PHOTOGRAMMETRIC 3D BUILDING RECONSTRUCTION FROM THERMAL IMAGES

E. Maset[a],[*] A. Fusiello[a], F. Crosilla[a], R. Toldo[b] and D. Zorzetto[c]

[a] DPIA, University of Udine, Via delle Scienze, 208 - 33100 Udine, Italy -
maset.eleonora@spes.uniud.it, (andrea.fusiello, fabio.crosilla)@uniud.it
[b] 3Dflow srl, Strada Le Grazie, 15 - 37134 Verona, Italy - roberto.toldo@3dflow.net
[c] AIRMAP Flight Analysis, Loc. Padriciano, 272 - 34149 Trieste, Italy - davide.zorzetto@airmap.it

KEY WORDS: Thermal Infrared Images, 3D Building Reconstruction, Unmanned Aerial Vehicles, Photogrammetry, Computer Vision

ABSTRACT:

This paper addresses the problem of 3D building reconstruction from thermal infrared (TIR) images. We show that a commercial Computer Vision software can be used to automatically orient sequences of TIR images taken from an Unmanned Aerial Vehicle (UAV) and to generate 3D point clouds, without requiring any GNSS/INS data about position and attitude of the images nor camera calibration parameters. Moreover, we propose a procedure based on Iterative Closest Point (ICP) algorithm to create a model that combines high resolution and geometric accuracy of RGB images with the thermal information deriving from TIR images. The process can be carried out entirely by the aforesaid software in a simple and efficient way.

1. INTRODUCTION

Nowadays, thermal imaging systems are used in a wide range of applications, such as archaeological heritage documentation (Brumana et al., 2013), landslide hazard assessment (Teza et al., 2015), crop health monitoring (Mangus et al., 2016) and photovoltaic plants inspection (Tsanakas et al., 2017). Thermographic cameras are increasingly becoming important also in buildings diagnostics. Indeed, thermography is a non-contact testing technology that allows to capture thermal irregularities caused by flaws or damages localized in buildings. Thanks to the representation through digital images, thermography avoids destructive interventions aimed at identifying the problem sources (Balaras and Argiriou, 2002). Inspections of building envelopes with thermal infrared (TIR) images can be used to detect heat losses, cracks, thermal bridges, air leakages, insulation problems in walls and roofs, moisture sources and other issues with thermal signature.

Moreover, the recent development of Unmanned Aerial Vehicle (UAV) platforms represents a great advantage for the data collection process, introducing a low-cost solution that is able to quickly deliver high temporal and spatial resolution information (Nex and Remondino, 2014). Thermal imaging cameras mounted on UAV systems thus allow an easy inspection not only of the facades, but also of the building roof, that cannot be captured by terrestrial images.

The analysis of TIR images can be complicated, since it is mostly done manually and images are often interpreted individually, without having a comprehensive view of the building, thus limiting in this way the full potential of this technique. In order to monitor big structures, therefore, a 3D building model becomes essential (Hoegner and Stilla, 2016). However, exploiting existing Photogrammetry and Computer Vision algorithms - developed to process RGB images - for the 3D reconstruction from TIR images is a topic that has not been fully explored yet. The geometric calibration of a thermographic camera has been studied in some works (Luhmann et al., 2013, Lagüela et al., 2011). (Hoegner and Stilla, 2015) propose a method for automatic texturing of building

facades from terrestrial thermal infrared image sequences, while (Pech et al., 2013) analyze the problem of using a TIR camera on an UAV for acquiring multi-temporal thermal images and generating thermal orthophotos. In (Khodaei et al., 2015) it is demonstrated that Digital Surface Models (DSM) generated from aerial thermal imagery can have comparable accuracy with respect to the one derived from visible images, while in (Hoegner et al., 2014) a fusion of time-of-flight depth images and TIR images is used to create an accurate 3D point cloud for subsequent scene segmentation and people detection. Moreover, the co-registration of images taken with both RGB and TIR cameras can give advantages in terms of accuracy and density of the resulting 3D point cloud, as highlighted in (Hoegner and Stilla, 2016, Hoegner et al., 2016).

The aim of this paper is first to show that a high quality 3D model reconstruction of a building can be obtained in a fast and automatic way *from TIR images only*, using the commercial Computer Vision software 3DF Zephyr, developed by 3Dflow[1]. Secondly, we also propose an automatic method for the integration of TIR and RGB images. The process is based on the registration with Iterative Closest Point (ICP) algorithm of dense point clouds, extracted from both datasets, and can be efficiently carried out entirely within the software. This approach allows to combine high resolution and geometric accuracy of RGB images with the thermal information deriving from TIR images.

The paper is organized as follows. In Sec. 2 the algorithms for reconstructing a 3D model directly from TIR images are described, while in Sec. 3 the procedure to automatically integrate TIR and RGB images is illustrated. Experimental validation of the two methods are reported in the respective sections, and discussed in Sec. 4.

2. AUTOMATIC 3D RECONSTRUCTION FROM THERMAL INFRARED IMAGES

Three-dimensional models of a subject can be obtained directly from unordered, uncalibrated TIR images using state-of-the-art

---

[*] Corresponding author

[1] www.3dflow.net

Photogrammetry and Computer Vision techniques. In the following sections, we will review the 3D reconstruction pipeline implemented in 3DF Zephyr, that is composed by three main modules, *Samantha* (Toldo et al., 2015), *Stasia* (Toldo et al., 2013) and *Sasha*. Starting from a set of images, *Samantha* is able to automatically recover the orientation (position and angular attitude) of the images. The robust auto-calibration algorithm that characterizes *Samantha* makes it possible to work with any digital camera, including TIR ones. *Stasia* extracts dense and accurate point clouds by multi-views stereo, while *Sasha* finally produces a texture mapped triangular mesh.

2.1 Structure-and-motion

The first step of *Samantha* is key-points extraction through the detector proposed by (Lindeberg, 1998), where blobs with associated scale levels are identified from scale-space extrema of the scale-normalized Laplacian. As for the descriptor, *Samantha* implements a 128-dimensional radial descriptor (similar to the log-polar grid of GLOH described by (Mikolajczyk and Schmid, 2005)), based on the accumulated response of steerable derivative filters. This combination of detector/descriptor performs in a comparable way to SIFT (Lowe, 2004), which is proved to be well suited for features detection also in low resolution thermal imagery (Pech et al., 2013), and at the same time avoids patent issues. Only a given number of key-points with the overall strongest response are retained. This number is a multiple of n (number of images), so as to fix the average quota of key-points per image.

As the images can be unordered, one must then recover the *epipolar graph*, i.e., the graph that tells which images overlap (or can be matched) with each other. This must be done in a computationally efficient way, without trying to match key-points between every image pair. In this broad phase only a small constant number of descriptors for each image is considered, and in particular the key-points with the higher scales, since their descriptors are more representative of the whole image content. So, each key-point descriptor is matched to its approximated nearest neighbors in feature space. A 2D histogram is then built that registers in each bin the number of matches between the corresponding images. Instead of directly picking pairs with the highest score in the histogram, that tends to create in the graph cliques of very similar images with weak inter-clique connection, an alternative approach based on taking maximum spanning trees is adopted, which creates a strongly connected epipolar graph (Toldo et al., 2015).

Subsequently, key-point matching follows a nearest neighbor approach with "ratio test" (Lowe, 2004), namely rejection of those key-points for which the ratio of the nearest neighbor distance to the second nearest neighbor distance is greater than a threshold. Matches that are not injective are discarded. Homographies and fundamental matrices between pairs of matching images are then computed using M-estimator SAmple Consensus (MSAC), a variation of RANSAC proposed by (Torr and Zisserman, 2000), and outliers are rejected. In the end, if the number of remaining matches between two images is less than 20% of the total number of matches before MSAC, they are discarded. The rationale is that if an excessive fraction of outliers has been detected, the original matches are altogether unreliable (Brown and Lowe, 2003). After that, key-point matching in multiple images are connected into *tracks*: consider the undirected graph key-points are the nodes and edges represent matches; a track is a connected component of that graph. Vertices are labeled with the image the

key-points belong to: an inconsistency arises when in a track a label occurs more than once. Inconsistent tracks and those shorter than three frames are discarded. A track represents the projection of a single 3D tie-point imaged to multiple exposures.

In the next step, images are organized into a tree (or dendrogram) with agglomerative clustering, using a measure of overlap as the affinity. The structure-and-motion computation follows this tree from the leaves to the root, where images are stored in the leaves and partial models (sets of exposures and 3D points expressed in a local reference frame) correspond to internal nodes.

The whole hierarchical structure-and-motion approach relies on four basic photogrammetric procedures, that are implemented in *Samantha* with a special attention to outliers resilience. *Intersection* (a.k.a. triangulation) is the procedure of computing 3D point coordinates from corresponding points in multiple images; *Resection* consist in recovering the camera matrix (or the exterior parameters only) from known 3D-2D correspondences; *Relative orientation* is the task of retrieving the relative position and attitude of two cameras from corresponding points in the two images; Finally, *absolute orientation* requires to compute the rigid (or similarity) transformation that brings two models that share some tie-points into a common reference frame. The hierarchical algorithm can be summarized as follows:

1. Solve many independent relative orientation problems at the leaves of the tree, producing many independent stereo-models.

2. Traverse the tree; in each node one of these operations takes place:

    (a) Update one model by adding one image with resection followed by intersection;

    (b) Merge two independent models with absolute orientation. Once the models are registered, tie-points are updated by intersection, and the new model is refined with bundle block adjustment.

If the tree reduces to a chain, the algorithm performs only steps 1. and 2.(a) which is tantamount to the classical resection-intersection sequential pipelines.

If the tree is perfectly balanced, only steps 2.(b) are taken, and the resulting procedure resembles the photogrammetric Independent Models Block Adjustment (IMBA) (Kraus, 1997) (where for each pair of overlapping images, a stereo-model is built and then all these independent models are simultaneously transformed into a common reference frame with absolute orientation), besides the fact that the models are disjoint and are recursively merged in pairs.

Compared to the standard sequential approach, this framework has a lower computational complexity, is independent of the initial pair of images, and copes better with drift problems, typical of sequential schemes. In any case, the final step is a bundle block adjustment.

The auto-calibration method of *Samantha* (Gherardi and Fusiello, 2010) is able to compute the interior parameters of the images without providing any control point. It is based on the enumeration of the inherently bounded space of the interior parameters of two cameras in order to find the collineation of space that upgrades a given projective reconstruction to Euclidean. Each sample of the search space (which reduces to a finite subset of $\mathbb{R}^2$ under mild assumptions) defines a consistent plane at infinity. This

in turn produces a tentative, approximate Euclidean upgrade of the whole reconstruction which is then scored according to the expected intrinsic parameters of a Euclidean camera.

2.2 Model georeferencing

Images taken by a camera mounted on a UAV are often provided with GNSS/INS information, that can be exploited as initial values for the unknown estimated orientation parameters. However, their availability is not guaranteed. In any case, *Samantha* does not require any ancillary information (GNSS/INS measures or calibration parameters); the resulting camera orientation and 3D points coordinates are expressed in a local reference frame and defined up to a scale.

In order to transfer the model coordinates into a global coordinate system and to remove the scale ambiguity, a similarity transformation must be computed. This can be done as soon as at least three Ground Control Points (GCPs) are available, whose coordinates are measured independently (by GNSS or other techniques) and are expressed in the global reference frame. These control points are identified manually in the images and their position in the 3D space is estimated by intersection. Correspondences between true and estimated 3D coordinates are used to transform the model with a similarity that aligns the control points in a least-squares sense. Please note that GCPs can also be used as a constraint to optimize the 3D reconstruction.

2.3 Multi-view stereo

The aim of multi-view stereo is to recover a dense point cloud representing the surface of the imaged object, given images position and attitude. In 3DF Zephyr this procedure is performed through *Stasia* (Toldo et al., 2013).

The first step of the algorithm is the extraction of depth hypothesis. The goal of this phase is to estimate a number of *candidates depths* for each pixel $\mathbf{m}$ and for each image I_i. These hypothesis will be later used as labels in a Markov Random Field (MRF) that extracts the final depth map $\delta_i(\mathbf{m})$. Similarly to many multi-view stereo algorithms, a pixel-level matching along epipolar lines is used, with Normalized Cross Correlation (NCC) as the matching metric, which gives a good trade-off between speed and robustness to photometric nuisances. Every depth map is created independently from the others. The extraction of candidate depths is performed by considering the reference image I_i and three neighboring views $\eta(I_i)$, chosen on the basis of the sparse structure and the visibility information provided by *Samantha*. The candidate depths for each pixel are searched along the optical ray, or equivalently, along the epipolar line of each neighboring image using block matching and NCC. In this way, a correlation profile $C_j(\zeta)$, parameterized with the depth ζ, is computed for every pixel $\mathbf{m}$ and every neighbor image $I_j \in \eta(I_i)$.

Candidate depths correspond to local peaks of the correlation (peaks with a NCC value lower than 0.6 are discarded). Since the search range of each pixel depth can heavily impact the performance of the algorithm, information coming from the structure-and-motion is used to limit the search range.

The final depth map δ_i is generated from the depth hypothesis using a discrete MRF optimization technique over the image grid. Depth maps are then lifted in 3D space to produce a *photoconsistency* volume ϕ, represented by an octree that accumulates the scores coming from each depth map δ_i. In order to avoid any

loss of accuracy, a moving average approach is used inside each bin. At the end of the lifting process, each cell $\mathbf{x}$ contains a 3D point position – which can be shifted with respect to the cell center – and a photoconsistency value $\phi(\mathbf{x})$ given by the sum of the correlation scores of the points that fall in that bin. The photoconsistency volume at this stage contains a lot of spurious points, which do not belong to a real surface. They are characterized by two features: i) their photoconsistency is generally lower than actual surface points, and ii) they usually occludes actual surface points. This observation leads to an iterative strategy where the photoconsistency of an occlusor is decreased by a fraction of the photoconsistency of the occluded point. Points with negative photoconsistency are eventually removed.

2.4 Mesh generation and texture mapping

At the end of the process, a surface is generated by the module called *Sasha*. It employs the Poisson algorithm (Kazhdan et al., 2006) starting with normals computed, at each point, by fitting a plane to the closer neighbors. Normal direction is disambiguated with visibility.

In addition, the surface can be further optimized inside the software with an optimization algorithm based on photoconsistency. As the initial surface reconstruction method is interpolatory and since the point cloud may contain a decent amount of noise, the obtained initial mesh is normally noisy and may fail to capture fine details. By using all the image data, this mesh is refined with a variational Multi-view stereo approach: the initial mesh is used as the initial condition of a gradient descent of an adequate energy functional with an algorithm similar to (Faugeras and Keriven, 2002, Vu et al., 2012).

Finally a texture (TIR) map is built by wisely making use of both the visibility and the view angle information.

2.5 Experiments

To assess the applicability of the proposed method, a dataset of 130 TIR images was acquired by a Optris PI450 LW thermal infrared camera (wavelength range from 7 μm to 13 μm), with a detector of 382×288 pixels and a temperature resolution of 0.4 mK.

Figure 1. UAV platform used for the survey.

The camera was mounted on a octorotor UAV with a total payload of 4 kg, roll and pitch axis stabilization, and a flight time of 20 minutes (see Fig. 1). The flight was planned for the images to have a forward and a side overlap of 80% and an average Ground Sampling Distance (GSD) of 94 mm.

The subject is a large building that houses an air conditioning and heating plant. On the roof, many solar panels are installed. Due to the size of the building, each image reproduces only a limited portion, as shown in Fig. 2. Hence, it is highly recommended in
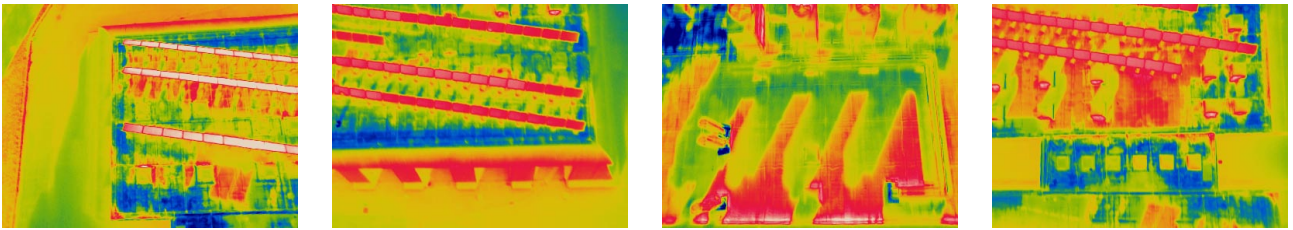
Figure 2. Sample images from the TIR dataset.
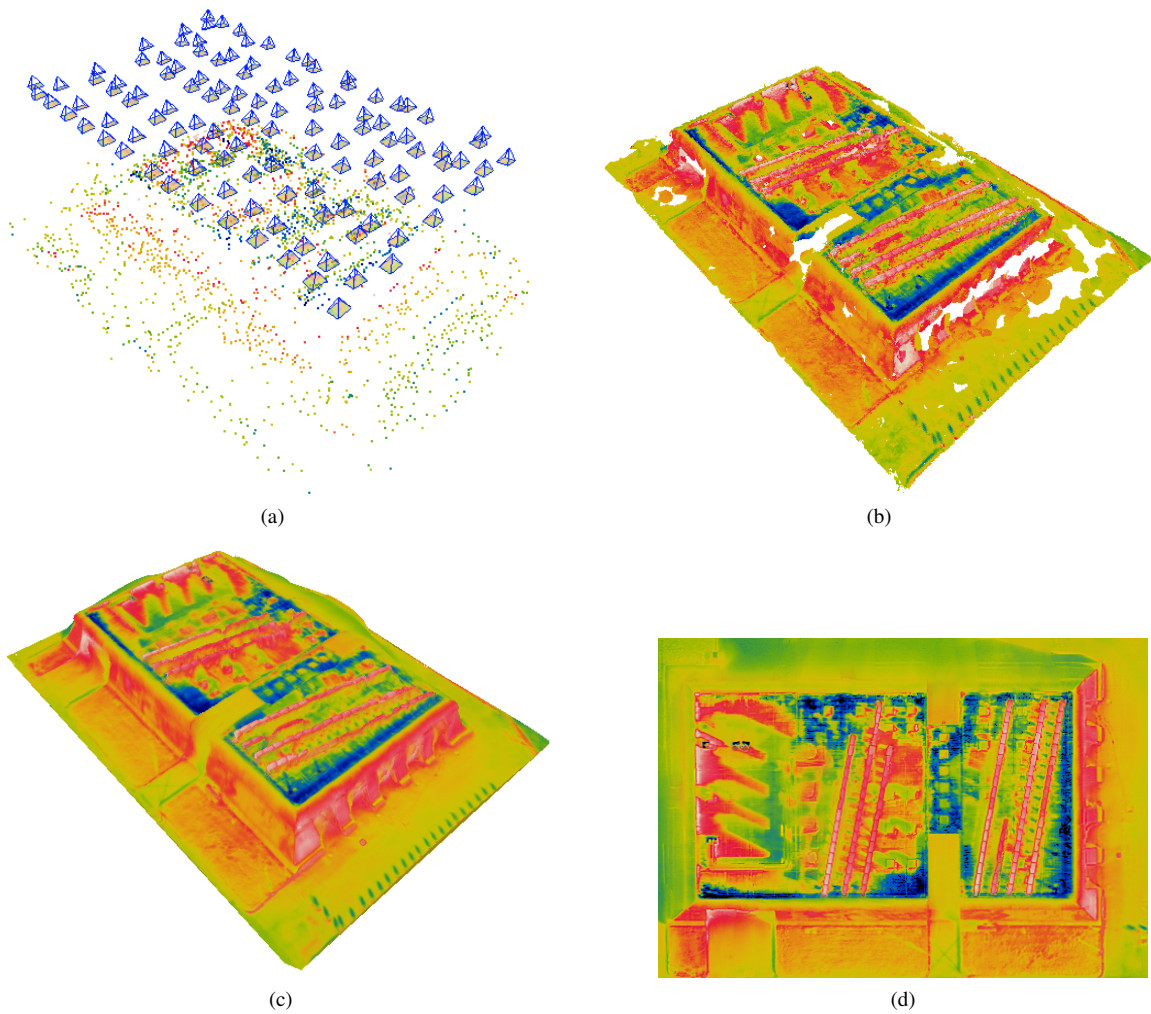

(a)

(b)

(c)

(d)

Figure 3. 3D reconstruction from TIR images: (a) sparse point cloud, (b) dense point cloud, (c) mesh with texture, (d) orthophoto.

this situation to create a 3D model and an ortophoto, in order to have a comprehensive view of the building and to make it easier the evaluation and interpretation of the thermal data.

Following the procedure described in Sec. 2, first the structure-and-motion algorithm was run to retrieve position and attitude of each image. After bundle adjustment, the RMSE was 0.31 pixel and the estimated reference variance was 0.35 squared pixel. The result, together with the generated sparse point cloud, is represented in Fig. 3(a). Then the model was georeferenced exploiting the coordinates of three GCPs, measured by highly accurate GPS differential positioning in a previous survey. This step was quite tricky, because the GCPs, placed on the edges of the building, were difficult to identify on the TIR images. In fact, the average residual was 0.8 m.

Subsequently, the dense point cloud (Fig. 3(b)) was created with the multi-view stereo algorithm (Sec. 2.3). Please note that, except for some portions of the building facades and the sidewalk around, the 3D point cloud has very high density and most of the details are reconstructed, including the solar panels (see Fig. 4).
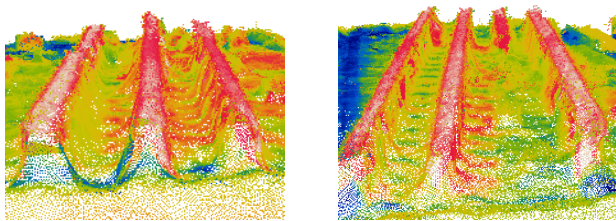

Figure 4. Details of the dense point cloud generated from the TIR dataset. The solar panels on the roof of the building were correctly reconstructed.

Finally, the mesh with texture was extracted (Fig. 3(c)) and a high quality thermal orthophoto was created (Fig. 3(d)). The entire process required less than 30 minutes on a PC with an Intel Core i5-4200M CPU @ 2.50GHz and 8GB RAM.

These results demonstrate how structure-and-motion and multi-view stereo algorithms can be successfully applied to create high detailed dense point cloud, 3D models and orthophoto directly from unordered, uncalibrated TIR images, without requiring additional RGB images or ancillary information. This makes TIR images of immediate use, reducing considerably the processing time.

However, due to low geometric resolution and low contrast of TIR images, 3D point clouds generated directly from them have a lower accuracy than those obtained from RGB images. Therefore, when RGB images of the same area are already available, it might be wise to try and use them, as we will discuss in the next section.

3. INTEGRATION OF TIR AND RGB IMAGES

In this section we will propose a procedure to combine TIR and RGB images, in order to create high detailed 3D models that at the same time maintain the thermal information. The method, briefly summarized in Fig. 5, can be carried out automatically by 3DF Zephyr on any set of TIR and RGB images, i.e., each RGB image does not need a corresponding TIR image taken from the same position and the same orientation, as required in (Hoegner et al., 2016).
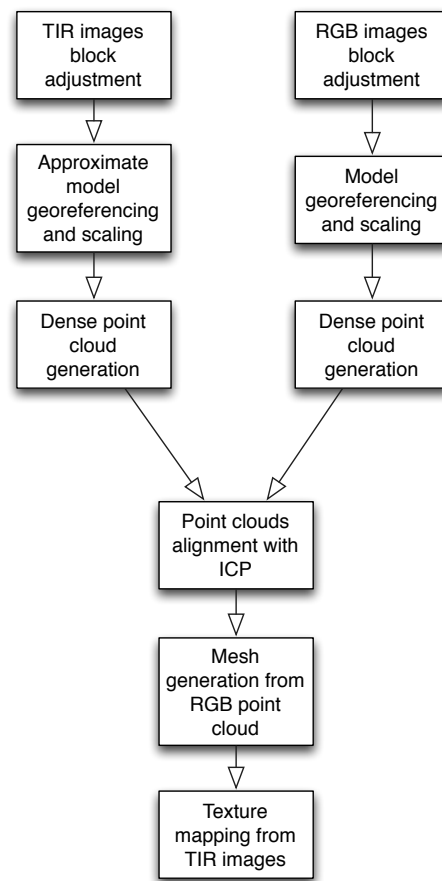

Figure 5. Simplified overview of the procedure to create a 3D building model from TIR and RGB images.

3.1 Point clouds generation

The first steps resemble the procedure described in the previous section. TIR and RGB images are oriented separately through the structure-and-motion algorithm (Sec. 2.1), without any a-priori knowledge about their position and attitude. After that, the obtained models are georeferenced and scaled using at least three GCPs (Sec. 2.2). Please note that in this step the model deriving from TIR images requires only a coarse georeferencing, since it will be subsequently registered onto the RGB point cloud. Then, the multi-view stereo algorithm (Sec. 2.3) independently generates TIR and RGB dense point clouds.

3.2 Alignment via Iterative Closest Point (ICP)

Thanks to georeferencing, RGB and TIR point clouds are close together in the 3D space and have approximately the same scale. Therefore, the TIR point cloud can be registered to the one generated from RGB images using the Iterative Closest Point (ICP) method (Rusinkiewicz and Levoy, 2001). This algorithm computes correspondences between the point sets given an estimate for the transformation, then updates the transformation based on the current correspondences, and iterates through these steps until convergence - to a local minimum - is reached.

3.3 Final model extraction

Orientation parameters of the TIR images are automatically transformed into the same reference system as the RGB images by the alignment of the TIR point cloud with the RGB one. This makes

it possible to associate the thermal information deriving from the TIR dataset to the 3D point clouds calculated from the RGB images. The final surface is generated from the RGB point clouds through the Poisson algorithm (Sec. 2.4), thus maintaining the geometric accuracy and the high level of detail arising from images acquired in the visible spectrum, whereas the texture map is built from TIR images. The texture coordinates are obtained projecting the vertices of the RGB mesh into the TIR images.

3.4 Experiments

To evaluate the procedure proposed for the integration of TIR and RGB images, an already available dataset of 27 RGB images taken over the same area was used. The images were acquired two years before with a Canon Power Shot S100 camera (1/1.7 CMOS sensor) with an image size of 3000×4000 pixels and an average GSD of 27 mm.

After the block adjustment with *Samantha* the estimated reference variance was 0.52 squared pixel. In this case, georeferencing the model was easier than for the TIR dataset because the GCPs, placed on building edges, road signs or edges of manholes nearby, were easily recognizable in the RGB images. The least-squares alignment was performed on 12 GCPs with a final average residual of 0.1 m. Then, the dense point cloud was generated from the RGB images and it was used as reference to align the TIR point cloud previously extracted, computing a similarity transformation. The orientation of the TIR images were updated accordingly. After the registration via ICP, the mean distance of corresponding 3D points of the two point clouds was 0.15 m, with a standard deviation of 0.07 m. To further evaluate qualitatively the obtained results, TIR and RGB point clouds were compared (using "CloudCompare" software[2]). From the distribution of the residual point-point absolute distance shown in Fig. 6, no deformations of the TIR block are appreciable.

The described process allowed the creation of an accurate surface of the building from the RGB point cloud (Fig. 7(a)), that was subsequently textured with the TIR images (Fig. 7(b)).

4. DISCUSSION

Results demonstrate that a Computer Vision software (3DF Zephyr), developed to process RGB images, can be used also for the 3D reconstruction from thermal infrared data. Structure-and-motion and multi-view stereo algorithms can be successfully applied to create high detailed dense point cloud, 3D models and orthophoto directly from TIR images, without requiring additional images in the visible spectrum. This makes TIR images of immediate use, reducing considerably the processing time. On the other hand, when RGB images are available in addition, the procedure proposed in Sec. 3 allows an automatic integration between TIR and RGB images.

Comparing the models generated from TIR and RGB datasets, the RGB one shows an increase in the number of details, thanks to the high resolution of the RGB images. This is due also by the fact that building edges are less visible in the thermal infrared with respect to the visual spectrum. As can be seen in Fig. 8, the edges of the building are sharper and better reconstructed in the RGB model. Furthermore, the distribution and the sizes of the polygons in the RGB mesh are more regular. Therefore, as anticipated, the integration of RGB and TIR guarantees the highest

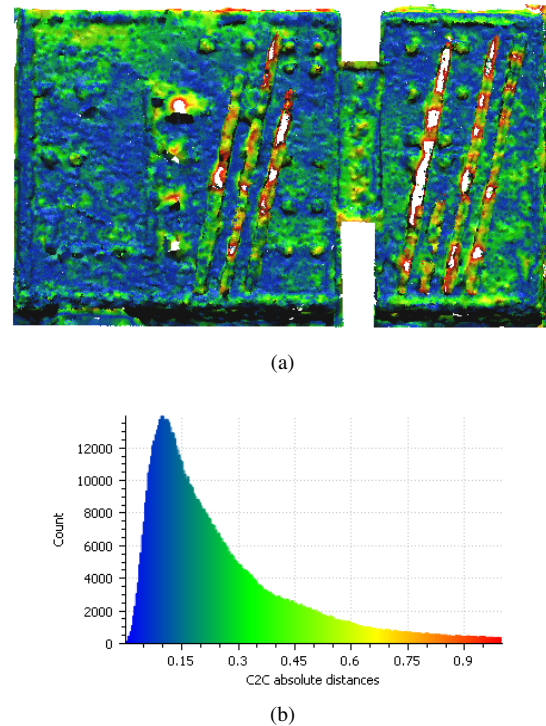[2]http://www.danielgm.net/cc/

(a)

(b)

Figure 6. Results of the comparison between TIR and RGB point clouds. Colors encode the residual absolute distance (units are in meters).

achievable geometric accuracy, maintaining at the same time the thermal information.

Finally, it is important to highlight the differences with comparable methods previously proposed in the literature (Hoegner and Stilla, 2016, Hoegner et al., 2016). Our procedures allow to perform the images block adjustment even without GNSS/INS data, which in some cases can be unavailable or, more often, have low accuracy. The model georeferencing and scaling can be done subsequently, exploiting at least three GCPs identified in the images and whose 3D coordinates are known. Moreover, if it is not necessary to express the model coordinates in a particular global reference frame, the use of GCPs can be avoided and only a control distance is required to scale the reconstructed model.

In our methods, the geometric calibration of TIR images is not requested, thanks to the robust auto-calibration algorithm implemented in *Samantha* (Toldo et al., 2015) that works even with TIR datasets, as demonstrated by the experimental evaluation.

Please note also that, unlike the procedure described in (Hoegner et al., 2016), we do not require that every RGB image has a corresponding TIR image taken from the same position and with the same orientation. TIR and RGB datasets can be totally independent and, if geometric variations do not occur in the area over time, RGB images can be acquired in a different time period. Moreover, this allows an automatic integration of UAV and terrestrial datasets.

To conclude, the procedures proposed in this paper for the 3D model reconstruction of a building from TIR images can provide an helpful support to make the interpretation of thermal data faster, more automatic and objective.
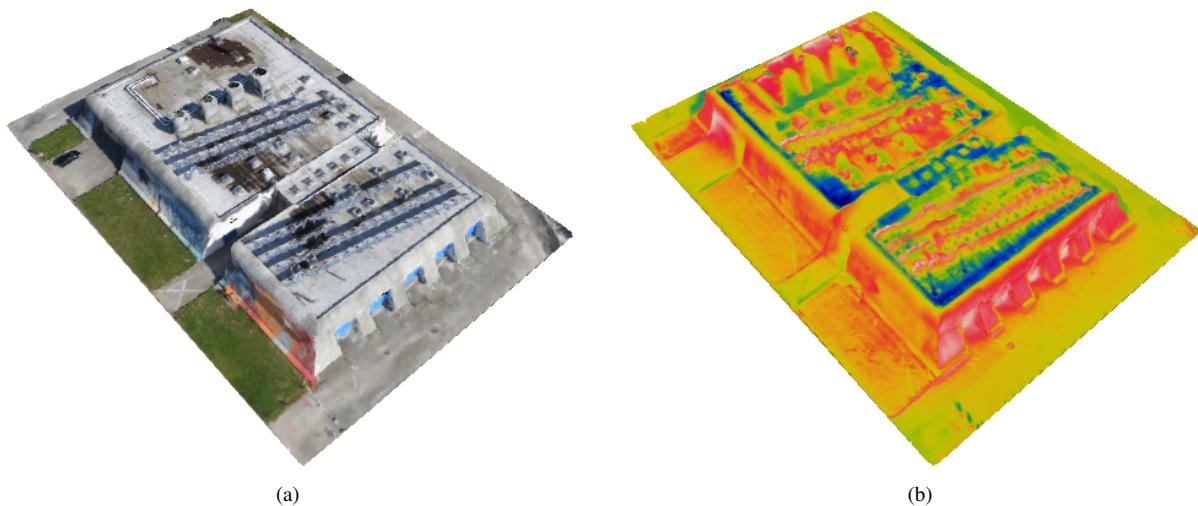
(a) (b)

Figure 7. Integrating TIR and RGB images: (a) 3D model from RGB images, (b) texture from TIR images.

REFERENCES

Balaras, C. and Argiriou, A., 2002. Infrared thermography for building diagnostics. *Energy and buildings* 34(2), pp. 171–183.

Brown, M. and Lowe, D., 2003. Recognising panoramas. In: *Proceedings of the 9th International Conference on Computer Vision*, Vol. 2, pp. 1218–1225.

Brumana, R., Oreni, D., Hecke, L. V., Barazzetti, L., Previtali, M., Roncoroni, F. and Valente, R., 2013. Combined geometric and thermal analysis from uav platforms for archaeological heritage documentation. In: *XXIV International CIPA Symposium, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 02–06.

Faugeras, O. and Keriven, R., 2002. Variational principles, surface evolution, pde's, level set methods and the stereo problem. In: *IEEE International Summer School on Biomedical Imaging*.

Gherardi, R. and Fusiello, A., 2010. Practical autocalibration. In: *Proceedings of the European Conference on Computer Vision*, Lecture Notes in Computer Science, pp. 790–801.

Hoegner, L. and Stilla, U., 2015. Building facade object detection from terrestrial thermal infrared image sequences combining different views. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(3), pp. 55–62.

Hoegner, L. and Stilla, U., 2016. Automatic 3d reconstruction and texture extraction for 3d building models from thermal infrared image sequences. *Proc. of QIRT'16*.

Hoegner, L., Hanel, A., Weinmann, M., Jutzi, B., Hinz, S. and Stilla, U., 2014. Towards people detection from fused time-of-flight and thermal infrared images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40(3), pp. 121.
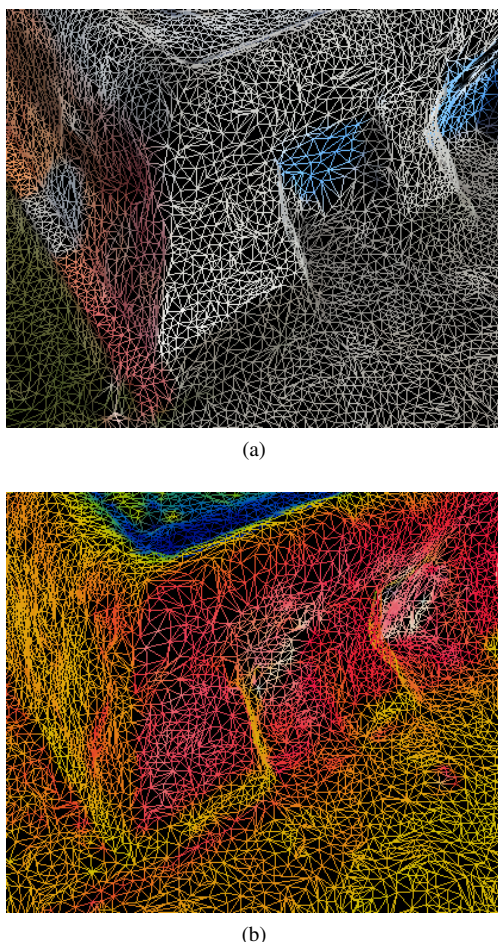
(a)

(b)

Figure 8. A detail of the RGB model (a) and of the TIR one (b). Please note that building edges are less visible in the TIR model.

Hoegner, L., Tuttas, S., Xu, Y., Eder, K. and Stilla, U., 2016. Evaluation of methods for coregistration and fusion of rpas-based 3d point clouds and thermal infrared images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* pp. 241–246.

Kazhdan, M., Bolitho, M. and Hoppe, H., 2006. Poisson surface reconstruction. In: *Eurographics symposium on Geometry processing*, pp. 61–70.

Khodaei, B., Samadzadegan, F., Javan, F. D. and Hasani, H., 2015. 3d surface generation from aerial thermal imagery. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40(1), pp. 401.

Kraus, K., 1997. *Photogrammetry: Advanced methods and applications*. Vol. 2, Dümmler.

Lagüela, S., González-Jorge, H., Armesto, J. and Arias, P., 2011. Calibration and verification of thermographic cameras for geometric measurements. *Infrared Physics & Technology* 54(2), pp. 92–99.

Lindeberg, T., 1998. Feature detection with automatic scale selection. *International Journal of Computer Vision* 30, pp. 79–116.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.

Luhmann, T., Piechel, J. and Roelfs, T., 2013. Geometric calibration of thermographic cameras. In: *Thermal Infrared Remote Sensing*, Springer, pp. 27–42.

Mangus, D. L., Sharda, A. and Zhang, N., 2016. Development and evaluation of thermal infrared imaging system for high spatial and temporal resolution crop water stress monitoring of corn within a greenhouse. *Computers and Electronics in Agriculture* 121, pp. 149–159.

Mikolajczyk, K. and Schmid, C., 2005. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 27(10), pp. 1615–1630.

Nex, F. and Remondino, F., 2014. Uav for 3d mapping applications: a review. *Applied Geomatics* 6(1), pp. 1–15.

Pech, K., Stelling, N., Karrasch, P. and Maas, H.-G., 2013. Generation of multitemporal thermal orthophotos from uav data. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1(2), pp. 305–310.

Rusinkiewicz, S. and Levoy, M., 2001. Efficient variants of the ICP algorithm. In: *Proceedings of the International Conference on 3-D Digital Imaging and Modeling*, pp. 145–152.

Teza, G., Marcato, G., Pasuto, A. and Galgaro, A., 2015. Integration of laser scanning and thermal imaging in monitoring optimization and assessment of rockfall hazard: a case history in the carnic alps (northeastern italy). *Natural Hazards* 76(3), pp. 1535–1549.

Toldo, R., Fantini, F., Giona, L., Fantoni, S. and Fusiello, A., 2013. Accurate multiview stereo reconstruction with fast visibility integration and tight disparity bounding. In: *Proceedings of the ISPRS Workshop 3D-ARCH 2013: 3D Virtual Reconstruction and Visualization of Complex Architectures*, ISPRS Archives, Vol. XL-5/W1, pp. 243–249.

Toldo, R., Gherardi, R., Farenzena, M. and Fusiello, A., 2015. Hierarchical structure-and-motion recovery from uncalibrated images. *Computer Vision and Image Understanding*.

Torr, P. H. S. and Zisserman, A., 2000. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding* 78, pp. 2000.

Tsanakas, J. A., Ha, L. D. and Shakarchi, F. A., 2017. Advanced inspection of photovoltaic installations by aerial triangulation and terrestrial georeferencing of thermal/visual imagery. *Renewable Energy* 102, pp. 224–233.

Vu, H., Labatut, P., Pons, J. and Keriven, R., 2012. High accuracy and visibility-consistent dense multi-view stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34(99), pp. 889–901.