

ROBUST BUILDING-BASED REGISTRATION OF AIRBORNE LIDAR DATA AND OPTICAL IMAGERY ON URBAN SCENES

Thanh Huy Nguyen^{*†}, Sylvie Daniel^{*}, Didier Guériot[†], Jean-Marc Le Caillec[†] and Christophe Sintès[†]

^{*}Université Laval, Québec City, QC G1V 0A6, Canada

[†]IMT Atlantique, UMR 6285 LabSTICC, 29238 Brest Cedex 3, France

{thanh.nguyen, didier.gueriot, jm.lecaillec, christophe.sintes}@imt-atlantique.fr,
sylvie.daniel@scg.ulaval.ca

ABSTRACT

The motivation of this paper is to address the problems of registering airborne LiDAR data and optical aerial or satellite imagery, which are acquired from different platforms at different times, as well as having different points of view and reflecting different levels of detail of an observed scene. In this paper, we present a robust registration based on building regions, which are extracted from optical image using mean shift segmentation, and from LiDAR data by a 3D point cloud filtering process. The matching of the extracted building segments are then carried out by using Graph Transformation Matching (GTM) which allows to determine a common pattern of relative positions of segment centers. Thanks to this registration the relative shifts between datasets are significantly reduced, which enables the possibility towards an accurate building extraction and furthermore a high-quality data fusion between these datasets.

Index Terms— Airborne LiDAR, aerial imagery, satellite imagery, heterogeneous registration, urban scene, building regions, mean shift segmentation, graph transformation matching.

1. INTRODUCTION

Existing works in the domain of fusion of data from aerial or satellite imagery and airborne LiDAR over the years have addressed very specific acquisition contexts, in which, the respective image and LiDAR 3D point cloud are already registered and/or they are acquired from the same platform at identical or very close dates. For example, works participating in the 2013 GRSS Data Fusion Contest [1] focused on performing a fusion between LiDAR data and hyperspectral imagery that have the same spatial resolution and are acquired on two consecutive days. Within the same contest in 2015 [2], extremely high resolution LiDAR data and RGB imagery were collected from the same airplane, so that the sensors would be rigidly fixed to each other. In other words, the proposed solutions from these contests have never been intended to respond to the inherent obstacles of a very important acquisition context. This context involves the datasets which are collected from different platforms with different acquisition configuration (e.g. flying track, height, orientation, etc.) at different moments and even in different seasons, and have different spatial resolutions and levels of detail. The need for a relevant registration in such context is exemplified in the work undertaken by Cura et al. [3]. However, a solution that is versatile enough to satisfy this difficult context still remains an unsolved research problem.

Accurate registration of LiDAR data and optical imagery is the inevitable prerequisite for any data fusion applications that use them

[4]. The majority of existing automatic methods for registration of these datasets can be classified into two categories, namely intensity-based and feature-based methods. Feature-based methods establish correspondence between datasets based on distinguishable features that are available on them, derived by a feature extraction algorithm and a feature matching strategy [5]; whereas intensity-based methods determine the optimal camera pose by minimizing a statistical similarity (e.g. Mutual Information) between the values of image pixels and of LiDAR-derived elevation and intensity images [4, 6].

However, an intensity-based registration between LiDAR data and optical imagery requires many conditions to succeed, beside their drawback of high computational cost. Indeed, the datasets need to be spatially close to each other, and have the same resolution, as well as having similarities in intensity characteristics [4, 6]. In this paper, we present a feature-based approach of registration, as it is more interesting and promising for a fusion of data in the aforementioned extreme conditions of this research context. We also have been focusing particularly on urban scenes because they are advantageous with very high availability of geometric primitives associated with physical objects such as buildings, which can lead to an accurate registration [7]. However, the extraction of buildings is usually carried out either by using conjointly co-registered LiDAR data and imagery [8], or manually [9]. Therefore, a registration with separate automatic extractions of buildings from LiDAR data and imagery will be more suitable.

2. PROPOSED METHOD OF REGISTRATION

Fig. 1 illustrates the main steps of the proposed registration based on building segments. On each dataset we perform different processes with the purpose of extracting buildings from the observed urban scene. On one hand, we rely on 3D coordinates of points and an altitude thresholding to filter building points out of a LiDAR point cloud. On the other hand, mean shift segmentation is performed with a carefully chosen bandwidth parameter to segment the optical image, followed by a refinement to remove unwanted segments and preserves building-like segments. Also, this proposed registration will be tested on three different pairs of datasets, as described by Table 1.

2.1. Feature extraction

2.1.1. Building extraction from LiDAR data

The extraction of building regions from LiDAR point cloud is effectuated through a chain of processes, as follows,

No.	Data type	Spectral resolution	Spatial resolution	Acquisition time (season)	Geometry/Properties	Estimated relative shift
1	Aerial optical imagery	8 bits (RGBI)	15 cm	June 2016 (summer)	• Orthorectified • Georeferenced	1 - 2 m
	LiDAR	8 bits (Intensity)	8 points/m ²	May-Jun 2017 (summer)	Classified	
2	Aerial optical imagery	8 bits (RGBI)	15 cm	Jul-Aug 2013 (summer)	• Central perspective • No georeferencing	2.5 - 10 m
	LiDAR	8 bits (Intensity)	2 points/m ²	Oct-Nov 2011 (winter)	Classified	
3	Satellite imagery	Panchromatic	50 cm	July 2015 (summer)	• No orthorectification • Georeferenced	25 - 40 m
		Multispectral (4 bands)	2 m			
	LiDAR	8 bits (Intensity)	2 points/m ²	Oct-Nov 2011 (winter)	Classified	

Table 1: Details of LiDAR datasets (© Ville de Québec), aerial imagery datasets (© Communauté Métropolitaine de Québec), and satellite imagery dataset (© Centre National d’Études Spatiales, France).

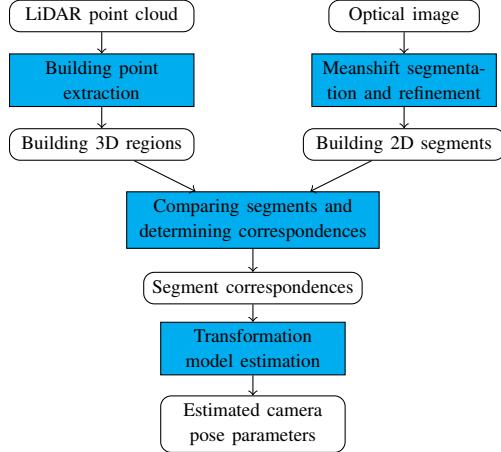


Fig. 1: Flowchart of the building segment-based registration between optical image and LiDAR point cloud.

Input | LiDAR 3D point cloud (X, Y, Z).

Step 1 | Altitude thresholding: separating non-ground points from ground points depending on their altitude value. This process is proposed by many existing works as an initial but necessary step [10]. The altitude threshold value is calculated as follows, $T_a = \text{mean}(z_G) + \max\{2.5, \text{std}(z_G)\}$, where z_G denotes the altitude of ground points.

Step 2 | Vertical projection: we then project all non-ground points vertically onto the plan $z = 0$, and create a 2D binary mask of non-ground points. The resolution of this binary mask is set accordingly to the point density of the input LiDAR point cloud to avoid null-value pixels, e.g. a resolution of 1 meter \times 1 meter for a point cloud of a density of 2 points/m².

Step 3 | Morphological opening is then applied on the binary mask to remove small regions as well as rounding up bigger ones. The morphological structuring element of choice is a diamond shape of a size of 5 to 7 pixels.

Step 4 | Connectivity labeling: connecting pixels into segments based on their connectivity, and then labeling these segments.

Step 5 | Then, a removal of small regions that are smaller than 20 square meters is effectuated, which results a labeled building mask.

Step 6 | Extracting building points: based on the labeled building mask, we go back to the non-ground points, and among these points we select only the regions that are seeded by labeled segments.

Output | Building 3D regions and their boundary.

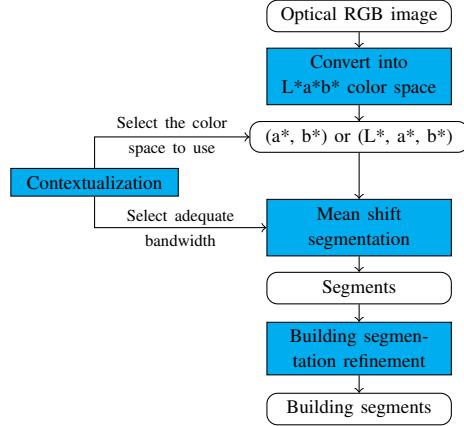


Fig. 2: Flowchart of the building segmentation from optical image.

2.1.2. Building segmentation from optical image using Mean shift

Mean shift is an unsupervised clustering method widely used in many areas of Computer Vision, including 2D shape extraction, and texture segmentation [11]. Compared to k -means clustering, mean shift doesn't require a prior number of classes, but a value of bandwidth corresponding to image color range and size of objects to be segmented. Moreover, in an urban area, a k -means fails to segment buildings because color of building roofs can vary a lot, and also building roofs and streets may have similar color sometimes.

Fig. 2 presents a flowchart of the building segmentation on optical image using mean shift algorithm. First, input data for the mean shift algorithm are converted into the CIE L*a*b* color space, as this color space allows better distinction of objects from images than RGB color space. In case of the satellite imagery, a pansharpening is carried out to merge 50-cm resolution panchromatic and 2-m resolution multispectral imagery to create a 50-cm color image, which will be segmented by mean shift. Also, determining the best bandwidth parameter for mean shift still remains difficult even though a number of approaches have been explored [12]. Thus, this bandwidth parameter should also be set manually according to the type of area (either residential, industrial, mixed, etc.), and the size of objects of interest, or, in other words, based on the contextualization of this scene, alongside with the choice between using either (a*, b*) values or (L*, a*, b*).

By virtue of the mean shift segmentation, many building regions have been segmented alongside with other regions related to trees, streets, or cars. Obviously, these unwanted non-building segments need to be removed before we compare them with the building segments extracted from LiDAR point cloud. To do that, first we can

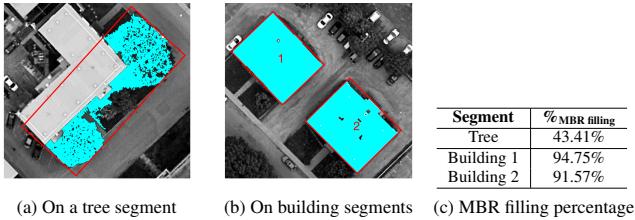


Fig. 3: Illustration of MBR (red rectangles) of a tree segment versus building segments (illustrated by cyan points).

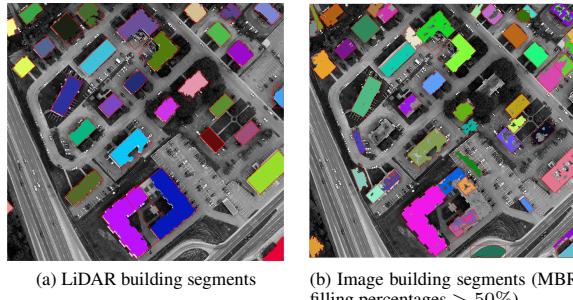


Fig. 4: Extracted building segments from optical image segmentation and from LiDAR point cloud.

rely on the number of pixels inside each segment to remove small segments as they usually correspond to trees and cars, as well as segments that are too large corresponding street regions. This filter is simple and efficient [13], but completely dependent on image resolution. Therefore, it needs a manual intervention to be set correctly. The authors of [13] also proposed two additional filters based on the length ratio of the segment major and minor axis, and the segment eccentricity to remove “false-detected” building segments and keep the segments that associate to rectangular and round building regions. However, they are not effective in the case of complex building segments. Besides, it is not clear how they determined the axes and the eccentricity of segments, and also, the thresholds used by these filters are highly subjective. In this paper, we present another approach to discriminate buildings apart from regions that relate to trees or streets. After effectuating the preliminary filter based on the number of pixels of segments, we identify the minimal bounding rectangle (MBR) of each segment. Based on this rectangle, we can calculate filling percentage of each segment $\%_{\text{MBR filling}} = \text{Area}(\text{segment})/\text{Area}(\text{MBR}) \times 100$, which can be evaluated to filter the unwanted segments, as filling percentage of a rectangle building segment should be higher than that of an unwanted segment, as compared on Fig. 3.

2.2. Feature comparison and matching

After extracting building segments from the datasets, the next step is to compare and match them. From the optical image, we select the segments that have the MBR filling percentages more than 50%. On the other hand, all building regions extracted from the LiDAR point cloud will be taken into consideration. Both sets of extracted segments are depicted on Fig. 4.

However, the facts that our datasets are relatively distant to each other (cf. Table 1), as well as a number of outliers can be anticipated (i.e. falsely detected segments may still exist after the MBR-based segment refinement) are also problematic for the segment compar-

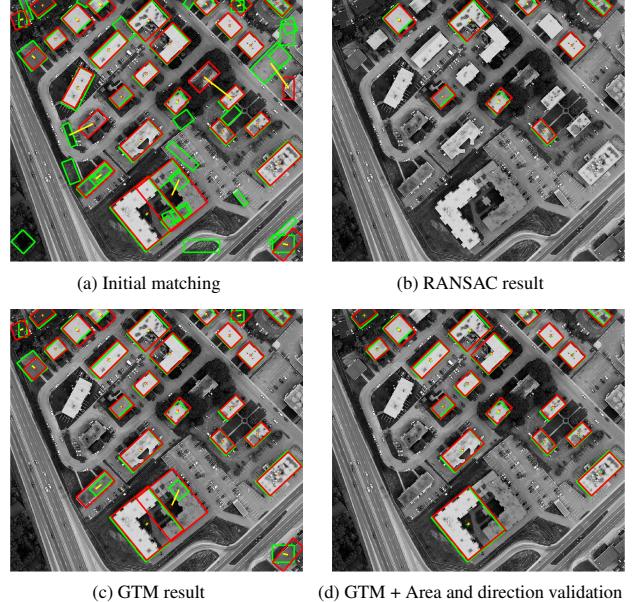


Fig. 5: Considering relative position of segment centers (green and red rectangles represent MBR of segments extracted from optical image and LiDAR point cloud).

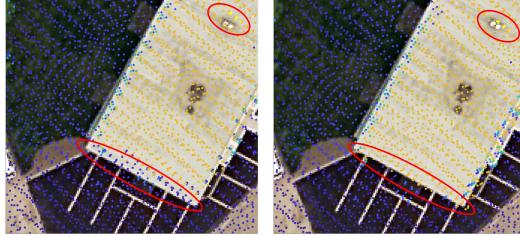
ison and matching step. Therefore, a matching of segments based on their spatial relation w.r.t. their neighbors is more relevant than comparing their individual values. Indeed, taking into account the position of segment center, a common pattern representing their relative spatial arrangement between datasets can be determined using GTM algorithm. GTM is a graph-based point matching algorithm designed for solving the registration between images with non-rigid deformations. This algorithm performs better than RANSAC in removing outliers on test image datasets [14], as well as in our work. In practice, both of them requires an initial one-to-one matching of segment centers, which can be carried out relying on the positions of vertically projected 3D building region centers onto plan $z = 0$ and the centers of 2D segments (extracted by mean shift segmentation). In case that the relative shifts are too big (e.g. dataset no. 3), this initial matching is added with a translation vector calculated based on the displacement of the largest segment. The result of the initial matching of segment centers, and then of RANSAC are shown on Fig. 5a and 5b; whereas Fig. 5c depicts GTM results, as well as the result after a refinement of false positives from GTM result based on the area value and the direction of segments on Fig. 5d.

2.3. Transformation model estimation

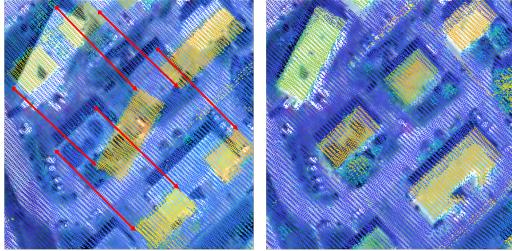
The coordinates of the matched segment centers are then used to estimate the transformation model composed of parameters of the imaging camera pose. They are exterior orientation parameters, which are the position (X_0, Y_0, Z_0) and orientation (ω, ϕ, κ) of the camera when the image was acquired. In this paper, this estimation is carried out using Gold Standard algorithm detailed in [15, p.187].

3. EXPERIMENTAL RESULTS

Table 2 summarizes the result of building extraction and matching on the selected region (Fig. 5), through the number of hits (good extraction of buildings), false alarms (wrongly extracted), and misses



(a) Overlapping of the back-projected LiDAR point cloud on the respective orthorectified optical image (dataset no. 1)



(b) Overlapping of the back-projected LiDAR point cloud on the respective satellite pansharpened image (dataset no. 3). Red arrows indicate displacements of some building corners for better understanding.

Fig. 6: Back-projected LiDAR point cloud overlapping on the respective image, before (left) and after registration (right).

(buildings exist but not extracted).

	Extracted from LiDAR data	Extracted from image by mean shift	Matching result by RANSAC	Matching result by GTM
H/FA/M	28/0/0	24/21/4	8/0/12	19/7/1
Precision	100%	53.33%	100%	73.08%
Recall	100%	85.71%	40%	95%

Table 2: Performance of building extraction and matching on a selected region (28 buildings in total).

Furthermore, the overall results of registration are presented through a reduction of relative shift between datasets at several manually selected control points by virtue of the registration summarized by Table 3, as well as through a visual assessment between datasets by overlapping them before and after the registration on Fig. 6. Full-scale color figures of this paper can be found on <https://github.com/nthuy190991/igarss2019>.

Dataset	Average estimated relative shift	
	Before	After
1	1.41 m	0.49 m
2	2.83 m	1.32 m
3	40.81 m	1.75 m

Table 3: Average relative shift before and after the registration.

4. CONCLUSIONS AND PERSPECTIVES

In this paper, we present a robust approach of registration between airborne LiDAR data and optical imagery in a difficult situation that they are not acquired from the same platform, not having the same point of view or the same spatial resolution. This approach centralizing on extracting and matching of building regions, allows reducing very significantly relative shifts between datasets, i.e. reducing more than 50% of displacements between LiDAR data and aerial imagery and 95.71% of displacements between LiDAR data and satellite imagery, as well as yielding better alignment when overlapping

the back-projected LiDAR point cloud on the optical image. This enables henceforth a fine registration between these datasets, which is necessary to align them at accuracy of 1-pixel level, in order to fully benefit the advantages of using them [4]. Furthermore, based on these results, we aim at a data fusion which yields a product that has a precision as high as the input data, but with a better completeness and a reduced uncertainty of the observed scene.

5. REFERENCES

- [1] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pižurica, S. Gautama, *et al.*, “Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2405–2418, 2014.
- [2] A.-V. Vo, L. Truong-Hong, D. Laefer, D. Tiede, S. d’Oleire Oltmanns, A. Baraldi, M. Shimoni, G. Moser, and D. Tuia, “Processing of extremely high resolution lidar and rgb data: outcome of the 2015 ieee grss data fusion contest—part b: 3-d contest,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 12, pp. 5560–5575, 2016.
- [3] R. Cura, J. Perret, and N. Paparoditis, “A scalable and multi-purpose point cloud server (pcs) for easier and faster point cloud data management and processing,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 127, pp. 39–56, 2017.
- [4] E. G. Parmehr, C. S. Fraser, C. Zhang, and J. Leach, “Automatic registration of optical imagery with 3D LiDAR data using statistical similarity,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 28–40, 2014.
- [5] R. M. Palenichka and M. B. Zaremba, “Automatic extraction of control points for the registration of optical satellite and lidar images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2864–2879, 2010.
- [6] A. Mastin, J. Kepner, and J. Fisher, “Automatic registration of LIDAR and optical images of urban scenes,” *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, pp. 2639–2646, 2009.
- [7] P. Rönnholm and H. Haggrén, “Registration of Laser Scanning Point Clouds and Aerial Images Using Either Artificial or Natural Tie Features,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, no. September, pp. 63–68, 2012.
- [8] M. Awrangjeb, M. Ravanbakhsh, and C. S. Fraser, “Automatic detection of residential buildings using lidar data and multispectral imagery,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 5, pp. 457–467, 2010.
- [9] M. Brell, C. Rogass, K. Segl, B. Bookhagen, and L. Guanter, “Improving sensor fusion: A parametric method for the geometric coalignment of airborne hyperspectral and lidar data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3460–3474, 2016.
- [10] M. Awrangjeb, C. Zhang, and C. S. Fraser, “Automatic extraction of building roofs using lidar data and multispectral imagery,” *ISPRS journal of photogrammetry and remote sensing*, vol. 83, pp. 1–18, 2013.
- [11] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [12] J. E. Chacón and P. Monfort, “A comparison of bandwidth selectors for mean shift clustering,” *arXiv preprint arXiv:1310.7855*, 2013.
- [13] N. L. Gavankar and S. K. Ghosh, “Automatic building footprint extraction from high-resolution satellite image using mathematical morphology,” *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 182–193, 2018.
- [14] W. Aguilar, Y. Frauel, F. Escalano, M. E. Martinez-Perez, A. Espinosa-Romero, and M. A. Lozano, “A robust graph transformation matching for non-rigid registration,” *Image and Vision Computing*, vol. 27, no. 7, pp. 897–910, 2009.
- [15] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.