

## Structure from Motion for aerial thermal imagery at city scale: Pre-processing, camera calibration, accuracy assessment

Paolo Conte, Valentina A. Girelli, Emanuele Mandanici\*

Dept. of Civil, Chemical, Environmental and Materials Engineering (DICAM), University of Bologna, viale del Risorgimento 2, 40136 Bologna, Italy



### ARTICLE INFO

#### Keywords:

Aerial thermography  
Geometric calibration  
Structure from Motion  
Orthomosaicking

#### 2018 MSC:

00-01  
99-00

### ABSTRACT

Airborne thermal cameras are a valuable source of information for energy analyses at city scale. The generation of accurate high-resolution thermal orthomosaics is a necessary but still challenging task, especially when a thermal camera is the only imaging sensor on-board, because of the peculiar characteristics of thermal imagery (i.e. low dynamic range and poor detail definition), large geometric distortions induced by the optical system and weak acquisition geometry. This paper discusses potentials and limitations of Structure from Motion approach for the automated generation of thermal orthomosaics, with the aim to define the best practices and assess the achievable accuracy. After processing with different strategies two thermal flights over a 10 km<sup>2</sup> area in Bologna city (Italy), it can be concluded that the absolute planimetric accuracy can be in the order of 3–4 pixels and the best results are obtained when computing camera calibration on a smaller subset of images, with a limited number of ground control points and an adaptive fitting algorithm. The analysis of generated point clouds (compared with reference LiDAR data) and calibration reports, in addition to check point residuals, proved to be crucial for a proper accuracy assessment.

### 1. Introduction

The use of thermal remote sensing for the study of different energy related aspects of urban environments has been growing in the last decades (Weng, 2009; Bechtel et al., 2016). Considering the relatively low amount of energy emitted by the Earth's surface in the long-wave thermal window (8–14 microns), combining a high radiometric resolution with a high spatial resolution, which are both essential for energy studies of cities, is still a technological challenge. Given the limited spatial resolution of the thermal sensors installed on-board the operating satellite platforms (the best one is ETM+ with a pixel size of 60 meters), aerial thermography has been representing the more suitable solution to image large areas (as large as a city) with a pixel size smaller than one meter. Considering the fact that the major space agencies have no immediate plans for new high-resolution satellite thermal sensors (Udelhoven et al., 2017), aerial surveys are unlikely to be outdated in the short term, even though the solutions offered by UAVs is experiencing an impressive growth (Colomina and Molina, 2014). The use of drone platforms, indeed, is still limited by battery autonomy and restrictive regulations concerning safety and security over built-up zones; therefore, drones are a suitable solution only for areas of limited extent.

Through aerial surveying, it is possible to capture a synoptic view of

an entire city and to derive useful indicators of energy efficiency of buildings, provided that an accurate calibration of the thermal images is accomplished (Bitelli et al., 2015). Calibration involves two main aspects: radiometry and geometry. The first refers mainly to the computation of reliable surface temperature values by compensating for atmospheric effects, surface emissivity and other factors which contribute to the radiance impinging on the detector (Mandanici et al., 2016). This paper, instead, is focused only on geometric aspects of aerial thermal images, in other words, the generation of thermal orthomosaics and the assessment of their metrical accuracy.

Although the attention for geometrical aspects of thermal imagery is raising (Maset et al., 2017; Iwaszczuk and Stilla, 2017), camera developers still show little interest in photogrammetric techniques, since their focus is mainly on radiometric accuracy (Luhmann et al., 2013). Accurate orthorectification, however, is an essential pre-processing step for energy analyses based on thermal images, because these images must always be integrated in GISs and overlaid with further ancillary data coming, for example, from technical cartographies. Generally speaking, orthorectification of aerial image blocks is a mature and well-established technique for digital cameras operating in the ordinary wavelengths of visible and near-infrared radiation. When coming to thermal cameras, however, some distinctive characteristics of thermal images may sometimes cause the failure of commonly used algorithms

\* Corresponding author.

E-mail address: [emanuele.mandanici@unibo.it](mailto:emanuele.mandanici@unibo.it) (E. Mandanici).

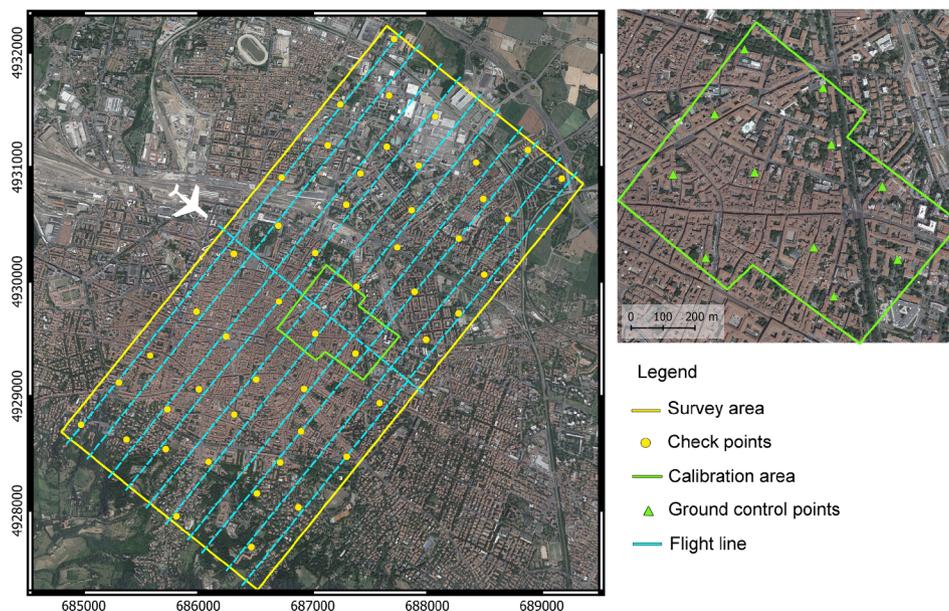


Fig. 1. Location of the study area and the smaller area used for calibration. Cartographic grid is UTM-ETRF2000 zone 32T.

or, more often, a significant drop in their accuracy (Khodaei et al., 2015; Maes et al., 2017). Thermal images, in fact, show lower dynamic range and geometric resolution when compared to the classic images in the visible wavelengths, together with poor definition of the discontinuities and small details (e.g. blurred edges), since digital number variations are related to surface temperature or emissivity differences.

In order to increase the accuracy of thermal orthomosaics, it becomes fundamental investigating the geometric image model of the thermal camera; also because in many practical cases (e.g. when thermal surveys are performed at nighttime) this camera is the only imaging sensor installed onboard and the whole mosaicking and orthorectification process must rely only on thermal images themselves.

Previous studies, conducted almost exclusively in close-range applications, highlighted that thermal cameras generally have worse metrological characteristics than visible ones (Lagüela et al., 2011) and radial distortion deriving from variations in refraction is the main error source. Luhmann et al. (2013) showed that, even though results vary significantly from camera to camera, all lenses show high distortion values, large shift of principal point and deviations from orthogonality of the image coordinate axes. In some cases, focal lengths computed within the block adjustment are significantly different from the values published in camera datasheets. Also in several works that analysed the generation of 3D models from thermal images acquired by UAV platforms (González Aguilera et al., 2013; Westfeld et al., 2015; Maset et al., 2017), the need of a pre-flight calibration is highlighted. In all these studies, in order to perform the thermal camera calibration in laboratory, new test fields were designed, the main challenge being the construction of special targets which can guarantee enough contrast in thermal images.

When working with airborne platforms flying hundreds of meters above the ground, however, the infinite focus makes it difficult to apply such laboratory calibrations. This task can be accomplished using the traditional field calibration photogrammetric approach or the newer Structure from Motion algorithms, but considering that a classical aerial survey does not present the optimal network geometry for interior parameters solution (James et al., 2017); in fact, some specifications of a typical calibration process are not met: convergent angles, high number of images “seeing” the same points, images with orthogonal roll angles (Remondino and Fraser, 2006). These factors force to find different solution strategies, identify an accurate camera model and carefully control the obtained results.

Structure from Motion (SfM) together with Multi View Stereo (MVS) algorithms is today the most popular photogrammetric approach to generate dense point clouds, 3D models and orthomosaics by images in many close-range and aerial applications (Gómez Gutiérrez et al., 2015). An in-depth understanding of principles and processes underlying the creation of 3D models with SfM/MVS algorithms is worthwhile for the proper assessment of model settings and the evaluation of error sources and their magnitude. In this case study, especially intended to produce accurate thermal orthomosaics, this approach was chosen for different factors: high automation in processing, speed in the elaboration of large image blocks, ability to recover the camera interior orientation (IO) parameters, robustness also in the treatment of “problematic” images, characterized by limited dynamic range, sensor noise and projective distortion (Lingua et al., 2009). In general, main factors affecting the effectiveness of SfM and quality of final products are acquisition distance and scheme, control measurements (exterior orientation parameters and/or ground control points) and image radiometry. In addition, the rolling shutter mechanism, which is implemented in many optical and thermal imaging sensors, can produce sensible geometric distortions when there is a relative movement between the camera and the observed objects (Vautherin et al., 2016). As pointed out by Peeters et al. (2017), the working principle of nearly all uncooled microbolometers is based on the rolling shutter principle.

In the light of these specific problems and considering the importance of obtaining reliable orthomosaics, this paper presents a set of experiments aiming at:

- the exploitation of SfM potentials for thermal camera calibration when other data are not available,
- the individuation of the best practices for the geometric correction of thermal images acquired from airborne platforms,
- the assessment of the metrical accuracy that can be achieved in realistic operational conditions.

## 2. Materials

Two aerial surveys were performed over an area of approximately 10 km<sup>2</sup> located in the city of Bologna (Italy). The area encompasses a portion of the city centre and some residential and productive districts, in order to include different kinds of urban texture (Fig. 1). In fact, the city centre is mainly composed of a dense pattern of historical buildings

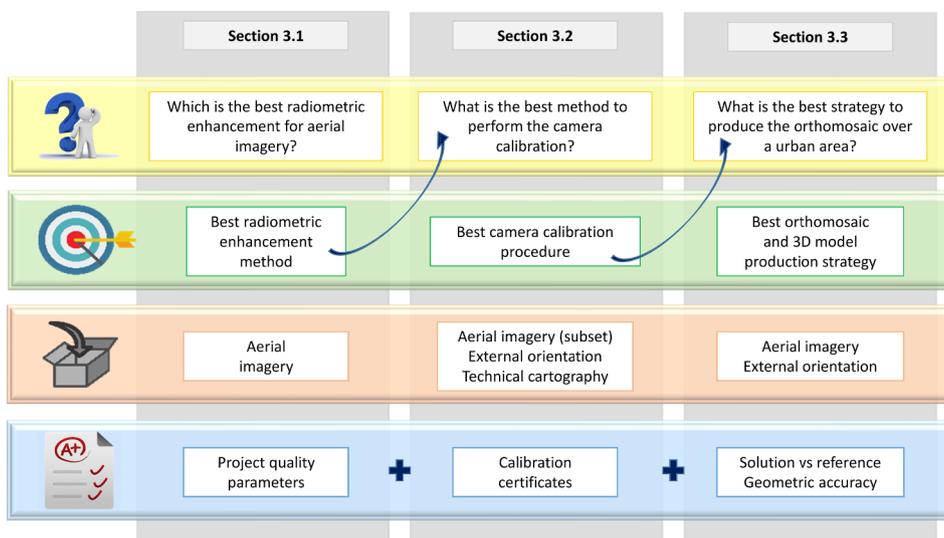


Fig. 2. Conceptual scheme of the performed tests.

with complex shapes and narrow streets, while in the peripheral districts a wider variety of building types can be found, together with a higher number of open spaces. The terrain has a flat topography, with an average ground elevation of 60 m a.s.l., except for the southern part of the study area which includes a hillside, with moderate slopes, higher vegetation coverage and far lower building density.

The flights were executed on 14 March 2016 and 8 March 2017 respectively (both approximately at midnight). A NEC TS9260 thermal camera was used, operating in the 7.5–13 μm infrared interval, with a resolution of 640 × 480 pixels and a noise equivalent temperature difference of 0.06 °C. Considering the narrow field of view of the optics (21.7 × 16.4°) and the desired ground sampling distance of 0.5 m, the flight height was set at about 850 metres above the ground and about 800 frames were acquired for each flight (average base 60 m). The aeroplane was also equipped with a high precision GPS receiver, an inertial measuring unit (IMU) and a video encoding/decoding device. Lever arms between GPS antenna, IMU and thermal camera were measured with a total station. GPS positioning was derived by a post-processing kinematic method, using as master a permanent GNSS antenna, called BOL1, located in the investigated area. The flight plan was designed to guarantee a minimum longitudinal and transversal overlap of 60% and 20% respectively. These overlaps are a compromise with the necessity to limit the duration of the flights (and thus the number of strips), in order to avoid excessive variations of surface temperature. For the same reason, only one transversal strip was acquired in the middle of the study area.

According to classical photogrammetry rules, the a priori root mean square errors (RMSE) of the object coordinates, X, Y and Z, can be predicted on the basis of the well-known formula, valid for the “normal case”, when image axes are parallel to each other and perpendicular to the base:  $\sigma_z = Z^2/f/B \cdot \sigma_\xi$ , where  $f$  is the focal length,  $Z$  is the acquisition distance,  $B$  is the camera base and  $\sigma_\xi$  is the precision in the horizontal parallax measurement. For these surveys all the quantities are known, with the exception of the accuracy in the measurement of the parallax, which in turn directly derives from the measurement of the image coordinates of two homologous points in a stereopair; using image matching techniques and SfM approach this measurement can reach sub-pixel accuracy, theoretically up to 0.1–0.2 pixels, for RGB images (Rosenholm, 2006; Remondino and El Hakim, 2006). However, with thermal imagery the precise collimation of a point in two adjacent frames is more difficult due to reduced contrast and blurring effects at the edges of objects (Weinmann et al., 2012), so it appears reasonable to use for  $\sigma_\xi$  a precautionary value of half a pixel. Using for  $Z$  and  $B$  the values set in the flight plan and the nominal camera parameters (focal

length 39.2 mm, pixel size 23.5 μm), the corresponding  $\sigma_z$  is equal to about 4 m.

The framework for the orthorectification of aerial images was provided by the technical cartography of the Municipality of Bologna, which comes with a nominal scale of 1:2,000 and a tolerance for planimetric coordinates of about 0.5 meters. For comparison and reference, a LiDAR survey performed on 26 January 2009 was used. The laser scanner was an Optech ALTM 3033 with a Field Of View of +/– 11 degrees. Given the operational altitude of 1250 m above the ground level and a distance between strip axis of 322 m, the horizontal accuracy is 0.6 m and the vertical one 0.2 m. The survey is framed in ETRF2000. The LiDAR cloud has a density of 1 point/m<sup>2</sup> in average, higher where two strips overlap (Africani et al., 2013).

Among all the software packages that use digital photogrammetry techniques enforced with SfM-MVS algorithms, Agisoft Photoscan was used. This software has nowadays become a standard in applications that require the generation of 3D spatial data because of some important characteristics: it is low cost, smart and so user friendly to be suitable also for non-experts, but on the other hand it offers to specialists the possibility to set and control the workflow, in order to adapt itself to the specific characteristics of each case study. This software supports the classic frame camera model, by implementing the well-known Brown’s non-linear distortion model (Brown, 1996). In camera calibration certificate the used parameters are, as usual, the focal length  $f$  (in pixels), principal point offset ( $c_x, c_y$ ) in pixels, four radial distortion coefficients ( $k_1, k_2, k_3, k_4$ ), four tangential distortion coefficients ( $p_1, p_2, p_3, p_4$ ) and the affinity and non-orthogonality (skew) coefficients ( $b_1$  and  $b_2$ ).

The open source and free software Cloud Compare was used to compute the distance between the point cloud provided by the LiDAR survey and the dense point cloud extracted by thermal images within Photoscan, in order to evaluate its accuracy.

### 3. Methods

This paper analyses mainly three aspects of the orthomosaic production: the pre-processing of the images, the geometric calibration of the infrared camera and the geometrical accuracy assessment. A conceptual scheme of the performed tests is reported in Fig. 2.

#### 3.1. Pre-processing

Thermal images are collected in a single spectral channel and store temperature data (or radiance) that are comprised in a very narrow

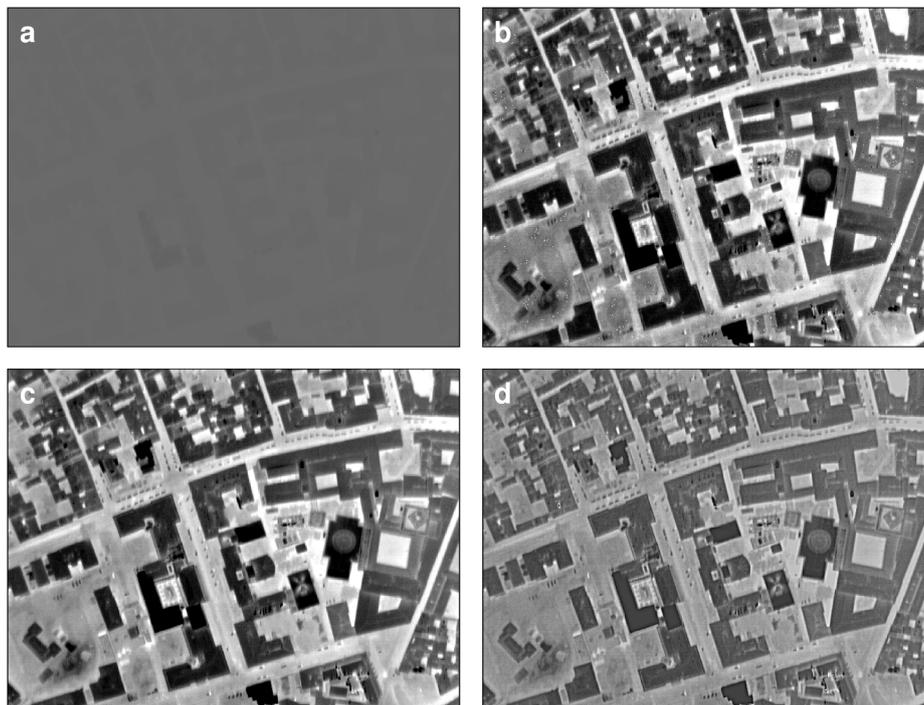


Fig. 3. Comparison of enhancement algorithms on a single frame. (a) unstretched 16-bit image; (b) 8-bit digital detail enhancement; (c) 16-bit linear stretching; (d) 16-bit Wallis filter.

numerical range, resulting in very poor contrast. Therefore, this section describes some pre-processing options that were tested in order to determine the best enhancement method, with particular emphasis on the bit depth of the output thermal images and the application of adaptive filters or other contrast stretching algorithms (Fig. 3).

One set of 44 thermal images from the NEC TS9260 camera, covering an area of about 0.5 km<sup>2</sup>, was firstly exported in ASCII format using the proprietary software provided with the camera, preserving the original temperature values for each pixel and avoiding the application of a standard false colour palette. In order to transform this dataset into a standard image format, a dedicated script was applied to produce 16-bit grey level TIFF images containing the apparent absolute temperatures in Kelvin multiplied by 100 and truncated.

After this preliminary phase, several datasets were derived using different transformation algorithms.

The first exploits the optimisation algorithm called Digital Detail Enhancement (DDE), developed by FLIR, which produces 8-bit images and can be considered a state-of-the-art algorithm for thermal imagery (YongJie et al., 2010).

The second, named “linear 2%”, is a linear stretching which preserves the 16-bit depth. It is derived by computing for the whole dataset an overall frequency histogram representing the number of occurrences of each temperature value among all the frames. Then, the tails of the calculated distribution containing 2% of the pixels were cut out, and all the images were exported applying the same linear contrast stretching between the defined extremes.

The third set is obtained by applying Wallis filter to the output of the previous “linear 2%” dataset. Wallis filter uses a local estimation of mean and standard deviation to calculate output images and has been suggested in several works about SfM as one of the best algorithms to emphasize details in images (Ribeiro Gomes et al., 2017). Different tests were carried out with Wallis filter varying the main parameters in order to define the optimal set. It is to be noticed that this kind of transformation alters the radiometric values of images as regard original values, thus making it necessary, for quantitative analysis, the replacement of the transformed dataset with the original one in the final orthorectification phase (this replacement is straightforward, once calculated

the external orientation, in most SfM software).

Finally, the last two datasets were obtained by converting to 8 bits the previous “linear 2%” and “Wallis” datasets, maintaining the relative distribution frequency, in order to compare the performance of SfM-MVS algorithms using images with different bit depths.

Each dataset was used to produce a 3D model, providing the SfM algorithm with neither external orientation parameters nor GCPs or manual tie points, and using as internal orientation parameters the nominal values obtained from the datasheet of the sensor. This way, only the capabilities to perform the relative orientation of the image block were tested, in terms of point number of both the sparse and dense clouds, reprojection error (RE) in pixels, marker residual maximum and RMSE. The full list of software settings is provided in the [supplementary materials \(SM1\)](#). Considering the nature of the thermal images, some parameters (e.g. tie point limit on every image) were forced in order to find more points for the alignment.

In this phase, the optimal number and type of internal calibration parameters for this kind of camera is unknown, thus two operational procedures supported by the software were compared: the first calculates simultaneously all the IO parameters, while in the latter (adaptive camera model) the software automatically selects the camera parameters to be included into the adjustment on the basis of their reliability estimates. For datasets with weak camera geometry, as for thermal aerial imagery, this procedure should help in avoiding divergence of some calibration parameters.

Before performing the optimization of camera parameters, all the sparse clouds were cleaned removing all the points that did not meet all the following criteria: RE lower than 0.5 pixels, reconstruction uncertainty lower than 50 and projection accuracy less than 10 pixels. These parameters express the accuracy of the automatic tie point positioning in the sparse cloud.

### 3.2. Geometric camera calibration

Once identified the pre-processing method that guarantees the best performances in relative orientation of images, a subset of the corresponding dataset was used for the determination of the best procedure



Fig. 4. 3D model computed for the smaller calibration area (CAL4 solution).

for camera optimization, which is the assessment of all the internal orientation parameters that define the optical system (focal length, principal point coordinates, radial and tangential distortion parameters, skew and affinity). The calibration area (Fig. 1) is located in the middle of the surveyed zone, where also the transversal strip is available, and has an extent of 0.5 km<sup>2</sup>. It is covered by 44 images from five strips (four parallel and one transversal) of the 2016 survey.

Several aspects were considered: the use of the exterior orientation measurements derived by GPS and IMU equipment, whether introducing ground control points (GCPs) and how many, the presence or absence of a transversal strip crossing the longitudinal ones, the modelling of the rolling shutter effect and, finally, the optimal configuration for the camera model in terms of number and type of parameters. Some of these aspects were investigated extensively for optical cameras, but poorly for thermal ones.

Several computations were run by creating different projects within the software and providing them with different sets of measurements before the alignment of thermal frames (one example is shown in Fig. 4). Some projects use only the external orientation (EO) parameters, others use GCPs only, or both GCPs and EO. The incorporation of external control measurements is generally helpful not only to define the project datum but also to mitigate or prevent systematic errors in 3D reconstruction (James et al., 2017). Even though good quality EO parameters are available, their adjustment is always necessary because of some inaccuracies in the synchronisation of camera on IMU clocks, which are difficult to eliminate completely (Stilla et al., 2009). For an easier reading, all the different combinations tested are reported in Table 1, where each record corresponds to a Photoscan project in which some of the factors described above were modified.

GCPs coordinates were derived from the building vector layer of the technical cartography (CTC) provided by the Municipality of Bologna and manually collimated on images. As already done in other studies (Mandanici et al., 2016), edges of building roofs were used as GCPs, to facilitate the identification of points in the thermal frames. With aerial thermal imagery, the identification of GCPs on images is quite challenging. Only features showing an adequate thermal contrast can be

Table 1  
Summary of the calibration projects.

Project	EO	GCP number	Transv. strip	Rolling shutter	Computed IO parameters
CAL1	x	None	x		$f, cx, cy, b1, b2, p1, p2, k1, k2$
CAL2		11	x		$f, cx, cy, b1, b2, p1, p2, k1, k2$
CAL3		11			$f, cx, cy, b1, b2, p1, p2, k1, k2$
CAL4	x	11	x		$f, cx, cy, b1, b2, p1, p2, k1, k2$
CAL5	x	11			$f, cx, cy, b1, b2, p1, p2, k1, k2$
CAL6	x	11	x		$f, cx, cy, p1, p2, k1, k2$
CAL7	x	11	x		$f, cx, cy, b1, b2, p1, p2, k1$
CAL8	x	8	x		$f, cx, cy, b1, b2, p1, p2, k1, k2$
CAL9	x	5	x		$f, cx, cy, b1, b2, p1, p2, k1, k2$
CAL10	x	11	x	x	$f, cx, cy, b1, b2, p1, p2, k1, k2$
CAL11	x	11	x	x	$f, cx, cy, p1, p2, k1, k2$

identified, and this happens mostly at the edge of building roofs. Still, the precise collimation remains difficult because edges often appear blurred, probably due to heat transfer effects at the edges and the influence of vertical elements on the detected infrared radiance.

It is to be noted that the mentioned CTC layer contains the orthometric heights of both the roof eaves and the base of buildings. To be correctly used in the project, where the external orientation data are derived from GNSS observation and thus referred to the WGS84 ellipsoid, the height values of points were transformed into the ellipsoidal datum by using the official geoid undulation model for the region.

The influence of GCP number on the quality parameters resulting from the software was assessed, i.e. the number of points were progressively reduced and the resulting project parameters were compared (CAL7–9 in Table 1).

Also the possible benefit of using the additional strip acquired with a flight direction transversal to the others was evaluated. This transversal strip was acquired primarily to perform a radiometric normalization between all the strips: images are acquired sequentially along the strips, thus there are sensible radiometric differences due to both thermal variations of the observed surfaces and the variable atmospheric influence during the aerial survey. In the camera calibration process, the benefit may derive from the local increase in the redundancy of the points and also the strengthening of the acquisition geometry. Indeed, the acquisition of images with a variable orientation of the camera is often suggested (Remondino and Fraser, 2006). In this work, two solutions were recomputed (CAL3 and CAL5) after removing the twelve images belonging to the transversal strip, while maintaining all the other settings of solutions CAL2 and CAL4 unchanged.

Another important aspect which is to be evaluated is the number and type of distortion parameters. With SfM techniques (and traditional photogrammetry as well), it is generally advisable to minimise the number of parameters to be calculated, in order to avoid an over-parameterisation of the model that can produce further errors; often this aspect is neglected when using SfM-MVS algorithms (James et al., 2017). Also the rolling shutter modelling adds further variables (related to the EO) in the bundle adjustment procedure and may therefore contribute to a possible over-parameterisation.

For the present work, different solutions were compared, progressively reducing the number of unknowns (Table 1). CAL4 used the most complete model, including radial and tangential distortions, affinity and non-orthogonality coefficients. CAL6, instead, did not consider affinity and non-orthogonality coefficients; CAL7 excluded also the computation of the k2 factor of the radial distortion.

Finally, two additional tests were carried out in order to evaluate the effectiveness of the rolling shutter correction. CAL10 and CAL11 have the same parameterisation of CAL4 and CAL6 respectively, but with the introduction of rolling shutter modelling during the optimisation phase. In rolling shutter cameras, image rows are captured at different time moments; consequently, when the camera is moving, a

different set of exterior orientation parameters should be computed for each row. Estimating EO separately for each image row is not feasible, as the number of unknowns would be very high and the problem would be highly underdetermined. To overcome this issue, the implemented algorithm assumes that during the capture of an entire image the camera is moving with constant velocity (both linear and angular). Under this simplification, the unknowns to be estimated for each image are the usual six EO parameters referred to the middle of the exposure, plus their constant rates of change during the acquisition (additional six parameters).

The IO parameters resulting from the best solution (according to the criteria described in Section 3.4) were used as a fixed pre-calibration for the thermal orthomosaics creation over the whole urban area of Bologna in the third phase of the processing (Section 3.3).

### 3.3. Thermal orthomosaicking of the urban area of Bologna

The final phase of the experimentation consists in the production of several thermal orthomosaics of the entire surveyed area in Bologna. In particular, it aims at comparing solution strategies based on a self-calibrating bundle adjustment where IO parameters are calculated simultaneously to EO parameters, with strategies based on a pre-calibrated camera where IO parameters are preliminarily fixed at the values obtained from the procedure described in the previous section.

Basically, the idea is to determine whether, and under which conditions, an internal calibration obtained with the processing of a small area and a few thermal images can be expanded to an entire aerial thermal dataset acquired with the same device but covering a whole city, and whether such pre-calibration is accurate enough to produce a thermal orthomosaic without using GCPs. Furthermore, the present necessity of implementing the rolling shutter correction in the SfM-MVS workflow was evaluated.

Overall, a total of ten thermal orthomosaics covering an area of about 10 km<sup>2</sup> were produced (Table 2). Solutions labelled with an -R suffix consider also the rolling shutter effect.

The solutions named STR1 and STR1R in Table 2 adopted the calibration certificates computed from the projects CAL4 and CAL11 respectively, in which both EO parameters and eleven GCPs were used as input data, the transversal strip was included and all the distortion parameters were considered (except *b1* and *b2* for STR1R). Two sets of images were processed using these fixed calibration certificates: one from the survey performed in 2016 (the same used for the computation of IO), the other from the survey of 2017, performed in the same period of the year and with similar characteristics in term of flight planning and atmospheric conditions. A rigorous approach would presuppose the execution of the camera calibration before each flight, but in the presented case study mainly oriented to production, it was preferred to test the methodology using a set of IO parameters calculated once for all the flights, also with the aim to save time and resources.

The solutions described above for the two surveys are then compared with the corresponding ones where only the nominal values of the main internal orientation parameters were provided to the software as approximated values for the self-calibration least-squares adjustment process (STR2 and STR2R in Table 2). This procedure is often used with

**Table 2**  
Summary of the solution strategies tested to produce the final orthomosaic of the entire study area.

Solution strategy	2016	2017	IO parameters
STR1	x	x	Fixed (from CAL4)
STR1R	x	x	Fixed (from CAL11)
STR2	x	x	Unknown
STR2R	x	x	Unknown
STR3		x	Fixed (from STR2-2016)
STR3R		x	Fixed (from STR2R-2016)

SfM software because, on the one hand, it is fast, does not require calibration certificates, is independent from other imagery sets and allows the calculation of all the unknown parameters at once. On the other hand, the number of unknown parameters is higher if compared with projects with a fixed camera calibration (especially for STR2R), thus the numerical solution may diverge during the iterative procedure for some parameters. Fortunately, Photoscan software provides an option, called “adaptive camera”, that automatically limits the number and type of parameters to be considered as unknown by selecting the most appropriate set.

For completeness, two further projects were processed for the 2017 survey (STR3 and STR3R). Here, a fixed calibration approach was adopted again, but this time the certificates were derived from the STR2 and STR2R solutions for the 2016 survey. These projects aim to simulate the calculation of the internal calibration parameters purely based on an image dataset, and its subsequent application to another set of images acquired in the same area, with the same camera and similar conditions. It is to be noted that the number of thermal frames covering the interested area slightly varies from year 2016 to 2017 (781 and 834 thermal images, respectively), due to the different spacing of the parallel strips acquired.

### 3.4. Accuracy assessment criteria

The overall evaluation of the proposed solutions is grounded on the following main aspects: quality parameters computed by the SfM software itself, similarity of intermediate 3D models with reference LiDAR data, planimetric accuracy of the final mosaics. A complete list is provided in Table 3, indicating also the experimentation phases to which they apply.

As for the evaluation of the intermediate 3D models of the SfM processing, some analyses were conducted on the dense point clouds calculated in Photoscan (after image orientation and camera optimization) and on the derived DSMs. In both cases, the reference data are derived from the LiDAR survey described in Section 2. For point cloud comparisons, original LiDAR data were filtered to remove second echoes and to obtain a homogeneous density over the study area. For the DSM comparisons, these data were transformed into a regular grid with 0.5 spacing (the same resolution of the thermal images).

The comparison of DSMs was performed through Dem of Differences (DoD) approach (Lague et al., 2013), in which frequency histograms and relevant statistical values are considered. In addition, six transects were defined on both the LiDAR and Photoscan generated DSMs, in correspondence of large areas with limited or no changes in elevation (flat roofs or plain ground), in order to evaluate the noise in the solutions. For each transect, the altimetric profiles derived from DSMs were compared in terms of minimum and maximum distances, average distance and standard deviation.

In order to directly compare the 3D models, the open-source

**Table 3**  
Summary of the accuracy evaluation criteria adopted in the present study and the phase of the processing where they were used.

Criterion	Pre-processing	Calibration	Final products
Sparse point number and density	x	x	x
Reprojection error (RE)	x	x	x
Dense cloud point number and density	x	x	x
Residuals on markers	x	x	x
Residuals on GCPs		x	
Calibration certificate analysis		x	x
Residuals on check points			x
C2C distance from LiDAR			x
M3C2 distance from LiDAR			x
DSM difference with LiDAR			x
DSM transect analysis			x

CloudCompare software was used, as it permits complex analysis on numerous three-dimensional data formats. For each project, the point cloud was exported in LAS format and a buffer of 30 m was applied in order to exclude border effects from the comparison. All the interested point clouds were neither manipulated nor co-registered to the LiDAR cloud; the latter process can be helpful to find temporal variations between different point clouds of the same object (Gómez Gutiérrez et al., 2015), but it is not recommended when testing the overall accuracy of reconstructed models as relative to a benchmark model. It is to be noted, here, that both LiDAR and thermal aerial surveys share the same reference frame (ETRF2000), derived from analogous GNSS kinematic positioning procedures.

Two approaches were used for these comparisons between point clouds: the Cloud-to-Cloud (C2C) distance and the Multiscale Model to Model Comparison (M3C2). A detailed explanation of these methods can be found in Lague et al. (2013). The main difference between the two methods is that M3C2 calculates positive or negative distances, permitting to analyse in which parts of the models the reference cloud is above or below the compared one. Furthermore, M3C2 is described as the most adequate descriptor of point cloud quality for scenes with complex topography (Gómez Gutiérrez et al., 2015).

For the present work, the point clouds calculated through SfM-MVS techniques were used as a reference because they store a number of points sensibly higher if compared to the LiDAR one and it is advisable using the denser cloud as a reference. Basic statistics on distances were used as indicators of the solution quality. Rasterized maps of M3C2 distance were also visually interpreted to identify patterns and characteristics of the errors.

Finally, to assess the accuracy of the final products, the thermal orthomosaics produced with the different approaches were analysed. A set of 50 Check Points (CP) uniformly spread over the test area was defined using the technical cartography as a reference (Fig. 1); the points were positioned on the corners of buildings easily identifiable on the thermal images and their planimetric coordinates were stored. The corresponding points were measured on all the thermal orthomosaics and the residuals from the reference coordinates were computed. The impact of subjectivity in the collimation process was estimated in 1 pixel, after the repetition of the measures on one mosaic by three independent operators. Again, the comparison between the different solutions is based on basic statistics on the CP residuals.

## 4. Results and discussion

In this section, the results of the experimentation are reported. For a better reading, the following sub-sections reflect the division operated in the Methods section.

### 4.1. Pre-processing

The application of the different radiometric enhancement algorithms, although not strictly related to the performances of SfM algorithms, is necessary for a proper elaboration of the project and the precise collimation of the GCPs on images.

Table 4 compares the results obtained from each set of enhanced images using two different strategies to fit the camera model parameters. The “default” one computes simultaneously all the considered parameters, whilst the “adaptive” one selects the parameters to be adjusted in each iteration to avoid divergence.

Several considerations can be made on these results. Firstly, it can be noted that in every test the use of the adaptive camera model is effectively helpful, because it produces an improvement on quality parameters in terms of both point cloud number and density, and above all reprojection error. The average point densities for the dense clouds obtained with these approaches is 50,334 and 50,995 points/image for the bundle adjustment and the adaptive model respectively, while the average values for sparse clouds are 461 and 474 points/image

**Table 4**

Basic statistics on the point clouds obtained from different sets of images which had been enhanced with the algorithms described in Section 3.1 (best results in bold). More details can be found in supplementary materials (SM2).

Enhancement	Fitting	Sparse cloud			Dense cloud	
		pts/m <sup>2</sup>	pts/image	RE (pixels)	pts/m <sup>2</sup>	pts/image
None	Default	0.035	395	0.53	4.51	51,253
	Adaptive	0.037	424	0.46	4.54	51,602
Linear	Default	0.033	370	0.53	4.39	49,929
	Adaptive	0.037	415	0.46	4.47	50,802
DDE	default	0.032	362	0.54	4.13	46,956
	Adaptive	0.040	452	<b>0.45</b>	4.44	50,442
Wallis	Default	0.038	434	0.51	4.54	51,549
	Adaptive	<b>0.041</b>	<b>463</b>	<b>0.45</b>	<b>4.57</b>	<b>51,912</b>

respectively (that become 411 and 454 after outlier removal).

Unexpectedly, the linear contrast stretching generates results slightly worse compared to the use of original imagery. More specifically, the number of elements in both point clouds are lower, while the RE is approximately the same. This fact is in agreement with the findings of Gómez Gutiérrez et al. (2015), who found that a higher dynamic range in the input images does not improve the solution sensibly. However, in the experiments presented here, this may be also due to the cut of the tails of the histogram, which reduces the total number of pixels to be processed.

The Digital Detail Enhancement (DDE) gives quite different results when working with or without the adaptive camera model: without it, the results of all the criteria are worse if compared to the ones obtained with both the original images and the linear stretching. Conversely, with the adaptive model they show better values for all the parameters considered except the size of the dense point cloud.

Anyway, the solutions that provide the best results are those based on the Wallis filter, with an increment in sparse cloud density of about 9%. Even though Table 4 shows only the optimal solution, several combinations of Wallis filter parameters were tested and results of the SfM procedure are every time better if compared to the other enhancements, in terms of sparse point number and density and RE. This improvement is less pronounced if compared with other works about SfM with images in the visible band (Ribeiro Gomes et al., 2017).

When using 8-bit images, results are every time worse than those obtained from their 16-bit counterparts; the difference is negligible for linear stretching alone, while becomes remarkable with Wallis filtered images (sparse and dense cloud sizes decrease by 25% and 2% respectively).

### 4.2. Camera calibration

In Table 5 the main quality parameters resulting from the Photoscan reports for all the calibration projects described in Section 3.2 (Table 1) are summarised.

First of all, considering the presence or absence of a transversal strip, it is evident how the number of points of the sparse cloud is strictly related to the number of images: without a transversal strip the sparse cloud is sensibly smaller (CAL3, CAL5). The MVS step does not seem to be affected by the absence of the transversal strip; in fact, the reconstructed dense point cloud is fully comparable in terms of number of points with projects involving the whole set of images, and the point/image density is even better than the other methods with a value of about 71,500 points/image instead of an average of about 52,000 points/image.

Outwardly, the approaches without a transversal strip seem to give better results not only considering point density per image, but also for

**Table 5**  
Summary of the results of the calibration projects (best results in bold). More details can be found in [supplementary materials \(SM4\)](#).

	Sparse cloud		Dense cloud	RMSE GCPs (m)			
	pts/m <sup>2</sup>	RE	pts/m <sup>2</sup>	East	North	Height	Total
CAL1	0.053	0.412	4.576				
CAL2	0.053	0.413	<b>4.584</b>	0.21	0.38	1.25	1.33
CAL3	0.041	<b>0.365</b>	4.583	0.22	0.39	1.18	1.26
CAL4	0.053	0.413	4.580	0.24	0.37	1.43	1.50
CAL5	0.041	0.366	4.582	0.23	0.40	1.36	1.44
CAL6	0.051	0.459	4.486	0.49	0.73	3.02	3.14
CAL7	0.053	0.413	4.575	0.24	0.37	1.46	1.52
CAL8	0.053	0.413	4.581	0.19	0.41	1.24	1.32
CAL9	0.053	0.413	4.571	0.23	0.23	1.08	1.13
CAL10	0.053	0.387	3.327	0.23	0.36	1.32	1.39
CAL11	0.053	0.401	4.513	<b>0.14</b>	<b>0.20</b>	<b>0.53</b>	<b>0.58</b>

what concerns the reprojection error, maximum residual on markers and residuals calculated for GCPs. It is likely that the presence of fewer images and thus fewer equations in the system induces an easier adaptation of the parameters to the model and consequently quality indicators appear better in numerical terms but are less representative of the actual quality of the model. This can be confirmed when analysing the difference between CAL5 and CAL3 projects, both performed without the transversal strips but with or without the approximated external orientation in input: no relevant differences are found between them except for GCP residuals. In fact, the project with further lower redundancy (CAL3, without EO) gives a lower RMSE on positions (especially on Z values) and a pixel RMS value slightly higher. This could indicate an under-parametrisation of the model, that is solved anyway by the software but without enough constraints to properly assess the quality of the project.

The same behaviour is observed when reducing the number of GCPs used for the solution of the adjustment model. As a matter of fact, if comparing CAL4, CAL8 and CAL9, varying the number of GCPs provided to the software, the quality parameters assume practically identical values for all the criteria, but GCPs residuals. In particular, both metric and pixel RMS residuals become lower if using fewer GCPs, giving the impression of an improvement in project quality; actually, the project with less redundancy should be worse while keeping all other variables unchanged. This effect is even more evident when passing from 8 to 5 GCPs.

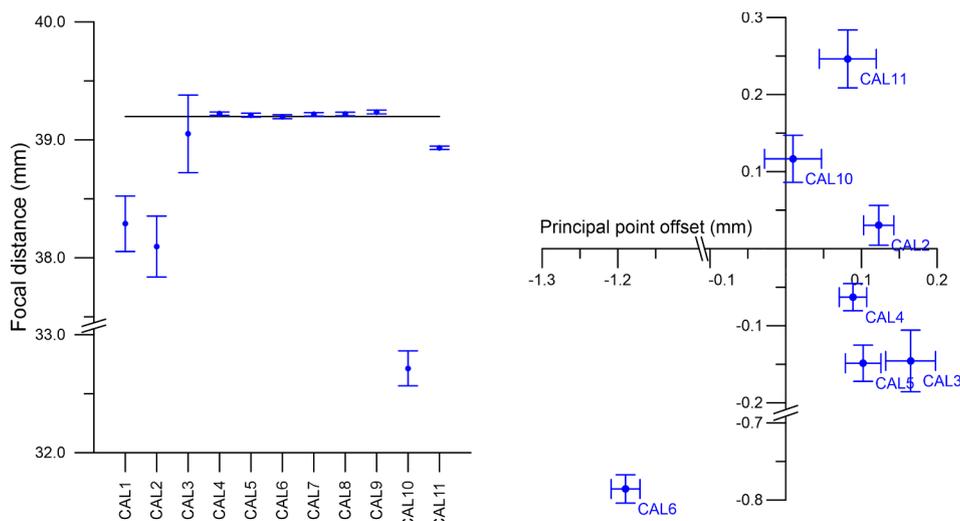
As for the analyses of the results obtained with different control

measurements, it is interesting to note that no significant difference was found between calibration project with only GCPs (CAL2), only external orientation (CAL1) or both input datasets (CAL4); the only noticeable difference is in the metric residual of GCPs, that also in this case results higher for the system with higher redundancy.

Finally, when comparing different camera models, it is evident how the hypothesis of square pixels with orthogonal axis (CAL6) is not adequate with this kind of thermal infrared cameras, as confirmed from previous studies (Luhmann et al., 2013), at least when rolling shutter is not considered. In fact, even if this approach gives a sparse point cloud slightly denser if compared to other camera models, all the other quality parameters are significantly worse, in particular the RE, the residuals on GCPs and the dense cloud size. Conversely, when considering the rolling shutter effect in the process, the best results are obtained when skew and affinity factors are neglected. In fact, CAL10 shows results that are fully comparable to the other approaches except for the number of points of the dense cloud (which are sensibly fewer). CAL11, where  $b_1$  and  $b_2$  factors are set equal to 0 prior to the optimization, shows instead very good results especially as for the GCP RMSE that is the lowest among all the approaches tested (about 1 pixel). The difference between these results might be due to the shape of the distortions induced by the rolling shutter in the case of translational camera motion (Hong et al., 2010), which are partially analogue to the ones produced by the skew and affinity parameters.

The project that neglects the second term of radial distortion  $k_2$  (CAL7) shows less evident differences: the number of points of sparse clouds is practically the same (compared to CAL4), as well as marker residuals in pixel units. Small differences can be found on the number of points in the dense cloud and on the metric residual on GCPs after the removal of points with high residuals and the optimization of the parameters. For the determination of the most suitable method, a deeper analysis of values, associated errors and correlations between parameters for each calibration certificate was performed. From this analysis, the best camera model results the one with the presence of  $k_2$ , not only because it provides overall quality parameters slightly better than the others, but also because the correlation value of  $-0.93$  between  $k_1$  and  $k_2$  does not justify its exclusion from the model (as done in CAL7). Furthermore, all the parameters always pass the significance test based on the t-Student distribution assuming a 95% confidence interval (Gruen, 1978).

When rolling shutter is not considered, further aspects can be highlighted from the comparison of the calibration certificates: for example, they confirm the unsuitability of camera models with  $b_1$  and  $b_2$



**Fig. 5.** On the left, focal lengths of the different calibrations listed in Table 1 with error bars (horizontal black line represents the nominal value); on the right, offsets of the principal point (CAL1 and CAL7–9 are very close to CAL4 and are omitted for clarity).

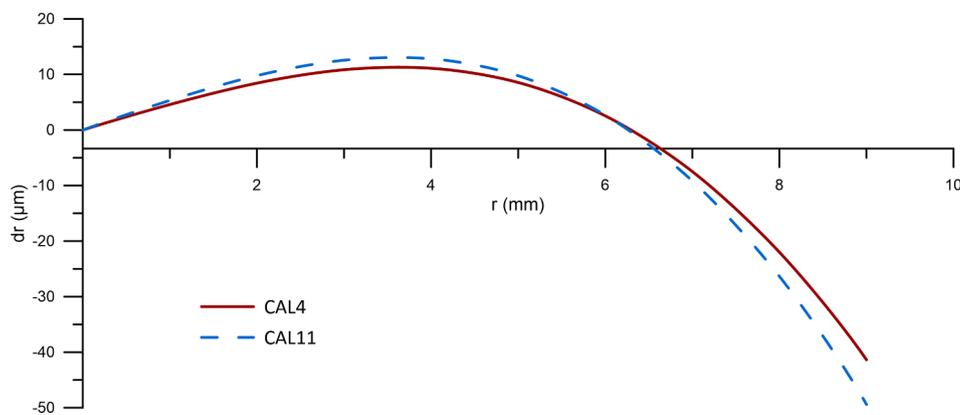


Fig. 6. Radial distortion curves for NEC TS9260 derived from CAL4 and CAL11 solutions.

equal to 0; in fact, when neglecting the computation of these parameters, a strong mispositioning of the principal point is introduced (Fig. 5). In addition, solutions of projects with poor input datasets (CAL1-2-3) are characterized by a focal distance considerably different from the others, and with an associated error of a higher order of magnitude.

The same behaviour is partially replicated for the CAL10 project, which shows a focal distance greatly different from the nominal value and the errors of focal distance, principal point coordinates and skew affinity factors sensibly higher if compared to the other calibration projects. Furthermore,  $b_1$  and  $b_2$  assume exaggerated and unrealistic values; this confirms the instability of the algorithm when solving the rolling shutter model together with skew and affinity factors, at least in this case study. Differently, when correcting the rolling shutter effect but setting  $b_1$  and  $b_2$  to 0 (CAL11), the divergence of the focal distance is much lower, as well as its associated error. Principal point and distortion parameters errors instead remain substantially stable.

In the light of all these considerations, CAL4 and CAL11 appear the most reliable solutions and were thus chosen for the following processing phases. Fig. 6 shows the obtained radial distortion curves.

#### 4.3. Orthomosaic generation

Several criteria were followed with the aim to determine the more suitable approach for the production of thermal orthomosaics over large urban areas. A summary of the most relevant results is reported in Tables 6 and 7 for all the solution strategies described in Section 3.3. More complete statistics can be found in the supplementary materials

Table 6

Result summary of the solution strategies for the whole study area (Section 3.3): sparse clouds, dense clouds, check points (best results in bold). More details can be found in supplementary materials (SM5).

		Sparse cloud			Dense cloud		CP residuals (m)	
		pts/m <sup>2</sup>	pts/image	RE (pixels)	pts/m <sup>2</sup>	pts/image	mean	STD
STR1	2016	0.046	589	0.382	4.819	61,706	1.56	0.89
	2017	<b>0.061</b>	<b>733</b>	0.332	5.195	62,290	1.96	0.92
STR1R	2016	0.046	584	0.363	4.637	59,377	10.63	5.72
	2017	<b>0.061</b>	729	<b>0.325</b>	5.030	60,318	10.84	5.76
STR2	2016	0.046	589	0.381	4.821	61,730	<b>1.29</b>	<b>0.62</b>
	2017	<b>0.061</b>	<b>733</b>	0.332	5.194	62,277	1.83	0.76
STR2R	2016	0.046	589	0.365	4.645	59,472	11.47	5.57
	2017	<b>0.061</b>	<b>733</b>	<b>0.325</b>	5.016	60,145	11.84	6.11
STR3	2017	<b>0.061</b>	<b>733</b>	0.332	<b>5.196</b>	<b>62,306</b>	1.89	0.84
STR3R	2017	<b>0.061</b>	730	0.326	5.028	60,290	11.54	5.94

Table 7

Comparisons between intermediate 3D products of the adopted solution strategies and the reference LiDAR data (best results in bold).

		DoD (m)		C2C (m)		M3C2 (m)		
		mean	STD	mean	STD	mean	abs. mean	STD
STR1	2016	1.18	5.38	2.04	<b>2.20</b>	0.63	3.42	5.25
	2017	<b>0.28</b>	5.13	<b>2.03</b>	2.34	<b>0.21</b>	<b>3.17</b>	5.12
STR2	2016	10.82	5.35	6.33	3.60	9.96	10.30	5.37
	2017	39.07	5.17	29.66	7.69	36.58	36.78	<b>4.65</b>
STR3	2017	9.89	<b>5.09</b>	6.31	3.43	9.12	9.47	5.18

(SM5).

First of all, it can be noted that the number of images influences the size and density of the point clouds. An increment of about 7% in the number of images (834 for the 2017 survey vs 781 for the 2016 one) induces a large increment in the density of homologue points found by the SfM algorithm (33%), while the number of points reconstructed in the dense clouds by MVS increases in average of about 1% only. This implies that, when using fewer frames to cover the same area, the reconstructed point clouds can have almost the same size but they are calculated with a sensibly lower redundancy, and consequently a lower accuracy. This is confirmed by the reprojection error, which has an average value of 0.38 for 2016 data and a practically constant value of 0.33 in 2017, with a 15% increase.

When comparing the different strategies of orthomosaic production,

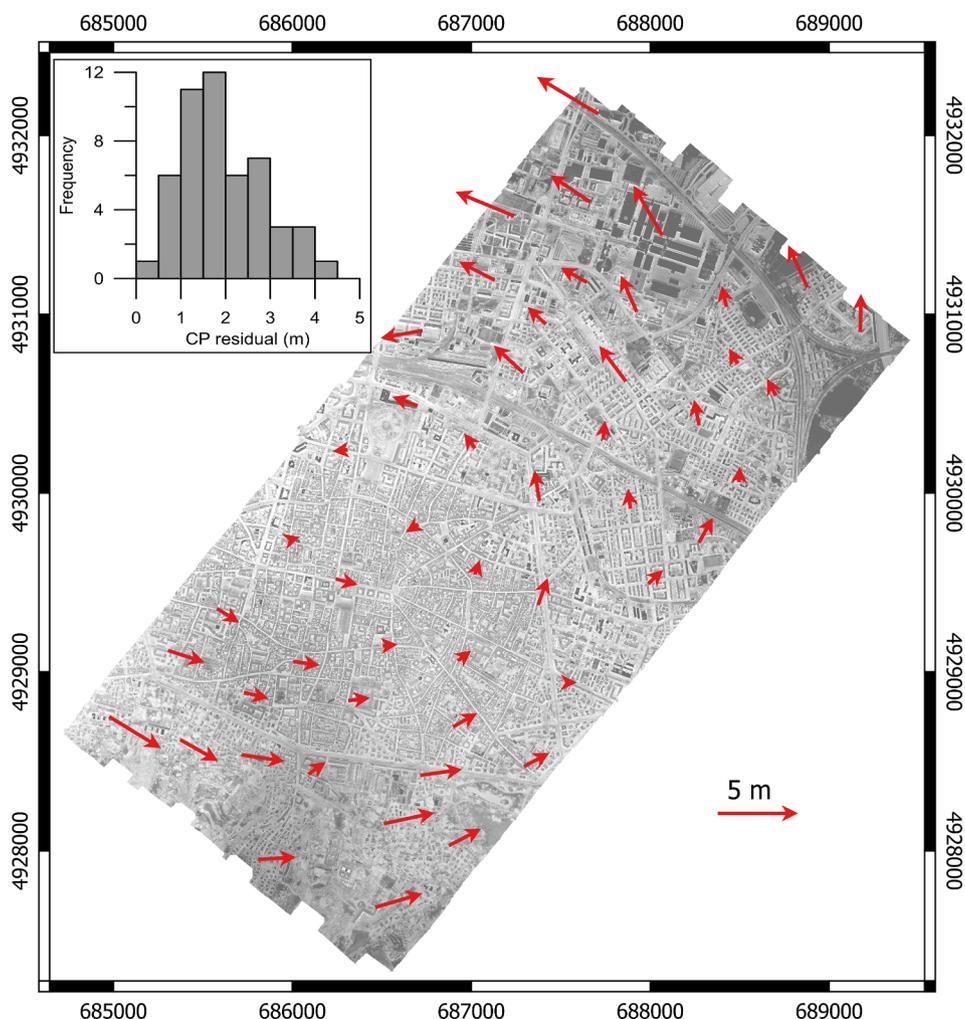


Fig. 7. Residuals on the check points for the STR1-2017 solution.

no sensible differences can be noticed in terms of quality parameters except for slight variations in the point number of clouds. Solutions with rolling shutter modelling show a slightly better RE.

The most important quality criteria for the final thermal orthomosaics is the planimetric accuracy, which was assessed on 50 check points. A first observation can be done on coherency of 2016 and 2017 mosaics: on average, positions differ by about 1.5 m (3 pixels) and 3 m (6 pixels) without and with rolling shutter correction respectively. Instead, average residuals with reference coordinates from cartography are reported in Table 6. Furthermore, single CP residuals for the STR1-2017 solution strategy are plotted in Fig. 7 as an example.

Recent works about SfM with optical imagery acquired in a nadiral geometry reached similar or slightly better results. For example, James et al. (2017) (dataset “Taroudant”) found an average residual on CP coordinates of about two pixels, using a fixed-wing drone flying over a bare soil area of about 0.07 km<sup>2</sup> and acquiring images with a Ground Sample Distance (GSD) of 2.3 cm, ensuring large overlaps. Peppas et al. (2016), instead, performed six flights during two years over the same landslide area (0.06 km<sup>2</sup>) and found an RMSE on CPs varying between 1.2 and 3.8 cm in planimetry, having hundreds of images with GSDs in the range of 2.8 and 3.8 cm and lateral overlaps varying from 40% to 70%.

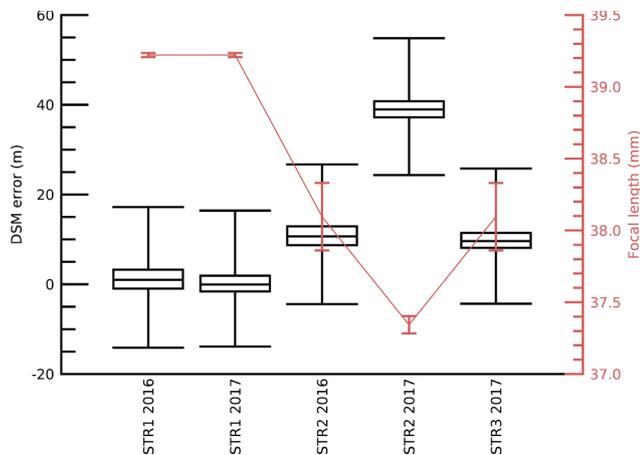
Looking at Table 6, both solutions for the 2016 survey show lower values of residuals for the set of check points used; this replays in part the behaviour observed in Section 4.2, where projects with lower redundancy showed in general lower GCP residuals.

All the solutions including rolling shutter optimisation,

unexpectedly, show far worse planimetric accuracy if compared to the others. CP RMSE, in fact, results greater by an order of magnitude, and also standard deviations and maximum residuals are dramatically larger. Such high values demonstrate once more how, in this case study, the algorithm suffers when dealing with these additional parameters. Since the main goal of the present work is the generation of thermal orthomosaics, the planimetric accuracy achieved including the rolling shutter effect in SfM workflow cannot be considered acceptable. Therefore, these solutions (STR1R, STR2R and STR3R) were discarded and excluded from subsequent analyses.

Among the remaining strategies, STR2 (the processing without a fixed camera calibration) seems to perform better if compared to the use of a calibration certificate in terms of CP residuals. However, the comparison of the DSMs derived from the dense point clouds with the reference DSM derived from LiDAR totally overturns this partial outcome (Table 7). For both years, the strategy that shows better results is STR1 (the one using a fixed internal orientation). In fact, DoD averages are sensibly lower (1.2 vs 10.8 m in 2016, 0.3 vs 39.1 m in 2017), meaning that STR2 solutions are likely to come with a strongly biased estimation of some orientation parameters. The correlation between DSM errors and focal length estimates is investigated in Fig. 8, where it is apparent that solutions with large height difference from the reference are those providing a focal length highly divergent from the nominal value.

Also the third strategy (STR3), where the calibration of STR2 for 2016 dataset is used as a calibration certificate, shows results very similar to the solution that originates this certificate (9.9 vs 10.8 m in



**Fig. 8.** Comparison of height differences between the DSMs generated by SfM solutions and LiDAR reference. Larger differences are related with more divergent focal lengths.

DSM error). Thus, only strategies that use the best fixed internal orientation calculated in Section 3.2 produce DSMs with elevation values comparable to the LiDAR ones, while all the other strategies produce systematic errors that imply a relevant vertical translation of DSMs, even though the produced orthomosaics show slightly better CP residuals. Standard deviations are all in the order of 5 m, suggesting that the main discrepancy among DSMs is a general shift along z-axis.

Direct comparisons between 3D models generally confirm the outcomes of DSM analyses. Looking at C2C distances, STR1 provides the best results, while STR2 (2016) and STR3 (2017) are almost equivalent. Also for this criterion, differences seem to depend mainly on differences between the focal distances; the more the focal distance diverges from the original value, the higher C2C distances become. Furthermore, it can be noted that STR1 produces practically identical results for both the years considered. Compared to DSM analyses, some discrepancies can be appreciated, probably due to both the method of computation (DSMs distances are referred to the elevation values of each pixel, while C2C are calculated in a spherical neighbourhood of reference points) and the further elaboration of data when passing from a 3D point cloud to a 2.5D image.

M3C2 distances are closer to DSM differences than C2C distances. This is probably a consequence of the choice of calculating M3C2 distances along Z direction only (Lague et al., 2013), and it seems to confirm the findings by Gómez Gutiérrez et al. (2015) who suggested that M3C2 method is more suitable than C2C to compare 3D models.

C2C distances are always lower than the corresponding absolute values of M3C2 ones; probably, when the topography is as complex as in urban environments, the calculation of distances in each direction instead of along Z-axis only causes the underestimation.

An example of M3C2 distance, referred to the project with overall best results (STR1-2017), is reported in Fig. 9. As already noticed (Weinmann et al., 2012), more prominent areas are better resolved by the SfM-MVS software, while the elevation of more hidden areas, such as urban canyons in the city center, are often overestimated. In addition, vegetated areas, with trees or other green elements with a relevant height, are always flattened during the processing chain. Some authors highlighted the dependence of height accuracy on the land cover, also for images in the visible band (Ressl et al., 2016). However, such a pronounced effect is characteristic of thermal images, because the reduced temperature contrast in vegetated areas makes it impossible to the software to identify corresponding features between overlapping images.

The analyses of the six transects extracted from DTMs further confirm the evaluation of these intermediate products. Again, STR1 produces the best results followed by STR3-2017, STR2-2016 and finally

from STR2-2017. Average differences on transects closely reflect the difference between DSMs, but on roofs the standard deviations are always higher than the ones calculated on ground transects. In addition, while on roofs their values are comparable for each approach considered, at the ground level stronger differences appear, without a clear behaviour. As an example, two height profiles of plain surfaces (one on the ground, the other on a roof) are showed in Fig. 10. All the solution strategies produce very noisy altimetric profiles (oscillations of few meters over plain surfaces), and only STR1 provides an acceptable overall distance from the LiDAR reference.

The accuracy of the STR1 3D model compares well with the outcomes of many empirical studies based on nadir optical imagery, even if a direct comparison is not immediate, because of the differences in terms of acquisition platforms, image constellation, GSD and number of GCPs. For example, Haala (2013) (dataset “Vaihingen/Enz”) surveyed an area of about 22 km<sup>2</sup> with different land covers, collecting 36 images with an overlap higher than 60% (both along and across the flight line) and a GSD of 0.2 m, and achieved a vertical accuracy of the final 3D model of 0.2 m. Similarly, Ressler et al. (2016) (dataset “Elbe/Klößen”) produced a model of a 1 km<sup>2</sup> area with an accuracy of 9 cm, using 98 images with an overlap of 80% and 70% along and across the flight line respectively, and a GSD of 6 cm. In general, it has been concluded that height accuracies in the order of one image GSD or slightly lower can be achieved (Remondino et al., 2014; Zhang et al., 2017).

Returning to the discussion of the obtained solutions, the comparison between calibration certificates produced with the different strategies (see supplementary materials SM3) highlights a large variability of almost all the parameters except for the values of radial and tangential distortions. The use of a fixed camera calibration avoids a marked divergence from the nominal value of the focal distance, which proved to be the most influential IO parameter. As already pointed out, this divergence leads to an erroneous reconstruction of the 3D topography of the scene.

Furthermore, the calculation of the focal distance on a small calibration area with the use of GCPs (on which STR1 is based) seems to be more reliable than its computation carried out by using far more images but without GCPs (which is the rationale behind STR2). Indeed, the error associated to focal distance is about five times lower with the first approach. On the contrary, the errors associated to the other parameters are always lower when performing the calibration on a larger area with more images rather than on a smaller area but with the help of GCPs. This suggests that the use of GCPs can improve the camera calibration mainly for what concern the focal distance, while the computing of the other IO parameters primarily benefits from the use of a larger dataset.

The STR3 approach shows intermediate results between STR1 and STR2: this approach surely benefits of the availability of more thermal imagery in the image matching (as evident if comparing reconstruction errors), but the residual RMSE for CPs is higher if compared to the use of the same calibration certificate in the year 2016.

The described procedures relate only to the geometrical aspects of thermal image correction. In order to derive accurate surface temperature estimates, also radiometric calibration is to be performed. As stated in the introduction, a discussion about the methods for radiometric calibration is beyond the scopes of the present paper. However, it is worthwhile to mention briefly that the described procedure can be used to produce mosaics of apparent temperature. The contrast stretching and filtering applied in the processing phase alter temperature values but they are necessary only for the alignment and the point cloud generation. After these phases, the original unfiltered images can be restored for the production of the final orthomosaics (Ribeiro Gomes et al., 2017). Then, apparent temperature mosaics can be corrected for the effects of atmosphere, emissivity and sky-view factor with several methods available in the literature (Meier et al., 2011; Mandanici et al., 2016), provided that no automatic colour balancing is performed during the mosaicking process. When necessary, a balancing among the

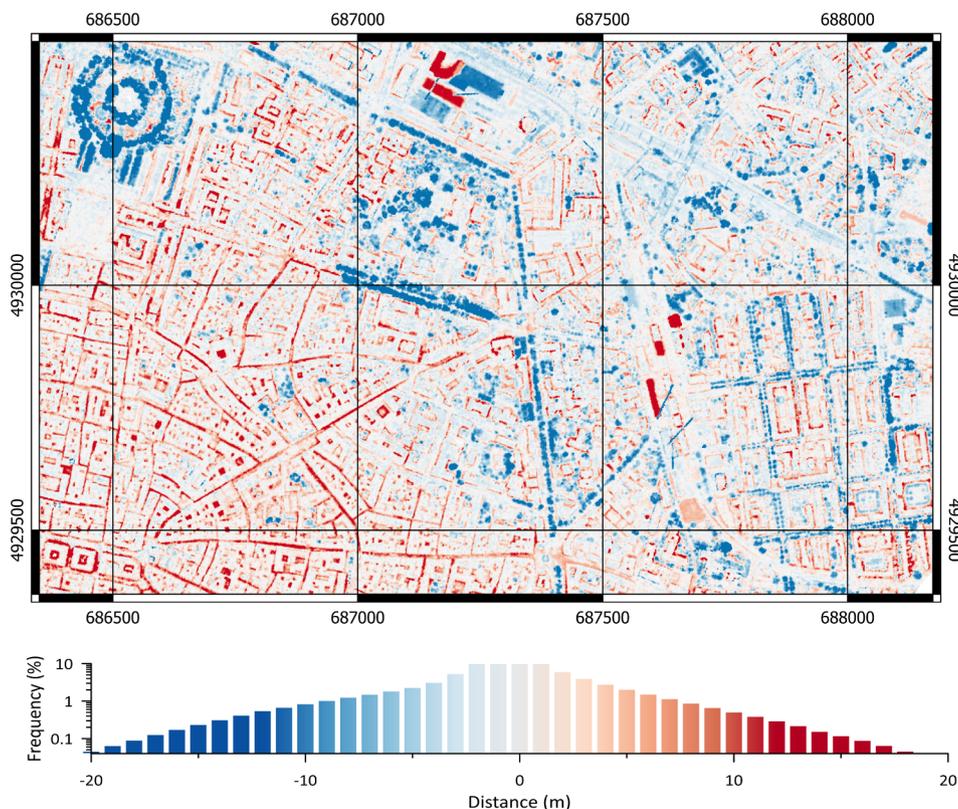


Fig. 9. Map of the M3C2 distance between the STR1-2017 point cloud and reference LiDAR cloud (top) and histogram of the values (bottom).

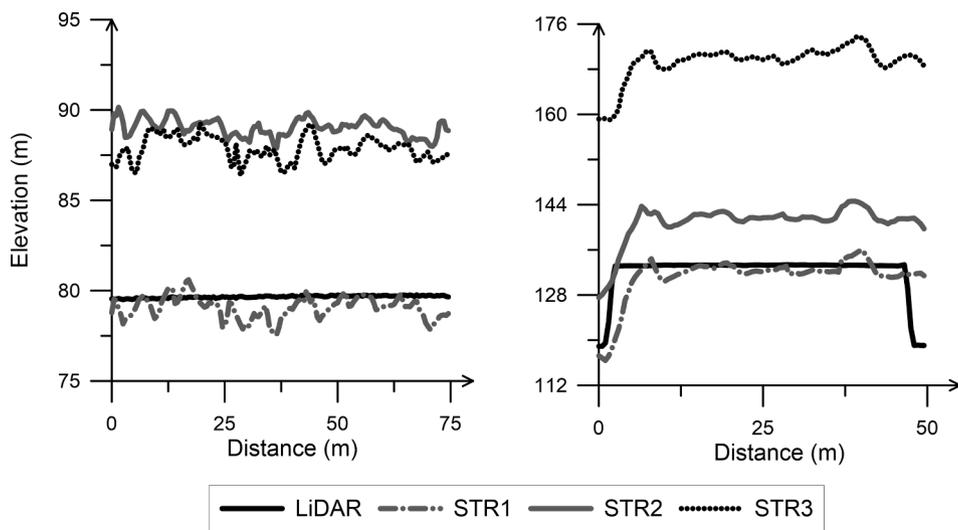


Fig. 10. Elevation profiles extracted from the DSMs generated by the three solution strategies for the 2017 survey.

different strips can be accomplished before the 3D modelling, by exploiting the transversal strip.

As a final remark, some authors (Lin et al., 2017; Meißner et al., 2017) hypothesized an influence of radiometric alterations – such as non-uniformity, vignetting and thermal drift (even if they are partially handled by the internal firmware of many devices) – over the geometric correction process, especially the feature matching algorithm used by SfM; further experiments are required to investigate and quantify these impacts.

### 5. Conclusions

The generation of accurate orthomosaics, which are necessary for

further quantitative analyses targeted to energy applications in urban environments, is particularly challenging with high-resolution thermal images, because of their limited dynamic range (which means a low contrast), narrow field of view (which means a huge number of frames over an entire city to preserve sub-metric spatial resolution), and the weak geometry that often characterize traditional aerial surveys. This paper discussed some solutions for the peculiar problems of thermal imagery, by exploiting two sets of nadir thermal images acquired from an airborne platform during two distinct surveying campaigns over Bologna city, the first in March 2016 and the second one year later.

Given the lack of reliable geometric calibration certificates for most thermal cameras, the Structure from Motion approach proved to be effective in reconstructing the 3D model and producing thermal

orthomosaics with a planimetric accuracy lower than four pixels (2 m with a 0.5 m GSD), which can be considered adequate for energy analyses at building block level, although it is still worse than the accuracy achievable with images in the visible band (up to one pixel or slightly lower).

Even though the rolling shutter correction is often recommended when using thermal uncooled microbolometers, in the presented experiments it proved to worsen the quality of the final products, probably because it introduced a large number of additional unknowns and thus an over-parameterisation of the model.

Enhancements of the raw thermal images improved the effectiveness of the process with the adopted SfM algorithm; in particular, adaptive filtering after a contrast stretching increases the number of tie points automatically matched. Most importantly, the use of a fixed camera calibration, previously determined on a small subset including a transversal strip, appears to be the best strategy toward an automated processing, especially when using parallel and flat acquisition geometry. On the one hand, it reduces dramatically the number of required ground control points (which would be very high for hundreds or thousands of frames); on the other hand, the planimetric accuracy results slightly lower (0.5 pixels only) if compared to a solution in which both interior and exterior orientation parameters are calculated at once. In this last case, however, focal length is likely to assume unrealistic values and, consequently, the altimetric accuracy of the 3D model to be compromised.

In this perspective, the performed tests highlighted the importance of considering many different indicators when evaluating the quality of the final products. In particular, the analyses of the computed calibration certificates and of the 3D point clouds is advisable (even though they are only intermediate products), because the use of planimetric check points only does not allow a full control over the alignments.

A further aspect to be pointed out is that larger image datasets produce not only lower reprojection errors and denser point clouds, but also more accurate interior orientation parameters and reduced roughness of the reconstructed surface. For this reason, increasing the overlap between images might be useful, even though this would increase the duration of the acquisitions, which is a problem for thermal surveys because of the temperature dynamics of observed surfaces.

As a final remark, this study confirms that thermal cameras are often affected by large distortions, also because the interest of producers and customers is more focussed on radiometric aspects than on geometric ones. However, when including thermal data in geospatial analyses, the importance of appropriate geometric corrections becomes evident and an accurate camera calibration is required to infer reliable information.

## Acknowledgements

This work is part of the ChoT project, funded by the Italian Ministry of Education, University and Research in the framework of the Scientific Independence of young Researchers (SIR) 2014 programme (Grant No. RBSI14SYES). This project aims at analysing the potential of aerial thermal imagery to produce large scale datasets for energetic efficiency analyses in urban environments. The authors would like to thank Franco Coren, Rita Blanos and Alessandro Pavan of the National Institute of Oceanography and Experimental Geophysics (OGS) for the execution of the thermal flights, and Andrea Minghetti, Paola Africani and Elisa Paselli of the “SIT” division of the Municipality of Bologna, for providing cartography and LiDAR data. The Authors are also grateful to Prof. Luca Vittuari (DICAM - University of Bologna) for his precious advices.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.isprsjprs.2018.10.002>.

## References

- Africani, P., Bitelli, G., Lambertini, A., Minghetti, A., Paselli, E., 2013. Integration of lidar data into a municipal GIS to study solar radiation. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci. XI-1/W1 1–6*. <https://doi.org/10.5194/isprsarchives-XI-1-W1-1-2013>.
- Bechtel, B., Keramitsoglou, I., Kotthaus, S., Voogt, J.A., Zakšek, K., 2016. Special issue “The application of thermal urban remote sensing to understand and monitor urban climates”. *Remote Sensing URL: <http://www.mdpi.com/journal/remotesensing/special\_issues/tirurbcli#published>*.
- Bitelli, G., Conte, P., Csoknyai, T., Franci, F., Girelli, V.A., Mandanici, E., 2015. Aerial thermography for energetic modelling of cities. *Remote Sensing* 7, 2152–2170. <https://doi.org/10.3390/rs70202152>.
- Brown, D.C., 1996. Decentering distortion of lenses. *Decent. Distort. Lenses* 32, 444–462.
- Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* 92, 79–97. <https://doi.org/10.1016/j.isprsjprs.2014.02.013>.
- Gómez Gutiérrez, A., de San José Blasco, J., Lozano Parra, J., Berenguer Sempere, F., de Matías Bejarano, J., 2015. Does HDR pre-processing improve the accuracy of 3D models obtained by means of two conventional SfM-MVS software packages? the case of the Corral del Veleta rock glacier. *Remote Sensing* 7, 10269–10294. <https://doi.org/10.3390/rs70810269>.
- González Aguilera, D., Lagüela, S., Rodríguez González, P., Hernández López, D., 2013. Image-based thermographic modeling for assessing energy efficiency of buildings façades. *Energy Build.* 65, 29–36. <https://doi.org/10.1016/j.enbuild.2013.05.040>.
- Gruen, A., 1978. Accuracy, reliability and statistics in close-range photogrammetry. In: *Inter-Congress Symposium of ISP Commission V*.
- Haala, N., 2013. The landscape of dense image matching algorithms. In: *Fritsch, D. (Ed.), Photogrammetric Week 13, Wichmann, Berlin/Offenbach, Germany*. pp. 271–284.
- Hong, W., Wei, D., Batur, A.U., 2010. Video stabilization and rolling shutter distortion reduction. In: *2010 IEEE International Conference on Image Processing, IEE*, pp. 3501–3504. <https://doi.org/10.1109/icip.2010.564959>.
- Iwaszczuk, D., Stilla, U., 2017. Camera pose refinement by matching uncertain 3D building models with thermal infrared image sequences for high quality texture extraction. *ISPRS J. Photogramm. Remote Sens.* 132, 33–47. <https://doi.org/10.1016/j.isprsjprs.2017.08.006>.
- James, M., Robson, S., d’Oleire Oltmanns, S., Niethammer, U., 2017. Optimising UAV topographic surveys processed with structure-from-motion: ground control quality, quantity and bundle adjustment. *Geomorphology* 280, 51–66. <https://doi.org/10.1016/j.geomorph.2016.11.021>.
- Khodaei, B., Samadzadegan, F., Dadras Javan, F., Hasani, H., 2015. 3D surface generation from aerial thermal imagery. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci. XI-1-W5 401–405*. <https://doi.org/10.5194/isprsarchives-xi-1-w5-401-2015>.
- Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: application to the Rangitikei canyon (N-Z). *ISPRS J. Photogramm. Remote Sens.* 82, 10–26. <https://doi.org/10.1016/j.isprsjprs.2013.04.009>.
- Lagüela, S., González Jorge, H., Armesto, J., Arias, P., 2011. Calibration and verification of thermographic cameras for geometric measurements. *Infrared Phys. Technol.* 54, 92–99. <https://doi.org/10.1016/j.infrared.2011.01.002>.
- Lin, D., Maas, H.G., Westfeld, P., Budzier, H., Gerlach, G., 2017. An advanced radiometric calibration approach for uncooled thermal cameras. *Photogram. Rec.* 33, 30–48. <https://doi.org/10.1111/phor.12216>.
- Lingua, A., Marenchino, D., Nex, F., 2009. Performance analysis of the SIFT operator for automatic feature extraction and matching in photogrammetric applications. *Sensors* 9, 3745–3766. <https://doi.org/10.3390/s90503745>.
- Luhmann, T., Piechel, J., Roelfs, T., 2013. Geometric calibration of thermographic cameras. In: *Kuenzer, C., Dech, S. (Eds.), Thermal Infrared Remote Sensing*. Springer Netherlands. volume 17 of *Remote Sensing and Digital Image Processing*. chapter 2, pp. 27–42. [https://doi.org/10.1007/978-94-007-6639-6\\_2](https://doi.org/10.1007/978-94-007-6639-6_2).
- Maes, W., Huete, A., Steppe, K., 2017. Optimizing the processing of UAV-based thermal imagery. *Remote Sensing* 9, 476. <https://doi.org/10.3390/rs9050476>.
- Mandanici, E., Conte, P., Girelli, V.A., 2016. Integration of aerial thermal imagery, LiDAR data and ground surveys for surface temperature mapping in urban environments. *Remote Sensing* 8, 880. <https://doi.org/10.3390/rs8100880>.
- Maset, E., Fusiello, A., Crosilla, F., Toldo, R., Zorretto, D., 2017. Photogrammetric 3D building reconstruction from thermal images. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inform. Sci. IV-2/W3 25–32*. <https://doi.org/10.5194/isprs-annals-iv-2-w3-25-2017>.
- Meier, F., Scherer, D., Richters, J., Christen, A., 2011. Atmospheric correction of thermal-infrared imagery of the 3-D urban environment acquired in oblique viewing geometry. *Atmospheric Measur. Techn.* 4, 909–922. <https://doi.org/10.5194/amt-4-909-2011>.
- Meißner, H., Cramer, M., Piltz, B., 2017. Benchmarking the optical resolving power of UAV based camera systems. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci. XLII-2/W6 243–249*. <https://doi.org/10.5194/isprs-archives-xlii-2-w6-243-2017>.
- Peeters, J., Louarroudi, E., De Greef, D., Vanlanduit, S., Dirckx, J.J.J., Steenackers, G., 2017. Time calibration of thermal rolling shutter infrared cameras. *Infrared Phys. Technol.* 80, 145–152. <https://doi.org/10.1016/j.infrared.2016.12.001>.
- Peppas, M.V., Mills, J.P., Moore, P., Miller, P.E., Chambers, J.E., 2016. Accuracy assessment of a UAV-based landslide monitoring system. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci. XLI-B5 895–902*. <https://doi.org/10.5194/isprsarchives-xli-b5-895-2016>.

- Remondino, F., El Hakim, S., 2006. Image-based 3D modelling: A review. *Photogram. Rec.* 21, 269–291. <https://doi.org/10.1111/j.1477-9730.2006.00383.x>.
- Remondino, F., Fraser, C., 2006. Digital camera calibration methods: consideration and comparisons. In: Maas, H.G., Schneider, D. (Eds.), *Proceedings of the ISPRS Commission V Symposium 'Image Engineering and Vision Metrology'*, ISPRS, Dresden. URL: <[http://www.isprs.org/proceedings/XXXVI/part5/paper/REMO\\_616.pdf](http://www.isprs.org/proceedings/XXXVI/part5/paper/REMO_616.pdf)>.
- Remondino, F., Spera, M.G., Nocerino, E., Menna, F., Nex, F., 2014. State of the art in high density image matching. *Photogram. Rec.* 29, 144–166. <https://doi.org/10.1111/phor.12063>.
- Ressl, C., Brockmann, H., Mandlbürger, G., Pfeifer, N., 2016. Dense image matching vs. airborne laser scanning – comparison of two methods for deriving terrain models. *Photogramm. – Fernerkundung - Geoinformation 2016*, 57–73. <https://doi.org/10.1127/pfjg/2016/0288>.
- Ribeiro Gomes, K., Hernández López, D., Ortega, J., Ballesteros, R., Poblete, T., Moreno, M., 2017. Uncooled thermal camera calibration and optimization of the photogrammetry process for UAV applications in agriculture. *Sensors* 17, 2173. <https://doi.org/10.3390/s17102173>.
- Rosenholm, D., 2006. Least squares matching method: some experimental results. *Photogram. Rec.* 12, 493–512. <https://doi.org/10.1111/j.1477-9730.1987.tb00598.x>.
- Stilla, U., Kolecki, J., Hoegner, L., 2009. Texture mapping of 3D building models with oblique direct geo-referenced airborne IR image sequences. In: Heipke, C., Jacobsen, K., Müller, S., Sörgel, U. (Eds.), *ISPRS Hannover Workshop 2009 – High-Resolution Earth Imaging for Geospatial Information*, ISPRS.
- Udelhoven, T., Schlerf, M., Segl, K., Mallick, K., Bossung, C., Retzlaff, R., Rock, G., Fischer, P., Müller, A., Storch, T., Eisele, A., Weise, D., Hupfer, W., Knigge, T., 2017. A satellite-based imaging instrumentation concept for hyperspectral thermal remote sensing. *Sensors* 17, 1542. <https://doi.org/10.3390/s17071542>.
- Vautherin, J., Rutishauser, S., Schneider Zapp, K., Choi, H.F., Chovancova, V., Glass, A., Strecha, C., 2016. Photogrammetric accuracy and modeling of rolling shutter cameras. *ISPRS Ann. Photogramm. Remote Sens. Spatial Informat. Sci.* 3.
- Weinmann, M., Hoegner, L., Leitloff, J., Stilla, U., Hinz, S., Jutzi, B., 2012. Fusing passive and active sensed images to gain infrared-textured 3D models. *ISPRS - International Archives of the Photogrammetry. Remote Sens. Spatial Informat. Sci.* XXXIX-B1 71–76. <https://doi.org/10.5194/isprsarchives-xxxix-b1-71-2012>.
- Weng, Q., 2009. Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS J. Photogramm. Remote Sens.* 64, 335–344. <https://doi.org/10.1016/j.isprsjprs.2009.03.007>.
- Westfeld, P., Mader, D., Maas, H.G., 2015. Generation of TIR-attributed 3D point clouds from UAV-based thermal imagery. *Photogrammetrie – Fernerkundung - Geoinformation 2015*, 381–393. <https://doi.org/10.1127/1432-8364/2015/0274>.
- Yong Jie, F., Wei Qi, J., Bin, L., Chong Liang, L., 2010. An analysis of digital detail enhancement (DDE) technology developed by FLIR. *Infrared Technol.* 3.
- Zhang, Z., Gerke, M., Peter, M., Yang, M.Y., Vosselman, G., 2017. Dense matching quality evaluation - an empirical study. In: 2017 Joint Urban Remote Sensing Event (JURSE). IEEE. <https://doi.org/10.1109/jurse.2017.7924579>.