3D City Models completion by Fusing Lidar and Image Data

L. Grammatikopoulos*[a], I. Kalisperakis[a], E. Petsa[a], C. Stentoumis[b]
[a]Laboratory of Photogrammetry, Department of Civil Engineering and Surveying & Geoinformatics Engineering, Technological Educational Institute of Athens, GR-12210 Athens, Greece;
[b] Laboratory of Photogrammetry, Department of Surveying, National Technical University of Athens, GR-15780 Athens, Greece

ABSTRACT

A fundamental step in the generation of visually detailed 3D city models is the acquisition of high fidelity 3D data. Typical approaches employ DSM representations usually derived from Lidar (Light Detection and Ranging) airborne scanning or image based procedures. In this contribution, we focus on the fusion of data from both these methods in order to enhance or complete them. Particularly, we combine an existing Lidar and orthomosaic dataset (used as reference), with a new aerial image acquisition (including both vertical and oblique imagery) of higher resolution, which was carried out in the area of Kallithea, in Athens, Greece. In a preliminary step, a digital orthophoto and a DSM is generated from the aerial images in an arbitrary reference system, by employing a Structure from Motion and dense stereo matching framework. The image-to-Lidar registration is performed by 2D feature (SIFT and SURF) extraction and matching among the two orthophotos. The established point correspondences are assigned with 3D coordinates through interpolation on the reference Lidar surface, are then backprojected onto the aerial images, and finally matched with 2D image features located in the vicinity of the backprojected 3D points. Consequently, these points serve as Ground Control Points with appropriate weights for final orientation and calibration of the images through a bundle adjustment solution. By these means, the aerial imagery which is optimally aligned to the reference dataset can be used for the generation of an enhanced and more accurately textured 3D city model.

Keywords: Image orientation, Lidar, Registration, 3D city models

1. INTRODUCTION

Today, 3D city models are of course an essential tool in a variety of applications, such as urban planning, tourism, real estate as well as asset and utility management, since they provide a more realistic representation of urban environments compared to traditional 2D maps and orthophotos. A crucial task in generating geometrically accurate and visually detailed 3D models of buildings, and urban areas in general, is the acquisition of high fidelity 3D data, which is usually performed by airborne Lidar scanners. However, in densely built areas these elevation data are often incomplete, due to failures, adverse geometries or occlusions, which are more severe on building facades. Hence, an interesting task is the fusion of Lidar data with those generated from image-based procedures in order to enhance or complete them. The work presented here is part of an ongoing research project which aims to create accurate 3D city models by combining various types of geospatial data, such as Lidar, orthomosaics along with vertical and oblique aerial imagery.

For the last years, the registration of Lidar with optical data has been a crucial issue that has attracted the interest of many researchers in the fields of Photogrammetry and Computer Vision. Among the first approaches, Delara et al.[1] and Habib et al.[2] collect manually control points and linear features on the Lidar surface and incorporate them in a final bundle adjustment solution. Böhm and Becker[3] extract and match SIFT[4] features among photographs and intensity images of scan data in order to align them. Smith et al.[5] use also SIFT features to register multiple scans. Following an automatic scheme, Liu and Stamos[6] exploit vanishing points and establish correspondences among 2D and 3D linear features extracted on the image and range data respectively, to further refine their orientation. Armenakis et al.[7] propose a semi-automatic approach for the alignment of photogrammetric and Lidar data based on the identification of corresponding buildings in the two datasets. An initial co-registration is achieved via anchor planar features and a final refinement is obtained by the ICP[8] algorithm between building points. Starting with a least squares surface matching between photogrammetric tie points and Lidar point cloud, Gneeniss et al.[9] optimize their registration by extracting automatically Lidar control points and integrating them into a final self-calibrating aerotriangulation. Similarly, Palenichka and Zaremba[10] extract SIDs (Salient Image Disks) features to determine control points for satellite image-to-Lidar registration. On the other hand, Yang and Chen[11] combine mini-UAV images with Lidar data. In their approach

*lazaros@teiath.gr; http://photogram.tg.teiath.gr/?page_id=21

images are first oriented via a Structure from Motion (SfM) algorithm and the obtained point cloud is georeferenced based on a building matching scheme and a subsequent ICP procedure.

In this contribution, we focus initially on the automatic registration of the images with the existing Lidar data and digital orthophotos. Then, dense matching algorithms allow the generation of a highly detailed point cloud and its integration with the Lidar model, thus finally leading to a more complete 3D mesh model. Our dataset covers a densely built neighborhood of Athens, Greece, and includes an orthomosaic with spatial resolution of 20 cm, provided by the National Cadastre & Mapping Agency of Greece, a DSM generated from Lidar with spatial resolution 1 m and aerial overlapping images (both vertical and oblique) with average resolution of 15 cm.

The image-to-Lidar registration is carried out in multiple steps. First, all available images are relatively oriented through a hierarchical Structure from Motion[12-14] scheme implemented by our team. Image pairs are identified among unordered sets of images; sparse matching is then performed employing state-of-the-art features with their descriptors (SIFT[4], SURF[15]) at multiple image scales. Image matches are thus established across different stereo-pairs leading to multi-image point correspondences. By means of closed-form algorithms, image orientations are initialized, and finally exterior and interior orientation parameters are refined through a typical self-calibrating bundle adjustment solution, in a suitably chosen arbitrary system that resembles the official reference system (Lidar and orthophoto). In a consequent step, state-of-the-art dense stereo matching algorithms are applied to selected image pairs, and results are combined into a unified DSM. The latter is used for the production of a *new orthomosaic (n-ortho)*, in which 2D SIFT and SURF keypoints are detected along with their corresponding descriptors.

On the other hand, features with descriptors are additionally detected on the existing *reference orthomosaic (r-ortho)* and assigned to 3D points via height interpolation on the Lidar DSM. Descriptor matching with robust outlier detection (by applying RANSAC to 2D affine or 3D similarity transformation) is performed among these two point sets. Inliers are subsequently compared against 2D features extracted on the oriented aerial images; this procedure is restricted on image areas defined by the backprojection of corresponding 3D points. In this way, 3D points are assigned to image features and can be treated as control point observations with appropriately tuned weights in a final bundle adjustment solution. Thus, an optimal registration of all data is achieved. At this point, the dense matching step can be repeated in order to complete the Lidar data. In addition the oriented image set can be used for a more detailed texture mapping of the final 3D city model.

The paper is organized into 4 sections. Section 2 describes the individual steps of the proposed approach for the integration of Lidar and image data towards the completion of 3D city models. Section 3 reports some preliminary results of the currently implemented process applied to a dataset from the Kallithea district of Athens, Greece. Finally, conclusions and a discussion of future tasks follow in Section 4.

2. FUSION OF LIDAR AND AERIAL IMAGE DATA

In this section we present an approach for the integration of existing geospatial data (Lidar surface model and orthomosaic) with newly acquired aerial image data of higher resolution. The general workflow consists of four individual steps; (a) an automatic aerial image orientation via Structure from Motion procedure, (b) a preliminary DSM and orthophoto creation from the oriented aerial images, (c) a mutual registration of Lidar and image data by exploiting a feature matching scheme between orthophotos and (d) the production of an enhanced photorealistic 3D city model.

2.1 Automatic aerial image orientation

In this phase, a structure from motion scheme on successive image scales has been implemented, to facilitate the use of a large amount of high resolution images. In a first step, stereo pairs are identified among the unordered set of aerial images. For this purpose, all images are subsampled to a low resolution; SIFT and SURF features are extracted and a matching scheme with outlier detection (RANSAC using fundamental matrix) is applied to all possible stereo image combinations. Valid stereopairs are defined based on the number of inliers, as well as the percentage of estimated outliers after RANSAC. In case the interior orientation of the camera is unknown, an initial estimation of a common camera constant may be computed as the median of all camera constant values extracted from the fundamental matrices of all valid stereopairs (assuming that the principal point coincides with the image center)[16].

Once stereopairs have been selected, SIFT and SURF features are extracted at a higher image scale and matched via RANSAC based on the five point algorithm[17] for the estimation of the essential matrix. Image matches are thus established across different stereopairs leading to multi-image point correspondences. A bucketing algorithm is then performed to reduce the number of tie points, without affecting their distribution on images.

For the initialization of all image exterior orientations, a stereopair is selected as reference; for every new stereopair, relative orientation is estimated from the essential matrix, tie points are reconstructed in 3D space through triangulation and a 3D similarity transformation allows inserting the current stereopair into the reference system. Local bundle adjustment solutions are held for every N successive images to ameliorate the exterior orientation accuracy, and a full self-calibrating bundle adjustment is performed among all available images.

Following the hierarchical scheme, new feature points are collected at successively higher image scales. Matching is restricted by the known image orientations (epipolar constraint) and by a rough 3D reconstruction of the object surface that is obtained from the tie points of previous image scale. This approach is repeated up to the full image resolution, leading to final bundle adjustment. Finally, a rotation is applied in order to approximate a reference plane vertical to the average image axis.

2.2 Preliminary DSM and orthophoto generation

Our approach for registering the aerial image data to Lidar surface takes advantage of the existence of a reference orthophoto (*r-ortho*), which is already aligned with the Lidar. For this purpose, a DSM is reconstructed along with a new orthophoto (*n-ortho*) from the oriented aerial image set. For the DSM generation several state-of-the-art dense stereo and multi-image reconstruction algorithms have been tested[18-21], which, either combine different disparity maps into a single model, or compute directly an optimal 3D surface from all images. The DSM presented in Section 3 has been generated using the CMPMVS algorithm of Jancosek and Pajdla[20]. An orthomosaic is then produced by a multi-image algorithm based on the automatic visibility checking and texture blending technique developed by our team[21].

2.3 Lidar to aerial image registration

This task can be divided into three main steps: (a) a 2D feature extraction and matching process between the two orthophotos (*r-ortho* and *n-ortho*) and the assignment of 3D coordinates to valid point matches; (b) the identification of these 3D points on the original aerial images by exploiting image descriptors; (c) a final refinement of image exterior and interior orientation with a bundle adjustment solution, in which extracted 3D points are treated as weighted ground control points (GCPs).

First, SIFT and SURF 2D features are extracted on both orthophotos, and then matched based on the Euclidean distance between their respective descriptors (Figure 1). Matched points are assigned with two sets of X, Y, and Z coordinates in the two different projection systems, by interpolation on the Lidar and the photogrammetric surfaces. Although these features are supposed to be invariant against scale, rotation and translation transformations, many false matches are expected, in particular due to differences in the projection system of the two orthophotos (local and geodetic reference system), possible changes of the depicted object (moving cars, new buildings and trees, etc) and radiometric differences due to variable weather conditions or different acquisition time. These outliers can be eliminated by applying the RANSAC algorithm based either on a 3D similarity transformation between the two 3D point sets, or a 2D affine transformation considering only their X and Y coordinates. Although the 3D similarity describes the exact underlying geometry, the simpler and faster 2D affine approximation gave equivalent results, eliminating most of the outliers. It should be noted here that there are cases with severe differences among the two orthoimages (mainly because of large scale and rotation differences) where matching may fail. In such a case, GPS information or a rough manual initialization can overcome this problem.

Since the identified 3D points will be used as ground control in a final bundle adjustment, they need to be detected on the aerial image set (Figure 2). An area of interest is defined by back-projecting every 3D point in the local photogrammetric system onto all aerial images in which it is visible. Then, a corresponding 2D image feature is sought by comparing image to orthophoto descriptors. Appropriate filtering, based on the matching score, the number of valid image observations and the space intersection error of corresponding rays, is applied to eliminate possible outliers. It should be also noted that points close to Lidar surface edges are discarded. In this way only points on planar surfaces are kept. Eventually, a bucketing procedure is applied in order to avoid unfavourable distribution of control points.

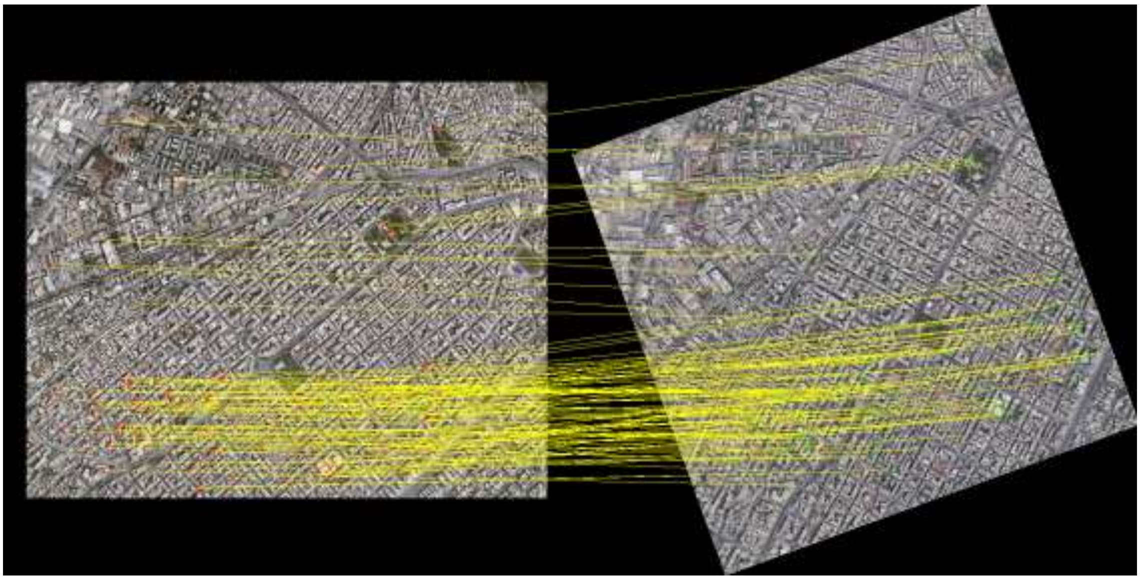

Figure 1. Inlier matches after RANSAC algorithm between the reference orthomosaic *r-ortho* (left) and the one generated from aerial images (*n-ortho*, right).

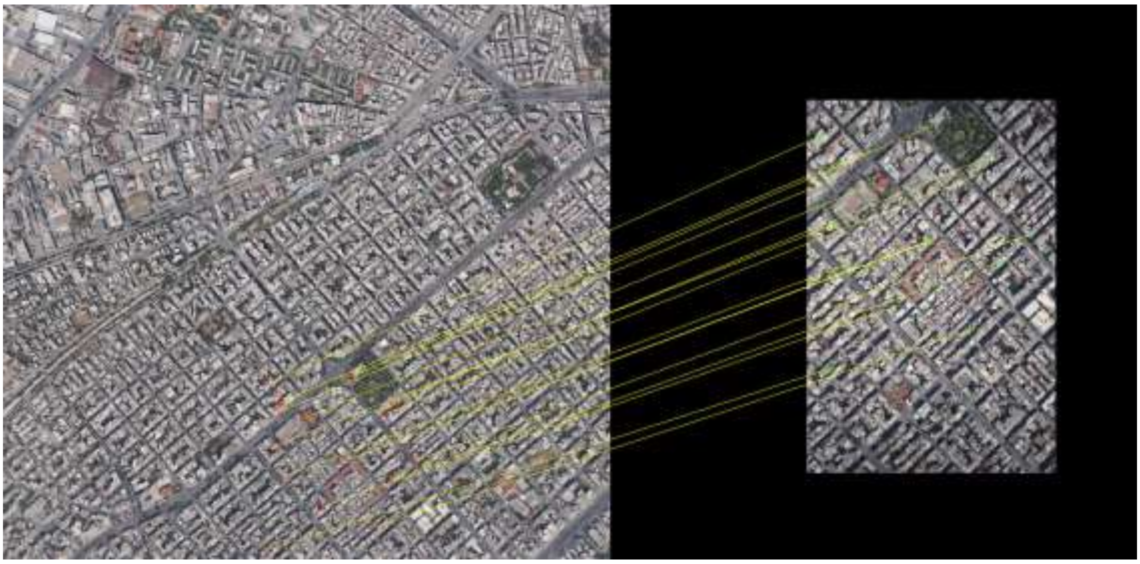

Figure 2. GCPs extracted automatically on the orthophoto (left) are identified on one of the aerial images (right).

Through the abovementioned procedure, points with known 3D coordinates on the Lidar reference system are identified on the available aerial images. These are introduced as ground control observations in a self-calibrating bundle adjustment solution weighted according to the vertical accuracy of the Lidar surface and the planimetric accuracy of the reference orthophoto. A final filtering based on the residual errors of the GCPs allows the elimination of any remaining outlier.

Compared to other approaches where the final imagery to Lidar registration is achieved by an optimal uniform 3D similarity transformation, the implementation proposed here allows the correction of small, non-uniform geometric deformations that are known to appear in the photogrammetric image network after SfM solutions without ground control information.

2.4 Final 3D model generation and texture mapping

At this stage, all available geospatial data, i.e. Lidar surface, reference orthomosaic and aerial imagery, are optimally registered. Due to the refined image orientations after the last bundle adjustment solution, the DSM that has been generated as an intermediate product needs to be updated. The combination of the updated surface with the Lidar data allows the construction of an optimal 3D city model with complementary surface information even on the building facades. This enhancement is supported by the exploitation of the oblique images present in the image network. Additionally, the oriented image set can be employed for texture mapping of the final 3D city model or even the initial Lidar point cloud/DSM.

3. RESULTS

In order to check the effectiveness of the proposed approach, a preliminary test was carried out on a dataset from the Kallithea neighborhood of Athens, Greece, which included the following:

i. An orthomosaic with spatial resolution of 20 cm generated from color aerial imagery. Image acquisition had taken place in 2007.

ii. Lidar-based DSM data with spatial resolution of 1m, acquired in 2003 (Figure 3 left).

iii. Vertical and oblique aerial imagery with average resolution of 15 cm acquired in 2012.

The first two data items had been already registered to the Greek Geodetic Reference System 1987 (GGRS87) and were used as reference for the registration of the aerial imagery. In a first step the aerial images were oriented automatically using a Structure from Motion approach (Section 2.1) in an arbitrary reference system initialized by GPS data. The standard error of this bundle adjustment solution was 0.5 pixel.

Next, a DSM (Figure 3 right) and an orthomosaic were generated, as outlined in Section 2.2, with resolution similar to that of the reference data.

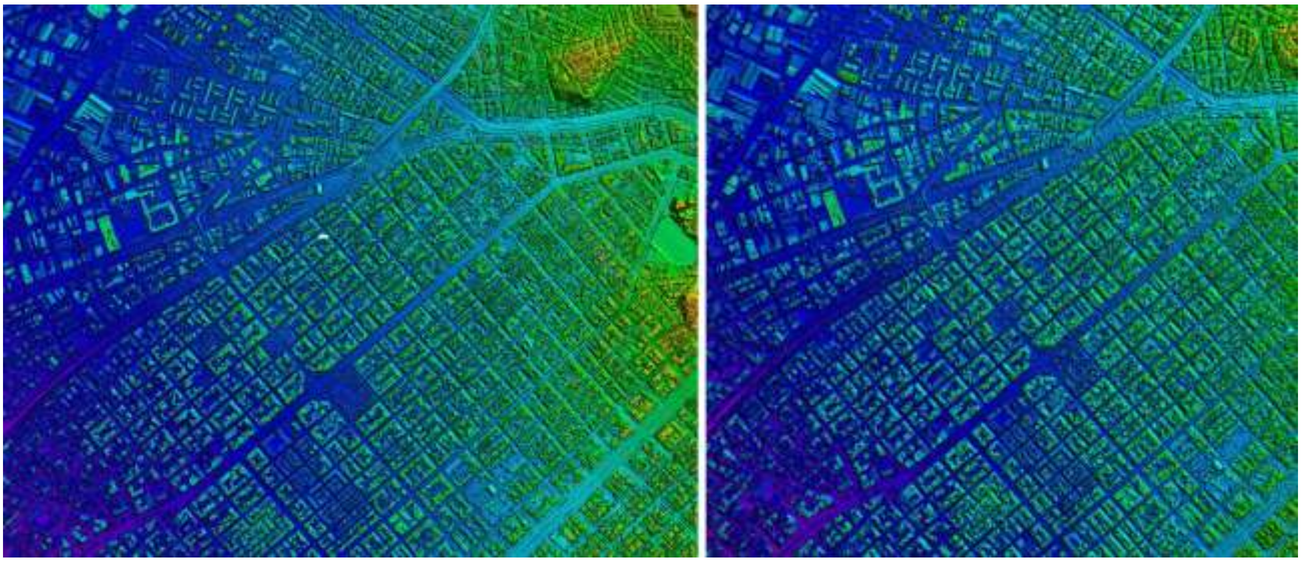

Figure 3. Lidar DSM (Left) and DSM generated from the aerial imagery (Right).

In a final step, the photogrammetric image block was aligned to the Lidar data and the reference orthomosaic through automatic extracted GCPs as presented in Section 2.3. The RMS error of the control points in the final bundle adjustment was around 40 cm, i.e. 2 times larger than the pixel size of the reference orthoimage and half the size of the Lidar DSM resolution. A separate solution with manually measured control points yielded an RMS of 20 cm. The introduction of these control points as check points in our solution gave an RMS error of 30 cm. These results are regarded as satisfactory, especially if one takes into consideration the different resolution of the three datasets, along with the significant difference of their acquisition time.

After the registration of the two datasets, the photogrammetric DSM generation was repeated in order to incorporate the refined orientations of the aerial images. The aerial imagery was also utilized for the texture mapping of the initial Lidar 3D model (Figure 4). This was performed by a multi-view algorithm that can compensate for different orientations, scale and resolution of the images involved, using an appropriately weighted blending scheme similar to the orthomosaic generation approach[22].

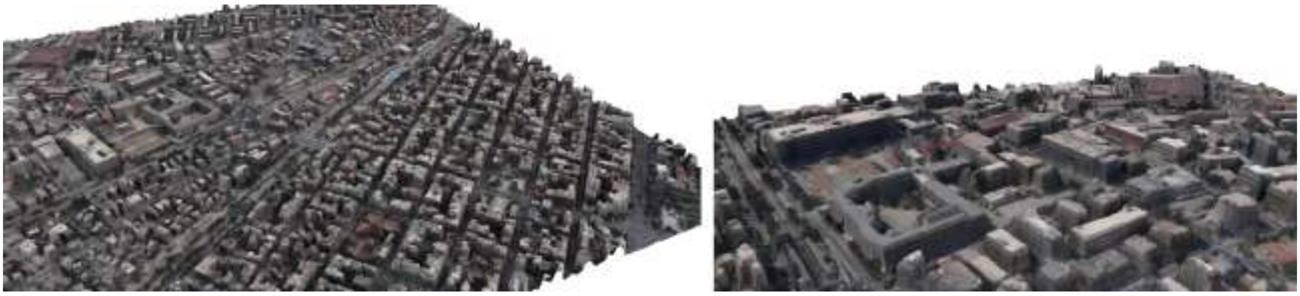

Figure 4. Texture mapping of the original Lidar surface from vertical and oblique aerial imagery.

4. CONCLUSIONS – FUTURE TASKS

The work presented here is part of an ongoing research project on 3D city models that focuses on multimodal representations of urban areas. A framework was proposed for the automatic orientation of aerial imagery and their co-registration with a reference Lidar surface and an RGB orthomosaic. First results have shown the effectiveness of the suggested approach. Therefore, the alignment of 3D data with digital imagery allows the generation of an enhanced and updated 3D surface representation as well as a detailed texture mapping of the final 3D city model.

Aspects for the improvement on the accuracy of the procedure could include (a) the integration of additional feature points such as Harris corners detected on planar surface areas, due to their more precise localization power compared to SIFT and SURF points, and (b) the introduction of a further constraint into the bundle adjustment solution that forces the elevations of the automatically extracted tie points to follow the Lidar surface.

ACKNOWLEDGEMENTS

This research has been co-funded by the European Union (European Social Fund) and Greek national resources under the framework of the Archimedes III: Funding of Research Groups in T.E.I. of Athens project of the Education & Lifelong Learning Operational Programme.

We would also like to thank GeoIntelligence S.A. for providing the DSM elevation data as well as the National Cadastre & Mapping Agency of Greece for supplying us with the high resolution aerial photographs.

REFERENCES

[1] Delara, Jr. R., Mitishita, E. A. and Habib, A., “Bundle Adjustment of Images From Non-Metric CCD Camera Using Lidar Data As Control Points,” International Archives of XXth ISPRS Congress, Commission III, 13–19, (2004).

[2] Habib, A., Ghanma, M., Morgan, M. and Al-Ruzouq, R., “Photogrammetric and Lidar data registration using linear features,” Photogramm. Eng. Remote Sensing 71 (6), 699–707, (2005).

[3] Böhm, J. and Becker, S., “Automatic marker-free registration of terrestrial laser scans using reflectance features,” 8th Conference on Optical 3D Measurement Techniques, (2007).

[4] Lowe, D., “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vision 60 (2), 91–110, (2004).

[5] Smith, E., King, B., Stewart, C. and Radke, R., “Registration of combined range-intensity scans: Initialization through verification,” Computer Vision and Image Understanding, 110(2):226–244, (2008).

[6] Liu, L. and Stamos, I., “A systematic approach for 2D-image to 3D-range registration in urban environments”, Computer Vision and Image Understanding. 116 (1), 25–37, (2012).

[7] Armenakis, C., Gao, Y. and Sohn, G., “Semi-Automatic Co-Registration of Photogrammetric and Lidar Data Using Buildings,” ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences I-3, 13–18, (2012).

[8] Besl, P. J. and McKay, N. D., “A Method for Registration of 3-D Shapes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 14, no. 2, pp. 239–256, (1992).

[9] Gneeniss, A. S., Mills J. P. and Miller, P. E., “In-Flight Photogrammetric Camera Calibration and Validation via Complementary Lidar,” ISPRS Journal of Photogrammetry and Remote Sensing 100, 3–13 (2014).

[10] Palenichka, R.M. and Zaremba, M.B., “Automatic extraction of control points for the registration of optical satellite and LiDAR images,” IEEE Trans. Geosci. Remote Sensing 48 (7), 2864–2879, (2010).

[11] Yang, B. and Chen, C., “Automatic registration of UAV-borne sequent images and Lidar data,” ISPRS J. Photogramm. Remote Sens., 101, 262–274, (2015).

[12] Hartley, R. and Zisserman, A., “Multiple View Geometry in Computer Vision,” Cambridge University Press, (2003).

[13] Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J. and Koch R., “Visual modeling with a hand-held camera,” International Journal of Computer Vision, 59(3), 207-232, (2004).

[14] Snavely, N., Seitz, S. and Szeliski, R., “Modeling the World from internet photo collections,” Int. J. Comput. Vision 80 (2), 189–210, (2008)

[15] Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., "SURF: Speeded Up Robust Features," Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346–359, (2008).

[16] Sturm, P., “On focal length calibration from two views,” IEEE Int. Conference on Computer Vision & Pattern Recognition, 145–150, (2001).

[17] Nistér, D., “An Efficient Solution to the Five-Point Relative Pose Problem,” Proc. Computer Vision and Pattern Recognition, 195-202, (2003).

[18] Hirschmüller, H., “Stereo processing by semiglobal matching and mutual information,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2), 328–341, (2008).

[19] Furukawa, Y. and Ponce, J., “Accurate, Dense, and Robust Multiview Stereopsis,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(8), 1362–1376, (2010).

[20] Jancosek, M. and Pajdla, T., “Multi-View Reconstruction Preserving Weakly-Supported Surfaces,” IEEE Conference on Computer Vision and Pattern Recognition, 3121-3128, (2011).

[21] Stentoumis, C., Grammatikopoulos, L., Kalisperakis, I. and Karras, G., “On accurate dense stereo-matching using a local adaptive multi-cost approach,” ISPRS Journal of Photogrammetry and Remote Sensing, Volume 91, May 2014, Pages 29-49, (2014).

[22] Karras, G., Grammatikopoulos, L., Kalisperakis, I. and Petsa, E., “Generation of orthoimages and perspective views with automatic visibility checking and texture blending,” Photogrammetric Engineering and Remote Sensing, 73(4), 403-411, (2007).