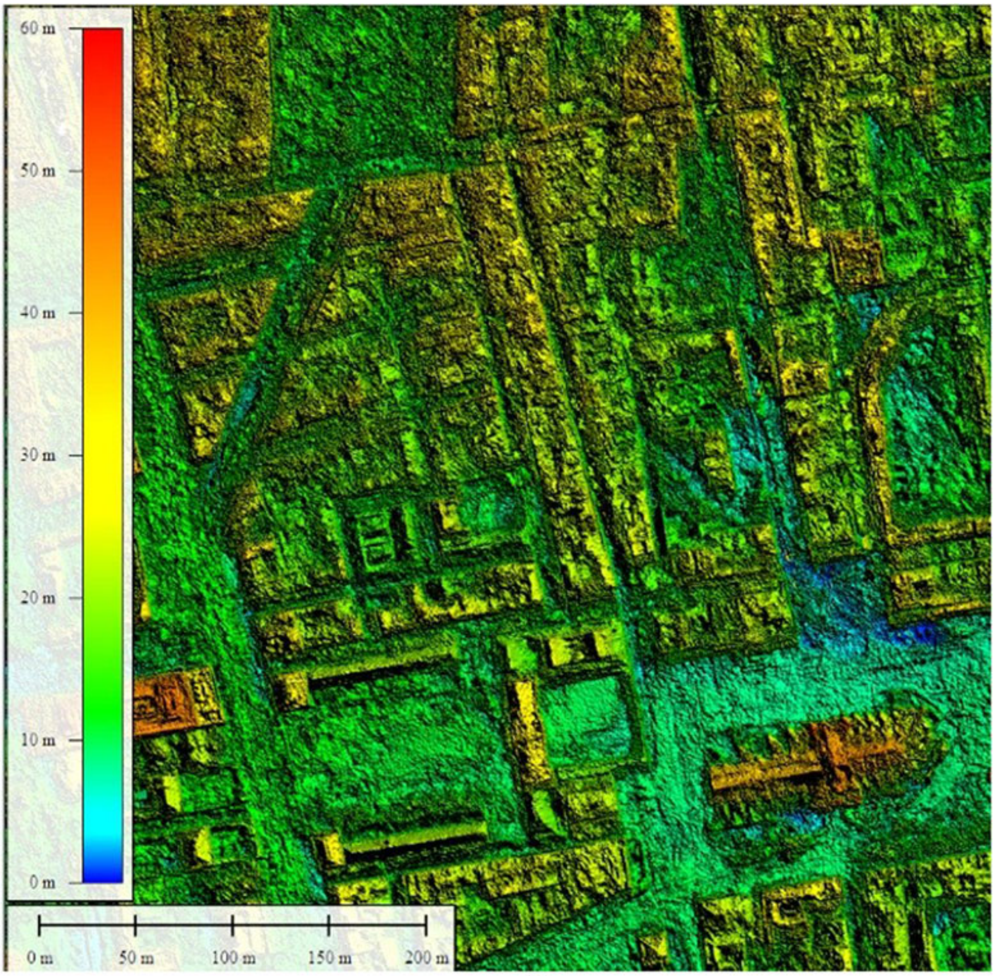
In fact, starting from the point cloud and using suitable tools developed in GIS software, it was possible to build a 2.5D map of the territory. So a DSM (Digital Surface Model) of the study area, where the altitude is encoded in the pixel value, was obtained.

In this case study, the point cloud generated by optical data and SfM-MVS approach was compared with the point cloud generated by ALS sensor. In this way, it was possible to identify the level of the detail of the photogrammetric point cloud. As shown in Figure 6, the map obtained by airborne images and photogrammetry approach is very detailed even if the elevation model is rather noisy if compared to that one obtained by the lidar technology. Especially in narrow streets, the model generated by optical sensor is quite noisy (see Figure 6(a)) while the model gener-ated from ALS sensor is precise and detailed over the entire study area. This means that the ALS is a very efficient sensor in order to build 2.5D map, as shown in Figure 6b.

From the analytical point of view, using ArcMap v 10.4.1 software, the DSM generated by SfM approach was compared with one generated by lidar sensor. The difference between the two raster showed a mean value of 0.13 m and a standard deviation of 1.35 m. The greatest differences occurred in the vegetated areas, along the edges of some buildings and in relation to the density of the points. This difference is due to the ability to lidar sensor to produce a point cloud denser than the photogrammetric one and consequently, obtaining a DSM with more details.

Use of SfM-MVS photogrammetry approach to build the orthophoto of the old town

After that the dense point cloud was reconstructed according photogrammetric method, it was possible to generate polygonal mesh model based on the dense cloud data. Unwanted faces were removed from the model and areas with lack of geometry was improved by the use of apposite tools developed within the software. With a geometric resolution of 5 cm, the orthophoto of the old town of Bordeaux was built. Since the exported area is quite large, the orthophoto was split in blocks feature of the dimension of 1000 m $\times$ 1000 m. A part of the entire orthophoto, where the two buildings are visible under investigation, is shown in Figure 7.



(a)

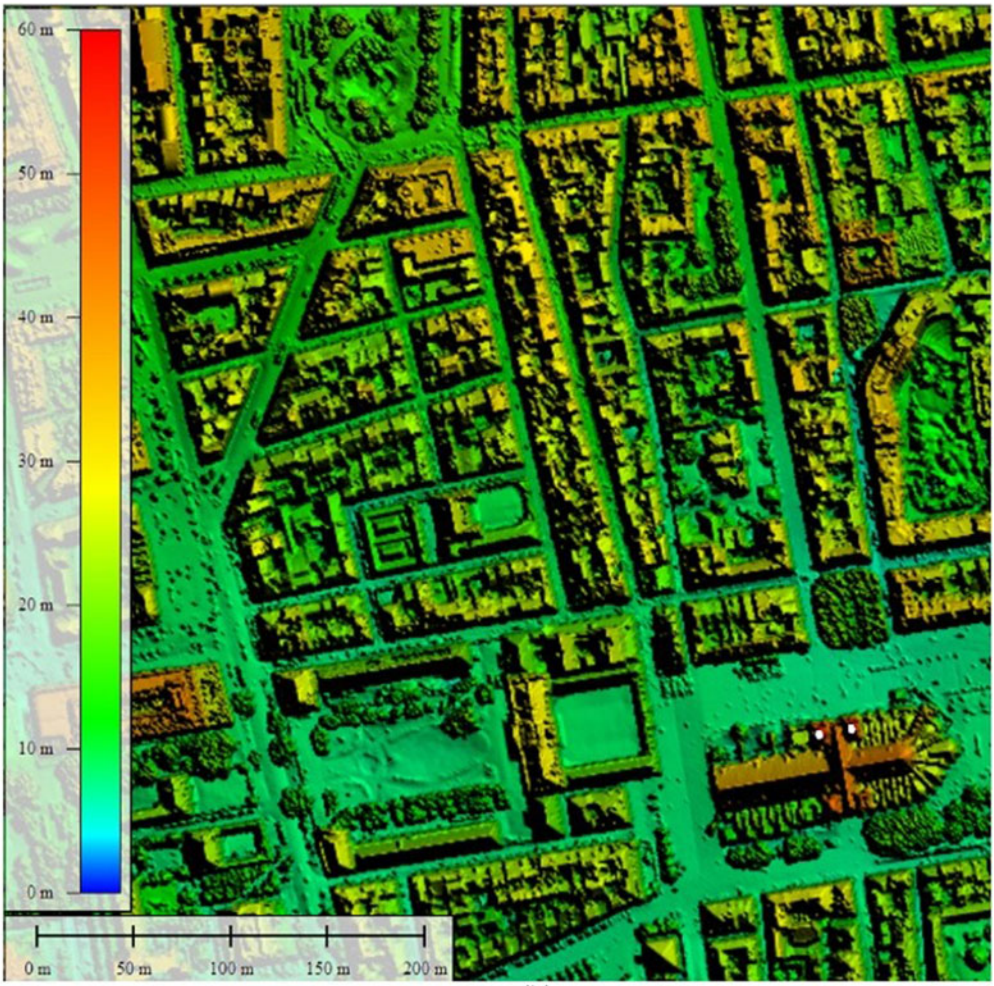
Figure 6. DSM of the part of the historic centre of Bordeaux. a) Photogrammetry DSM; b) lidar DSM.

Use of SfM-MVS photogrammetry approach to obtain 3D building models

The case study of the Grand Théâtre de Bordeaux

In this study about the Grand Théâtre de Bordeaux aerial images were used only. In order to build the 3D model of the *Grand Théâtre de Bordeaux*, from the dataset were chosen the images where the building of interest was visible. For this reason, the dataset was reduced to 39 images for a total of 12.3GB in *TIFF* format. Once selected these images and using EO parameters, a new project in SfM/MVS was created in order to take into account the TIFF format of the images. In this case study, all the images were correctly aligned and the quality of the matching was evaluated as the minimum and maximum value of number projections and error, as shown in Table 2.

The evaluation about the accuracy of the model was obtained comparing the real dis-tance between 2 points easily recognizable (even on the photos) with the measurable one of the model. Therefore, six ground control points were identified on the images. The real



(b)

Figure 6. Continued

measurement was performed through a direct survey with a laser distance. As a consequence, three measurements along the structure were carried out. The results of the comparison are shown in Table 3.

Once verified that the images were positioned in the right way, it was possible to build the dense cloud using the 'Ultra High' setting. Subsequently, starting from the dense point cloud, the mesh was built. This way, it was possible to obtain the 3D model of the theatre, as shown in Figure 8a.

The model can be exported in a special format and used in other software for further analysis and representation, such as Google Earth KMZ, Wavefront OBJ, 3DS models, and Adobe PDF. Since PhotoScan tends to produce 3D models with excessive geometry resolution, it could be better to decimate mesh before exporting, avoiding performance decrease of the external program. Recently, in the world of 3D models, the use of Sketchfab is becoming increasingly popular. Sketchfab is a platform to publish, share, discover, buy and sell 3D models. It provides a viewer based on the WebGL (Web Graphics Library) technology that allows users to display 3D models on the web. These models can

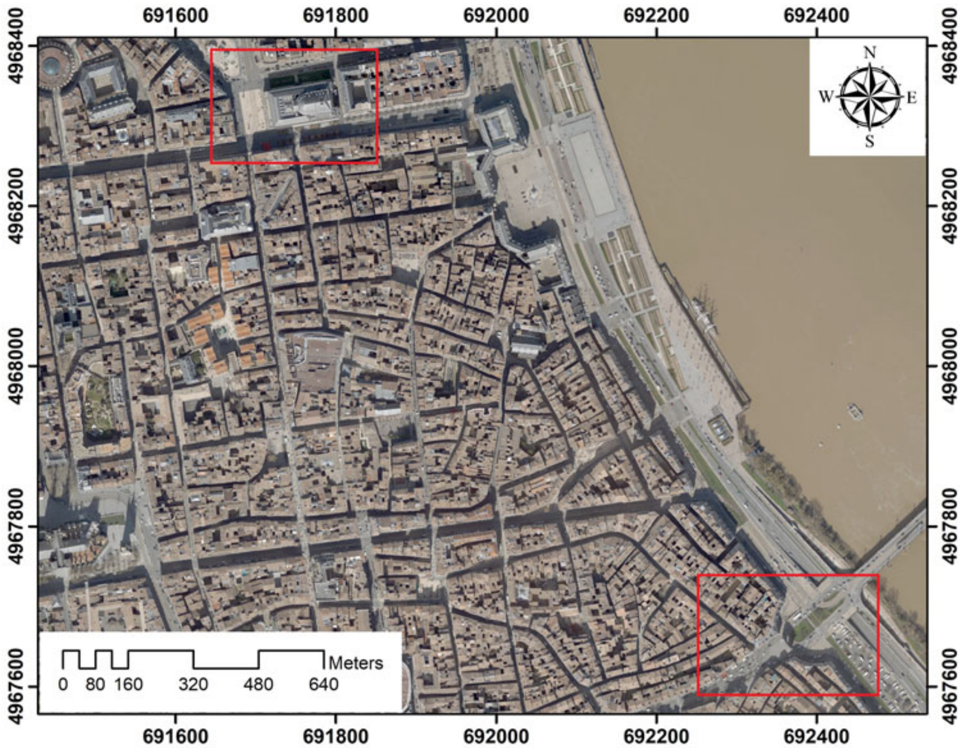


Figure 7. Part of orthophoto of the old town of Bordeaux with identification of the buildings under investigation in UTM 30N WGS84 coordinates. The orthophoto was generated by aerial dataset with a geometric resolution of 5cm: at the top left of the picture is located the Grand Théâtre de Bordeaux while in the lower right is located the Porte de Bourgogne.

Table 2. Quality of the alignment.

	Number of Projections	Error (pixel)
min	8465	0.435
max	14983	0.814

Table 3. Accuracy of the photogrammetric block of the images.

Scale bar	Real dimension (m)	Distance on the model (m)	Error (m)
1–2	1.30	1.27	−0.03
3–4	2.22	2.30	0.80
5–6	1.30	1.24	−0.06

be viewed on any mobile browser, desktop browser or Virtual Reality headset (Sketchfab 2019). Therefore, the final 3D model can be uploaded (for example in OBJ format) and managed on Sketchfab platform.

In addition, once obtained the 3D model in real dimension, it was possible to generate the orthophoto of the plan and the single façade of the building (2D) according to the traditional approach (see Figure 8(b)). Indeed, by choosing appropriate local reference systems it was possible to construct the orthophoto of the individual façades.

To realize this task, on a façade of the model were chosen three points (GCPs) easily recognizable on each image. So PhotoScan software performs automatically a roto-

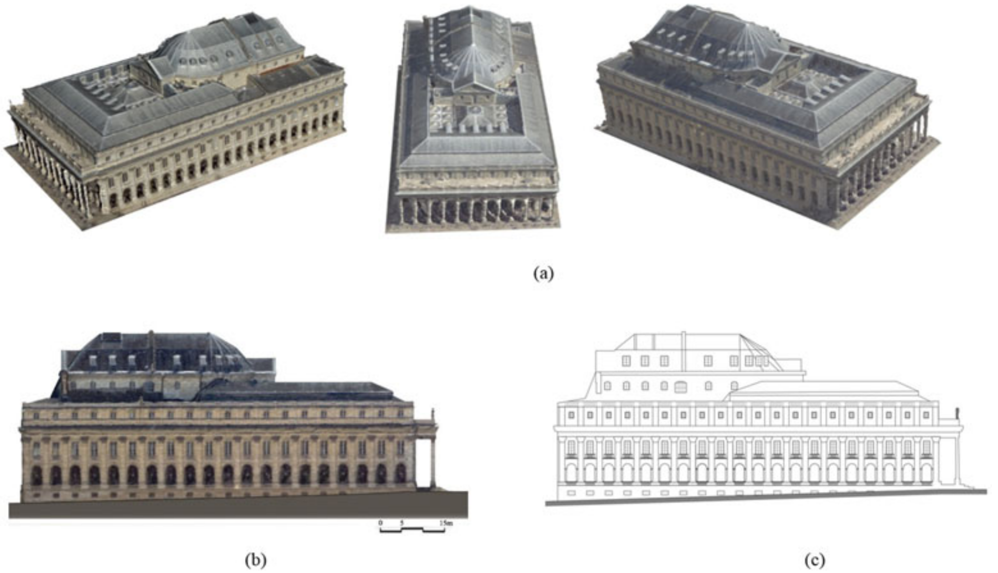


Figure 8. Representation of the Grand Théâtre de Bordeaux. a) 3D model obtained using only aerial images; b) 2D representation (orthophoto) of the lateral façade of the theatre c) 2D CAD representation of the lateral façade of the theatre.

translation of the model for these points and identifies a specific plane where building the orthomosaic. In this environment, rotating the object every 90°, it is possible to build the orthomosaic of the single façade of the theatre. To improve the quality of the orthophoto, a special editing (correction of brightness, exposure, contrast and tone) in Adobe Photoshop environment was carried out. For example, Figure 8b shows the orthophoto of the lateral side of the *Grand Théâtre de Bordeaux* while in Figure 8c a traditional repre-sentation in Computer-Aided Design (CAD) environment is showed.

Compared to the first case study that involved images of the entire historic centre, it was preferred to use only a part of the dataset and keep the TIFF format. In fact, the compression of the original file in JPEG format causes a degradation of the image quality. The greater the compression, the greater the loss of image quality. In order to identify the impact of compression in the construction of the 3D model, four scenarios were built with varying compression levels. The image compression operation was performed auto-matically in Adobe Photoshop. This way, beyond the previous project that use images in TIFF format, three projects in the SfM software were created using images with a level of compression of 1, 5 and 10. Therefore, the original size image (332MB) in TIFF format becomes 24.07MB for a compression level of 10, 8MB for a value of 5 and 3.84MB for a value of 1. The values of size dimension for JPEG format represent the average of the dataset. Processing the several datasets in SfM software, it was possible building three dif-ferent models of the theatre according a point clouds representation. Subsequently, the use of Cloud Compare software, an open source software for 3D point cloud processing, it has been possible to calculate the distances between the reference model, that is the model generated by TIFF images and the other models generated by JPEG images. In order to improve the accuracy of the point clouds, a 'cleaning' operation to identify and exclude outlier points that can negatively influence the results were performed. This activ-ity was carried out using a special filter implemented within Cloud Compare called "S.Q.R. filter". This filter computes the average distance of each point to its neighbour

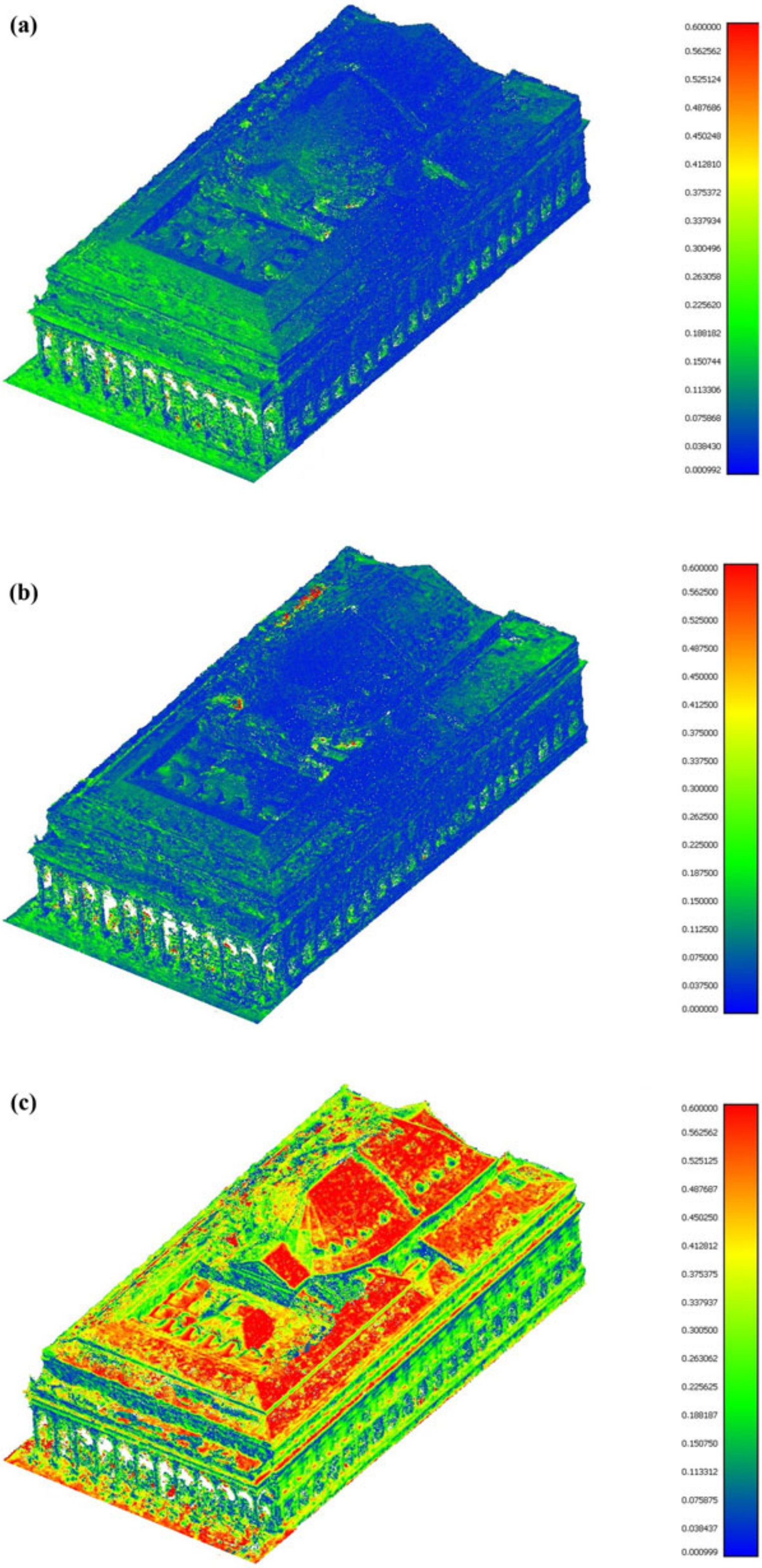


Figure 9. Cloud-to-cloud distance model: comparison between 3D models generated by images in TIFF format and JPEG with a level of compression of 10 (a), 5 (b) and 1 (c).

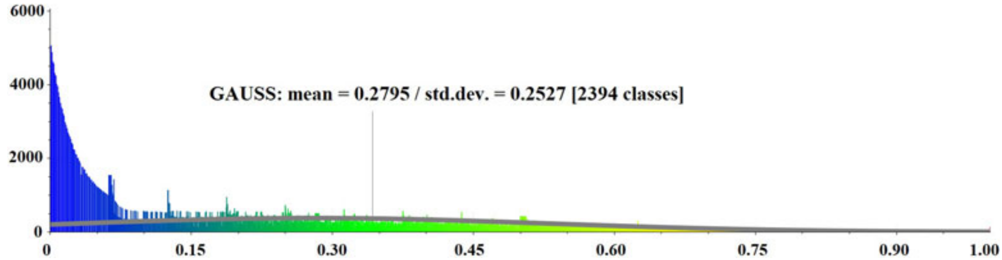


Figure 10. Gaussian distribution of the comparison between the dense point cloud (generated by original dataset) and the reduced point cloud performed in Cloud Compare software.

and then it rejects the points which distance exceeds the average distance plus a number of times the standard deviation value. The comparison between the several point clouds is shown in Figure 9.

As shown in Figure 9, increasing the level of compression the distance between the point clouds increases. Indeed, the mean distance between the reference model and one generated by JPEG images was 0.307 m using a level of compression equal to 1, 0.080 m using a level of compression equal to 5 and 0.039 m using a level of compression equal to 10. This means that, in order to obtain accurate 3D models, it is preferable to use images in TIFF format or, if it is necessary, to minimize compression in the JPEG format. The compression value of 10 represents a good compromise.

The influence of the geometric configuration was evaluated reducing the number of images and, consequently, the number of homologous points. This task was done by removing an image between two consecutive images from the original dataset. The comparison between the point cloud generated by the dataset of 39 images and that reduced to 21 images was performed in Cloud Compare. The results of the comparison, repre-sented in the form of error histograms, are shown in Figure 10.

The comparison shows how the reduction of images leads to an average reduction in accuracy of about 30 cm and, in some points, a maximum value of almost 1 meter. Furthermore, the reduction of images led to a significant decrease in the number of points, both in *tie points* and in *dense point cloud* of about one-third compared to the ori-ginal dataset.

The case study of the *Porte de Bourgogne*

The *Porte de Bourgogne* was visible in 15 aerial images. For this reason, these 15 images were selected from aerial dataset (4 nadirs and 11 oblique). Subsequently, always in PhotoScan software, the images were correctly aligned (see Figure 11(a)–left) and the dense point cloud was built in accordance with the SfM-MVS pipeline.

Some parts of *Porte de Bourgogne* were not visible by aerial images (intrados and the lower parts of the arch) the 3D model of the entire structure was carried out by the integration of aerial and terrestrial images.

The integration between aerial oblique imagery and terrestrial imagery was analysed, recently in several papers. Hidayat and Cahyono (2016) discussed about the integration of the UAV aerial and terrestrial images in order to build a 3D model of *Singosari Temple* using SfM approach. Arza-García et al. (2019a) wrote about the survey of an important archaeological site where several cameras and platforms were used according to the hybrid photogrammetry approach. Wu et al. (2018) discussed about the integra-tion of aerial oblique imagery and terrestrial imagery for optimized 3D modelling in

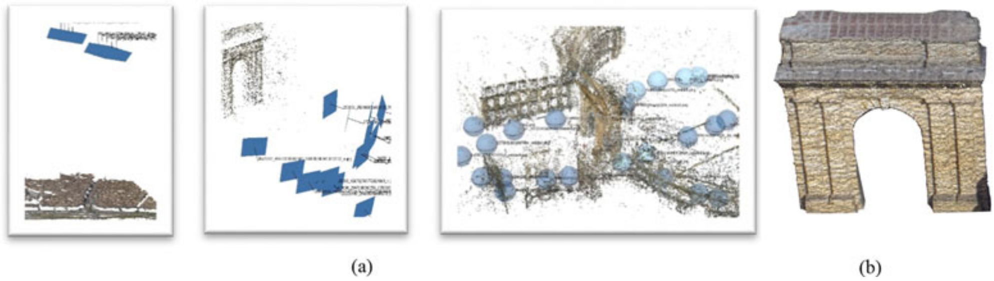


Figure 11. 3D models of the Porte de Bourgogne. a) Cameras pose and representation of the sparse point cloud gen-erated in alignment processing in the case of aerial photo (left), free frame images of a façade (centre) and panorama GSV image (right); b) 3D point cloud with top down view.

urban areas. In this latter manuscript, more case studies are shown; aerial images were obtained from aerial multi-cameras while the terrestrial dataset was obtained by an accurate survey performed for a specific area. Arza-García et al. (2019b) wrote about the data integration between UAV data and Google Earth^{TM} applications using Sputnik GIS software.

Compared to the papers listed above, this one uses an aerial dataset that consists of numerous images where each image takes large dimensions. In addition, the terrestrial dataset was obtained by the download of free images available on Internet or on social networks (such as Facebook and Instagram) and Google Street View (GSV) images.

Before using the free images available on Internet or on social networks, it was necessary to cover some elements that were not part of the scene, such as people, cars, etc., through the use of masks implemented in the software. After making appropriate masks, it was possible to align the images and build the dense point cloud of the terrestrial dataset. Most of the images in the free terrestrial dataset concern the two main façades of the structure only. Since there were no images on the side of the structure, this dataset was divided into two separate projects: a dataset which uses the images facing the river and another one using images in the direction of the historic city centre. In Figure 11a–center, it is possible to see the alignment of the images concerning one of free dataset.

To overcome the lack of terrestrial images on the sides of the structure, images obtained from Google Street View (GSV) were used. Indeed, GSV is a technology that provides pano-rama images at street level in many parts of the world. The images were acquired by Street View camera system, which is a *Rosette* (*R*) of more small, outward-looking cameras using (CMOS) image sensors and custom, low-flare, controlled-distortion lenses. The number of camera and the resolution of everyone is changed over the time. More versions of *R* system were developed. For example, the *R7* (seventh generation of a series of system that Google has developed in-house) Street View camera system is constituted of 15 cameras able to acquire with a resolution of 5-megapixel (Anguelov et al. 2010).

In the case study, 25 panorama images that observe the structure under investigation were used. The panorama images have a resolution of 13312 × 6656 and within of the calibration parameters of PhotoScan software, spherical image setting was configured. The processing of the images within SfM software generated a sparse point cloud where 23 of 25 images were correctly aligned (see Figure 11a–right). This result is connected to the geometric characteris-tics of the photogrammetric block and the quality of the single panorama images.

Therefore, several projects were elaborated according the SfM-MVS pipeline, which main parameters are reported in the table 4.

Table 4. Summary of input and output parameters within of the SfM/MVS software.

Features	Dataset 1	Dataset 2		Dataset 3
				
Aerial/ terrestrial dataset	Aerial dataset	Terrestrial dataset: images from website and social network		Terrestrial dataset: images from Google Street-View
No. of images	15	57		25
No. of images aligned	14 of 15	14 of 31	12 of 26	23 of 25
Min. no. of Projections	7165	402	357	1884
Max no. of Projections	10514	782	547	2717
Error (min)	0.93	0.86	0.89	0.95
Error (max)	0.99	1.14	1.20	2.01
Key points limit	40,000	40,000	40,000	40,000
Tie points	38,705	12,541	11,037	88,045
Point clouds	296,726	216,693	481,194	1,519,085

Table 5. Coordinates of the GCPs taken into consideration and their accuracy.

Id point (#)	*X* (m)	*Y* (m)	*Z* (m)	Error (m)
1	0.000	0.000	0.000	0.039
2	18.217	0.011	−0.010	0.033
3	12.900	−0.623	−0.006	0.019
4	17.553	−0.617	−0.008	0.035
5	12.901	−0.603	1.470	0.016
6	9.126	−0.002	12.365	0.043
7	0.695	−0.626	0.008	0.046
8	5.355	−0.617	0.003	0.052

In order to obtain a unique 3D model, the several projects were connected by common points. The final point cloud model contains about 2 million points and it is shown in Figure 11b.

The evaluation of the model accuracy was carried out taking into consideration 8 GCPs distributed on the façade of the *Porte de Bourgogne*. The spatial coordinates meas-ured on the structure and their relative errors are shown in the table 5. The local refer-ence system is defined as follows: the origin of the axes is in point 1, the x axis starts from point 1 and is directed towards point 2, the y axis is perpendicular to this direction, the z axis is directed upwards (vertical axis). As shown in the table 5, the root mean square error for x, y, z is resulted less than 5 cm.

Discussion

The oblique and nadir images generated through airborne sensor with multi-camera view made possible to obtain a high quality of 3D models. Complex structures, such as the roof or the columns on the main façade of the *Grand Théâtre de Bordeaux*, were correctly represented too. In the place where it is forbidden to overfly by UAV platform, the multi-camera sensors mounted on airborne represents a valid tool in order to obtain geo-spatial data of the upper part of the historical building. If parts of the building are not visible from the aerial images, the aerial survey can be integrated by terrestrial images. As shown in this case study about the *Porte de Bourgogne* the aerial images have been integrated with images generated by free dataset. This way a unique 3D model was obtained. Anyway it must be emphasized that the terrestrial dataset was not acquired for photogrammetric purposes. As a consequence the lighting and stereoscopic positioning conditions may not be ideal; it may be necessary to perform a specific terrestrial survey.

Using SfM and MVS approaches to aerial images, two main aspects must be considered: the low side-lap/over-lap values of the images and time of processing. About the first aspect, it is desirable an increased the percentage of side-lap in the photogrammetric block. In fact, as the paper shows, images with a few homologous points involved in a weak geometric configuration. This issue has been partly solved using external orientation parameters obtained from GNSS/INS sensors and adequate post-processing. Therefore, the geometry configuration may be improved necessary paying more attention to the design of the flight planning and the following aspects:

- increase the degree of image side-lap: this aim can be achieved by increasing the value of the side-lap, i.e. it means increasing the number of flight lines of the photogrammetric block;
- to design a flight planning with orthogonal flight lines between them in order to increase the views of the city. This strategy is already successfully adopted in lidar sensors in order to increase the density of the point cloud.

As regards the time processing, the generation of the point cloud over the entire historic centre was obtained in a time quite elevated. Considering the increasing and constant performance of the PCs, the processing times are destined to decrease. In addition to manage and process images more quickly the images have been reduced to JPEG format. It was shown, in the case study of the 3D model of Grand Théâtre de Bordeaux how, that the excessive compression of images leads to the construction of a point cloud that differs significantly from that generated by images in the tiff format. Therefore, as far as possible, it is desirable to keep the original format.

At the end, as shown in the 2.5D case study, the point cloud generated by ALS sensor is more detailed than to the point cloud obtained from images and obtainable in more restricted times. This means that the ALS sensor represents a valid tool to produce in a rapid way geo-referencing point cloud and useful to generate detailed Digital Terrain Models and Digital Surface Models.

Conclusions

The use of aerial hybrid sensors, i.e., digital optical multi-view cameras and lidar sensor, allowed to build orthophoto of the city (2D), detailed DSM (2.5D), solid and textured building model (3D). In particular, the use of SfM-MVS approach was applied with success in order to mapping the historic centre of Bordeaux (France) and to build realistic 3D building models.

It is possible providing to designers, planners, restorers but also simple users, a lot of information about the city with particular focus on the buildings. The processes of conservation, restoration and maintenance of the buildings within the historic centres can be realized only if a detailed and precise survey of the state of the architectures and urban heritage was carried out.

As shown this survey, carried out by high quality sensors and performing a post-processing with adequate photogrammetry software based on SfM-MVS approach, allowed to obtain a modern, accurate and realistic representation of the city.

Acknowledgements

We want to thanks Mr Felix Rohrbach of Leica Geosystems for providing the raw aerial datasets. In addition, we want to thanks the reviewers for their careful reading of the manuscript and their constructive remarks.

Disclosure statement

No potential conflict of interest was reported by the author(s).

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. 43(6):32–38.

Arza-García M, Gil-Docampo M, Ortiz-Sanz J. 2019. A hybrid photogrammetry approach for archaeological sites: block alignment issues in a case study (the Roman camp of A Cidadela). J Cult Herit. 38: 195–203.

Arza-García M, Gil-Docampo M, Ortiz-Sanz J, Martínez-Rodríguez S. 2019. Virtual globes for UAV-based data integration: Sputnik GIS and Google Earth™ applications. Int J Digital Earth. 12(5):583–593.

Costantino D, Angelini M G, Settembrini F. 2017. Point cloud management through the realization of the intelligent cloud viewer software. Int Arch Photogramm, Remote Sens Spatial Inf Sci. 42: 105–112.

Cramer M. 2001. On the use of direct geo-referencing in airborne photogrammetry. Proceedings of the 3rd International Symposium on Mobile Mapping Technology; Jan 3–5; Cairo, Egypt. Calgary: University of Calgary. p. 1–13.

Dold J, Flint D. 2007. Leica geosystems photogrammetric sensor and workflow developments. In: Fritsch D, editor. Photogrammetrische Woche 2007. Vol. 1. Heidelberg: Wichmann Verlag; p. 3–14.

Everaerts J. 2008. The use of unmanned aerial vehicles (UAVs) for remote sensing and mapping. Int Arch Photogramm. Remote Sens Spatial Inf Sci. 37(2008):1187–1192.

Falchi U. 2017. Spatial data: from cartography to geodatabase. Geodesy and Cartography. 43(4):142–146.

Fischler M, Bolles R. 1987. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In Readings in computer vision: issues. Problems, principles, and paradigms. p. 726–740.

Gruen A, Spiess E. 1987. Point positioning and mapping with large format camera data. Geocarto Int. 2(2):3–15.

Hidayat H, Cahyono AB. 2016. Combined aerial and terrestrial images for complete 3D documentation of Singosari Temple based on Structure from Motion algorithm". In. IOP Conf Ser: Earth Environ Sci. 47(1) :012004. IOP Publishing.

Höhle J. 2013. Oblique aerial images and their use in cultural heritage documentation. Proc. Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Volume XL-5/W2, XXIV International CIPA Symposium, 2 – 6 September 2013, Strasbourg, France.

Höhle J. 2011. DEM generation by means of new digital aerial camera. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 38 (3/W22):185–190.

Jensen JR, Cowen DC. 1999. Remote sensing of urban/suburban infrastructure and socio-economic attributes. Photogramm Eng Remote Sens. 65:611–622.

Lowe D. G. 2004. Distinctive image features from scale-invariant keypoints. IJCV. 60(2):91–110. Luhmann T, Robson S, Kyle S, Harley I. 2007. Close range photogrammetry. Wiley.

Morgan D, Falkner E. 2001. Aerial mapping: methods and applications. Boca Raton (FL): CRC Press.

Neumann K, Welzenbach M, Timm M. 2016. CMOS imaging sensor technology for aerial mapping cameras. Int Arch Photogramm Remote Sens Spatial Inf Sci. 41(B1):69–72.

Olsen RC, Olsen RC. 2007. Remote sensing from air and space (Vol. 1). Bellingham: SPIE Press.

Pavlidis G, Koutsoudis A, Arnaoutoglou F, Tsioukas V, Chamzas C. 2007. Methods for 3D digitization of cultural heritage. J Cult Herit. 8(1):93–98.
Peña-Villasenín S, Gil-Docampo M, Ortiz-Sanz J. 2017. 3-D modeling of historic façades using SFM photogrammetry metric documentation of different building types of a historic center. Int J Architect Herit. 11(6):871–890.
Pepe M. 2018. CORS architecture and evaluation of positioning by low-cost GNSS receiver. Geodesy Cartography. 44(2):36–44.
Pepe M. 2018a. Image-Based Methods for metric surveys of buildings using modern optical sensors and tools: from 2D approach to 3D and vice versa. Int J Civil Eng and Technol. 9(9):729–745.
Pepe M, Fregonese L, Scaioni M. 2018. Planning airborne photogrammetry and remote-sensing missions with modern platforms and sensors. Eur J Remote Sens. 51(1):412–436.
Pepe M, Prezioso G. 2016. Two approaches for dense DSM generation from aerial digital oblique camera system. Proceedings of the 2nd International Conference on Geographical Information Systems Theory, Applications and Management; April 26–27; Rome, Italy. p. 63–70.
Pepe M, Prezioso G, Santamaria R. 2015. Impact of vertical deflection on direct georeferencing of airborne images. Surv Rev. 47(340):71–76.
Petrie G, Walker A. S. 2007. Airborne digital imaging technology: a new overview. Photogramm Record. 22(119):203–225.
Skarlatos D, Kiparissi S. 2012. Comparison of laser scanning, photogrammetry and SFM-MVS pipeline applied in structures and artificial surfaces. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci. 3: 299–304.
Sketchfab 2019. Your 3D content on web, mobile, AR, and VR. [accessed 2019 Aug 12]. https://sketchfab.com/.
Snavely N. 2008. Scene reconstruction and visualization from internet photo collections [Unpublished PhD thesis]. Washington (DC): University of Washington.
Tran G, Nguyen D, Milenkovic M, Pfeifer N. 2015. Potential of full waveform airborne laser scanning data for urban area classification-Transfer of classification approaches between missions. Int Arch Photogramm. Remote Sensd Spatial Inf Sci. 40(7):1317.
Wagner R, Lieckfeldt P, Rothe R, Markram J. 2013. The Leica geosystems citymapper solution. In: Fritsch D, editor. Photogrammetric Week '13. Stuttgart: Wichmann; p. 89–99.
Wang Y, Steve S, Frank G. 2008. Pictometry's proprietary airborne digital imaging system and its application in 3D city modelling. Int Arch Photogramm. Remote Sens Spatial Inf Sci. 37(Part B1):1065–1070.
Westoby MJ, Brasington J, Glasser NF, Hambrey MJ, Reynolds JM. 2012. Structure-from-Motion' photogrammetry: a low-cost, effective tool for geoscience applications. Geomorphology. 179:300–314.
Wu B, Xie L, Hu H, Zhu Q, Yau E. 2018. Integration of aerial oblique imagery and terrestrial imagery for optimized 3D modeling in urban areas. ISPRS J Photogramm Remote Sens. 139:119–132.