AUTOMATIC INTEGRATION OF LASER SCANNING AND PHOTOGRAMMETRIC POINT CLOUDS: FROM ACQUISITION TO CO-REGISTRATION

T. Partovi[1]*, M. Dähne[2], M. Maboudi[1], D. Krueger[2], M. Gerke[1]

[1] Institute of Geodesy and Photogrammetry, Technische Universität Braunschweig, Germany
(t.partovi, m.maboudi, m.gerke)@tu-bs.de
[2] Gesellschaft zur Förderung angewandter Informatik e. V. (GFaI), Berlin- (mdaehne, krueger)@gfai.de

Commission II, II/1

KEY WORDS: Terrestrial Laser Scanning (TLS), Hand-held Camera Imagery, Synthetic Image, Occlusion Detection, Co-registration

ABSTRACT:

Laser scanning systems have been developed to capture very high-resolution 3D point clouds and consequently acquire the object geometry. This object measuring technique has a high capacity for being utilized in a wide variety of applications such as indoor and outdoor modelling. The Terrestrial Laser Scanning (TLS) is used as an important data capturing measurement system to provide high quality point cloud from industrial or built-up environments. However, the static nature of the TLS and complexity of the industrial sites necessitate employing a complementary data capturing system e.g. cameras to fill the gaps in the TLS point cloud caused by occlusions which is very common in complex industrial areas. Moreover, employing images provide better radiometric and edge information. This motivated a joint project to develop a system for automatic and robust co-registration of TLS data and images directly, especially for complex objects. In this paper, the proposed methods for various components of this project including gap detection from point cloud, calculation of initial image capturing configuration, user interface and support system for the image capturing procedures, and co-registration between TLS point cloud and photogrammetric point cloud are presented. The primarily results on a complex industrial environment are promising.

1. INTRODUCTION

Recently, the rapid development of laser scanning (LS) technology makes it possible to obtain a wide range of 3D spatial data directly in a short time. These active sensors allow to capture the object geometry accurately. Besides, they are almost independent of lighting conditions. LS offers new possibilities for many applications such as Building Information Modelling (BIM), documentation of cultural heritage, infrastructure inspection, and additive manufacturing in construction (Aryan et al., 2021; Mattia Previtali et al., 2020; M. Maboudi et al., 2020).

The point clouds can be captured with static, mobile LS (MLS) or the combination of them in the sensor and point cloud levels, both indoor and outdoor. The static LS or Terrestrial Laser Scanner (TLS) captures the surrounding object geometry from one viewpoint in the so-called panorama mode, while using MLS measurements are carried out when the platform is moving. Static LS systems capture point cloud with higher-density and with less noise than MLSs and consequently more detailed information from the object (Mehdi Maboudi et al., 2018). Therefore, TLS is commonly utilized for small areas containing detailed information, for instance, cultural heritages or construction sites. MLSs can scan a large area or a network of transport corridors in a short time but currently with lower quality.

TLS is a line-of-sight instrument with specific angular resolution. Accordingly, it typically cannot provide a full scene coverage by only one scan in complicated scenes. Therefore, the captured dataset may contain gaps, and/or in some areas with low density. To overcome these deficiencies, there are two main approaches: 1) the trivial solution is to collect the partially overlapping point clouds from different locations through multiple scans, 2) covering the incomplete area of interest (AOI) using other sensors such as very high-resolution cameras (Meierhold et al., 2010; Moussa, 2014). The object can be captured by hand-held digital cameras from different orientations and positions with a higher resolution while the position and practical field of view of TLS are restricted and even using multiple scans may not guarantee full coverage of the scene. Moreover, details in camera images are better identifiable than TLS data in a sub-pixel accuracy level. On the other hand, for the large or hard to access objects (buildings, planes or ships) using merely the digital camera is technically difficult in practice. The images can be used as complementary dataset together with TLS data where either a more detailed information is required or in occluded areas.

To combine TLS data and images, the co-registration should be performed in which TLS coordinate system is considered as the reference coordinate system. A fully automatic and robust co-registration of TLS and photogrammetric based point cloud has been still active and interesting research topic. This motivated a collaboration of the Institute of Geodesy and Photogrammetry of Technical University of Braunschweig (Brunswick) and the Society for the Promotion of Applied Computer Science in Berlin in a project called LaScaBi to develop an application for registering hybrid data sets including TLS point clouds and image -based point clouds, efficiently, and accurately. The main innovation of our approach is the end-to-end-combination of partly existing methods, from initial TLS-data capturing to camera view planning, user guidance, data co-registration. In this paper, the preliminary results of this project are presented and discussed. Different research and development components of the project are briefly overviewed in the following sections. In Section 2, the related works for each project component are reviewed. Section 3 explains the main steps of the research and development components including AOI detection, synthetic image generation from TLS point clouds, co-registration of TLS and hand-held camera images, and user guidance and support system. In section 4, we show the primarily results for each project component and discuss them. Finally, the closing discussion and conclusion are presented in section 5.

2. RELATED WORKS

Our project constitutes different components such as gap detection, user guidance for capturing datasets, and fully automatic co-registration of TLS and camera imagery. State-of-the-art research papers for each component of the system are briefly explained in this section. Generally, the methods of the gap detection in the previous research papers are categorized into two main groups depending on the processing of point cloud in a direct or indirect manner using the ray-tracing algorithm. The direct methods are applied to the unorganized point clouds, finding the geometrical relation between neighbouring points by generating either the voxels in a volumetric representation or the triangulation mesh surface (Guo et al., 2018; Kazhdan et al., 2006)

In the ray-tracing based methods, the sensor positions and raw data (range and angles instead of the point cloud) is required. In these methods, the visibility of the point cloud from any viewpoints are investigated. The visibility of the point clouds can be mathematically checked using the intersection of line of sight of the rays with the voxel or surface triangles or analysing the triangulated surface normal (B. Alsadik et al., 2014). The Voxel based approaches are usable for the visibility detection and consequently gap detection in computer graphics applications, since they present the points as voxels which are efficient in term of computer memory. The voxel-based approaches are based on two different techniques including voxel-ray intersection, voxel distance buffering. Adan and Huber, (2011) applied ray-tracing and occlusion labelling in the point clouds using support vector machine classifier to detect opening areas from the gaps. Previtali et al., (2014) also used segmentation of planes and ray-tracing based method to detect the openings and occluded areas for indoor modelling. For the gap detection in TLS point clouds of the outdoor environment, Alsadik et al., (2014) proposed a method based on volumetric space representation and ray-tracing followed by a classification into gap or opening area.

Using AOI detection, the user can be guided for taking the image block in the appropriate camera positions. The image block is later employed to generate dense point cloud using Structure from Motion (SfM). Subsequently, the photogrammetric point cloud is combined with TLS point cloud to fill the gaps and to provide more information in the AOI. Therefore, the determination of suitable positions of the cameras assists the user to capture image block with enough overlap and proper configuration. Alsadik et al., (2013) designed an optimum camera network from the video frames of which a rough point cloud is generated. Furthermore, the visibility of the object is considered as an important factor for designing the network in such a way that one point of object should be visible from at least three images. Schindler and Förstner, (2012) developed and designed a software for real time guidance of unexperienced user to reconstruct 3D scene. Captured images are immediately integrated into the bundle adjustment and help the user to find the next best camera position. Wenzel et al., (2013) determined the camera position and image scale based on the accuracy requirement.

The registration between digital camera imagery and laser scanning are performed using different features such as points, corner, blobs, line, and plane. (Urban & Weinmann, 2015) used Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Oriented Fast and Rotated BRIEF (ORB), Binary Robust Independent Elementary Feature (BRIEF), and AKAZE detectors and descriptors for feature matching between two panoramic intensity images generated from reflectivity

values of the TLS point cloud. They also suggested the combination of feature detector and descriptor for obtaining higher accuracy and better performance in the feature matching. Hussnian et al., (2016) suggested a method for feature matching between aerial images and mobile laser scanning using the modified Harris corner detector and Learned Arrangement of Three Patch Codes (LATCH) descriptor. Their proposed method finds the correct correspondences if two images have the similar scale and rotation. The feature matching between intensity image generated from TLS and camera image is investigated by Forkuo and King (2004). Meierhold et al., (2010) used SIFT detector and descriptor for matching between an intensity image obtained from central perspective projection of TLS point clouds and camera images. Moussa (2014) used Affine SIFT (ASIFT) for feature matching between RGB images obtained from TLS and camera images with wide viewing angles. Most of the mentioned feature detector and descriptors work well if two images have the similarity in geometry and intensity properties, otherwise the results are not promising.

Moussa, (2014) estimated the exterior camera orientation relative to the TLS point cloud by solving Perspective-n-Point (PnP) problem which is based on space resection. Urban and Weimann, (2015) suggested the robust method for registering the point clouds obtained from two scan positions. In this method, 3D to 3D sparse point clouds obtained from 2D corresponding keypoints of the panoramic images are registered indirectly using 3D to 2D co-registration. One of the non-reference sparse point clouds are back projected into the virtual image. Then, the PnP algorithm combined with RANSAC the so-called Efficient PnP (EPnP) is used to estimate the transformation matrix between 2D and 3D feature correspondences. This coarse registration is followed by the iterative closest points (ICP) to perform fine registration and to increase the accuracy of the transformation. Meierhold et al., (2010) extracted 3D coordinates of corresponding keypoints from TLS. The space resection is applied to estimate the exterior parameters of the four images and interior parameters of the camera using 2D and 3D corresponding keypoints.

The focuses of our work are ray tracing and voxel intersection-based gap detection, the adequate camera positions determination, and the keypoint-based point cloud registration. The image matching between the reflectance image of TLS point cloud and images of digital camera is sensitive to the illumination and changing of viewing directions. The potentiality of Brisk feature detector and descriptor in image matching between intensity image obtained from TLS point cloud and camera images is investigated, which has not been performed in the previous works to the best of our knowledge. In our own approach we employ a combination of Affine transform and Brisk detector and descriptor (ABrisk), which will be explained in detail in section 3.4. Inspiring the suggested methods of (Urban and Weimann, 2015) and (Meierhold et al., 2010), we proposed a 3D to 3D co-registration method to estimate the transformation between TLS and image block-based point clouds.

3. RESEARCH AND DEVELOPMENT COMPONENTS

3.1 AOI Detection

We define AOI as gaps or some other regions where the user needs more information from those parts. These regions might be for example an object or a gauge for the heating pipe in the industrial environment. The position of the TLS and the corresponding sampling density of the measurements may cause some imperfections such as gaps in the point clouds. These

imperfections are recognized by processing the TLS point clouds in the primarily step. In indoor and outdoor reconstruction, the gaps of point clouds are mostly due to either windows/doors or existence of the other objects closer to the scanner, respectively. According to the TLS data capturing principle, we employ a gap detection method based on ray tracing and voxel ray intersection. The main steps in the presented approach are the 3D voxel representation of point clouds, ray tracing, detecting visible and occluded voxels, and voxel labelling.

3.1.1 Voxelization of 3D Point Cloud

The first step of the proposed gap detection algorithm is the volumetric representation of the point cloud based on the voxels. The goal of the 3D voxel representation of a space is to define an environment with imposed topology. A reduced and discrete 3D space is created through voxelization in which each voxel could be labelled as empty or occupied.

3.1.2 Ray Tracing

The ray is emitted from a sensor to the objects which is defined by line parameters in spherical coordinate system with $R = (\rho, \varphi, \theta)$ which ρ is maximum range in the 3D point cloud and $\varphi \in [\varphi_{min}, \varphi_{max}]$ is the vertical angle and $\theta \in [\theta_{min}, \theta_{max}]$ is the horizontal angle of the laser beam , which are changed with an angle interval. The domain of changing of φ and θ are defined according to the AOI of which the gap area to be detected. Points on a ray are sampled using Bresenham algorithm (Bresenham, 2010) with a precision determining the interval between the sequential points on the line. The voxels containing the line points are selected as candidate voxels for the next step and labeled as either visible or occluded.

3.1.3 Detecting and Labelling Visible and Occluded Voxels

The closest candidate voxel to the sensor position that contains some points is labeled as a visible voxel. We proposed a new method to avoid labeling neighboring voxels of a visible voxel as occluded voxels. It is assumed that a vertical plane passes through the center of the visible voxel. Afterwards, the distance of other centers of candidate voxels except for the center of visible voxel to this plane are compared with a threshold determined based on the diagonal of the voxel to consider them as visible or not. The farthest candidate voxels which are not in the list of visible voxels are chosen. The rest of the candidate voxels on the ray including occupied and non-occupied are labeled as the occluded voxels. For each new generated ray, the algorithm searches for the visible and occluded voxels.

3.2 Calculation of Camera Positions for User Guidance

After AOI detection, the preliminary camera positions configuration should be designed to guide the unprofessional user for capturing the images. A typical network is designed, which is a sequence of normal images. The designed camera positions are affected by many parameters, such as camera properties (e.g. focal length), distance to the AOI, the distance between stations, the position and size of AOI, and the available data capturing zone around the AOI. The image acquisition is commonly planned according to the precision and resolution requirements (Wenzel et al., 2013). In the stereo normal imaging, the relation between depth and other configuration parameters can be approximated as:

$$-Z = \frac{fB}{dp} \quad \Rightarrow \quad d = -\frac{fB}{Zp} \qquad (1)$$

where Z is the depth or distance from the object, B is the baseline, f is the focal length, p and d are the pixel pitch and disparity. The depth precision (σ_Z) is calculated based on the propagation of the variance in depth in respect to the variance in image.

$$\sigma_Z = \frac{pZ^2}{fB} \sigma_d \qquad (2)$$

where σ_Z is the depth precision and σ_d is the precision of disparity.

According to (2), the depth precision mainly depends on two components: pZ/f which represents Ground sample distance GSD and Z/B which affects the intersection angle of the rays. The later component is related to occlusions. It means that the point of the object is visible if the absolute difference between the angle between optical axis of the camera and surface normal in a point is less than a threshold (e.g. < 90°) (Wenzel et al., 2013; Alsadik et al., 2013).

According to the required GSD and depth precision, the distance from an object and the baseline are to be determined. An appropriate completeness including high image similarity as well as high performance of the matching procedure is achieved based on small baseline and large distance. However, the weak geometrical condition leads to poor depth precision. In contrast, a large baseline, and close distance achieve a higher precision. Its drawback is related to corresponding lower image similarity. Therefore, a trade-off between the baseline and distance from the object should be considered to satisfy the required depth precision and matching performance. Our suggested approach is firstly to consider a fixed focal length of the camera, in order to guarantee a stable estimation of interior camera parameters. The maximum possible distance from an object is calculated considering the specified GSD (e.g. =1mm). In close-range photogrammetry, the overlap between two neighbouring images is considered high >85% in frontal and 65% side directions. The baseline is calculated based on overlapping part and footprint.

The depth precision is calculated using (2) and compared with the required depth precision ($depth_{precision}$). If it is not satisfied, the distance from the object will be reduced. After that, the baselines are calculated for each strip of the image. We repeat these steps till we can get a $\sigma_Z \le depth_{precision}$. For convergent image configuration, we consider circular configuration especially for cylindrical objects such as monuments. In this design, the intersection angle of two conjugant rays which depends on Z/B is an important parameter. It can be considered between 5° and 10° resulting in the appropriate overlap between the images and noise reduction in the dense point cloud (Wenzel et al., 2013), according to depth precision which has been already explained.

3.3 Calculation of Synthetic Images from TLS point cloud

Synthetic image can be generated using two approaches based on imaging of polar scanner coordinates called reflectance panoramic image and central perspective representation (Meierhold et al., 2010). In the former approach, each 3D point is assigned to one pixel based on the scan resolution. The maximum and minimum horizontal and vertical angles of TLS determine the width and height of the image, respectively. The later approach projects the TLS data to a virtual image plane based on the collinearity equation. This method is more applicable than former one since it is possible to generate the virtual images with perspective geometry. Using the approximate interior orientation parameters of the camera which is going to be

used for data capturing, a block of distortion free synthetic images could be generated.

To generate synthetic image from the 3D point cloud of the TLS, position and orientation of the camera are required to be retrieved in the specific alignment of TLS. The TLS rotates around its vertical axis with a horizontal angle (AZ) to measure 3D point cloud (Al-Manasir & Fraser, 2006; Omidalizarandi et al., 2019). To align TLS coordinate system to the virtual camera coordinate system, the TLS measurements are rotated around the Z-axis of the TLS by using a 3×3 rotation matrix ($R_3(AZ)$). Fig. 1 shows the relation between two coordinate systems from top view.
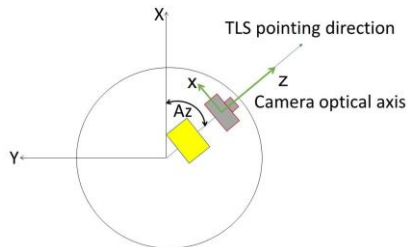
Figure 1. Relation between TLS coordinate system and virtual camera coordinate system.

In addition, the Z-axis of the digital camera (i.e. its viewing direction) is in direction of Y-axis of TLS. Therefore, a rotation of 90° around X-axis needs to be imposed to bring TLS coordinate system to camera coordinate system.

The exterior orientation parameters of the virtual camera, with respect to the TLS coordinate system is determined based on the detected AOI. The transformation between two Cartesian coordinate system is described as follows:

$$\begin{bmatrix} r \\ s \\ q \end{bmatrix} = R_{\kappa\varphi\omega} \times \left[R_3(AZ) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} \right] \qquad (3)$$

where (r, s, q) are the 3D point coordinates in the camera coordinate system, (X, Y, Z) are the 3D point cloud coordinates in the object space (i.e. TLS coordinate system), (X_o, Y_o, Z_o) are the translations vector between TLS and digital camera, $(\kappa, \varphi, \omega)$ are the rotation angles between TLS and digital camera, and AZ is horizontal angle measurement of the TLS. Transforming the virtual camera coordinate system (r, s, q) to the image coordinate system (x, y) is performed as follows:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{q} \left(\begin{bmatrix} f & 0 & x_p \\ 0 & f & y_p \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} r \\ s \\ q \end{bmatrix} \right) \qquad (4)$$

where f is focal length and x_p, y_p are the principle point shifts ($x_p = w/2$ and $y_p = h/2$ are defined on the basis of interior orientation of the hand-held camera. To convert (x, y) from metric to pixel (u, v) coordinates, each metric coordinate is divided by pixel size. Since the density of the TLS data could be coarser than the resolution of the camera, the pixel size of the virtual image is defined as a factor of the camera's pixel pitch. To enhance the synthetic image visually, normalization and gamma correction are applied on the reflectivity values from TLS data.

3.4 Co-registration of TLS Data and Hand-held Camera Images

3.4.1 Feature Detection, Description and Matching

In this paper, Affine Binary Robust Invariant Scalable Keypoints (ABrisk) feature detector and descriptor is used for feature matching. ABrisk generates a set of images with different view of direction from the original one by varying the two camera axis orientation parameters (Yu & Morel, 2011). Then it applies Brisk to all generated images. ABrisk as complementary algorithm of Brisk covers all six parameter which are needed for affine invariant matching. It simulates three parameters of scale, the camera's longitude and latitude angles, and normalizes the other three parameters of translation and rotation.

Brisk is a feature point detection and description algorithm with scale, rotation invariance, developed by (Leutenegger et al., 2011). The keypoint detection methodology of Brisk is inspired by AGAST algorithm (Mair et al., 2010) to extract the keypoints. The algorithm searches for the maxima not only in the image plane but also in scale space using FAST score, which can achieve the scale invariance properties (Leutenegger et al., 2011). Brisk descriptor is represented by a binary bit string. This descriptor employs fixed neighbourhood sampling pattern to describe the feature points. Firstly, a pattern is used for sampling the neighbourhood of keypoint. Four concentric circles containing 40×40 pixels are centred around the keypoint. N ($N = 60$) points with uniform distribution and equally spaced are sampled from the four concentric circles. The smoothed intensity values of the pixels of pair points on the pattern are used to estimate the local gradients. Accordingly, the overall characteristic pattern direction of keypoint is estimated. To build a scale and rotation invariance, the sampling pattern is then rotated around the keypoints based on the rotation angle calculated from previous step. To supply the illumination invariance properties, the results of brightness tests employed on the pair points of rotated sampling pattern are concatenated to generate the descriptor as a binary string. The length of bit string depending on the sampling pattern and the distance thresholds might be changed from 64 to 512 bits.

The feature matching strategy uses nearest neighbor distance ratio to find the corresponding keypoints in the two images. In this matching scheme, the distances of a keypoint descriptor are compared in the first image to the nearest neighbor and to the second nearest neighbor in the second image (Weinmann et al., 2011). The ratio of these distances must be less than a given threshold $t \in (0,1)$. As the ABrisk describes the keypoints in a binary string, the Hamming distance is utilized to calculate the distance between the descriptors of the keypoints. Hamming distance calculation is implemented based on a bitwise XOR operation. The two bits with equal values result in "0", otherwise it is "1" in the final bit string. Therefore, the large "1" numbers shows the dissimilarity of the two descriptors (Liu et al., 2018).

Images captured from a scene using different cameras with different viewpoints are related by epipolar geometry. The fundamental matrix representing interior and exterior orientation parameters of two camera is determined under the epipolar constraint which maps a point in one image to the line (epiline) in another image. To find the correct matches and their fundamental matrix, PROSAC algorithm (Chum & Matas, 2005) is used in this paper. PROSAC algorithm is an improved algorithm based on RANSAC. The sampling process is on the progressively larger sets of top-ranked correspondences while RANSAC considers all correspondences equally and draws the

random samples uniformly from the full set. Under this assumption, PROSAC can predict the correct correspondences based on the similarity rather than random guessing (Chum & Matas, 2005).

3.4.2 Feature based 2D to 3D Registration

Feature based 2D to 3D registration intends to find the camera calibration parameters in TLS coordinate system which is considered as the reference. The updated coordinate system of an image block results in point cloud obtained from dense matching in TLS system. To this end, the 3D coordinate of the correct correspondences between the synthetic image and at least three camera images can be considered as Ground Control Points (GCPs) in the process of dense image matching using SfM. Due to radiometric and geometrical differences between the synthetic image and the camera image, finding the correct matches between the synthetic image and a block of camera images might be difficult. Therefore, an indirect transformation is suggested for finding the transformation between image and TLS coordinate systems. In this step, a sparse 3D point cloud is generated from SfM. The corresponding 3D coordinates of the 2D keypoints of the camera image which has maximum correct correspondences with synthetic image is identified from the sparse 3D points. Finally, the transformation between the image block and TLS including rotation, translation, and scale is solved by calculating the similarity transformation matrix of 3D coordinate of the corresponding keypoints between synthetic image and camera image. The coarse registration can be followed by ICP to do a fine registration and find the transformation parameters more precisely.

3.5 User Interface and Support System Software Design for Demonstrator Application

To join all the separate working packages with their functions and algorithms an application has to be designed for the LaScaBi project. The purpose is to give the user the ability to get an overview of the scanned data, interact with them and use the developed functions on selected data sets. Figure 2 shows the software design to achieve this goal.
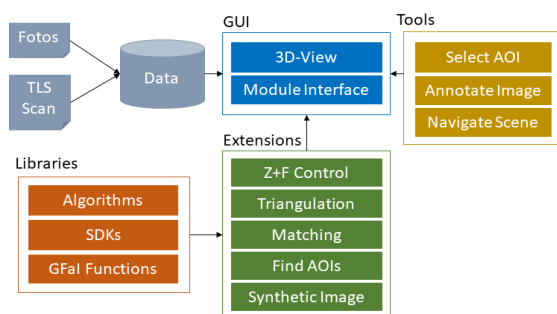

Figure 2. Design of demonstrator application.

To group the functions of the separate working packages of this project, four modules was defined. The first one is the Graphic User Interface (GUI) module which includes the embedded 3D-View, where the loaded data are displayed, and an interface for all other modules to add elements to the GUI. The Tools module is the second, which adds functions to the application where algorithms are needed. The third type includes the extensions, which are more complex functions that load and manipulate data sets. This module uses implemented algorithms to calculate new data, provided SDKs to communicate with hardware and already implemented functions from the internal GFaI library. With this software design, the application can be extended during the

project duration when a new algorithm or function is completed without major changes to already finished modules. That means every module has to be independent from other modules, so an agile and stable software development is possible.

4. RESULTS AND DISCUSSION

4.1 Data set

The room where test data was acquired, contains many pipes and facilities as well as two columns which could serve as AOIs and may lead to different gaps in the dataset. The room was scanned from one position using the TLS of type Z+F IMAGER® 5010X in a high-quality mode with an angular resolution of 0.036°. The RGB images are captured using Canon EOS 5D mark IV with the fixed focal length of 28 mm and pixel pitch of 4.24 μm. Figure 3 (a) shows the 3D point cloud of the room after separation of the corresponding roof part of the data set. The AOI is the gap areas located behind the column in the right corner of the room. The point cloud and one of the camera images are illustrated in Fig. 3 (b) and (c), respectively.
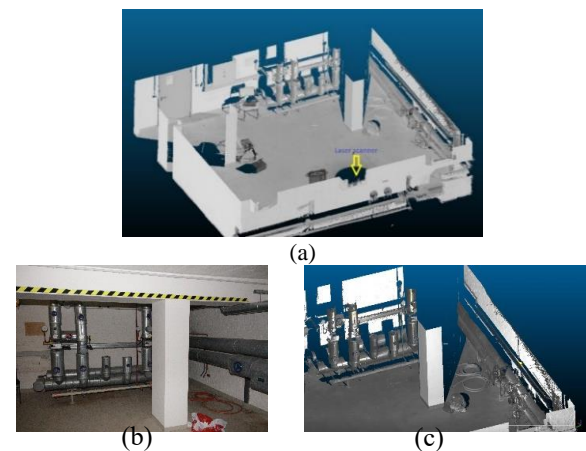

Figure 3. Test data set: (a) 3D point cloud, (b) camera image of AOI, (c) 3D point cloud of AOI.

4.2 Experiments

Our experiments focus on the gap detection, feature extraction and matching between each single camera image and synthetic image of one scan to find the correct keypoint correspondences between them. Additional images at different perspective and distances considering depth accuracy of 1 mm and sub-millimetre GSD are captured from the aforementioned AOI and its surrounding to build an image block. Therefore, it allows to generate sparse point cloud using SfM as well as dense point cloud. Finally, the transformation matrix between sparse point cloud obtained from image block and TLS point cloud is calculated using the 3D coordinates of correct corresponding keypoints resulted from feature matching.

4.2.1 Gap Detection

The algorithm explained in subsection 3.1 is employed to detect the gap area shown in Fig. 3 (a). The position of the TLS is depicted by the yellow arrow. From this position, the column on the right side is an obstacle for scanning a part of the room in behind and causes the gap in the point cloud. The proposed algorithm which is based on ray-tracing and voxel intersection method requires some parameters as shown in Table 1. These parameters depend on the point cloud density and the complexity of the environment. To detect the gap area, the horizontal and

vertical angle ranges are set to $\theta \in [270, 360]$ and $\varphi \in [0, 135]$, respectively.

Parameters	Value
Number of voxels	128x128
The minimum number of occupied voxels along the ray	2
The maximum distance of the neighbouring voxel from visible voxel	0.2 m
The increment of θ angle	0.3°
The increment of φ angle	0.3°

Table 1. Parameters of the gap detection algorithm.

The result of the gap detection is visualized in Fig. 4. This figure shows a volumetric representation of the point clouds using occluded and visible voxels. The detected gap areas (red voxels) of the AOI behind the column and pipes are recognizable from the other visible occupied voxels (yellow voxels).

— voxels in the gap areas
— visible voxels occupied by point cloud
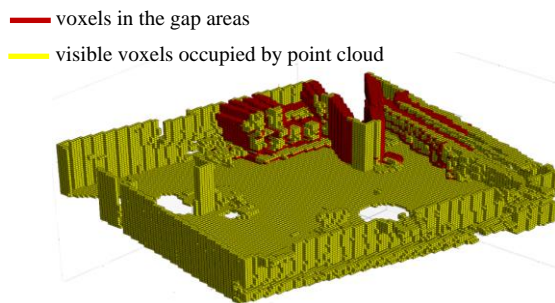
Figure 4. Voxelization of the test data set and detected gap areas

4.2.2 Feature Matching

To generate the central perspective-based synthetic image from AOI, the horizontal angle of TLS is considered as $AZ = 128°$. The synthetic image is generated with the same properties as camera image consisting of the focal length ($28\ mm$) and sensor size ($24 \times 36\ mm$). Since the density of the point cloud is lower than the camera images, the resolution of the synthetic image is considered as 5×pixel size (i.e. $5 \times 4.23\mu m = 21.19\ \mu m$) of the camera image. Therefore, it results in an image with size of 1158×1737 pixels compared to the raw the camera image size of 5792×8688 pixels. The generated synthetic image is shown in Figure 5.
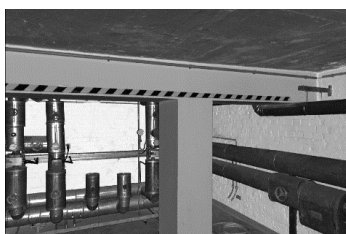

Figure 5. The central perspective synthetic image.

The OpenCV library is used in the python scripting language for applying Brisk feature detector and descriptor with affine transform (ABrisk) on the synthetic and camera images. To obtain Brisk feature detector and descriptor parameters, without loss of generality, the default values of the corresponding functions in the OpenCV library are considered. Figure 6 depicts the feature keypoints extracted from synthetic image (a) and camera image (b).
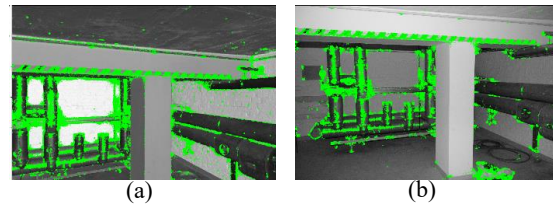

Figure 6. The ABrisk feature keypoints extracted from (a) synthetic image and (b) camera image.

After extracting the keypoints using ABrisk feature detector, the algorithm searches for reliable point correspondences by employing a distance ratio test using a threshold value of $t = 0.66$ which is empirically obtained. Figure 7 shows the results of this primary feature matching.
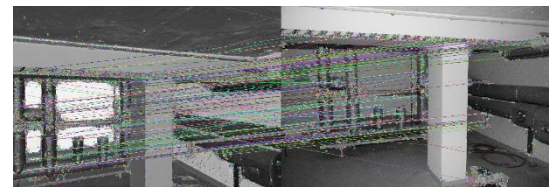

Figure 7. Feature matching using distance ratio test.

As shown in this figure, there are some outliers which are removed after applying the estimated fundamental matrix with PROSAC. Figure 8 shows the result of feature matching after outlier removal.
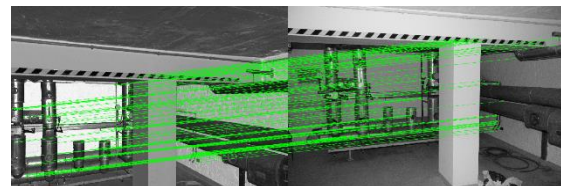

Figure 8. Feature matching after outlier removal

The number of extracted feature points from synthetic image and camera image equals to 75057 and 57211, respectively. This difference might be due to either noise or low spectral resolution of the synthetic image generated from reflectivity values.

The number of correspondences found by ABrisk feature descriptor with imposing the ratio test is 297 and after estimating fundamental matrix using PROSAC is 198. From this result, it is obvious that about 66% of the ABrisk correspondences is remained after using fundamental matrix and PROSAC, which is adequate for co-registration between TLS point cloud and image block.

4.2.3 Co-registration between TLS and Image Block

As illustrated in Fig. 9, 52 camera images are calibrated to generate a sparse 3D point cloud using SfM algorithm in Pix4D software. The 3D coordinates of correct corresponding keypoints generated by matching between a TLS synthetic image and one of the undistorted camera images (with the closest perspective to the synthetic image) are obtained from TLS point cloud and 3D sparse point, respectively. Afterwards, the rigid-body transformation matrix including 3D rotations and translations, as well as scale is calculated using the 3D coordinates of correct correspondences.
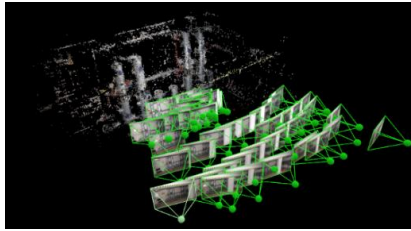
Figure 9: The generated sparse point cloud using SfM

To exclude the possible outliers and having a robust estimation of the transformation parameters, the 3D residuals of the corresponding keypoints are computed after transformation with respect to their corresponding TLS coordinates. The outliers are eliminated by performing 2σ test in five iterations. The L2-norms of the residuals are computed for each correct corresponding point and illustrated in Fig. 10. As shown in this figure, the L2-norms of the residuals are mostly close to zero and only a few of them are around 2 mm, which means the fine registration might be neglected.
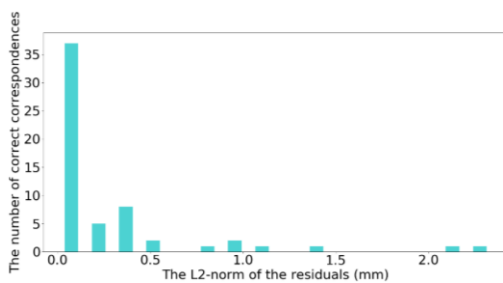

Figure 10. L2-norms of the residuals for correct corresponding keypoints.

Although the results in Fig. 10 look very good, it is worth mentioning that these are the residuals of the 58 points which are used in 3D rigid body transformation between the point clouds and should not be considered as the determinant criteria for evaluation of the whole process. Therefore, we will also compute and report the deviation of the photorammetric point cloud (after transformation) from TLS Point cloud.

Using final transformation matrix, the exterior orientation parameters of the camera images are calculated in the TLS coordinate system. The generated dense photogrammetric point cloud is then integrated with TLS point cloud to complete the gap areas as shown in Fig. 11.
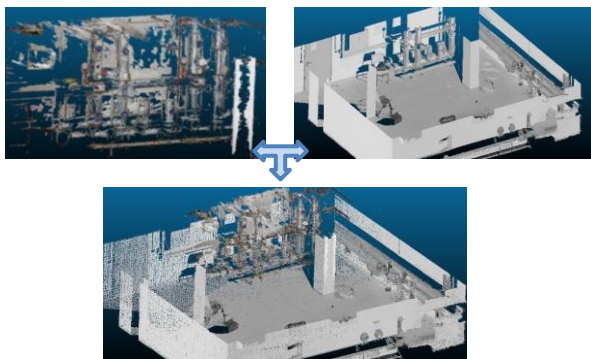

Figure 11: The entire point cloud of the experimental room including TLS point cloud and SfM-based point cloud after transformation. Colorized point cloud is generated from SfM, and TLS point cloud is shown with grey colour.

The Cloud to Cloud (C2C) distances between transformed dense photogrammetric point cloud and TLS point cloud are computed. To avoid the effect of non-correspondent points in our analysis, we use a cut-off distance r which is the largest acceptable distance between TLS point cloud and the closest point in the photogrammetric point cloud. Hence, the C2C distances larger than the cut-off distance r will be excluded from the computations. Moreover, to further avoid the outliers, we use the median absolute distance to measure the deviation of the photogrammetric point cloud from TLS data. The median absolute distance of the point clouds as a function of cut-off distances are illustrated in Fig. 12 (a). For the cut-off distances lower than 1.5cm, the medians of absolute C2C distances are almost equal to cut-off distances. However, it is almost constant and equal to 13 mm for the cut-off distance of 15 mm and higher. The results reveal that the calculated similarity transformation from corresponding keypoints can give the promising results for the robust registration of the image-based dense point cloud to TLS data.
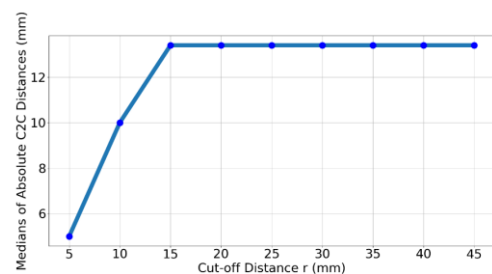

Figure 12. Median of absolute distances between SfM-based dense point cloud and TLS point cloud for each cut-off distance.

5. CONCLUSIONS

The LaScaBi project aims at a combination of TLS and photogrammetric point clouds in an automatic manner. In this paper, different components of the project as well as related methodologies are discussed. The result of the gap detection module shows that the proposed method can detect gap areas successfully. The geometrical parameters of the gap areas such as size and volume can be estimated in the volumetric space. This can be utilized as a constraint for determining the position of the cameras in the user guidance step. The results of feature matching module show that applying a binary detector and descriptor like Brisk with affine transform (ABrisk) can handle the different characteristics of camera and synthetic images derived from TLS. The proposed coarse registration of photogrammetric point cloud to the TLS point cloud using 3D coordinates of the 2D feature correspondences lead to the reliable and precise results for which the fine registration can be neglected. As future work, the proposed algorithm will be validated with on more images and different data sets. To design an optimum camera positions, we plan to use an optimization algorithm to prone a large amount of video frames considering the constraint obtaining appropriate depth precision. The 3D correspondences can be used as control points in a bundle adjustment integrated with robust estimation method to approximate the interior and exterior orientation of the cameras of the image block simultaneously. The use of a self-calibrating algorithms is suggested for bundle block adjustment in locally acquired overlapping image data and extracted object coordinate.

ACKNOWLEDGEMENTS

REFERENCES

Adan, A., & Huber, D. (2011). 3D reconstruction of interior wall surfaces under occlusion and clutter. *Proceedings - 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2011*, 275–281. https://doi.org/10.1109/3DIMPVT.2011.42

Al-Manasir, K., & Fraser, C. S. (2006). Registration of Terrestrial Laser scanner Data using Imagery. *The Photogrammetric Record*, *21*(115), 255–268. https://doi.org/10.1111/j.1477-9730.2006.00379.x

Alsadik, B., Gerke, M., & Vosselman, G. (2014). Visibility Analysis of Point Cloud in Close Range Photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*(5), 9–16. https://doi.org/10.5194/isprsannals-II-5-9-2014

Alsadik, Bashar, Gerke, M., & Vosselman, G. (2013). Automated Camera Network Design for 3D Modeling of Cultural Heritage objects. *Journal of Cultural Heritage*, *14*(6), 515–526. https://doi.org/10.1016/j.culher.2012.11.007

Aryan, A., Bosché, F., & Tang, P. (2021). Planning for Terrestrial Laser Scanning in Construction: A review. In *Automation in Construction* (Vol. 125, p. 103551). Elsevier B.V. https://doi.org/10.1016/j.autcon.2021.103551

Bresenham, J. E. (2010). Algorithm for Computer Control of a Digital Plotter. *IBM Systems Journal*, *4*(1), 25–30. https://doi.org/10.1147/sj.41.0025

Chum, O., & Matas, J. (2005). Matching with PROSAC - Progressive Sample Consensus. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, *I*, 220–226. https://doi.org/10.1109/CVPR.2005.221

Forkuo, E. K., & King, B. (n.d.). *Automatic Fusion of Photogrammetric Imagery and Laser Scanner Point Clouds*.

Guo, X., Xiao, J., & Wang, Y. (2018). A Survey on Algorithms of Hole Filling in 3D Surface Reconstruction. *Visual Computer*, *34*(1), 93–103. https://doi.org/10.1007/s00371-016-1316-y

Hussnain, Z., Oude Elberink, S., & Vosselman, G. (2016). Automatic Feature Detection, Description and Matching from Mobile Laser Scanning Data and Aerial Imagery. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B1*, 609–616. https://doi.org/10.5194/isprs-archives-xli-b1-609-2016

Kazhdan, M., Bolitho, M., & Hoppe, H. (2006). Poisson Surface Reconstruction. In *Eurographics Symposium on Geometry Processing*.

Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011). BRISK: Binary Robust invariant scalable keypoints. *Proceedings of the IEEE International Conference on Computer Vision*, 2548–2555. https://doi.org/10.1109/ICCV.2011.6126542

Liu, Y., Zhang, H., Guo, H., & Xiong, N. (2018). A FAST-BRISK Feature Detector with Depth Information. Sensors, 18(11), 3908. https://doi.org/10.3390/s18113908

Maboudi, M., Gerke, M., Hack, N., Brohmann, L., Schwerdtner, P., & Placzek, G. (2020). Current Surveying Methods for the Integration of Additive Manufacturing in the Construction Process. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B4-2*, 763–768. https://doi.org/10.5194/isprs-archives-xliii-b4-2020-763-2020

Maboudi, Mehdi, Bánhidi, D., & Gerke, M. (2018). Investigation of Geometric Performance of an Indoor Mobile Mapping System. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(2), 637–642. https://doi.org/10.5194/isprs-archives-XLII-2-637-2018

Mair, E., Hager, G. D., Burschka, D., Suppa, M., & Hirzinger, G. (2010). Adaptive and Generic Corner Detection based on the Accelerated Segment Test. *Lecture Notes in Computer Science*, *6312 LNCS*(PART 2), 183–196. https://doi.org/10.1007/978-3-642-15552-9_14

Meierhold, N., Spehr, M., Schilling, A., Gumhold, S., & Maas, H.-G. (2010). Automatic Fusion of Photogrammetric Imagery and Laser Scanner Point Clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII, Part 5*, 446–451.

Moussa, W. (2014). *Integration of Digital Photogrammetry and Terrestrial Laser Scanning for Cultural Heritage Data Recording*.

Omidalizarandi, M., Kargoll, B., Paffenholz, J. A., & Neumann, I. (2019). Robust External Calibration of Terrestrial Laser Scanner and Digital Camera for Structural Monitoring. *Journal of Applied Geodesy*, *13*(2), 105–134. https://doi.org/10.1515/jag-2018-0038

Previtali, M., Barazzetti, L., Brumana, R., & Scaioni, M. (2014). Towards Automatic Indoor Reconstruction of Cluttered Building Rooms from Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*(5), 281–288. https://doi.org/10.5194/isprsannals-II-5-281-2014

Previtali, Mattia, Brumana, R., & Banfi, F. (2020). Existing Infrastructure Cost Effective Informative Modelling with Multisource Sensed Data: TLS, MMS and photogrammetry. *Applied Geomatics*, 1–20. https://doi.org/10.1007/s12518-020-00326-3

Schindler, F., & Förstner, W. (2012). Real-time Camera guidance for 3D Scene Rreconstruction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *1*, 69–74. https://doi.org/10.5194/isprsannals-I-3-69-2012

Urban, S., & Weinmann, M. (2015). Finding a Good Feature Detector-Descriptor Combination for the 2D Keypoint-based Registration of TLS Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*(3W5), 121–128. https://doi.org/10.5194/isprsannals-II-3-W5-121-2015

Weinmann, M., Weinmann, M., Hinz, S., & Jutzi, B. (2011). Fast and Automatic Image-based Registration of TLS Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(6 SUPPL.), S62–S70. https://doi.org/10.1016/j.isprsjprs.2011.09.010

Wenzel, K., Rothermel, M., Fritsch, D., & Haala, N. (2013). Image Acquisition and Model Selection for Multi-View Stereo. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-5/W1*, 251–258. https://doi.org/10.5194/isprsarchives-xl-5-w1-251-2013

Yu, G., & Morel, J.-M. (2011). ASIFT: An Algorithm for Fully Affine Invariant Comparison. *Image Processing On Line*, *1*, 11–38. https://doi.org/10.5201/ipol.2011.my-asift