Coarse-to-Fine Registration of Airborne LiDAR Data and Optical Imagery on Urban Scenes

Thanh Huy Nguyen [ID] , *Student Member, IEEE*, Sylvie Daniel [ID] , *Senior Member, IEEE*, Didier Guériot, *Senior Member, IEEE*, Christophe Sintès, *Senior Member, IEEE*, and Jean-Marc Le Caillec [ID] , *Senior Member, IEEE*

Abstract—Applications based on synergistic integration of optical imagery and LiDAR data are receiving a growing interest from the remote sensing community. However, a misaligned integration of these datasets fails to fully profit from the potential of both sensors. An optimum fusion of optical imagery and LiDAR data requires an accurate registration. This is a complex problem since a versatile solution is still missing, especially when data are collected at different times, from different platforms, under different acquisition configurations. This article presents a coarse-to-fine registration method of optical imagery with airborne LiDAR data acquired in such context. First, a coarse registration involves processes of extraction and matching of building candidates from the two datasets. Then, a mutual-information-based fine registration is carried out. It involves a superresolution approach applied to LiDAR data to generate images with the same resolution as the optical image, and a local approach of transformation model estimation. The proposed method succeeds at overcoming the challenges associated with this difficult context. For instance, considering the experimented airborne LiDAR (2011) and orthorectified aerial imagery (2016) datasets, their spatial shift is reduced by 48.15% after the proposed coarse registration. Moreover, the incompatibility of size and spatial resolution is well addressed by the superresolution. Finally, a high accuracy of dataset alignment is also achieved, highlighted by a 40-cm error based on a check-point assessment and a 64-cm error based on a check-pair-line assessment. These promising results enable further researches for a complete fusion methodology between these datasets in this challenging context.

Index Terms—Aerial imagery, airborne LiDAR, building extraction, coarse-to-fine, heterogeneous sensors, mutual information (MI), optical imagery, registration, satellite imagery, superresolution (SR), urban scenes.

I. INTRODUCTION

THE PERCEPTION of an environment on the Earth's surface and follow-up exploitations require using multiple

Thanh Huy Nguyen is with the Department of Geomatics, Université Laval, Québec, QC G1V 0A6, Canada, and also with the IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France (e-mail: thanh.nguyen@imt-atlantique.fr).

Sylvie Daniel is with the Department of Geomatics, Université Laval, Québec, QC G1V 0A6, Canada (e-mail: sylvie.daniel@scg.ulaval.ca).

Didier Guériot, Christophe Sintès, and Jean-Marc Le Caillec are with the IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France (e-mail: didier.gueriot@imt-atlantique.fr; christophe.sintes@imt-atlantique.fr; jm.lecaillec@imt-atlantique.fr).

sensors to capture specific and complementary characteristics of this environment [1]. In many areas of remote sensing, observations from heterogeneous sources are coupled and jointly analyzed to achieve a richer description of a scene. This approach allows to mutually benefit from their strengths, as well as reducing the data uncertainty and incompleteness relating to each sensor [2]–[4]. As a matter of fact, the fusion of multisource data has become one of the mainstream research topics in the remote sensing community nowadays [1], [5].

Light detection and ranging (LiDAR) and photogrammetry systems are major sources for fast and reliable spatial data acquisition. They provide data that are complementary to each other while the two systems differ fundamentally in their operation and data collection principles. The first one is an active sensor while the second is passive. On the one hand, airborne LiDAR systems are widely used for providing accurate three-dimensional (3-D) surface information and 3-D geometry of objects and ground elements, in the modality of scattered point clouds (recorded according to range detection principle). On the other hand, aerial and satellite photogrammetry supplies rich semantic and texture information, in the form of multispectral images. By integrating the two technologies, many applications have been enabled such as building extraction [3], [6], city digital twin construction [7], land use and land cover classification [8], and so on [9], [10].

A. Motivation

Over the years, existing works in the domain of data fusion between optical imagery and airborne LiDAR data have addressed dedicated acquisition contexts, in which the respective image and the LiDAR point cloud are already registered and/or they are acquired from the same platform at identical or very close dates. For instance, solutions submitted to the 2013 Data Fusion Contest of the IEEE Geoscience and Remote Sensing Society [11] focused on the fusion between LiDAR data and hyperspectral imagery with the same spatial resolution, acquired on two consecutive days. The same contest in 2015 [12], [13] involved extremely high resolution LiDAR data and RGB imagery collected from the same aircraft with the sensors being rigidly fixed to the same platform. In other words, the solutions submitted to these contests, as well as many others [14]–[16], have not intended to cope with the inherent obstacles of the context where datasets are collected from different platforms with different acquisition configuration (i.e., different flying

TABLE I
SENSOR AND PLATFORM SPECIFICATIONS

	Optical imagery		LiDAR data	
	Aerial image (2016)	Satellite image (2015)	Airborne LiDAR (2017)	Airborne LiDAR (2011)
Principle	Passive	Passive	Active	Active
Device (Camera/LiDAR) - Platform	Vexcel UltraCAM Xp - Piper Navajo	Pléiades HR	Optech ALTM Galaxy - Piper Aztec	Optech ALTM Gemini - Piper Navajo
Sensor design	Time-Delay Integration Camera	Pushbroom	Whiskbroom	Whiskbroom
Acquisition dates (season)	**June 2016 (summer)**	**June 2015 (summer)**	**May-June 2017 (summer)**	**Oct.-Nov. 2011 (winter)**
Flying height	**2955 m**	**695 km**	**1300 m**	**950 m**
Swath width	**2597 m (cross-track) 1697 m (along-track)**	**20 km**	**946 m**	**620 m**
Field-Of-View (FOV)	**55° (cross-track) 37° (along-track)**		**20°**	**20°**
Instantaneous Field-Of-View (IFOV)	0.17 mrad (cross-track) 0.17 mrad (along-track)	1 µrad (PAN) 4 µrad (MS)	-	-
Laser repetition rate	-	-	350 kHz	100 kHz
Scan frequency	-	-	70 Hz	50 Hz
Laser beam divergence	-	-	0.25 mrad	0.25 & 0.8 mrad
Spectral bands/ Laser wavelength	R, G, B, NIR	470-830 nm (PAN) R, G, B, NIR (MS)	1064 nm	1064 nm
Number of returns per pulse	-	-	4	4
Point classification	-	-	U, G, LV, MV, HV, B[1]	U, G, LV, MV, HV
Point spacing (point density)	**-**	**-**	**35.4 cm (8 points/m^2)**	**70 cm (2 points/m^2)**
Ground sample distance	**15 cm**	**50 cm (PAN), 2 m (MS)**	**-**	**-**
Laser beam footprint size	-	-	33 cm	23 cm
Lateral overlapping	35%	-	65 %	30%
Longitudinal overlapping	60%	-	-	-
Theoretical horizontal accuracy (RMSE)	13-16.5 cm (at perspective centers), 4-5.5 cm (at control points) after Aerotriangulation	1 m (with ground control points) and 3 m (without ground control points)	1/7500×Altitude, i.e. ≈17 cm	1/5500×Altitude, i.e. ≈17 cm
Theoretical elevation accuracy (RMSE)	-	-	3-20 cm	5-35 cm

The boldface rows highlight the differences between datasets notably concentrated in this article.
[1] Classification of LiDAR point cloud: unclassified (U), ground (G), low vegetation (LV), medium vegetation (MV), high vegetation (HV), and building (B).

track, height, orientation, and so on) at different moments and even in different seasons, with different spatial resolutions and levels of detail.

This article aims to propose a relevant registration method in this unresolved context. Table I summarizes the specifications of the sensors and their platforms considered in this work. The need for a relevant registration in such a crucial context is exemplified in the work undertaken by Cura *et al.* [17]. It also relates to the rise of the availability of data captured by different heterogeneous sensors that requires an efficient integration [5]. However, a solution that is versatile enough to overcome this difficult context still remains an unsolved research problem [18].

B. Challenges

The development of a relevant registration approach in this unresolved context faces many challenges.

1) Spatial Shift Between Datasets: The first challenge relates to the differences between the dataset point of view and field of view, which lead to a significant spatial shift between them. For instance, a spatial shift exists approximately 1-2 m between the orthorectified airborne image (2016) and the LiDAR data (2011), or up to 40 m between the Pléiades image and the LiDAR data (2011). According to our literature review, a coarse registration, which is necessary to reposition the two datasets,

has not been rigorously studied by existing works. This step is often inadvertently bypassed using the dataset geospatial coordinates provided by a GPS/IMU system [7], [16], [19].

2) Uncertainty, Imprecision, and Incompleteness: Distortions in the information extracted from optical images can be caused by radiometric errors like sensor sensibility, illumination changes, atmospheric effects, and geometric errors such as relief displacement, occlusions, or shadows [20]. On the other hand, points may be missing in the LiDAR data due to occlusion or presence of water [21]. These errors, distortions and missing data from each of the two datasets induce incompleteness, imprecisions, and uncertainties within the registration and fusion processes of these data [22].

3) Spatial Resolution and Level of Detail: There are significant differences in spatial resolution and level of detail between the airborne or satellite imagery and LiDAR data. For example, as highlighted in Table I, the considered LiDAR datasets in 2011 and 2017, respectively, have a point spacing 70 cm and 35.4 cm. On the other hand, the aerial image (2016) has a ground sampling distance of 15 cm, whereas that of the Pléiades panchromatic (PAN) and multispectral (MS) images are 50 cm and 2 m, respectively. Such differences affect the appearance of the same scene elements to be different on the two datasets, making it difficult to determine and extract the corresponding features between them [23]. This issue has not always been addressed by existing multisource registration works. For instance, Li *et al.*

[24] and [25] proposed a registration framework between LiDAR image and optical image in a context where the datasets always have the same spatial resolution. Among the proposed solutions to overcome the spatial resolution and level of detail differences between the datasets, some involve a multiresolution approach [9], [26] or a resampling step [20], [27].

4) Relevance of Registration Features: The nature of a scene, either in urban or natural environment, conditions strongly the entities within the datasets that would be relevant to perform registration [28].

5) Accuracy of Dataset Registration: When performing the fusion of airborne LiDAR data and optical imagery, even a small misalignment between them can lead to an unfavorable impact on the quality of the integrated product, or a significant reduction of data information content [29]. Thus, an accuracy level of 1-pixel is recommended for the dataset registration [15]. As a matter of fact, a *subpixel level* of accuracy, assessed by measuring the distances between control points, is usually preferred for a *good* registration. However, such a qualitative criterion is difficult to achieve because the image pixel resolution can vary from several dozens of centimeters to several meters depending on the platform (i.e., airborne versus satellite). Current works in the literature involve resulting discrepancies between the registered datasets ranging from 45 to 50 cm [14], [30]. They state that such discrepancies are a decent and desirable registration accuracy.

C. Contribution

This article addresses the need for a versatile and relevant registration approach able to overcome the aforementioned challenges. The versatility of our proposed method is reflected through its capability of registering the datasets that are not acquired simultaneously, nor from the same platform and same acquisition configuration, nor having same spatial resolution. These assumptions are crucial to the existing works [14]–[16], [24], [25]. It should be noted that the proposed method does not aim to address every scene possible, as we focus on a registration on urban scenes. In this regard, we propose a coarse-to-fine registration approach.

1) First, a coarse registration is performed to reposition the datasets closer to each other. It addresses the challenge of spatial shifts between datasets which is problematic but usually overlooked [7], [16], [19], [30]. In this article, we present a coarse registration relying on the primitives that are buildings.

2) Second, a fine registration is carried out based on a local transformation model estimation. It is enabled by a superresolution (SR) approach applied to LiDAR data in order to generate images with the same resolution as the optical image. This approach is devoted to overcome the hindering caused by the spatial resolution difference between datasets.

Such a coarse-to-fine approach is necessary in order to register an airborne LiDAR dataset with an optical image. The mentioned coarse registration aims to reposition the two datasets in a fast but reliable manner. As a result, a global transformation model, composed of a set of coarsely estimated camera pose parameters, is determined. Even though the global transformation does not permit the dataset to be precisely registered, it narrows down the search space for optimal camera pose parameters from an initial set of values during the fine registration. However, the main drawback of this feature-based coarse registration is that the building primitives are not distributed evenly throughout the datasets. Hence, the global transformation has the tendency to prioritize a region exhibiting more primitives than others. Therefore, we propose a subsequent fine registration that focuses on determining the optimal parameters for each local region of the considered urban area. Such a local approach brings two benefits, namely a higher registration accuracy and a reduced computational cost of this fine registration. Then, we also propose a refinement of locally optimized transformation models, in order to avoid conflicts between them. Finally, the proposed method relies on tailored series of well-known processes and algorithms while avoiding complicated and labor-intensive processes.

The remainder of this article is organized as follows. A brief review of existing works related to the registration of optical imagery and airborne LiDAR data is provided in Section II. Then, Section III presents the proposed methodological approach, consisting of two parts: coarse registration, then fine registration. Then, multiple quantitative assessments involving different datasets are presented and discussed in Section IV. Finally, Section V concludes this article.

II. LITERATURE REVIEW

Accurate registration of LiDAR data and optical imagery is the crucial prerequisite to any data fusion applications using them [14]. The majority of automatic methods for registering such datasets can be classified into two categories, namely area-based and feature-based methods. On the one hand, area-based methods determine the optimal pose of the camera by maximizing a statistical similarity, e.g., mutual information (MI), between the values of optical image pixels and LiDAR-derived image pixels [7], [15], [26]. The LiDAR-derived image is either a digital surface model (DSM), an intensity image, or an image of *pdet* (probability of detection) attributes derived from the LiDAR point cloud [7]. Their main drawbacks, in addition to the high computational cost, are the necessities for the datasets to be spatially close to each other, as well as to have the same resolution and display similar intensity characteristics. For instance, the similarity of characteristics between 2-D images and normals to a 3-D surface has been shown to be of paramount importance for area-based registration methods [31].

On the other hand, feature-based methods establish correspondence between the datasets based on available distinguishable features. They involve feature extraction algorithms and feature matching strategy [32]–[35]. The employed features can be either from built environment, such as corner points, break lines, and planar surfaces found in man-made objects, or natural features like trees, bushes, and ground surface features. In general, features from built environment usually yield higher registration accuracy result than natural features [32].

Wong and Orchard [34] proposed a registration method between LiDAR data and optical image, assuming that they are two images of the same resolution. From the LiDAR data, it is an image of laser return intensity data. This method consists in using a modified Harris corner detector to extract control points from the two images. Then, an exhaustive search for correspondences, accelerated by the Fast Fourier Transform, among all extracted control points is carried out. However, this method fails to produce accurate registration result in the case of very high resolution images [28]. Palenichka and Zaremba [35] proposed a registration method between LiDAR-derived DSM and optical imagery. It involves an automatic extraction of salient points from both the DSM and the optical image that allows the discrimination of the objects of interest from the background. This method facilitates the automatic selection of control points that also works on natural scenes. According to [28], the high computational cost and the lack of concern for the relief displacement are the drawbacks of this method. Liu *et al.* [30] proposed a registration method between airborne LiDAR data and UAV (unmanned aerial vehicle) remote sensing imagery, based on 3-D and 2-D line segments extracted, respectively, from the LiDAR point cloud and the image. For each 3-D line segment, a number of 2-D line segments are extracted from the same location on the image. Then, a manual selection is carried out to yield the correspondences (i.e., the conjugated line segments). Such a manual approach is prone to human bias. Also, this method does not account for the potential spatial shift between the datasets. Therefore, it could work on the datasets with a small spatial shift, but fails for large spatial shifts.

Many studies have proposed to use different features to increase the registration accuracy. For example, Huang *et al.* [36] proposed a registration method using two features at two scales, i.e., a line network of roads extracted using k-means clustering at the first scale, and building corners at the finer scale. However, the use of k-means clustering as an unsupervised classification on aerial images is seemingly too simple to extract roads effectively. Ding *et al.* [33] performed a coarse-to-fine approach to register oblique aerial image and LiDAR data based on vanishing points estimated from parallel vertical building edges at the coarse level, and then based on building corners at the fine level. While the vertical vanishing points can be estimated using oblique images, this can hardly be done using vertical aerial and satellite images, as well as orthorectified images. A similar coarse-to-fine approach is also proposed by Brell *et al.* [14] to register hyperspectral image and LiDAR data simultaneously acquired from the same aircraft. First, scale-invariant feature transform [37] keypoint detector is used to determine tie points between the LiDAR data and the hyperspectral image. Then, an area-based optimization is carried out to find optimal camera pose parameters. Within a small range from the values coarsely estimated using the tie points, these parameters are then refined based on the minimization of a cost function. Such cost function is the zero-mean sum squared distances calculated between the pixels of the hyperspectral image and the image generated from LiDAR intensity data using a ray-tracing module. However, this method does not address the registration between the datasets acquired separately, in which the spatial shift between the two

datasets can be problematic to the tie-point-based registration. Also, there is a potential issue due to the spatial resolution of the LiDAR data for generating a suitable image for the area-based optimization. This issue was not addressed in their work.

In conclusion, all the methods reviewed in this section either assume that the airborne LiDAR data and the optical imagery data are spatially close to each other, have been recorded simultaneously (or on very close dates), and/or have similar spatial resolution and level of detail. These constraints have been previously discussed (see Section I-B) to be challenging to a registration method in the considered context. To the best of our knowledge, a method explicitly devoted to the registration of LiDAR and image datasets acquired from two different platforms, with different configurations at different times and even seasons, has not yet been proposed. In what follows, we present how our method is able to achieve such purposes.

III. PROPOSED METHOD

Fig. 1 presents the full flowchart of the proposed method. First, the coarse registration approach is presented. It aims to reposition the two datasets based on the extraction and matching of building candidates. Based on these primitives, a global transformation model is estimated, which is represented by a set of camera pose parameters, denoted by θ_{global}. Second, a fine registration based on an SR of LiDAR values and an area-based optimization is carried out. Such SR process takes into account a transformation model (i.e., θ_{global} at the first iteration) and generates high-resolution LiDAR-based images. Next, a statistical similarity measure, namely MI or NCMI, between these super-resolved images and the optical image is estimated. Thus, the estimated MI (or NCMI) value can be considered as a function of the transformation model. The maximum value of such measures is expected to be achieved when the involved images (i.e., the optical image and the super-resolved LiDAR-based images) are geometrically aligned [15]. As a result, an optimal transformation model associated to this maximum MI (or NCMI) value is determined. We describe the two registrations in the two following sections.

A. Coarse Registration

Fig. 2 sums up the proposed coarse registration, which has been originally introduced in our previous work [38]. Man-made structures in urban scenes like buildings are more suitable for accurate registration, compared to natural features [32]. In addition, they remain unchanged through a relatively long period of time (e.g., several years). However, in airborne LiDAR datasets, the point density around vertical surfaces like building facades can be low. Hence, the localization accuracy of features like building corners and edges is deficient. Therefore, our coarse registration method relies on region-based primitives namely buildings.

Different series of processing steps are carried out on the LiDAR and optical image datasets respectively in order to extract buildings. On the one hand, we apply a series of processing steps starting with an elevation thresholding on LiDAR point cloud. On the other hand, mean shift segmentation [39] is performed
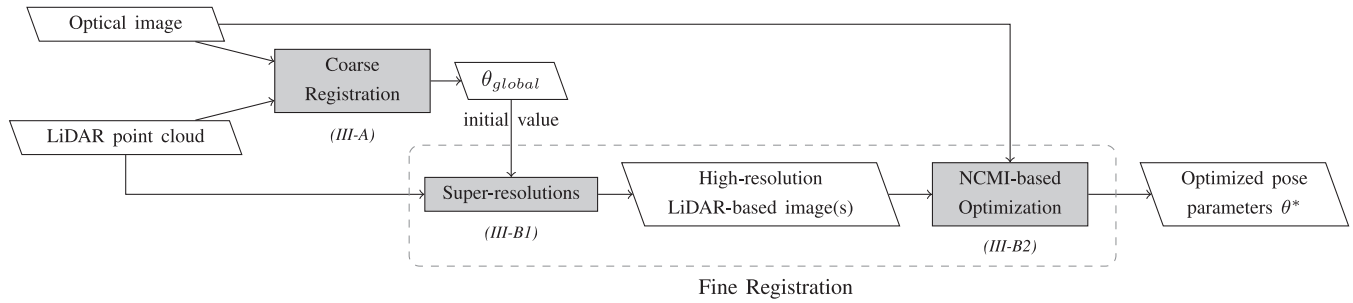
Fig. 1. Flowchart of the registration of optical imagery and airborne LiDAR data (NCMI: Normalized Combined Mutual Information). A parenthesis below each procedure block denotes their respective descriptive subsection.
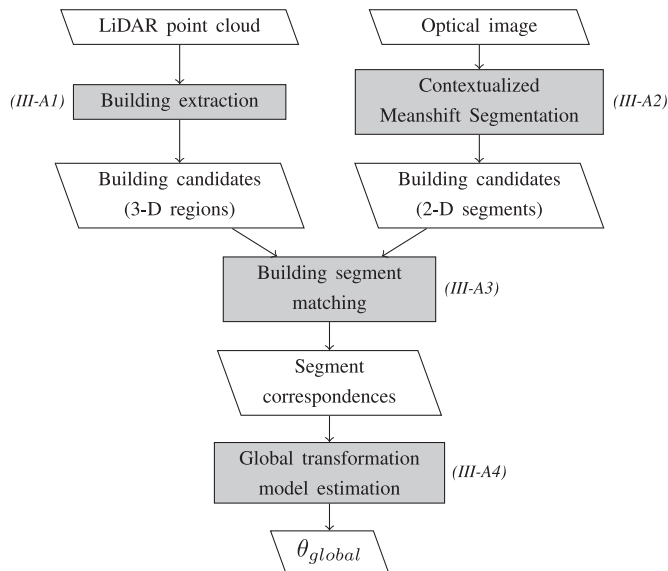

Fig. 2. Flowchart of the building-based coarse registration between optical image and LiDAR point cloud. A parenthesis next to each procedure block denotes their respective descriptive subsection.
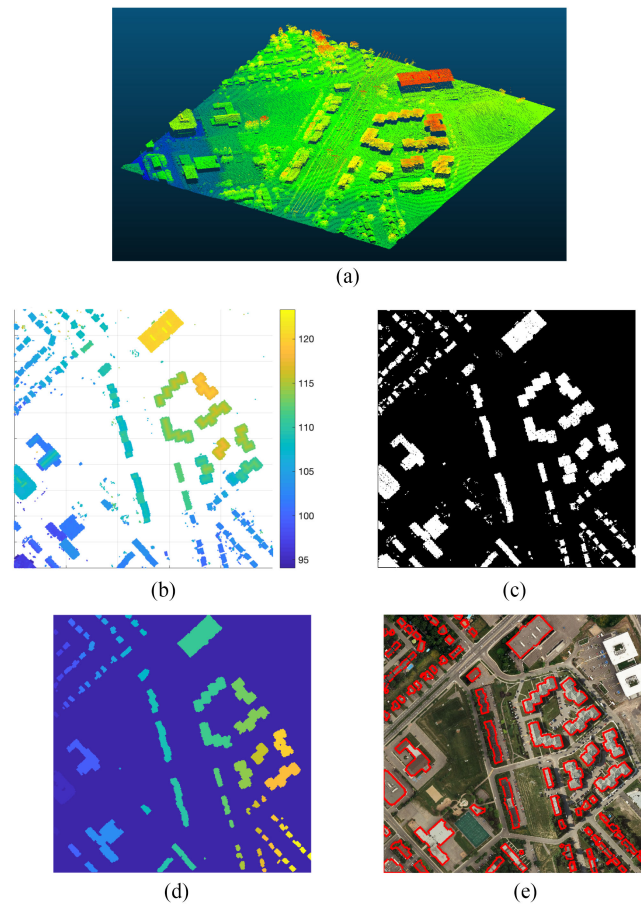

Fig. 3. Illustration of different steps of the building extraction from LiDAR data. (a) LiDAR 3-D point cloud (visualized by CloudCompare 2.9.1, GPL software). (b) Grid of non-ground points (color-coded by elevation). (c) Binary grid of non-ground points. (d) Labeled segments (distinguished by color). (e) Extracted building regions (overlapped on optical image).

on the optical image with a contextually chosen bandwidth parameter. Further processing is then applied to remove unwanted segments and preserve building-like ones. The respective process of building extraction from the LiDAR point cloud and the optical image are described in Section III-A1 and III-A2. Then, the building candidates from each dataset are matched and yield a set of correspondences (Section III-A3), which are then used to estimate the global transformation model (Section III-A4).

1) Building Extraction From LiDAR Data: The extraction of buildings from LiDAR point cloud is carried out through a series of steps. They are depicted in Fig. 3, whereas the input point cloud is shown by Fig. 3(a). First, non-ground points are separated from ground points using an elevation thresholding. This thresholding is proposed by many existing works as a necessary initial step [40]. The threshold T_e is set as follows, $T_e = H_g + T_{rf}$, where H_g denotes the ground elevation and T_{rf} is a relief factor. The first value H_g, as proposed by [40], can be determined from a Digital Terrain Model (DTM) generated from the LiDAR point cloud data, e.g., by performing [41]. This DTM generation method allows us to handle complex terrains, such

as combination of hills, steep slopes, and plateaus. Also, since the LiDAR point cloud can be classified as described in Table I, e.g., by using [42], we can measure H_g by the average elevation of ground points, i.e., $H_g = \text{mean}(z_g)$ where z_g represents the elevation of ground points. The second value T_{rf} is empirically set to $T_{rf} = 2.5$ m (usual minimum height of a building).

All non-ground points are then vertically projected onto the plane $z = 0$. A raster grid representing these projected points is

created [see Fig. 3(b)]. The resolution of the grid is set according to the LiDAR point cloud density, in order to avoid null-valued pixels. For instance, for the LiDAR data 2011 with a point spacing of 70 cm, the resolution of the grid is set to 1 m. A binary grid of the same resolution is also generated, shown by Fig. 3(c). Its cell value is set to 1 or 0 according to the presence or absence of projected non-ground points in the cell ("1": presence, "0": absence). Then, a morphological opening operator is applied in order to remove small artifacts on the binary grid. Remaining grid cells with value set to 1 are grouped into labeled segments based on their connectivity. Next, small segments (e.g., smaller than 10 square meters) are removed. The resulting grid consists of a number of relatively large labeled segments related to buildings [see Fig. 3(d)]. These segments are then used to select the building points in the LiDAR point cloud. A convex hull is calculated on each set of these 3-D building points, yielding a set of boundary points for each building. In Fig. 3(e), these building boundaries are shown overlapping on the orthorectified aerial image for visual assessment purpose.

2) Building Extraction From Optical Image: First, the optical visible image is converted into the CIE L*a*b* color space, since this color space allows a better distinction of objects than RGB color space [43]. In this article, we propose to use mean shift to segment building regions from optical image of an urban area. This technique is more efficient than a k-means clustering since the color of building roofs can vary a lot and some roofs have similar color with the surrounding areas or streets. There exist many other segmentation methods, for example methods based on graph-cut which require different priors, such as connectivity prior [44], shape prior [45], or priors about color of background and foreground pixels given by several brush strokes on an image. As a matter of fact, the graph-cut-based image segmentation methods require a high amount of user inputs in order to yield accurate results [46]. On the other hand, mean shift requires only a value of bandwidth corresponding to the image color range and size of objects to be segmented. This technique has been extensively used for many years in the field of computer vision and image processing. However, its relevance cannot yet be dismissed. Nevertheless, determining the best bandwidth parameter for mean shift still remains difficult despite many investigated approaches [47]. This parameter can be set adaptively according to the type of urban area (either residential, industrial, mixed, etc.), and the size of objects of interest. In other words, a contextualization is needed to set up the mean shift parameter. Such a contextualization is carried out based on the meaningful information in the observed area, such as an estimated number of buildings and their relative distance—this knowledge is derived from the building extraction process using LiDAR data—as well as the resolution and the color range of the optical image. Future works will investigate the automation of this step.

Once the mean shift segmentation is performed, a refinement of the extracted segments is carried out. First, we compute the size of the segments, and remove the small ones, since they usually correspond to trees and cars. Large segments corresponding to street regions are similarly removed. This filtering is simple and efficient [48], but depends on the image
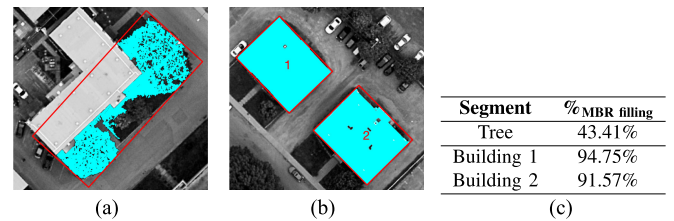

Fig. 4. Comparison of the MBR filling percentage between a tree segment versus building segments. The segment pixels are in cyan, whereas the MBR of each segment is in red. (a) On a tree segment. (b) On building segments. (c) MBR filling percentage.

Segment	$\%_{\text{MBR filling}}$
Tree	43.41%
Building 1	94.75%
Building 2	91.57%

resolution. Therefore, it needs a manual intervention to be set correctly. In this article, we propose to remove segments whose actual area is smaller than 20 square meters or larger than 2000 square meters, which are not the typical area of buildings. Second, we identify the minimal bounding rectangle (MBR) [49] of each of the remaining segments and calculate, using (1), the percentage of their area over the area of the MBR

$$\%_{\text{MBR\_filling}} = \frac{\text{area(segment)}}{\text{area(MBR)}} \times 100. \qquad (1)$$

This filling percentage aims to eliminate coarsely the irregular segments such as trees and grass, while retaining highly regular shape building segments. Fig. 4 depicts a comparison between the MBR filling percentages of two building segments and a tree segment, from which a clear margin between the two types of segment can be observed. In this article, a threshold of 50% for the MBR filling percentages is typically applied. However, on a scene with numerous irregular shape buildings, this threshold can be relaxed. It is worth noting here that this MBR-based refinement only acts as a preliminary filter. Although it cannot remove every non-building segment, it allows to effectively eliminate coarsely the irregular segments. Then, these extracted and refined segments, even with a number of potential outliers, will be fed into the graph-based matching step.

3) Graph-Based Matching of Extracted Segments: The two sets of building candidates extracted from the LiDAR and optical image datasets are taken into consideration and matched. Regarding the optical image, only the segments having a higher percentage than the fixed threshold are considered as stated in the previous point. On the other hand, all building regions extracted from the LiDAR point cloud are taken into consideration. The comparison and matching of these segments can be difficult due to several issues. First, several tree and grass segments wrongly extracted as buildings still remain after the MBR-based segment refinement. In addition, the datasets can be relatively distant to each other (as mentioned in Section I-B1), making a direct matching of segments based on their location is not suitable. Therefore, a matching of segments based on their relative position with respect to their neighbors is more relevant than comparing their individual values, such as location, area, shape similarity, and so on.

A common pattern connecting the centers of neighboring building segments representing their relative spatial arrangement on both datasets is determined using the graph
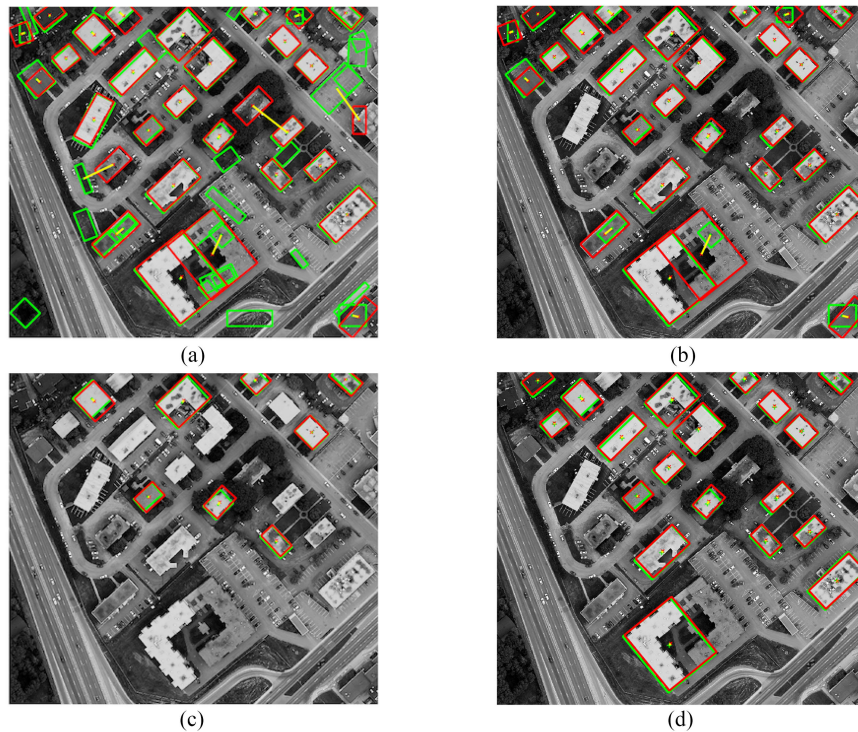
Fig. 5. Matching of building segment centers by considering their relative position. The green and red rectangles represent the MBR of the segments extracted from, respectively, optical image and LiDAR point cloud. The yellow lines connect the centers of the matched segments. (a) Initial matching. (b) GTM result. (c) RANSAC result. (d) GTM + Area and direction validation.

transformation matching (GTM) algorithm [50]. GTM is a graph-based point matching algorithm designed for nonrigid registration between images. Compared to a conventional method like RANSAC [51], this algorithm performs a better removal of *outliers*, i.e., wrongly paired buildings in this work.

In practice, both GTM and RANSAC require an initial one-to-one matching of segment centers, which can be carried out based on the positions of vertically projected 3-D building region centers onto the plane $z = 0$ and the centers of 2-D segments extracted by mean shift segmentation. In the specific case of satellite imagery and LiDAR data where the relative shifts are large (i.e., approximately up to 40 m), this initial matching is guided by a translation vector. It is calculated based on the shift of the largest segment in the area. The largest segment is determined relying on the segment absolute area value and its relative area value with respect to other segments.

Result of the segment matching is shown in Fig. 5, whereas Fig. 5(a) depicts the initial matching. As we could expect, a number of wrongly paired buildings (i.e., outliers) result from the initial matching. They are originated from the tree and grass segments extracted as buildings, or from the buildings that exist on one dataset but not on the other one. These outliers are then removed using GTM. Fig. 5(b) depicts the result of GTM, whereas Fig. 5(c) presents the result of RANSAC. As we have been considering only the relative position of the segment centers, a refinement of false positives from GTM result is carried out based on the area value and the direction of segments. Here, we allow some tolerance for the area value (i.e., a 15%

difference) and direction (i.e., a $2°$ difference) between paired segments provided by GTM. Such tolerance values are chosen empirically. Only the pairs of segments having area and direction differences smaller than the tolerances are preserved. With the selected tolerance values, we consider that only reasonable correspondences of buildings will remain. The result of this refinement is presented by Fig. 5(d).

The capability of GTM to cope with high amount of outliers—theoretically up to three times more numerous than the correct pairs [50]—is advantageous when handling the potential high number of outliers among the extracted segments from the optical image using mean shift. It is also anticipated to handle well the registration of datasets that were acquired within a large timespan, e.g., several years. This temporal variability can lead to significant changes in urban area, such as construction or deconstruction of buildings.

4) Global Transformation Model Estimation: Next, the co-ordinates of the matched building segment centers are used to determine the transformation model between the LiDAR and the imagery datasets. It involves in estimating the camera pose internal and external parameters. The internal parameters are consisted of the scale factors in the x- and y-coordinate direction, respectively α_x and α_y, the skew parameter s and the coordinates of the principal point (p_x, p_y) in terms of pixel dimensions. On the other hand, the external parameters are the position (X_0, Y_0, Z_0) and the orientation (ω, ϕ, κ) of the camera when the image was acquired. The set of all these parameters of the camera pose is denoted by a vector θ henceforth.

Even though the camera pose external parameters can be measured by a GPS/IMU system, it is still necessary to reestimate them, since the image can suffer from radiometric and geometric errors as well as undergone an orthorectification process. Based on the coordinates of the resulting matched building segment centers, a transformation model is estimated using the Gold Standard algorithm for finite projective camera model [52, p.187]. We denote the set of parameters associated with this estimated global transformation model by θ_{global}.

The transformation from 3-D homogeneous coordinates to 2-D homogeneous coordinates is given by the following 3×4 camera matrix

$$P = KR[I|-C] \tag{2}$$

where K is the camera calibration matrix, R stands for the rotation matrix describing the orientation of the camera, I is the identity matrix, and C denotes the coordinates of the camera center. The matrices from (2) are defined as follows:

$$K = \begin{bmatrix} \alpha_x & s & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3a}$$

$$C = \begin{bmatrix} X_0 & Y_0 & Z_0 \end{bmatrix}^T \tag{3b}$$

$$R = R_z(\kappa)R_y(\phi)R_x(\omega) \tag{3c}$$

where R_x, R_y, and R_z are the rotation matrices for rotations around x-, y- and z-axis. As presented by (3), the transformation model involves 11 degrees of freedom, related to the camera internal and external parameters.

B. Fine Registration

After coarsely repositioning the datasets, the next step is dedicated to register them precisely. An area-based optimization approach is relevant in the present context, in order to determine the optimal set of parameters that enables the most accurate registration [14], [15]. However, this approach involves several constraints, such as the datasets need to be spatially close to each other, as well as to have the same resolution and display similar intensity characteristics. As a result of the presented coarse registration and the SR process elaborated in the Section III-B1, these constraints are fulfilled.

We propose a fine registration method, summarized by Fig. 6, which involves an SR applied on the LiDAR data. Then, an estimation of local transformation models (Section III-B2) is performed based on the maximization of the NCMI or MI measured between the optical image and the *high-resolution* LiDAR-based images, resulted from the SR. The high-resolution term means that these images have the same resolution and size as the optical image. NCMI achieves its maximum values when the images are geometrically aligned [15], yielding an optimal set of camera pose parameters, denoted by θ^*. We describe these points in what follows.

1) SR of LiDAR Data: LiDAR point cloud is usually significantly subsampled compared to optical image. This subsample problem is usually addressed by a sparse reconstruction (e.g., for pansharpening [53]) or an SR of low-resolution depth maps
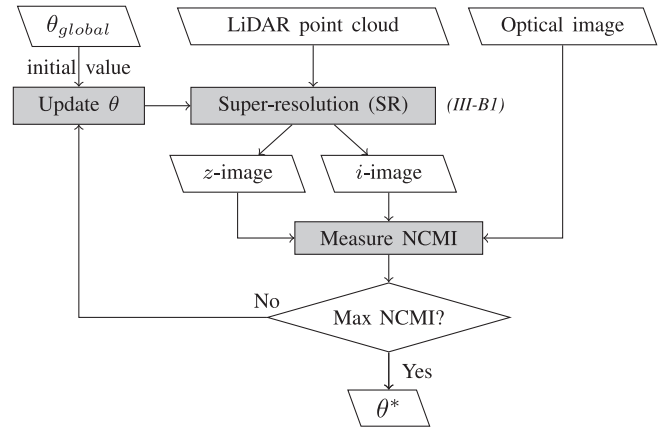

Fig. 6.　Flowchart of the proposed fine registration between optical image and LiDAR point cloud.

[54]. Thus, we propose a process of transferring and propagating values from LiDAR point cloud onto the frame of the optical image. Such process is to generate a rasterized dataset with the same size and spatial resolution as the optical image, thus it is called SR. Pixels of the super-resolved image contain the values derived from the LiDAR 3-D points, i.e., altitude values or laser return intensity values. The super-resolved image of LiDAR-derived altitude values is called the z-image, whereas the image of intensity values is called i-image. The purpose of the SR is also to provide an approach to neutralize the sampling density difference between the two datasets, thus facilitate the area-based fine registration of them.

a) Mathematical notation: The inputs of the SR process are the LiDAR point cloud, a transformation model, the frame of reference, and the size of the optical image. We denote the optical image by $u \in \mathbb{R}^{n_x \times n_y \times 3}$, where n_x and n_y are, respectively, the number of rows and columns. The LiDAR point cloud is represented by $\psi \in \mathbb{R}^{m \times 4}$, where m is the number of LiDAR points. Each point contains three spatial coordinates (x, y, z) and a laser return intensity value i. We also use $\psi^z \in \mathbb{R}^m_+$ to denote the column of altitude values of the LiDAR points; whereas $\psi^i \in \{0, 1, ..., 255\}^m$ stands for the intensity value of LiDAR points. For the sake of simplicity, we use the same notation ϕ to denote the result of the SR, i.e., the z-image and i-image. During the process, ϕ is vectorized into a column vector of $n = n_x \times n_y$ elements. Fig. 7 describes the principle of the proposed SR.

b) Transfer of LiDAR values: First, LiDAR values, i.e., altitude and laser return intensity values, are projected onto the optical image space using the transformation model associated with camera pose parameters θ. At the first iteration of the fine registration, θ is given by θ_{global} obtained from the coarse registration. Mathematically, the value transfer is presented by the following equation:

$$\phi_{\Omega^*} = H_{\Omega^*}\psi^z \text{ or } \phi_{\Omega^*} = H_{\Omega^*}\psi^i \tag{4}$$

where Ω^* and Ω denote, respectively, the subsets containing the indices of pixels from ϕ, having or not an associated altitude value (or intensity value) transferred from ψ. Thus, ϕ_{Ω^*} and ϕ_Ω, respectively, denote the subvector containing the pixels with and
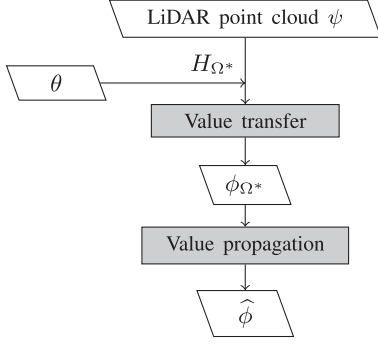
Fig. 7. Overview of the SR process, to generate a high-resolution LiDAR-based image (z- or i-image).

Algorithm 1 Solving (5) by FISTA Algorithm With Constant Step Size γ.

input:
- sparse image ϕ^{spa} ($\phi^{\mathrm{spa}}_{\Omega^*} = H_{\Omega^*}\psi^z$ or $\phi^{\mathrm{spa}}_{\Omega^*} = H_{\Omega^*}\psi^i$, $\phi^{\mathrm{spa}}_{\Omega} = 0$)
- a maximum number of iterations $k_{\max}$
- step size $\gamma > 0$
- soft thresholding parameter $\lambda > 0$
- a tolerance value ϵ for stopping criterion

set: $k \leftarrow 1, t_0 \leftarrow 1, y^{(0)} \leftarrow \phi^{\mathrm{spa}}$
repeat
$$x^{(k)}_{\Omega} = \mathcal{T}_{\lambda\gamma}(y^{(k-1)}_{\Omega} - \gamma H_{\Omega}\nabla f_{\mathrm{SSDG}}(y^{(k-1)}))$$
$$t_k = \frac{1}{2}(1 + \sqrt{1 + 4t^2_{k-1}})$$
$$y^{(k)}_{\Omega} = x^{(k)}_{\Omega} + (\frac{t_{k-1}-1}{t_k}) \times (x^{(k)}_{\Omega} - x^{(k-1)}_{\Omega})$$
$k \leftarrow k + 1$
until $k > k_{\max}$ **or** $|y^{(k)} - y^{(k-1)}| < \epsilon$
set: $\widehat{\phi}_{\Omega^*} \leftarrow \phi^{\mathrm{spa}}_{\Omega^*}$ **and** $\widehat{\phi}_{\Omega} \leftarrow y^{(k)}_{\Omega}$
output: dense image $\widehat{\phi}$

without a transferred altitude value; whereas ϕ denotes the vector containing all pixels. The dimension of ϕ_{Ω^*} and ϕ_{Ω}, respectively, are $m \times 1$ and $(n-m) \times 1$. The matrix H_{Ω^*} associated to the camera pose parameters θ, is an index matrix allowing selecting only the pixels whose values are transferred from the LiDAR point cloud. It is computed based on the projection related to θ of the LiDAR 3-D point cloud onto the 2-D optical image space. Next, the transferred values are propagated to their neighboring pixels.

c) Propagation of transferred LiDAR values: The propagation of transferred values is carried out through the minimization of a cost function $\mathcal{F}(\phi)$, defined by (5). It is composed of the sum of squared directional gradients (SSDG) of ϕ, and an $\mathcal{L}^1$-norm term to promote the sparsity of ϕ, subjecting to the values transferred from the point cloud [described by (4)].

$$\widehat{\phi} = \underset{\phi}{\arg\min}\, \mathcal{F}(\phi)$$
$$\text{with } \mathcal{F}(\phi) = \underbrace{\|\nabla_x\phi\|^2_2 + \|\nabla_y\phi\|^2_2}_{f_{\mathrm{SSDG}}(\phi)} + \lambda\,\|\phi\|_1 \quad (5)$$
$$\text{subject to } \phi_{\Omega^*} = H_{\Omega^*}\psi^z \text{ or } \phi_{\Omega^*} = H_{\Omega^*}\psi^i$$

where $\|\cdot\|_p$ stands for the $\mathcal{L}^p$-norm, ∇_x and ∇_y represent the directional gradient operators along the x- and y-axis, whereas the parameter λ controls the amount of the $\mathcal{L}^1$-regularization.

Our SR approach is inspired by the work of Castorena *et al.* [55]. However, they proposed a cost function that is solely defined by SSDG for the fusion of terrestrial LiDAR data with optical imagery. It is based on hypothetical characteristics of a depth map, namely that the magnitude and occurrence of depth discontinuities inside such depth map should be minimum. The advantage of using this cost function is its convexity and ease to compute. Castorena's method showed good results in propagating depth values across homogeneous regions. However, the mentioned hypothetical characteristics are not suitable in an airborne context, where off-terrain objects like buildings or trees always exhibit strong elevation discontinuities. By iteratively minimizing the SSDGs, these discontinuities will be gradually flattened, hence resulting in inaccurately estimated z-image at these elevation-transitioning regions. Such discontinuities should be preserved during the SR process. Thus, an $\mathcal{L}^1$-norm

term is additionally proposed in our approach to promote sparsity of the z-image, i.e., to preserve the elevation discontinuities stemming from buildings and trees.

d) Propagation algorithm: The optimization problem described by (5), containing the term $\|\phi\|_1$, is solved iteratively. Each iteration involves calculating the gradient descent of the SSDG term (i.e., ∇f_{SSDG}) followed by a shrinkage/soft-threshold step. The shrinkage operator $\mathcal{T}_{\alpha}: \mathbb{R}^n \to \mathbb{R}^n$ is defined as follows:

$$\mathcal{T}_{\alpha}(x) = (|x| - \alpha)_+ \times \mathrm{sign}(x) \quad (6)$$

where $(|x| - \alpha)_+ = \max(|x| - \alpha, 0)$, and α is a threshold value, which is set to $\alpha = \lambda\gamma$ in Algorithm 1.

Algorithm 1 presents the process of solving (5), using the fast iterative shrinkage-thresholding algorithm (FISTA) [56] with a constant step size. In this Algorithm, the superscript (k) of a vector denotes its state at the kth iteration. The subvector x_{Ω^*} (and y_{Ω^*}) contains only the values of pixels indexed by Ω^*, i.e., the pixels having a LiDAR transferred value. They remain unchanged during the propagation process. On the other hand, x_{Ω} (and y_{Ω}) represents the subvector containing the values of pixels indexed by Ω, i.e., the null-valued pixels before the value propagation. The vector ϕ without an index subscript is the vector contains all pixels, i.e., $\phi = \phi_{\Omega\cup\Omega^*}$. For instance, ϕ^{spa} represents the sparse image where pixels of index in Ω^* are transferred from LiDAR data, while other pixels (i.e., the one of index from Ω) are null-valued.

FISTA with its computational simplicity is adequate for solving large-scale problems. It also converges more quickly than ISTA, with a rate of $O(1/k^2)$ [56]. The convergence rate of the SRs is depicted in Fig. 8. Indeed, Fig. 8(a) and (c), respectively, depict the errors between the estimated z-images and i-images at two consecutive iterations, i.e., $\|\phi^{(k+1)} - \phi^{(k)}\|_2$. The values of the cost function $\mathcal{F}(\phi^{(k)})$ through iterations are also shown
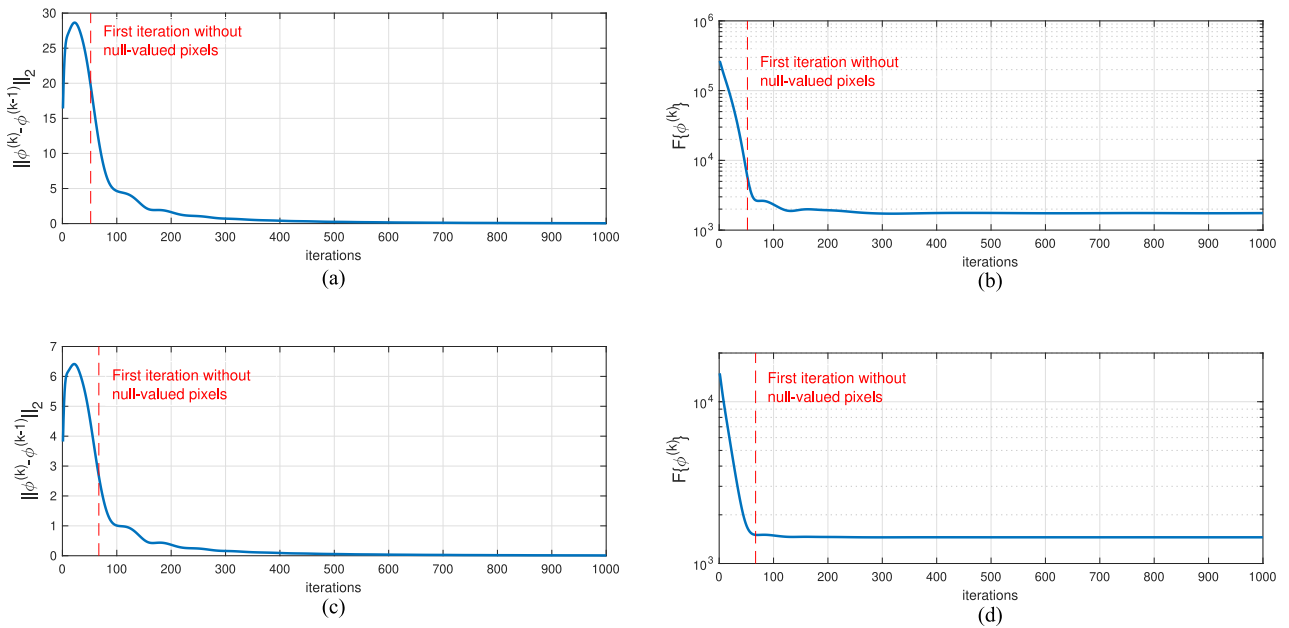
Fig. 8. Errors $\|\phi^{(k+1)} - \phi^{(k)}\|_2$ and the cost values $\mathcal{F}(\phi^{(k)})$ plotted as a function of iterations. The vertical red-dashed lines indicate the first iteration where every pixel of the estimated image is filled. First row: the plots from the SR process of generating the z-image; second row: the plots from the SR process of generating the i-image. (a) Error $\|\phi^{(k+1)} - \phi^{(k)}\|_2$. (b) Function value $\mathcal{F}(\phi^{(k)})$ (in log-scale). (c) Error $\|\phi^{(k+1)} - \phi^{(k)}\|_2$. (d) Function value $\mathcal{F}(\phi^{(k)})$ (in log-scale).
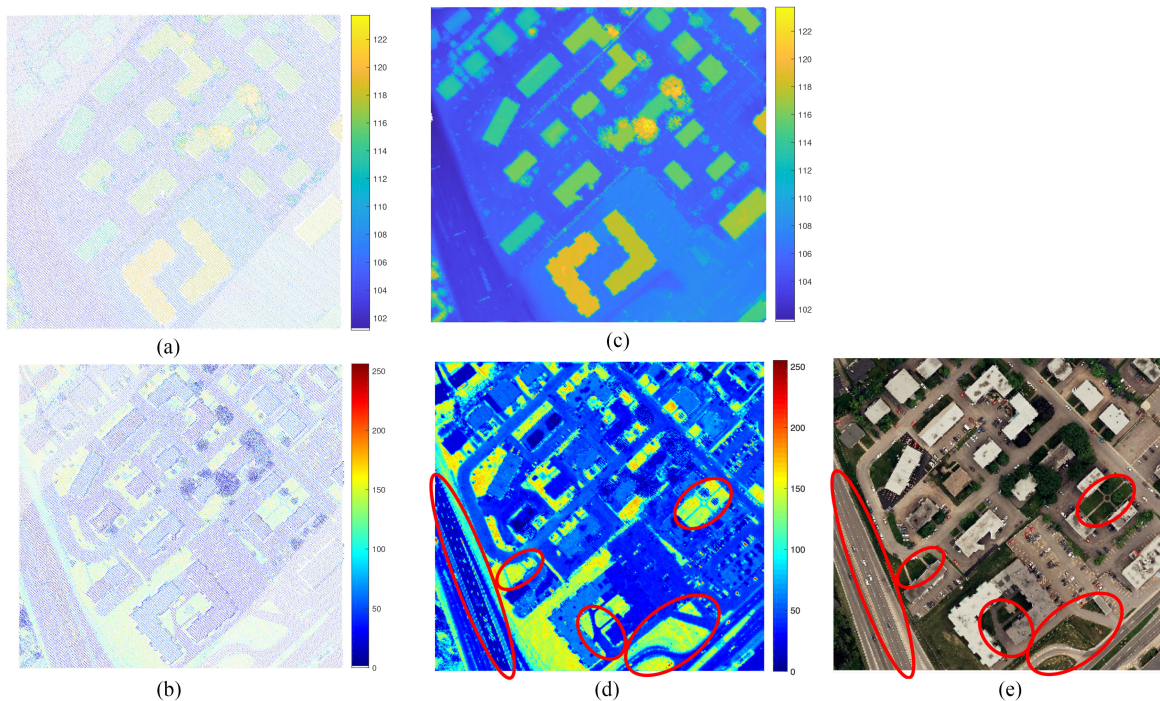

Fig. 9. Illustration of the SR results. First column: the sparse z-image and i-image from the value transfer process; second column: the dense z-image and i-image from the value transfer and propagation process; third column: the optical image of the same scene shown for the sake of comparison. (a) Sparse z-image. (b) Sparse i-image. (c) Dense z-image. (d) Dense i-image. (e) Optical image.

in Fig. 8(b) and (d). We can remark that after approximately 600 iterations, the estimated z-image and i-image have nearly converged into stable solutions.

Finally, Fig. 9 shows the results of a transfer and propagation of altitude and intensity values from the LiDAR data onto the

frame of the optical image. The value transfer results are depicted through the sparse images [see Fig. 9(a) and (b)], while the value propagation results are shown by the dense images [see Fig. 9(c) and (d)]. On the z-images, the pixel color represents the altitude in meters. In contrast, the pixel color on the i-images

represents the intensity value between 0 and 255. The reference optical image on the same urban scene [see Fig. 9(e)] allows a visual quality assessment of the superresolved images. On the one hand, we can observe that the elevation of buildings and other off-terrain objects (e.g., trees, power lines), as well as the relief of the urban scene are well presented on the dense z-image [see Fig. 9(c)], and correspond with the information in the optical image. On the other hand, different elements of the scene like buildings, grasses, or roads can be discriminated on the i-image, similarly on the optical image. For example, several regions with distinctive elements are highlighted by the red ellipses on Fig. 9(d) and (e).

2) Estimation of Local Transformation Model: As aforementioned, an MI-based registration method involves many constraints. One of them is that the to-be-registered datasets must have the same resolution. The solution to this problem is to use high-resolution LiDAR-based images generated by the presented SR approach.

a) Proposed approach: This article presents an MI-based registration method that relies on the MI measured between the optical image and the i-image. An NCMI-based registration method is also proposed. It involves measuring the NCMI between the optical image and both the z-image and i-image. Both similarity measures, MI and NCMI, are expected to achieve their maximum value when the images (i.e., the optical image and the LiDAR-based images) are geometrically aligned.

Moreover, the proposed fine registration also consists in a local approach of transformation model estimation. It involves dividing the study area into many patches of same size and estimating a local transformation model for each patch. Such a patch-based approach allows accelerating the MI maximization process. For each patch of the optical image u_t and the LiDAR data ψ_t ($1 \leq t \leq T$, with T is the total number of patches), the determination of the optimal set of camera pose parameters, denoted by θ_t^*, is carried out based on the maximization of MI or NCMI, as follows:

$$\theta_t^* = \arg\max_{\theta \in \Theta} \mathrm{MI}(f_{\mathrm{SR}}^i(\theta, \psi_t); u_t) \tag{7}$$

$$\theta_t^* = \arg\max_{\theta \in \Theta} \mathrm{NCMI}((f_{\mathrm{SR}}^i(\theta, \psi_t), f_{\mathrm{SR}}^z(\theta, \psi_t)); u_t) \tag{8}$$

Equations (7) and (8) present the maximizations based on, respectively, MI and NCMI. f_{SR}^i and f_{SR}^z represent the SR process that generates, respectively, the i-image and z-image (denoted by ϕ from the previous Section III-B1), given the camera pose parameters θ and the LiDAR data ψ_t.

Given two random variables A and B with marginal probability distribution functions (pdf), $p_A(a)$ and $p_B(b)$ and joint pdf $p_{AB}(a, b)$, the MI between A and B, denoted by $\mathrm{MI}(A; B)$, measures the degree of dependence of A and B by the distance between the joint pdf $p_{AB}(a, b)$ and the pdf associated with the case of complete independence $p_A(a).p_B(b)$. This entropic distance is expressed by the means of the Kullback–Leibler divergence measure [57], given by

$$\mathrm{MI}(A; B) = \sum_{a,b} p_{AB}(a, b) \log \frac{p_{AB}(a, b)}{p_A(a).p_B(b)} \tag{9}$$
$$= H(A) + H(B) - H(A, B)$$

where $H(X) = -\sum_x p_X(x) \log p_X(x)$ is the *Shannon entropy* of random variable X. Its estimation is proposed by Mokkadem [58]. The registration method based on the maximization of MI is originally introduced by [59]. Since then it has been extensively studied in many research areas, particularly to register an optical image with an image derived from LiDAR data. This image is either the LiDAR-derived DSM or the intensity image, which has the same resolution with the optical image [7], [19].

Another statistical similarity measurement used for the registration between LiDAR data and optical imagery (8) is the NCMI [15], given by

$$\mathrm{NCMI}((A, B); C) = \frac{H(A; B) + H(C)}{H(A; B; C)} \tag{10}$$

NCMI-based registration method relies on the similarity between the optical image and both LiDAR images, i.e., DSM and intensity image which are inherently registered. This combined similarity measurement is shown to be more informative than the conventional MI [60].

Mastin *et al.* [7] compared the three usages of LiDAR-derived images in the MI-based registration involving measuring its/their similarity with the optical image, i.e., first, using only the DSM image, second, using only the intensity image, third, using both images. They demonstrate that the usage of the intensity image yields more accurate registration result than using the DSM image. The usage of both images is also shown to yield more accurate result than the two individual usages [15].

b) Implementation: To resolve (7) and (8), we use Nelder–Mead simplex algorithm [61]. Such algorithm is derivative-free and also straightforward in terms of implementation. The initial values for the optimization are given by the θ_{global}, resulted from the coarse registration. In this article, the considered urban area is divided into equal patches, of which the size is chosen as 500×550 pixels. This patch size for the fine registration has been selected based on the study of [62]. It is not related to the building size. It only aims at reducing the computational cost while maintaining a sizable patch for a reliable MI calculation. The division of the area into equal patches is irrespective of the distribution of buildings, or in other words, independent of the distribution of correspondences used in the coarse registration.

3) Smoothing of Patch-Based Registration Results: A potential problem expected from this patch-based registration approach is the incoherence of local camera pose parameters θ_t^* between patches, especially in transition regions between patches. For instance, Fig. 10 illustrates two examples where the difference of transformation models of patches causes a conflict [see Fig. 10(a)] and a discontinuity [see Fig. 10(b)] of projected points on the transition regions between patches. In these regions, there are 3-D points which are spatially neighboring but they belong to two adjacent patches. Consequently, they are projected with two different transformation models, hence causing such conflict or discontinuity. However, this problem
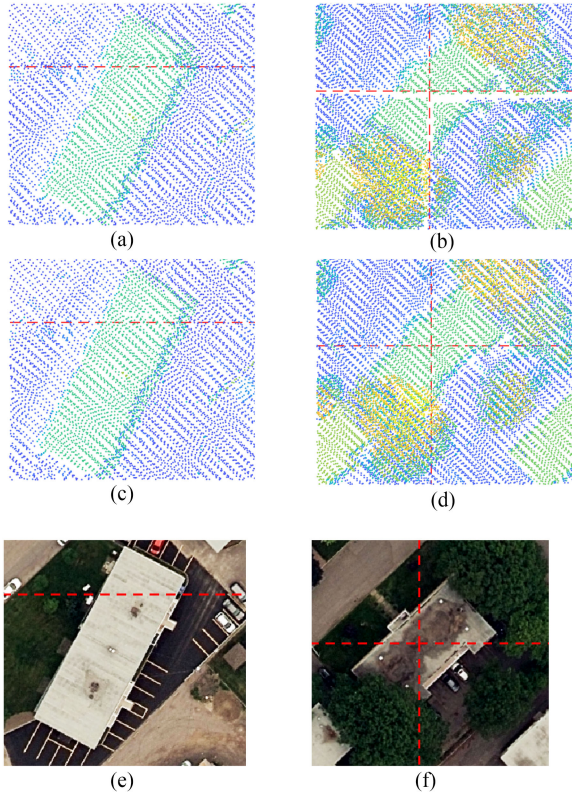
Fig. 10. Examples of (a) a conflict, and (b) a discontinuity of projected points between neighboring patches, and the results (c) and (d) on the same regions after the smoothing using IDW-based interpolation. The red-dashed lines represent the patch boundaries. The optical images (e) and (f) of the considered areas are shown for the sake of comparison. As such, the shapes of the exemplified buildings are well retrieved after the smoothing.

has not yet been addressed by existing work that presented a similar patch-based approach [19].

In this article, we present a smoothing of camera pose parameters from the patch-based transformation models. Instead of using the transformation model of a patch for the projection of every point in this patch, we propose to compute the projection of each 3-D point based on a weighted average of several neighboring local transformation models. We denote the center of neighboring patches by C_i, and the local transformation model of the patches by θ_i, $i = 1, .., N$, where $N = 9$ is the number of neighboring patches of a point p. The transformation model for the point p, denoted by $\theta(p)$, is interpolated using the inverse distance weighting (IDW) [63] average of neighboring local transformation models, given by

$$\theta(p) = \begin{cases} \dfrac{\sum_{i=1}^{N} w_i \theta_i}{\sum_{i=1}^{N} w_i}, & \text{if } d(p, C_i) \neq 0 \\ \theta_i, & \text{if } d(p, C_i) = 0. \end{cases} \quad (11)$$

The weights w_i are computed by the inverse squared Euclidean distance from the considered point p to the neighboring patch centers C_i, as follows:

$$w_i = \frac{1}{d(p, C_i)^2}, i = 1, .., N. \quad (12)$$

Fig. 10(c) and (d) depicts the outcomes of the resolved incoherence problem between patches, using the IDW-based interpolation of patch-based camera pose parameters.

IV. RESULTS AND DISCUSSION

A. Assessment Methodology

Experiments have been carried out to evaluate the quality of the registration and determine whether it is good enough to be beneficial for a subsequent data fusion or other applications. However, the lack of a ground truth, i.e., true values of the camera pose parameters, makes such an evaluation difficult. To overcome this problem, Mastin *et al.* [7] proposed to use expert-chosen control points to determine these values. Otherwise, without a ground truth, the registration quality of existing methods has been assessed in these following manners.

1) Using a subjective quality indicator or by a visual assessment: e.g., a *good assignment of 3-D point-to-pixel on the colorized point cloud* [13, Sec. 3], or assessing whether *the images are close enough for the projective texture mapping* [7], or based on how well the representations of objects (e.g., buildings, vegetation) align.
2) Using the average spatial discrepancy between datasets measured at manually determined check points, or using check pair lines.
3) Involving a determination of an optimal set of parameters that minimizes a cost function or maximizes a statistical dependence measurement. In other words, a registration is considered successful when the determined parameters are optimal. The cost function can be the MI or its variation between the optical image and the LiDAR intensity image [7], [15], [19]. It can also be defined by the pixel-wise distances calculated between the hyperspectral image and the LiDAR-derived image [14].

Since a thorough quality assessment of a registration method is still missing, we present multiple evaluations in this article. First, a visual assessment is carried out based on the alignments of scene elements. Second, an evaluation of building candidate extraction and matching steps from the coarse registration is carried out. Third, since our proposed registration method already involves a maximization of MI between datasets for determining optimal camera pose parameters, we perform subsequently two spatial discrepancy evaluations. They are based on first, check points which are the centroids of manually determined building roofs, and second, check pair lines manually sketched from the two datasets.

In this article, we perform and assess the proposed registration method on four pairs of LiDAR-optical image datasets. They consist of four datasets on Quebec City (QC, Canada), namely two optical imagery and two LiDAR datasets. It is worth reemphasizing that all four datasets were acquired from different platforms with different acquisition configurations at different times and have different spatial resolution. Their specifications have been presented in Table I. On the one hand, the first optical imagery dataset consists of the 15-cm resolution orthorectified aerial images acquired in 2016. The second dataset is provided by the Pléiades satellite involving 50-cm resolution images

Fig. 11. Two selected urban areas on which we evaluate the registration method. The first urban area is composed of 28 buildings; whereas the second urban area is composed of 12 buildings. Particularly in the area 2, there are two buildings highlighted in red-dashed ellipse which were not built in 2011, namely the year when one of the LiDAR dataset was recorded. (a) Urban area 1. (b) Urban area 2.
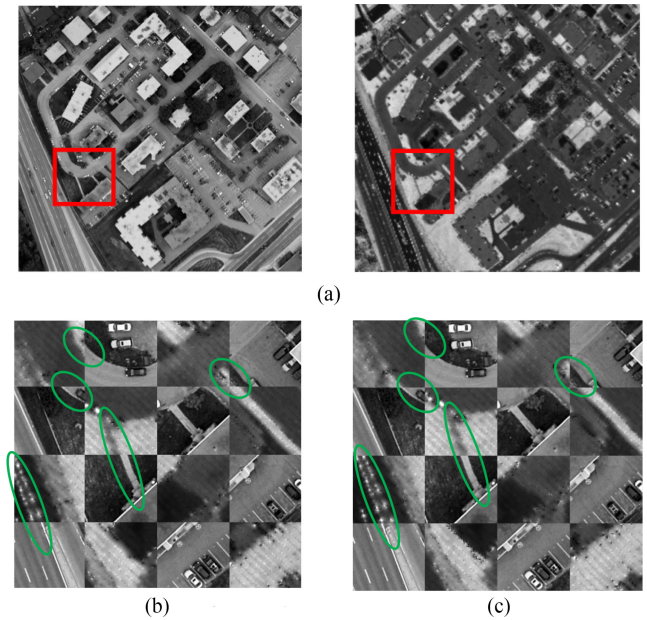

Fig. 12. Visual assessment of the proposed registration method, focusing on the subimage indicated by the red rectangle on the optical image and the i-image from LiDAR data. The two subimages are overlaid in a checkerboard-type display, in order to assess the alignment between multiple elements (green ellipses) yielded by the registration. (a) Optical image and i-image on the area 1. (b) Checkerboard overlay (before registration). (c) Checkerboard overlay (after registration).

resulted from a pansharpening [64] of 50-cm resolution PAN images and 2-m resolution MS images acquired in 2015. On the other hand, the LiDAR datasets involve the airborne LiDAR point clouds acquired in 2011 and 2017. Their point spacing are, respectively, 70 cm and 35.4 cm. We specifically measure the spatial discrepancy between LiDAR and imagery datasets, before and after carrying out our proposed registration method on two urban areas of representative characteristics in Quebec City.

The two selected urban areas are shown in Fig. 11. The first area is composed of mostly planar roof buildings, which are relatively near each other. In contrast, the second area is composed of big buildings and gable-roof houses. Particularly in the second area, there are two buildings located at the top left corner of Fig. 11(b) (in the red-dashed ellipse) that were not built before 2013, hence they did not exist in the LiDAR dataset recorded in 2011. This absence highlights a temporal variability presented on this area. As one can see, these two areas are composed of buildings of various types, sizes, and shapes. Many buildings have inhomogeneous color rooftops, or even have small objects on them. Several buildings are also surrounded by trees, or have casted very contrasting shadow regions (e.g., the gable houses on the second area). These elements show the complexity of the selected areas. The versatility of the proposed method can be demonstrated through such different datasets on these complex areas. Therefore, we are convinced that they are representative for the scenes which this research work aims to address.

B. Visual Assessment of the Registration

Fig. 12(a) depicts a pair of images of the first urban area; the first one is the orthorectified optical image in 2016 (in gray-scale) and the second one is the i-image from the LiDAR 2011 related to the proposed registration method. Since the two images are of the same resolution, we display them in a checkerboard overlay, in order to assess visually the accuracy of the registration focusing on the small region encompassed by the red rectangles.

Fig. 12(b) shows the checkerboard overlay between the optical image and the i-image before the registration, whereas Fig. 12(c) shows the one after the registration. It is worth noting a difference between these two images concerning the grayscale of different objects. Indeed, on the optical image roads, pathways and buildings have bright pixels, whereas the grasses and trees have darker grayscale. It is the opposite for the i-image where roads, pathways, and buildings are in darker grayscale than trees and grasses.

Let us focus on the alignment of multiple elements—continuous straight lines or curves from pathways and roads—on the region circled in green. When putting together the two images that are well registered, these continuous elements should be visually aligned. Indeed, being mindful of the grayscale differences between the two images, we can see that before the registration these circled elements are misaligned while all of them are rectified after the registration. It is also shown that the initial misalignments between the two datasets, even if relatively small, can still be reduced by the virtue of the proposed registration.

C. Evaluation of Building Candidate Extraction and Matching Steps From the Coarse Registration

The results of the building extraction from LiDAR data (2011) and from aerial optical image (2016) are shown in Fig. 13. Table II summarizes the quantitative results of building extraction and matching on the selected areas. The performance metrics are based on the number of true positives (TP), false alarms (FA), and misses (M). In the context of the building extraction,
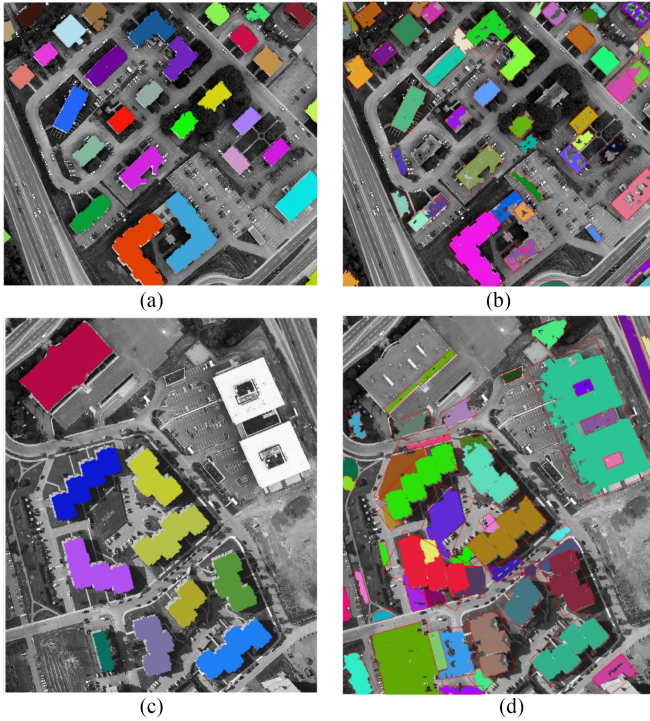
Fig. 13. Illustration of building extraction from LiDAR point cloud and from optical image. First column: LiDAR building segments, overlapped on optical image for visual comparison; second column: Image building segments with MBR filling percentage greater than 50%.

TABLE II
PERFORMANCE OF BUILDING EXTRACTIONS AND MATCHING PROCESSES ON THE SELECTED AREAS

	Extracted from LiDAR data	Extracted from image by mean shift	Matching result by RANSAC	Matching result by GTM
TP/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%

(a) On the area 1 with 28 buildings in total.

	Extracted from LiDAR data	Extracted from image by mean shift	Matching result by RANSAC	Matching result by GTM
TP/FA/M	10/0/0	11/37/1	7/1/2	9/1/0
Precision	100%	22.92%	87.5%	90%
Recall	100%	91.67%	77.78%	100%

(b) On the area 2, there are in total 12 buildings (after 2013), but there were only 10 (before 2013).

a TP indicates that a building is correctly extracted, whereas an FA represents a nonbuilding segment incorrectly extracted as a building, and M means that a building exists but is not extracted. On the other hand, in the context of segment matching, a TP indicates a good match, an FA represents an incorrectly matched pair of segments, whereas M means a pair of segments that should be matched are not paired. The Precision and Recall metrics are computed as follows, Precision $=$ TP/(TP $+$ FA) and Recall $=$ TP/(TP $+$ M).

As shown by Table II, the building extraction from LiDAR data (both 2011 and 2017) has achieved 100% of precision and

recall thanks to the contribution of 3-D information. Similar performance of extracting buildings from LiDAR data has been found consistently, as presented by [6] when conducting tests on the ISPRS Vaihingen benchmark dataset [65].

On the other hand, the mean shift segmentation and subsequent segment refinement have yielded a high number of TPs but also a high number of FAs (i.e., 21 segments are incorrectly extracted as buildings). This number is even higher on the second area, i.e., 37 FAs. However, as one can see, the number of undetected building (i.e., M) is relatively low, namely four buildings on the first area and only one building on the second. This results in the high Recall metric value from mean shift segmentation on both areas, 85.71% and 91.67%.

Considering the matching step, despite yielding relatively high precision (i.e., 100% on the first area and 87.5% on the second one), RANSAC provides a very low number of TPs and a high number of misses. GTM outperforms RANSAC on both areas, yielding more correct matches of building segments. However, on the first area, GTM yields a 95% of recall, with a relatively high number of FAs (i.e., seven segment pairs are wrongly matched). These FAs are then eliminated by the subsequent validation based on segment area and direction (cf., Fig. 5).

D. Patch-Based Transformation Model Estimation

The division of a considered urban area into equal patches, and the local transformation model estimation are shown by Fig. 14. On each patch, the maximized value of MI between the optical imagery and LiDAR data is displayed, as well as the variations of θ_t^* compared to θ_{global}. It should be noted that there is no relationship between the maximized MI values and the number of correspondences in each patch. For example, the patch $(2,1)$ with no correspondence can have a higher maximized MI value results than the patch $(1,3)$ with four correspondences. The relative difference among the maximized MI values stems from the different content of each patch. These variations of θ_t^* compared to θ_{global} are different from one patch to another without any noticeable common pattern. Thus, potential incoherences between them can be expected. Such incoherences are resolved as a result of the IDW-based smoothing presented in Section III-B3.

Table III summarizes the evolution of the resulting MI measurements between the global transformation model (i.e., outcome of the coarse registration) and the local transformation model (i.e., outcome of the fine registration) on each individual patch. Taking into account the number of correspondences among patches [as shown in Fig. 14(a)], we remark lower MI gains (not maximized MI values) for the patches with three to four correspondences than for the patches with fewer correspondences. In fact, the higher MI gains on the patches with few or no correspondence (the bold rows in Table III) shows the interest of the fine registration. Indeed, the coarse registration on these patches is less effective than on the patches with many correspondences, requiring the fine registration step to compensate for more data misalignment, hence resulting in higher MI gains.
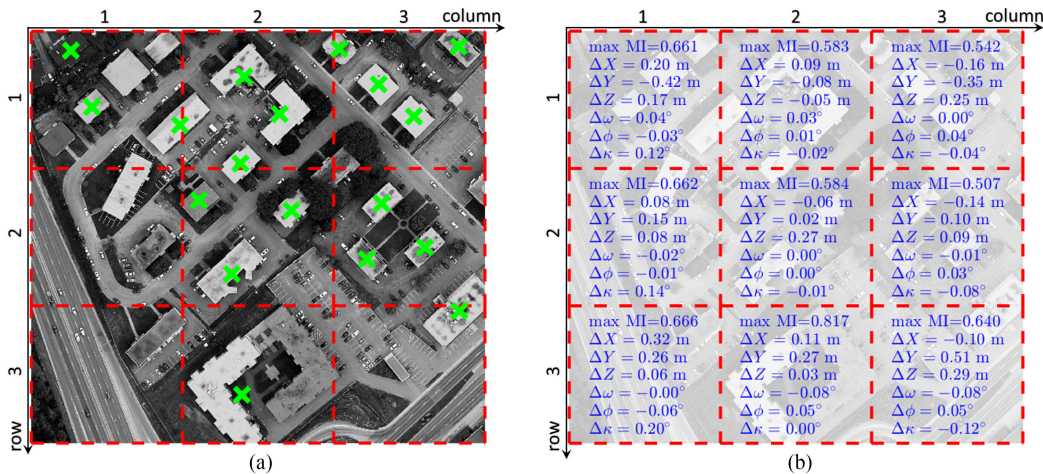
Fig. 14. Illustration of the local transformation model estimation. (a) Division of a considered urban area into equal patches. The red-dashed lines depict the patch boundaries, whereas the green "x" represent the correspondences (i.e., matched primitives) from the coarse registration. (b) Values displayed in each patch are: first row: the maximized MI value; other rows: variation of the camera external parameters (i.e., $\theta_t^* - \theta_{\text{global}}$).

TABLE III
EVOLUTION OF MI FROM THE COARSE REGISTRATION
TO THE FINE REGISTRATION

Patch (row, column)	# of correspondences	$\mathrm{MI}(\theta_{global})$	$\mathrm{MI}(\theta_t^*)$	$\dfrac{\mathrm{MI}(\theta_t^*)}{\mathrm{MI}(\theta_{global})} - 1$
(1,1)	3	0.615	0.661	7.4%
(1,2)	3	0.546	0.583	6.7%
(1,3)	4	0.479	0.542	13.1%
(2,1)	**0**	**0.572**	**0.662**	**15.8%**
(2,2)	3	0.532	0.584	9.6%
(2,3)	3	0.409	0.507	23.9%
(3,1)	**0**	**0.393**	**0.666**	**69.4%**
(3,2)	**1**	**0.700**	**0.817**	**16.7%**
(3,3)	**1**	**0.329**	**0.640**	**94.5%**
Average	-	0.504	0.629	24.8 %

E. Spatial Discrepancy Evaluation

The third evaluation focuses on assessing the registration based on the spatial discrepancy between the datasets. In this regard, we propose to use manually determined check points (i.e., centroids of building roofs) and check pair lines (i.e., mainly building straight boundaries). They are determined from the optical image, and from the high-resolution z-image generated from LiDAR point cloud. Indeed, in order to estimate the spatial discrepancy of datasets before the registration, a z-image is generated using the presented SR process with a vertical projection of LiDAR points. For the coarse registration, another z-image is generated using the global transformation model given by θ_{global}. Finally, in order to assess the spatial discrepancy after the fine registration (either MI or NCMI), the smoothed patch-based transformation model is used to generate the z-image. In the two following assessments, we propose to use z-image instead of i-image for these assessments, since it allows a better manual determination of building roofs and building boundary line segments [cf., Fig. 9(c) and (d)].

1) Based on Centroids of Building Roofs: The first evaluation is carried out based on check points which are the centroids of manually delineated building roofs on the two selected areas.

The distance (in meters) between the centroids of each pair are measured. A smaller distance indicates a more accurate registration. Each column of Table IV presents the spatial discrepancy evaluation between one LiDAR dataset and one optical imagery dataset, among the four datasets described in Table I. The evaluation is presented by the mean and standard deviation of the measured distances. Indeed, the assessments on the registration between the LiDAR data 2011 and then the LiDAR data 2017 with the orthorectified aerial imagery 2016 are given by the column one and two of Table IV. Then, the ones between these LiDAR datasets with the Pléiades multispectral imagery data 2015 are provided by the column three and four. The spatial discrepancy values between the two considered datasets are averaged on all check points from the two selected areas. An insignificant difference of approximately 10–15 cm is obtained between the two areas. The gain values are computed based on average spatial discrepancy values after registration with respect to the values before registration.

a) Between orthorectified aerial image and LiDAR data: Considering the orthorectified aerial image (2016), as a result of the image orthorectification, the average discrepancy between this dataset and the airborne LiDAR dataset (both 2011 and 2017) is already relatively small, i.e., respectively 1.08 and 1.05 m. The results summarized by Table IV show that our proposed registration yields an even smaller discrepancy. Indeed, the proposed coarse registration method results in a reduction of these values by 48.15%. This reduction highlights the effectiveness of repositioning the datasets closer to each other. Then, a spatial discrepancy of 40 cm between the LiDAR data (2011) and the orthorectified aerial imagery (2016), and of 35 cm between the LiDAR data (2017) and the orthorectified aerial imagery (2016) are provided by the NCMI-based fine registration method.

It is worth noting that both the LiDAR datasets acquired in 2011 and 2017 involve a horizontal accuracy of approximately 17 cm; whereas the horizontal accuracy of the orthorectified aerial imagery is 16.5 cm (cf., Table I). It means that the resulting discrepancy values, respectively 40 and 35 cm, are only

TABLE IV
BUILDING REGION CENTROIDS-BASED SPATIAL DISCREPANCY EVALUATION

|	LiDAR data (2011) and aerial imagery (2016)			LiDAR data (2017) and aerial imagery (2016)			LiDAR data (2011) and Pléiades imagery (2015)			LiDAR data (2017) and Pléiades imagery (2015)		
	Mean	(Std)	Gain	Mean	(Std)	Gain	Mean	(Std)	Gain	Mean	(Std)	Gain
Before registration	1.08	(0.52)	-	1.05	(0.68)	-	42.89	(1.47)	-	44.43	(1.73)	-
Coarse registration	0.56	(0.30)	48.15%	0.54	(0.55)	48.57%	2.06	(1.24)	95.20%	1.39	(0.44)	96.87%
MI-based fine registration	0.46	(0.29)	57.41%	0.43	(0.32)	59.05%	1.41	(0.78)	96.71%	1.20	(0.58)	97.30%
NCMI-based fine registration	0.40	(0.27)	62.96%	0.35	(0.31)	66.67%	0.99	(0.45)	97.69%	0.82	(0.45)	98.15%
Gain of using NCMI over MI	13.04%			18.60%			29.79%			31.66%		

slightly bigger than the combination of horizontal accuracy of the considered datasets.

Also, the reported average discrepancy between the LiDAR data 2011 and the orthorectified aerial imagery 2016 after the registration (i.e., 40 cm) is slightly bigger than 1/2 of the average point spacing of the considered LiDAR point cloud (i.e., 70 cm). On the other hand, regarding the registration between LiDAR data 2017 and the orthorectified aerial imagery 2016, the resulting average discrepancy (i.e., 35 cm) approximates the average point spacing of the LiDAR point cloud (i.e., 35.4 cm).

b) Between Pléiades image and LiDAR data: The discrepancy evaluation of the registration between the airborne LiDAR data (2011 and 2017) and the Pléiades imagery (2015) can be analyzed similarly from the results presented at the third and fourth columns of Table IV. The resolution of the Pléiades optical image is 50 cm, and its horizontal accuracy is theoretically between 1 and 3 m, depending on the usage of ground control points on the considered area. These two characteristics of the Pléiades imagery data, especially the horizontal accuracy, are the major factors causing its registration with the LiDAR data (both 2011 and 2017) to be not as accurate as the registration between the LiDAR data and the aerial image (2016). When regarding the resulting average discrepancy (i.e., 0.99 and 0.82 m) and taking into account the spatial resolution of the datasets (i.e., 50 cm for the Pléiades imagery data and 70 cm or 35.4 cm for the LiDAR point spacing), one may interpret that these results are not good enough. However, as the horizontal accuracy of the Pléiades imagery dataset varies between 1 to 3 m as aforementioned, we would argue that such average registration discrepancy results are highly desirable.

Comparing the MI-based and the NCMI-based fine registration, a gain averaging approximately 23% is achieved when using NCMI instead of MI. In other words, as it has already been highlighted by other studies [7], [15], we also demonstrated that using both z- and i-images yields more accurate result than using either one of them. However, it is worth noting the research context of this article is more complex than those of the previous works on this topic. Nevertheless, it is worth noting that the NCMI-based fine registration takes twice as long as the MI-based, since it involves performing the SR twice, for the z-image and the i-image. Therefore, the use of both statistical similarity measurements shows beneficial result but not seamlessly. And the decision that which one of them should be used is essentially a compromise between accuracy and computational cost of the registration method.

Overall, the proposed coarse-to-fine method yields a very significant reduction of the spatial discrepancy between datasets,
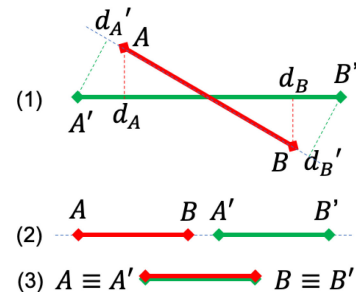
	Distance measure [16]	Modified distance measure
(1) Intersection	$(d_A + d_B)/2$	$\max\{d_A, d_B, d'_A, d'_B\}$
(2) Collinear	0	$\max\{|AA'|, |BB'|\}$
(3) Identical	0	0

Fig. 15. Illustration of the modified distance measure for the check pair line evaluation. Note the difference of the resulting distances between the collinear and the identical cases.

namely 63% to 67% for the registration of LiDAR data and orthorectified aerial imagery, and 98% for the registration of LiDAR data and Pléiades imagery.

2) Based on Check Pair Line Segments: Peng *et al.* [16] proposed an evaluation based on distances between line segments that are manually sketched from the two datasets. Compared with check point-based evaluation, this evaluation relies on a higher geometrical basis (i.e., line segments compared with points) to assess the registration. The distance proposed in [16] between two line segments p (with A and B are its end-points) and q is defined as follows:

$$d(p, q) = \frac{1}{2}(d_A + d_B) \qquad (13)$$

where d_A and d_B, respectively, denote the distances from A and B to the line segment q. However, this distance given by (13) is not quite relevant, because it yields small values when the two line segments are far away but collinear. Based on the literature review of line segment distances [66], the Hausdorff line segment distance [67] is more suitable for this evaluation. It measures the longest of all the distances from a point on one line segment to the other segment, and it equals zero only if the two line segments are identical, i.e., same two end-points. Thus, it reflects fairly the discrepancy between two line segments, even when they intersect or are collinear. Fig. 15 shows such improvement when using the Hausdorff distances instead of the Euclidean distances in three typical cases, namely intersection, collinear, and identical line segments.

TABLE V
CHECK PAIR LINE-BASED SPATIAL DISCREPANCY EVALUATION

	LiDAR data (2011) and aerial imagery (2016)			LiDAR data (2017) and aerial imagery (2016)			LiDAR data (2011) and Pléiades imagery (2015)			LiDAR data (2017) and Pléiades imagery (2015)		
	Mean	(Std)	Gain	Mean	(Std)	Gain	Mean	(Std)	Gain	Mean	(Std)	Gain
Before registration	1.19	(0.67)	-	1.08	(0.80)	-	44.61	(0.74)	-	44.68	(2.02)	-
Coarse registration	0.95	(0.83)	20.17%	0.81	(0.32)	25.00%	2.01	(0.39)	95.49%	2.18	(1.00)	95.12%
MI-based fine registration	0.64	(0.30)	45.82%	0.67	(0.29)	37.96%	1.95	(0.65)	95.63%	1.92	(1.25)	95.70%
NCMI-based fine registration	0.64	(0.29)	45.95%	0.63	(0.26)	41.67%	1.47	(0.61)	96.70%	1.21	(0.33)	97.29%
Gain of using NCMI over MI	2.33%			5.97%			24.62%			36.98%		

In this assessment, two sets of 72 line segments are manually sketched on the optical image and on the generated z-image of the two selected areas. They are then manually matched, yielding check pair line segments. The source for these line segments are mainly the building straight boundaries. Then, the Hausdorff distance between each pair is computed. A smaller distance indicates a more accurate registration. Table V summarizes the quantitative results of check pair line-based evaluation of the registrations between airborne LiDAR data (2011 and 2017) with orthorectified aerial imagery (2016), and with the Pléiades multispectral imagery data (2015).

As we can see from Table V, the discrepancy between the datasets measured based on manually sketched line segments is significantly reduced after each of the registration, i.e., the coarse and the fine registration. On the one hand, the proposed registration method ultimately yields an average discrepancy of 0.63 and 0.64 m between the LiDAR data (2017 and 2011) and the orthorectified aerial imagery (2016). On the other hand, for the registration of LiDAR data (2017 and 2011) and the Pléiades multispectral imagery (2015), the resulting check pair line-based discrepancy value is 1.21 and 1.47 m. However, it is important to remind the principle of Hausdorff distance in order to evaluate the discrepancy results. For example, we consider two nearly identical horizontal line segments (i.e., parallel to x-axis), having same first end-point, and the second end-point of one line segment is four pixels away from the second end-point of the other line segment (with a pixel size of 15 cm). Consequently, the resulting Hausdorff distance between them is 60 cm. Therefore, it should be noted that the mentioned discrepancy values yielded by our proposed method are relatively small. Overall, for all four registrations, the check pair line-based discrepancy varies between three and four pixels.

A discrepancy reduction of approximately 42% to 46% is achieved on the registration between the LiDAR data (2011 and 2017) and the orthorectified aerial imagery (2016). Similarly, a spatial discrepancy reduction of approximately 97% (96.70% and 97.29%) is benefited from the registration between the LiDAR data (2011 and 2017) and the Pléiades multispectral imagery (2015).

Finally, it can be noted that, in the registration between the LiDAR data and the Pléiades image, the benefit of using NCMI instead of MI is much more evident than in the registration between the LiDAR data and the orthorectified aerial image. Indeed, based on both check point-based and check pair line-based evaluation result, using NCMI instead of MI results in average gains of 30.7% and 30.8% of spatial discrepancy

reduction for the registration between the LiDAR data and the Pléiades image. The first percentage is computed from check point-based evaluation result (i.e., 29.79% and 31.66% from Table IV), whereas the second percentage is computed from check pair line-based evaluation result (i.e., 24.62% and 36.98% from Table V). On the other hand, for the registration between the airborne LiDAR data and the orthorectification aerial image, these average gains are only 15.8% and 4.15%.

Both spatial discrepancy assessments and all these mentioned elements show that the results yielded by our proposed method are relevant. These presented assessments have also shown and validated the versatility of the proposed method, through the differences between the registered datasets and the complexity of the test areas. However, it should be noted that it is virtually impossible for a registration method to perform well on any other scene without an adaptation or reparametrization. Notwithstanding, in another context, namely, European urban scenes, the same registration approach should be applicable without major difficulties.

V. CONCLUSIONS

This article has presented and evaluated a coarse-to-fine registration method between airborne LiDAR data and optical imagery. It is dedicated to overcome the challenges associated with the difficult context, where the two datasets are not acquired from the same platform, neither from the same point of view nor having the same spatial resolution and level of detail. In the literature, even one or several of these constraints have been shown problematic for carrying out a registration method (see Section II). To the best of our knowledge, there is currently no solution able to achieve the registration between airborne LiDAR and optical imagery under such constraints altogether. As a matter of fact, the proposed registration method has been evaluated according to its own quality, before and after the registration. Indeed, it is not compared with existing methods because they were not designed to address the considered context. Nevertheless, we reconsidered the subjective accuracy suggestion related to a subpixel level of accuracy for a registration (see Section I-B5). Instead, we rely on an objectively quantitative accuracy which is that, if the resulting spatial discrepancy less than 50 cm, then the registration will be considered accurate. In this regard, the proposed registration method has achieved a highly desirable accuracy.

The proposed method can be summarized as follows. First, a coarse feature-based registration is carried out based on the

extraction and matching of building candidates on the two datasets, reducing significantly the spatial shift between them. Here, it should also be noted that this building-based approach certainly does not limit the usability and versatility of our method, since urban scenes with buildings (even very sparse) are available most of the time [68]. Then, a fine registration based on the maximization of MI or NCMI (both measures have been performed separately) is carried out to determine the optimal camera pose, granting the datasets to be precisely aligned. It involves a process of SR of LiDAR data to generate high-resolution images of altitude and intensity values. This approach neutralizes the difference of spatial resolution and level of detail between datasets, enabling the MI-based and NCMI-based fine registration. The fine registration also involves in dividing the considered area into many equal patches. For each patch a local transformation model is estimated. This approach allows reducing significantly the computational cost of the fine registration. Finally, a smoothing of the patch-based transformation models is carried out to resolve the conflicts and discontinuities between them. It involves an IDW average of camera pose parameters from neighboring patches.

As one can realize, many elements of the proposed method are intended as the solution to the challenges associated with the considered context. First, in order to address the spatial shift between datasets caused by the differences of points of view and fields of view, a coarse registration is carried out. It relies on using buildings as primitives, which is a relevant choice of primitive considering the low density of airborne LiDAR point cloud around vertical surfaces. Then, the differences of spatial resolution and level of detail between datasets have been dealt with by the SR approach. An area-based fine registration using MI or NCMI measurement is carried out to finely tune the optimal local transformation model. Overall, as highlighted by the comprehensive spatial discrepancy assessments, the proposed method has achieved a very high registration accuracy. It is especially desirable when taking into account the difficulties of the considered context, and the horizontal accuracy of the datasets.

It is suggested that only one registration approach is not sufficient to register the data accurately from heterogeneous sensors, even when they are rigidly fixed to the same platform [14]. In this article, we proposed a coarse-to-fine registration method consisting of two steps of registration. It reinforces the relevance of a coarse-to-fine approach for registering an optical aerial/satellite imagery with an airborne LiDAR dataset. Nevertheless, it can be anticipated that the proposed approach could have limitations to operate in an environment lacking of man-made objects providing reliable primitives, such as forest and desert areas. Thus, a study on the relevance and reliability of primitives found on these environments is necessary for an effective solution therein. However, if we could carry out the coarse registration with manual control points, the proposed subsequent fine registration would not be limited and can be well carried out on these scenes. Finally, with these promising results, the reported research has established a basis for a comprehensive fusion of aerial/satellite optical imagery and airborne LiDAR data in future researches.

REFERENCES

[1] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.

[2] E. P. Baltsavias, "A comparison between photogrammetry and laser scanning," *ISPRS J. Photogramm. Remote Sens.*, vol. 54, no. 2/3, pp. 83–94, 1999.

[3] M. Awrangjeb, M. Ravanbakhsh, and C. S. Fraser, "Automatic detection of residential buildings using lidar data and multispectral imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 5, pp. 457–467, 2010.

[4] T. H. Nguyen, D. Gueriot, J.-M. Le Caillec, C. Sintes, and S. Daniel, "Heterogeneous data registration for 3D underwater scene reconstruction," in *Proc. IEEE OCEANS–Anchorage*, 2017, pp. 1–6.

[5] P. Ghamisi *et al.*, "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 6–39, Mar. 2019.

[6] T. H. Nguyen, S. Daniel, D. Guériot, C. Sintès, and J.-M. Le Caillec, "Unsupervised automatic building extraction using active contour model on unregistered optical imagery and airborne lidar data," *ISPRS—Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W16, pp. 181–188, 2019.

[7] A. Mastin, J. Kepner, and J. Fisher, "Automatic registration of LIDAR and optical images of urban scenes," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2639–2646.

[8] M. Khodadadzadeh, J. Li, S. Prasad, and A. Plaza, "Fusion of hyperspectral and lidar remote sensing data using multiple feature learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2971–2983, Jun. 2015.

[9] R. M. Lucas, A. C. Lee, and P. J. Bunting, "Retrieving forest biomass through integration of CASI and lidar data," *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1553–1577, 2008.

[10] N. Longbotham *et al.*, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 331–342, Feb. 2012.

[11] C. Debes *et al.*, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.

[12] M. Campos-Taberner *et al.*, "Processing of extremely high-resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS data fusion contest–part A: 2-D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5547–5559, Dec. 2016.

[13] A.-V. Vo *et al.*, "Processing of extremely high resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS data fusion contest—Part B: 3-D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5560–5575, Dec. 2016.

[14] 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 Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3460–3474, Jun. 2016.

[15] E. G. Parmehr, C. S. Fraser, C. Zhang, and J. Leach, "Automatic registration of optical imagery with 3D lidar data using statistical similarity," *ISPRS J. Photogramm. Remote Sens.*, vol. 88, pp. 28–40, 2014.

[16] S. Peng, H. Ma, and L. Zhang, "Automatic registration of optical images with airborne lidar point cloud in urban scenes based on line-point similarity invariant and extended collinearity equations," *Sensors*, vol. 19, no. 5, 2019, Art. no. 1086.

[17] 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 J. Photogramm. Remote Sens.*, vol. 127, pp. 39–56, 2017.

[18] J. Zhang and X. Lin, "Advances in fusion of optical imagery and lidar point cloud applied to photogrammetry and remote sensing," *Int. J. Image Data Fusion*, vol. 8, no. 1, pp. 1–31, 2017.

[19] E. G. Parmehr, C. S. Fraser, C. Zhang, and J. Leach, "Automatic registration of optical imagery with 3d lidar data using local combined mutual information," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. II-5/W2, pp. 229–234, 2013.

[20] B. O. Abayowa, A. Yilmaz, and R. C. Hardie, "Automatic registration of optical aerial imagery to a lidar point cloud for generation of city models," *ISPRS J. Photogramm. Remote Sens.*, vol. 106, pp. 68–81, 2015.

[21] S. O. Elberink, "Problems in automated building reconstruction based on dense airborne laser scanning data," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XXXVII, pp. 93–98, 2008.

[22] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[23] J. Castorena, U. S. Kamilov, and P. T. Boufounos, "Autocalibration of LIDAR and optical cameras via edge alignment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 2862–2866.

[24] Y. Ye, L. Bruzzone, J. Shan, F. Bovolo, and Q. Zhu, "Fast and robust matching for multimodal remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9059–9070, Nov. 2019.

[25] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2941–2958, May 2017.

[26] P. Bunting, F. Labrosse, and R. Lucas, "A multi-resolution area-based technique for automatic multi-modal image registration," *Image Vision Comput.*, vol. 28, no. 8, pp. 1203–1219, 2010.

[27] M. L. Uss, B. Vozel, V. V. Lukin, and K. Chehdi, "Multimodal remote sensing image registration with accuracy estimation at local and global scales," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6587–6605, Nov. 2016.

[28] R. K. Mishra and Y. Zhang, "A review of optical imagery and airborne lidar data registration methods," *Open Remote Sens. J.*, vol. 5, no. 1, 2012.

[29] G. P. Asner *et al.*, "Carnegie airborne observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion," *Remote Sens. Environ.*, vol. 124, pp. 454–465, 2012.

[30] S. Liu *et al.*, "A linear feature-based approach for the registration of unmanned aerial vehicle remotely-sensed images and airborne lidar data," *Remote Sens.*, vol. 8, no. 2, 2016, Art. no. 82.

[31] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.

[32] P. Rönnholm and H. Haggrén, "Registration of laser scanning point clouds and aerial images using either artificial or natural tie features," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. I-3, pp. 63–68, 2012.

[33] M. Ding, K. Lyngbaek, and A. Zakhor, "Automatic registration of aerial imagery with untextured 3D LiDAR models," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[34] A. Wong and J. Orchard, "Efficient FFT-accelerated approach to invariant optical-LiDAR registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3917–3925, Nov. 2008.

[35] R. M. Palenichka and M. B. Zaremba, "Automatic extraction of control points for the registration of optical satellite and LiDAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2864–2879, Jul. 2010.

[36] S. Huang, S. Chen, Y. Zhang, P. Guo, and H. Chen, "A multi-scale registration of urban aerial image with airborne lidar data," in *Selected Papers of the Photoelectronic Technology Committee Conferences Held June–July 2015*, vol. 9795, International Societyfor Optics and Photonics. Bellingham, WA, USA: SPIE, 2015, pp. 293–296.

[37] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[38] T. H. Nguyen, S. Daniel, D. Guériot, C. Sintes, and J.-M. Le Caillec, "Robust building-based registration of airborne LiDAR data and optical imagery on urban scenes," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 8474–8477.

[39] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[40] S. Gilani, M. Awrangjeb, and G. Lu, "An automatic building extraction and regularisation technique using lidar point cloud data and orthoimage," *Remote Sens.*, vol. 8, no. 3, 2016, Art. no. 258.

[41] A. S. Maguya, V. Junttila, and T. Kauranne, "Adaptive algorithm for large scale DTM interpolation from lidar data for forestry applications in steep forested terrain," *ISPRS J. Photogramm. Remote Sens.*, vol. 85, pp. 74–83, 2013.

[42] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Conditional random fields for lidar point cloud classification in complex urban areas," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 1, no. 3, pp. 263–268, 2012.

[43] B. Hill, T. Roger, and F. W. Vorhagen, "Comparative analysis of the quantization of color spaces on the basis of the CIELAB color-difference formula," *ACM Trans. Graph.*, vol. 16, no. 2, pp. 109–154, 1997.

[44] S. Vicente, V. Kolmogorov, and C. Rother, "Graph cut based image segmentation with connectivity priors," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[45] D. Freedman and T. Zhang, "Interactive graph cut based segmentation with shape priors," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 755–762.

[46] R. Szeliski, *Computer Vision: Algorithms and Applications*. 1st ed. Berlin, Germany: Springer-Verlag, 2010.

[47] J. E. Chacón and P. Monfort, "A comparison of bandwidth selectors for mean shift clustering," in *Theoretical and Applied Issues in Statistics and Demography*, C. H. Skiadas, Ed. Athens: International Society for the Advancement of Science and Technology (ISAST), 2014, pp. 47–59.

[48] N. L. Gavankar and S. K. Ghosh, "Automatic building footprint extraction from high-resolution satellite image using mathematical morphology," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 182–193, 2018.

[49] H. Freeman and R. Shapira, "Determining the minimum-area encasing rectangle for an arbitrary closed curve," *Commun. ACM*, vol. 18, no. 7, pp. 409–413, Jul. 1975.

[50] W. Aguilar, Y. Frauel, F. Escolano, M. E. Martinez-Perez, A. Espinosa-Romero, and M. A. Lozano, "A robust graph transformation matching for non-rigid registration," *Image Vision Comput.*, vol. 27, no. 7, pp. 897–910, 2009.

[51] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[52] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[53] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May 2013.

[54] J. Lu and D. Forsyth, "Sparse depth super resolution," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2015, pp. 2245–2253.

[55] J. Castorena, G. Puskorius, and G. Pandey, "Motion guided lidar-camera self-calibration and accelerated depth upsampling," *CoRR*, vol. abs/1803.10681, 2018. [online] Available: http://arxiv.org/abs/1803.10681.

[56] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[57] I. Vajda, *Theory of Statistical Inference and Information*. Norwell, MA, USA: Kluwer, 1989.

[58] A. Mokkadem, "Estimation of the entropy and information of absolutely continuous random variables," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 193–196, Jan. 1989.

[59] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[60] N. D. Cahill, "Normalized measures of mutual information with general definitions of entropy for multimodal image registration," in *Biomedical Image Registration*, B. Fischer, B. M.Dawant, and C. Lorenz, Eds. Berlin, Germany: Springer, 2010, pp. 258–268.

[61] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.

[62] E. G. Parmehr, C. S. Fraser, and C. Zhang, "Automatic parameter selection for intensity-based registration of imagery to LiDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7032–7043, Dec. 2016.

[63] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proc. 23rd ACM Nat. Conf.*, 1968, pp. 517–524.

[64] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS Pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, Oct. 2010.

[65] M. Cramer, "The DGPF-test on digital airborne camera evaluation–overview and test design," *Photogrammetrie-Fernerkundung-Geoinformation*, vol. 2010, no. 2, pp. 73–82, 2010.

[66] S. Wirtz and D. Paulus, "Evaluation of established line segment distance functions," *Pattern Recognit. Image Anal.*, vol. 26, no. 2, pp. 354–359, 2016.

[67] H. Alt, B. Behrends, and J. Blömer, "Approximate matching of polygonal shapes," *Ann. Math. Artif. Intell.*, vol. 13, no. 3/4, pp. 251–265, 1995.

[68] D. Vlahov and S. Galea, "Urbanization, urbanicity, and health," *J. Urban Health*, vol. 79, no. 1, pp. S1–S12, Mar. 2002.

Thanh Huy Nguyen (Student Member, IEEE) was born in Nha Trang, Vietnam, in 1991. He received the M.Eng. and M.Sc. degrees in information and image processing systems from the IMT Atlantique (formerly known as Telecom Bretagne), Brest, France, in 2016. He is currently working toward the Ph.D. degree in geomatics and remote sensing under a joint supervision between the Université Laval, Quebec, QC, Canada, and the IMT Atlantique.

During his study program, he did several engineering internships at multiple companies, namely Thales Underwater Systems, ECA Robotics, and Orange France. His research interests include remote sensing data registration and fusion in urban and underwater scenes, as well as machine learning applications.

Sylvie Daniel (Senior Member, IEEE) received the graduate degree in electrical engineering, and the Ph.D. degree in signal processing and telecommunications from the University of Rennes I, Rennes, France, in 1994 and 1998, respectively.

She has been working with imagery and geospatial data for more than 20 years, including in the industry where she led and managed international research and developments projects. Since 2004, she has been with the Département des sciences géomatiques, Laval University, Quebec, QC, Canada, where she is currently a Full Professor. Strongly committed to research, her interests include data acquisition (images, LiDAR, and bathymetric point clouds), image processing and artificial intelligence, 3-D modeling, data fusion, and augmented reality. Her research projects focus on 3-D modeling of infrastructure and urban environment, and on new autonomous solutions for hydrographic survey.

Prof. Daniel is a Member of IEEE and ACM, Principal Researcher of the Centre for Research in Geospatial Data and Information, Member of the Centre for studying distributed intelligent shared environments, and member of the Institute Intelligence and Data. In 2014, she was the Acting Director of the Institute for Information Technology and Society, Laval University, a key player in the field of smart cities.

Didier Guériot (Senior Member, IEEE) received the degree in computer science from the Institut National des Sciences Appliquées, Rennes, France, in 1991, and the Ph.D. degree in electronics from the University of Haute Alsace, France, in 1998.

He focused on extending genetic algorithms to variable length solutions for difficult problems, such as neural networks learning and symbolic sonar image registration. From 1993 to 2005, he had been a Research and Development Engineer for several research labs and companies within the sonar field, dealing with high resolution sonar imagery for the Groupe d'Etudes Sous-Marines de l'Atlantique, a French Navy Research Agency and multibeam echoesounders for RESON in Santa Barbara, CA, USA. In 2005, he joined the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, as an Associate Professor. His research interests include sonar data processing to artificial intelligence, including machine learning topics, mainly applied to underwater environment reconstruction using heterogeneous sensors.

Christophe Sintès (Senior Member, IEEE) received the M.Eng. degree in electronics engineering from the École Nationale Supérieure de Techniques Avancées Bretagne, Brest, France, in 1995, and the Ph.D. degree in electronics from the Université de Rennes I, Rennes, France, in 2002.

He spent seven years (1995–2002) as a Military R&D Engineer with the Délégation Générale pour L'Armement, working in underwater acoustics for hydrographic applications and mine warfare. In 2003, he joined the École Nationale Supérieure des Télécommunications de Bretagne, a French graduate engineering school in telecommunications, Brest. He is currently a Professor with research interests including interferometry and array processing.

Jean-Marc Le Caillec (Senior Member, IEEE) was born in 1968 in Paris, France. He received the engineering degree in telecommunications from the École Nationale Supérieure des Télécommunications de Bretagne (Telecom Bretagne), Brest, France, in 1992, and the Ph.D. degree in mathematics and signal processing from the University of Rennes I, Rennes, France, in 1997.

From 1997 to 1999, he was with Thomson AirSys (now Thales AirSys). He joined Telecom Bretagne, Brest, as an Associate Professor, in 1999, and became a Full Professor, in 2007. He is currently a Professor with the Department of Information and Image Processing (2IP), Telecom Bretagne. His main research interests include statistics, nonlinear system modeling, mathematics, and signal processing for applications in remote sensing and finance.

Prof. Le Caillec is in charge of the Division CID (Knowledge, Information, Decision), Laboratory of Sciences and Technics for Information, Communication and Knowledge, Lorient, France.