AUTOMATIC ORIENTATION ESTIMATION OF MULTIPLE IMAGES WITH RESPECT TO LASER DATA

Fangning He, Ph.D. student
Ayman Habib, Professor
Department of Geomatics Engineering
University of Calgary, 2500 University Dr. NW,
Calgary, Alberta, T2N 1N4, CANADA
hef@ucalgary.ca
ahabib@ucalgary.ca

ABSTRACT

In this paper, we present an automatic algorithm to estimate the orientation of multiple images with respect to terrestrial laser data. The proposed algorithm takes advantage of both conventional single-view and multi-view registration approaches. The algorithm consists of three steps. In the first step, intensity images are generated from the terrestrial laser data with intensity information, and the initial exterior orientation parameters (EOPs) of each camera image are then estimated from feature correspondences with the intensity images. In the second step, the initial EOPs obtained from previous step are refined through a bundle adjustment process. In the third step, the point cloud reconstructed from the bundle adjustment process is registered to the terrestrial laser point cloud to determine the final orientation of each image. The proposed algorithm is tested on experimental datasets, and the results demonstrate that the proposed algorithm can efficiently handle the orientation estimation of multiple images with respect to the terrestrial laser data.

KEYWORDS: registration, image matching, orientation estimation, point cloud

INTRODUCTION

Nowadays, 3D modeling of objects can be achieved through either active or passive remote sensing systems. Active sensors, such as laser scanners, are able to directly provide precise and reliable 3D information of scanned objects. However, the obtained 3D data (point cloud) usually lacks color information (especially when collecting data from mobile platforms). On the other hand, passive sensors, which commonly use digital frame cameras, can be incorporated for 3D reconstruction while providing spectral information of the mapped objects. This spectral information would allow for the derivation of more reliable semantic information when compared with active ranging sensors. However, the main challenge in deriving 3D information from imagery is the automated identification of conjugate features in overlapping images, which is known as the matching problem. Both active and passive approaches have their limitations, and cannot individually solve all the problems during 3D reconstruction of real world (González-Aguilera et al., 2009). Therefore, the integration of both active and passive approaches for 3D reconstruction can provide benefits for different applications.

Many approaches have been developed to register imagery with laser data. However, these approaches are not fully automated and time-consuming, mainly because of inefficient 2D-to-3D correspondences between imagery and laser data. This paper presents a fully automatic algorithm to estimate the orientation of multiple images relative to terrestrial laser data. The algorithm consists of three steps. In the first step, intensity images from laser data are created, and then the initial exterior orientation parameters (EOPs) of each camera image are estimated from feature correspondences with intensity images. In the second step, the initial EOPs obtained from previous step are refined through a bundle adjustment process. In the final step, the point cloud reconstructed from previous bundle adjustment process is registered to 3D laser points to determine orientation of each image with respect to laser data.

The remainder of this report presents the proposed algorithm in more details. First, a literature review of related works is given. Then, the proposed algorithm is presented. Finally, the experimental results and conclusions are shown respectively.

LITERATURE REVIEW

There has been a considerable amount of research efforts in registering images with laser scanning data. Considering the different strategies used for image registration relative to laser scanning data, current registration approaches can be divided into two different categories: single-view registration approach and multi-view registration approach. In the single-view registration approach (Stamos & Alien, 2001; Ding et al., 2008; González-Aguilera et al., 2009; Mastin et al., 2009; Sattler et al., 2011; Guan et al., 2013), the correspondences between 2D features from the image and 3D points from the laser data are first determined. Then, the orientation of each image is estimated separately. The single-view registration approach is easy and fast to implement. However, if there are multiple overlapping images, this approach doesn't consider the geometric constraints and feature correspondences among different images. On the other hand, instead of finding 2D-to-3D feature correspondences between each individual image and the laser points, in the multi-view registration approach (Liu et al., 2006; Mastin et al., 2009), a 3D model (3D point cloud) is usually generated from multiple images within a local coordinate system. Then, the 3D model is registered with the 3D laser points to determine the orientation parameters of involved images. The shortcoming of this approach is that generating a 3D model from multiple images can be very computationally expensive. Therefore, a new algorithm, which takes advantage of both single-view and multi-view registration approaches, is developed and presented in this paper. Reviews of the single-view and the multi-view registration approaches are respectively given in following sections.

Single-view Registration Approach

Since the registration of a single image to laser data can be formulated as a camera pose estimation problem, the pre-requisite of single-view registration approach is to establish feature correspondences between 2D image features and 3D laser points. One straightforward way to solve the feature correspondences problem is to extract features from both types of data and find direct 2D-3D correspondences. Sattler et al. (2011) developed a direct 2D-3D descriptor matching method using visual vocabularies quantization. In their method, both 2D point features in the image and 3D points in the point clouds are assigned visual vocabularies. Then, a features matching process is carried out based on a search through features with similar visual words. However, a prioritized correspondence search is required. Ding et al. (2008) proposed a fast and automated algorithm for registering oblique aerial imagery onto 3D geometric models obtained from LiDAR. In their algorithm, the vanishing points are used for initial angular estimation, and then 2D corners from the images are matched with orthogonal 3D structural corners in the 3D geometric models. However, this method requires GPS and compass measurements for coarse estimation of the image orientation, and it doesn't work for scenes without orthogonal 3D structural corners.

With the previous discussion, due to the characteristics of laser data, it's usually very hard to find exact 2D-3D point feature correspondences between 2D camera images and 3D laser data. Instead of direct 2D-3D point feature correspondences, structural features, such as building edges and other linear features, are used in many research efforts. For example, Stamos and Alien (2001) presented a semi-automatic method for image-to-model registration of urban scenes. In their method, 3D lines are extracted from point clouds of urban scenes, and matched with extracted edges from the images. Because this method involves parallelism and orthogonality constraints that exist in urban environments, this method only works for scenes containing linear features with strong geometry constraints.

One advantage of implementing direct 2D-3D feature correspondences is that the orientation parameters of each image can be directly computed through spatial resection. However, most direct 2D-3D registration algorithms only work for a certain type of images with special geometric constraints, which limit their ability to solve different registration problems.

Considering that current laser data usually contains intensity information, which could be very helpful for matching 3D point cloud with 2D camera images. Therefore, instead of the direct 2D-3D feature correspondences, 2D intensity images can be generated from 3D laser points, and 2D-2D feature correspondences can be applied between the camera images and the laser-based intensity images. As a result, an indirect 2D-2D-3D registration framework can be applied to determine the feature correspondences between the image and the laser points. Then, the orientation of each image can be estimated in the same way as the above-mentioned direct 2D-3D feature correspondences methods.

In the indirect 2D-2D-3D registration framework, similarity measures among different features, such as the normalized cross correlation, are usually used for 2D-2D feature correspondences between camera images and laser-based intensity images. For example, Mastin et al. (2009) presented a 2D-3D registration method based on mutual information. In their method, laser points are back-projected onto a 2D image plane, which is obtained from GPS/INS data or a user pre-defined image orientation. Then, mutual information between image point features and

the projected laser features are computed to update the estimated image orientation. González-Aguilera et al. (2009) proposed a robust 3-level feature matching strategy to match digital camera images with laser intensity images. In the proposed strategy, point features from different sources are initially matched in an image pyramid based on cross-correlation coefficient. Then, the initial matches are refined using Least Squares Matching. Finally, the fundamental matrix within the image stereo pair is estimated through a RANSAC algorithm to remove mismatches between the digital camera images and the laser intensity images. Similarly, Guan et al. (2013) developed an indirect feature matching algorithm to estimate the position of a hand-held camera with respect to terrestrial LiDAR data. In the method, the camera image is initially registered with a terrestrial LiDAR data using SIFT features. Then, the orientation of the camera is refined by iterative feature matching using Harris features.

The 2D-2D-3D registration methods, which usually convert 3D laser points into 2D intensity image, take the advantage of current 2D image processing techniques. However, the conversion from 3D laser point cloud to 2D laser intensity images may introduce additional errors and reduce the registration accuracy between images and laser points.

One common limitation of both the direct and the indirect 2D-3D single-view registration approaches (Sattler et al., 2011) is that they can't handle scenes lacking easily extracted points and lines. In addition, possible significant occlusions may also reduce the accuracy of estimated image orientation. Therefore, a multi-view registration approach, which uses multiple overlapping images, is usually an alternative option for registering images with laser data.

Multi-view Registration Approach

Different from the above-mentioned single-view registration approach, multi-view registration approach usually follows a 3D-3D registration framework. For example, Zhao et al. (2005) used iterative closest point algorithm (ICP) to align a point cloud derived from a sequence of video images to a 3D point cloud from a laser scanner. In their method, the 3D point cloud is first reconstructed from a sequence of video images relative to a local coordinate system. Then, the outcome from the image-based 3D reconstruction is matched to the geo-referenced laser points in order to estimate the transformation between the image-based local and the laser-based reference coordinate systems. Meanwhile, the authors also assume that the initial transformation parameters for the ICP algorithm are available from GPS measurements or manual geo-referencing process. Liu et al. (2006) also applied structure-from-motion algorithm on a group of images, and then implemented 3D-3D registration from the sparse point cloud, which was obtained from structure-from-motion, to the 3D range data.

There are several advantages of the 3D-3D registration framework. One is its capability to handle images with significant occlusions. For example, in the 2D-3D registration framework, if there isn't enough feature correspondences between the image and the laser points, the image orientation can be poorly estimated. However, in the 3D-3D registration framework, the image orientation can be determined through a relative orientation procedure. Another advantage is that the 3D-3D registration framework removes the demand for feature extraction, like corners and edges, from 3D laser points prior to the registration. However, the 3D-3D registration framework could fail for certain images and scenes. For example, for those images with weak intersection geometry, the reconstructed 3D point cloud may be distorted. As a result, the image orientation cannot be accurately estimated.

METHODOLOGY

The proposed algorithm takes advantages of both single-view and multi-view registration approaches. As shown in **Figure 1**, there are three steps in the proposed algorithm. In the first step, intensity images from laser data are generated, and then the initial EOPs of each camera image are estimated using feature correspondences with the laser-based intensity images. In the second step, the initial EOPs obtained in the first step are refined through a bundle adjustment process. In the third step, the Iterative Closest Patch (ICPatch) algorithm is applied to register the point cloud reconstructed from the bundle adjustment process to the 3D terrestrial laser points.

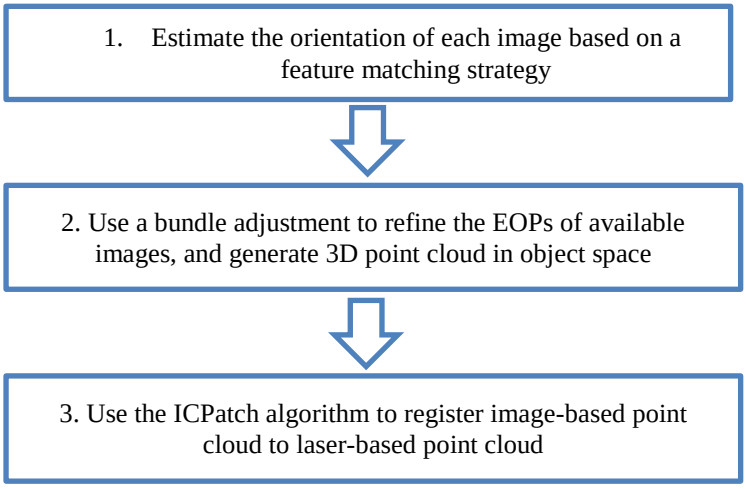


Figure 1. Proposed algorithm

Image Orientation Initialization

In this paper, the image orientation is estimated for distortion-free digital camera images. It means that the digital camera used for data collection is calibrated, and the distortion of each image is removed.

Synthetic Views of 3D Laser Data. As shown in Figure 2, virtual cameras with different viewing directions are placed around the 3D laser points. Then, laser points are projected onto the image plane of each virtual camera to generate intensity images. In this process, Z-buffering algorithm is implemented to handle the problem of occlusions.

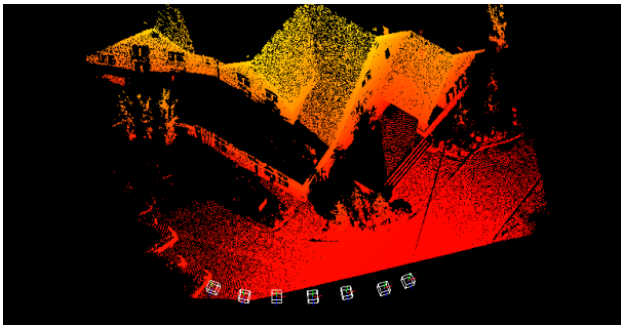


Figure 2. Virtual cameras can be placed around 3D terrestrial laser point cloud with different viewing directions and distances.

In different applications, the number of virtual cameras depends on the type of feature operators (like Harris operator, SIFT operator, and etc.) that is used for feature matching between camera and intensity images. If the extracted feature is able to handle rotation, scale and wide-baseline, less virtual cameras are needed. In contrast, if the extracted feature is neither rotation-invariant nor scale-invariant, more virtual cameras are needed.

Another problem of generating laser-based intensity images is the presence of holes and gaps in the intensity images, due to an insufficient density of laser points. To solve this problem, an appropriate pixel size has to be estimated and assigned to the virtual camera at each viewpoint (see **Equation (1)).**

$$pixel = \frac{fD}{H} \qquad (1)$$

Where:

$pixel$ is the pixel size of the virtual camera;
f is the principal distance of the virtual camera;
D is the point spacing in the point cloud;
H is the depth from laser point to the perspective center of the virtual camera.

However, this formula only considers the situation, in which all laser points fall on the same plane that is parallel to the image plane. Therefore, additional interpolation is then applied on each intensity image to fill gaps and reduce the influence of insufficient laser point density.

Feature Matching Process. In close-range photogrammetry applications, the baseline cannot be always kept constant and the rotation around each axis is usually significant. Therefore, the extracted features for matching digital camera images with laser-based intensity images should be invariant under different transformations.

In the proposed algorithm, Scale-Invariant Feature Transform (SIFT) features (Lowe, 2004), which are invariant to image scaling and rotation, are used for feature matching between the camera images and the laser-based intensity images. The proposed feature matching process is implemented in two different levels.

At the first level, the initial feature matching between the camera image and the laser-based intensity image is determined through a Euclidean-distance-based nearest neighbor matching approach. Then, matching error detection is carried out using forward/backward consistency check. As shown in **Figure 3(a)**, from the camera image to the intensity image, P1 matches Q1, and P2 matches Q2; from the intensity image to the camera image, Q1 matches P1, and Q2 matches P3. Through camera/intensity consistency check, (P1, Q1) is accepted as a correct, while (P2, Q2) and (Q2, P3) are discarded as mismatches.

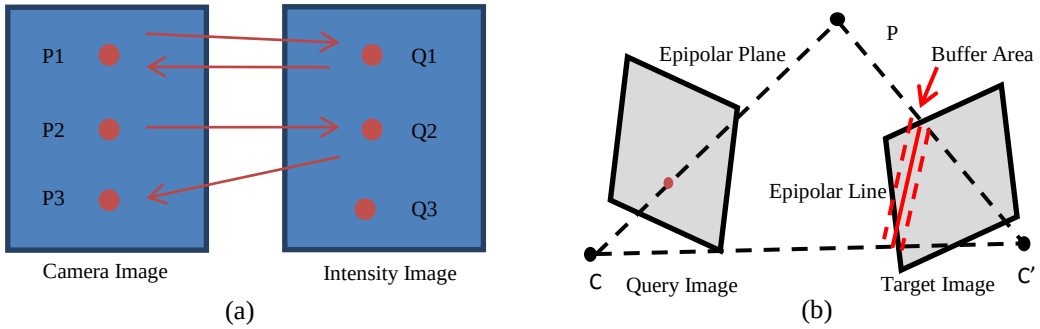


Figure 3. (a) Forward/backward consistency check; (b) feature matching along conjugate epipolar lines

At the second level, with the relative orientation parameters estimated from co-planarity model (Mikhail et al., 2001), we can generate more feature correspondences by limiting the feature corresponding search range along the epipolar line. As shown in **Figure 3.3(b)**, for each feature in the query image, its search space for corresponding features is reduced from the whole image to the epipolar line. A buffer area around the epipolar line is generated as the search space. In the search space, the feature with the minimum Euclidean-distance between feature descriptors is considered as the candidate match. The same forward/backward consistency check is also carried out at the second level to remove mismatches between the camera and the intensity images.

After applying the proposed feature matching process, the feature correspondences between camera image and laser-based intensity image can be determined.

Image Orientation Estimation.

- Relative Orientation Parameters (ROPs) Estimation

In general, the image orientation is directly estimated through a spatial resection process with corresponding 2D to 3D features. Instead, in this research, the relative orientation parameters (ROPs) between the camera image and the laser-based intensity image are initially estimated.

The well-known co-planarity model (e.g. Mikhail et al., 2001) is adopted to solve the relative orientation within an image stereo pair. The co-planarity model is presented in **Equation (2)**.

$$u_l^T \begin{bmatrix} 0 & Tz & -Ty \\ -Tz & 0 & Ty \\ Ty & -Tx & 0 \end{bmatrix} (R_c^l)^T u_c = 0 \qquad (2)$$

Where $u_l = (x_l, y_l, -c_l)^T$ represents the laser-based intensity image coordinates; $u_c = (x_c, y_c, -c_c)^T$ represents the camera image coordinates. The rotation matrix R_c^l is the rotation of the camera image with respect to the laser-based intensity image. Tx, Ty, and Tz are three translations between the perspective centers of the camera image and the laser-based intensity image. Based on the direction of the baseline, either Tx or Ty is assigned an arbitrary value. The co-planarity model can be solved through a least-squares adjustment. At least five corresponding point pairs within a camera/intensity image stereo pair are needed to recover the relative orientation parameters within the

same image stereo pair (e.g. three rotation angles, and two translations).

Due to projection and interpolation errors, the 3D coordinates of the extracted features in the laser-based intensity images are not accurate. As a result, the image orientation that is estimated through the single photo resection procedure would not be accurate either. In this paper, the relative orientation parameters estimated between the camera image and the laser-based intensity image only utilizes feature correspondences in 2D space. Therefore, the estimated image orientation is more accurate.

- Compatibility Analysis

After the ROPs estimation, the EOPs of a camera image are then estimated within a selected triplet, which consists of two intensity images and one camera images. There are two requirements for the selected triplet:

1. There should be sufficiently matched features within the selected triplet;
2. There should be a good geometry within the selected triplet.

For the first requirement, we only need to maximize the number of matched features within the selected triplet. For the second one, a compatibility analysis approach is proposed to evaluate the geometry within all possible triplets.

As shown in **Figure 4**, one possible triplet includes Intensity Image i, Intensity Image j, and Camera Image k. The relative orientation parameters for the stereos (k, i) and (k, j) are obtained through the SIFT feature matching process and relative orientation procedure, which have been described in previous sections. Since the ROPs of the stereos (k, i) and (k, j) are estimated with different scales, the scale factors for the stereos (k, i) and (k, j) can be determined through **Equation (3)**.

$$r_j^i = \lambda_2 r_k^i - \lambda_1 R_j^i r_k^j \qquad (3)$$

Where r_k^j and r_k^i are the two translation vectors that have been estimated from the relative orientation procedure; r_k^j describes the translation between Camera Image k and Intensity Image i, and r_k^i describes the translation between Camera Image k and Intensity Image j. The two scale factors λ_1 and λ_2 are then used to transform the two translation vectors r_k^j and r_k^i into the same scale. In the intensity image stereo (j, i), r_j^i describes the translation between Intensity Images j and i. The matrix R_j^i describes the rotation from Intensity Image j to Intensity Image i. Since the EOPs of both Intensity Images i and j are known, both r_j^i and R_j^i are computed from the known EOPs.

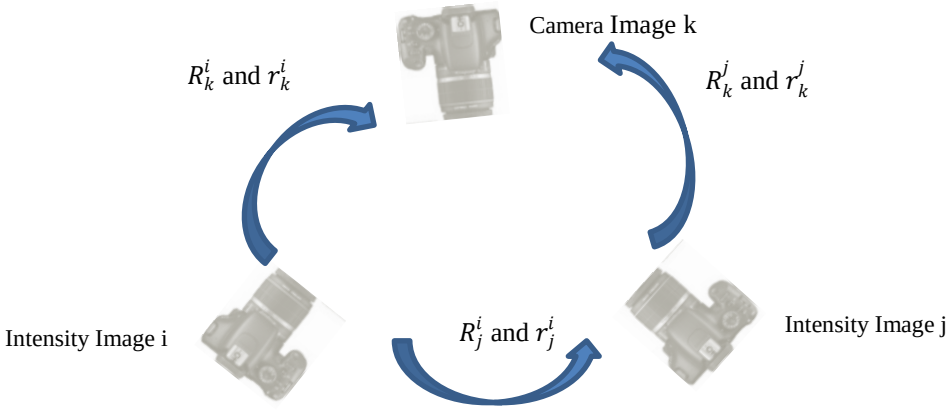


Figure 4. Compatibility analysis within a triplet of two intensity images and one camera image

As long as **Equation (3)** is solved, and the scale factors (λ_1 and λ_2) for the stereos (k, i) and (k, j) are determined, the approximate EOPs of the Camera Image k can be computed through **Equation (4)**.

$$\begin{cases} \begin{bmatrix} X_{0k} \\ Y_{0k} \\ Z_{0k} \end{bmatrix} = \begin{bmatrix} X_{oi} \\ Y_{oi} \\ Z_{oi} \end{bmatrix} + \left(\lambda_2 \mathrm{r_k^i} + \mathrm{r_j^i} + \lambda_1 \mathrm{R_j^i r_k^j}\right)/2 \\ R_k^{ref} = R_i^{ref} \cdot \left(\left(R_j^i \cdot R_k^j + R_k^i\right)/2 \right) \end{cases} \quad (4)$$

Where (X_{0k}, Y_{0k}, Z_{0k}) and (X_{0i}, Y_{0i}, Z_{0i}) represent the positions of Camera Image k and Intensity Image i within the laser point cloud reference frame. The two scale factors (λ_1 and λ_2) are computed from **Equation (3)**. The two vectors $\lambda_2 \mathrm{r_k^i}$ and $(\mathrm{r_j^i} + \lambda_1 \mathrm{R_j^i r_k^j})$ describe the translation from Camera Image k to Intensity Image i. Both r_k^i and r_k^j are estimated from the relative orientation procedure of the stereos (k, i) and (k, j), and r_j^i is computed from the EOPs of Intensity Images i and j. The two rotation matrices R_i^{ref} and R_k^{ref} describe the rotations from Intensity Image i and Camera Image k to the reference frame. The two matrices $R_j^i \cdot R_k^j$ and R_k^i describe the rotation from Camera Image k to Intensity Image i. All the three rotation matrices R_j^i, R_k^i , and R_k^j represent the rotations within the stereos (j, i), (k, i), and (k, j). Both R_k^i and R_k^j are estimated from the relative orientation procedure of the stereos (k, i) and (k, j), and R_j^i is computed from the EOPs of Intensity Images i and j.

As shown in **Figure 4**, the image orientation of Camera Image k can be estimated from either Intensity Image i or Intensity image j through a relative orientation procedure. Ideally, these two estimated image orientations (one is estimated from Intensity Image i, and the other one is estimated from the Intensity Image j) should be identical. However, due to the noise in the image, there would be a discrepancy between the two estimated image orientations. Therefore, a score function for compatibility analysis is defined as **Equation (5).**

$$\mathrm{S} = \sum_{i=1}^{3} abs(r_i) \cdot \sum_{m=1}^{3} \sum_{n=1}^{3} f_{mn}^2 \quad (5)$$

Where:

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \lambda_2 r_k^i - \left(r_j^i + \lambda_1 R_j^i r_k^j\right) \text{ and } \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} = R_j^i \cdot R_k^j - R_k^i$$

$\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$ describes the differences between the two estimated rotation matrices of Camera Image k with respect to Intensity Image i; (r_1, r_2, r_3) describes the differences between the two estimated positions of Camera Image k with respect to Intensity Image i. Ideally, both the rotation matrix and the position differences should be zero. In practice, good geometry within the selected triplet usually would lead to a small difference between the two estimated image orientations. Therefore, a small value of S computed from the score function (**Equation (5)**) indicates that there is good geometry within the triplet. By maximizing the product of the number of matches within the triplet and the reciprocal of the score value S defined in **Equation (5)**, the triplet with sufficient feature matches and good geometry can be selected. Then, the EOPs of Camera Image k can be estimated within the selected triplet through a bundle adjustment process. The initial approximations of EOPs are defined in **Equation (4)**.

In practice, due to occlusions, some camera images cannot be matched with an intensity image. In this situation, the EOPs of the unreferenced camera images are estimated through relative orientation procedure with the camera images, which have been referenced with the laser points.

Refinement by Bundle Adjustment

In the previous section, the camera images are matched with the laser-based intensity images, and the EOPs of each camera image are estimated with respect to the laser point cloud reference frame. Then, in this section, a bundle

adjustment process is carried out to refine the EOPs of each camera image. **Figure 5** illustrates the inputs and the outputs of the bundle adjustment process. As for the inputs, the initial EOPs of each camera image are obtained from the previous image orientation estimation process. Tie points are automatically extracted and matched through a proposed SIFT feature matching process. The proposed SIFT feature matching process here is similar to the above-mentioned feature matching process in the previous section. The only difference is that the proposed SIFT feature matching process in this section deals with the feature correspondences between two camera images. The approximate 3D object points are then computed through a spatial intersection process with the initial EOPs of each camera image and image coordinates of the tie points.

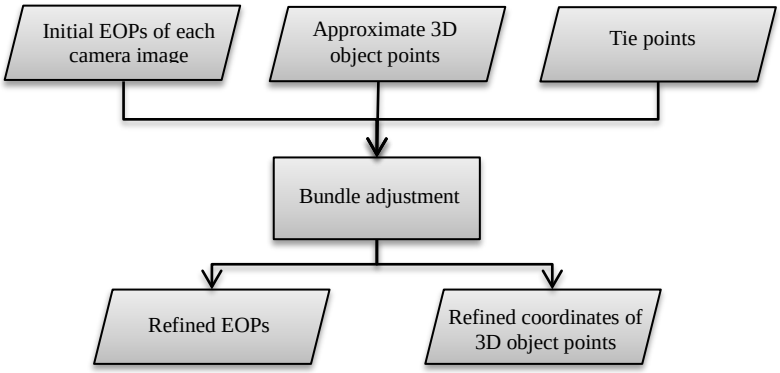


Figure 5. Inputs and outputs of the bundle adjustment process

Refinement by Iterative Closest Patch (ICPatch)

After the refinement through the bundle adjustment process, a 3D-3D Iterative Closest Patch (ICPatch) registration process is applied to register the 3D point cloud reconstructed from the previous bundle adjustment process to the 3D laser point cloud.

Due to the irregular nature of point clouds, exact point-to-point correspondence usually cannot be assumed. Instead of point-to-point correspondence, the geometric primitives chosen for the ICPatch registration are points and triangular patches (Habib, 2010). Therefore, for any two overlapping models, one of the point clouds is kept as it is, and the other one is converted to a Triangulated Irregular Network (TIN). In this paper, because the point cloud reconstructed from the bundle adjustment process is sparse, the image-based point cloud is kept as it is, and TIN is generated from the laser-based point cloud. Similar to the Iterative Closet Point (ICP) registration algorithm, the correct correspondences between the image-based points and the laser-based triangular patches is established through an iterative approach. To solve for the transformation parameters between the two different point clouds, co-planarity constraint is implemented.

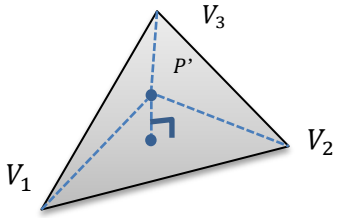


Figure 6. Conjugate primitives for ICPatch registration

In the co-planarity constraint model, points P', V_1, V_2, and V_3 are assumed to be coplanar. This means that the volume of the pyramid, which is created by P', V_1, V_2, and V_3 (as shown in **Figure 6)** should be zero. The co-planarity constraint is given in **Equation (6)**.

$$det\begin{bmatrix} X_{P'} & Y_{p'} & Z_{p'} & 1 \\ X_{V_1} & Y_{V_1} & Z_{V_1} & 1 \\ X_{V_2} & Y_{V_2} & Z_{V_2} & 1 \\ X_{V_3} & Y_3 & Z_{V_3} & 1 \end{bmatrix} = 0 \qquad (6)$$

In this paper, because the EOPs of each image are estimated with respect to the laser point cloud reference frame, the image-based point cloud is close enough to the laser-based point cloud. All the initial transformation parameters for the ICPatch registration process are set to zero. The final EOPs of each camera image are given in **Equation (7).**

$$\begin{bmatrix} X_i^{las} \\ X_i^{las} \\ X_i^{las} \end{bmatrix} = s \cdot R_{img}^{las} \begin{bmatrix} X_i^{img} \\ Y_i^{img} \\ Z_i^{img} \end{bmatrix} + r_{img}^{las} \quad and \;\; R_i^{las} = R_{img}^{las} \cdot R_i^{img} \tag{7}$$

Where:

$(X_i^{img}, Y_i^{img}, Z_i^{img})$ and $(X_i^{las}, Y_i^{las}, Z_i^{las})$ respectively represent the position of the perspective center of the camera image i within the image-based and the laser-based point cloud coordinate frame; the two rotation matrices R_i^{img} and R_i^{las} describe the rotations of the camera image i with respect to the image-based and the laser-based point clouds. The rotation matrix R_{img}^{las} and the translation vector r_{img}^{las} describe the rotation and the translation from the image-based point cloud to the laser-based point cloud. s is the scale factor between the two different point clouds. R_{img}^{las}, r_{img}^{las} , and s are all computed from the ICPatch registration process.

After the ICPatch registration process, the image-based and the laser-based point clouds are registered into the same reference frame. Then, by computing the normal distances between the image-based points and the corresponding laser-based triangular patches, the registration accuracy is evaluated.

EXPERIMENTAL RESULTS

Description of the Experimental Dataset

The test site involved in the experiment is the Ronald McDonald House in Calgary, Canada. 21 images around the Ronald McDonald House were captured by a Canon Rebel T3 digital camera. The Canon digital camera is calibrated, and the IOPs of the calibrated camera is given in **Table 1**. In the meantime, terrestrial laser points were acquired by a Faro Focus3D scanner. The resolution of the laser point cloud is up to 2 cm at the distance of 15 meters.

Table 1. The IOPs of the calibrated Canon Rebel T3 digital Camera

	Principal point coordinate *xp* (mm)	Principal point coordinate *yp* (mm)	Principal distance c (mm)	Pixel size (mm)	Image size (pixel^2)	K1 (mm^{-2})	K2 (mm^{-4})
Calibrated	-0.02233	-0.14730	22.32591	0.0052	4272×2848	-3.0637×10^{-4}	6.8414×10^{-7}

We place 12 virtual cameras around the obtained terrestrial laser point cloud. These virtual cameras are aligned along two different rows (see **Figure 7.(a)**). The minimum distance from the first row to the front façade of the Ronald McDonald House is about 20 meters; the minimum distance from the second row to the same façade is about 25 meters. All the virtual cameras on the same row are placed uniformly in viewing directions. The intersection angle between two adjacent virtual cameras is 15 degree, and the intersection angle between the first and the last cameras on the same row is 75 degree. The distance between two adjacent virtual cameras is 3 meters.

To simplify the feature matching process between the camera and the intensity images, the same image and pixel sizes of the Canon Rebel T3 camera are assigned to the virtual cameras. The principal distances of the virtual cameras are all set to 20 mm, which is close to the principal distance of the real camera. Meanwhile, to reduce the influence of insufficient point density, interpolation is applied to fill gaps and holes in the intensity images.

(a) (b)

Figure 7. (a) 12 virtual cameras are placed in front of the Ronald McDonald House; (b) one intensity image generated from the point cloud acquired by Faro Focus 3D laser scanner

Registration Results

Because the virtual cameras are placed around the front façade of the Ronald McDonald House (see **Figure 7(a)**), four out of 21 camera images, which are close to the virtual cameras, are successfully matched with the laser-based intensity images. For the four matched camera images, the initial EOPs are estimated through the proposed image orientation estimation process. For the remaining 17 camera images, the initial EOPs are estimated through the image feature matching and the relative orientation procedure with the four matched camera images. Then, the estimated image EOPs are sequentially refined through the bundle adjustment and the ICPatch registration processes. **Figure 8** illustrates the final orientation of each camera image after the refinements. Comparing it with the positions of 12 virtual cameras (see **Figure 7(a)**), we observe that the 17 unmatched camera images are far away from the 12 virtual cameras. It gives us an explanation why there are only four out of 21 camera images successfully matched with the laser-based intensity images.

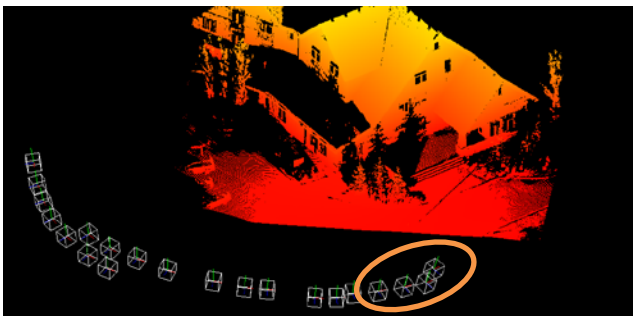
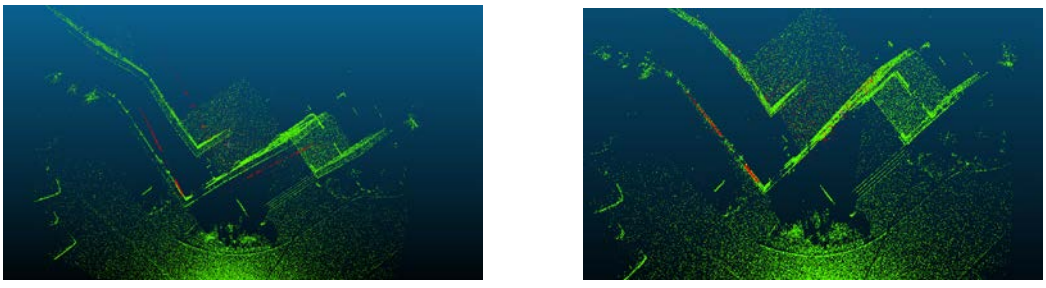


Figure 8. The final image orientations with respect to the terrestrial laser data: four matched camera images are within highlighted ellipse

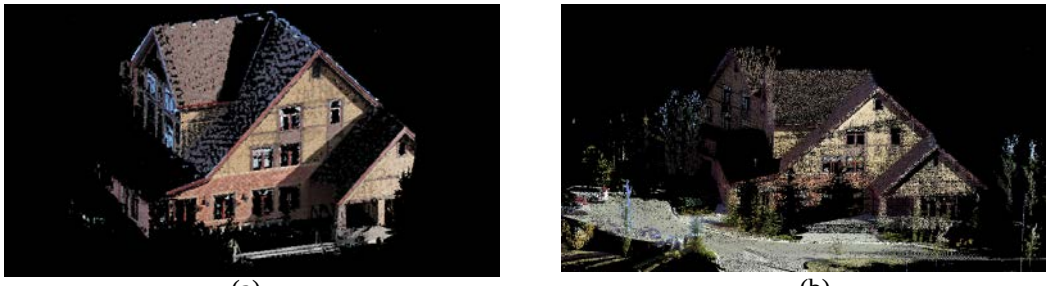
Figure 9(a) and **Figure 9(b)** illustrate the two point clouds before and after the ICPatch registration process. It's obvious that the discrepancies between the two point clouds disappear after the refinement of the ICPatch registration process. The normal distances from the image-based points to the corresponding laser-based triangular patches are then computed. The average normal distance of point-to-patch is 0.027 m.



(a) (b)

Figure 9. The image-based and laser-based point clouds before (a) and after (b) the ICPatch registration process; the point cloud with the red color is the image-based, and the point cloud with the green color is the laser-based.

To evaluate the accuracy of the proposed algorithm, a dense point cloud is reconstructed from the same experimental dataset using semi-global dense matching algorithm. Then, the normal distances from the dense-reconstructed points to the corresponding laser-based triangular patches are computed. The average normal distance is 0.025 m. This result is consistent with the average normal distance obtained from the proposed algorithm. **Figure 10** illustrates the dense reconstructed point cloud as well as the point cloud acquired by the terrestrial laser scanner.



(a) (b)

Figure 10. (a) Dense point cloud reconstructed from multiple images using the semi-global dense matching algorithm; (b) point cloud acquired by the Faro Focus3D terrestrial laser scanner; colors are obtained from the camera integrated with the laser scanner.

Meanwhile, the laser-based point cloud is back-projected onto the camera images to evaluate the accuracy of the registration. The result (See **Figure 11**) shows that the back-projected laser points are accurately matched with the camera image, which indicates that the estimated image EOPs are accurate.

(a) (b)

Figure 11. (a) The back-projected laser points with the camera image; (b) the details of the back-projected laser points

CONCLUSIONS AND RECOMMENDATIONS

This paper proposed a fully automatic algorithm to estimate the orientation of multiple images with respect to the terrestrial laser data. Virtual cameras are employed to generate the intensity images from terrestrial laser data. Then, the SIFT feature matching process is introduced in order to find the feature correspondences between the camera and the intensity images. Afterwards, the bundle adjustment and the ICPatch registration processes are sequentially applied to refine the estimated image EOPs. The proposed algorithm is tested on real datasets. The results show that the proposed algorithm is robust, and the estimated image EOPs are accurate.

The proposed algorithm utilizes the virtual cameras to generate the intensity images from the terrestrial laser data. However, the positions of the virtual cameras have to be carefully estimated. Another limitation of the proposed algorithm is that it can only deal with the terrestrial laser data with intensity information. Future work therefore could include improving the proposed algorithm for the image orientation estimation without using virtual cameras, and increasing the applicability of the proposed algorithm for laser data without intensity information.

ACKNOWLEDGEMENT

The author would like to thank CANTEGA, Tecterra and the Natural Sciences and Engineering Research Council of Canada (NESRC) for the financial support of this research work.

REFENCES

Ding, M., Lyngbaek, K., & Zakhor, A. (2008). *Automatic registration of aerial imagery with untextured 3d lidar models.* Paper presented at the Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on.

González-Aguilera, D., Rodríguez-Gonzálvez, P., & Gómez-Lahoz, J. (2009). An automatic procedure for co-registration of terrestrial laser scanners and digital cameras. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*(3), 308-316.

Guan, W., You, S., & Pang, G. (2013). *Estimation of camera pose with respect to terrestrial LiDAR data.* Paper presented at the Applications of Computer Vision (WACV), 2013 IEEE Workshop on.

Liu, L., Stamos, I., Yu, G., Wolberg, G., & Zokai, S. (2006). *Multiview geometry for texture mapping 2d images onto 3d range data.* Paper presented at the Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision, 60*(2), 91-110.

Mastin, A., Kepner, J., & Fisher, J. (2009). *Automatic registration of LIDAR and optical images of urban scenes.* Paper presented at the Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.

Mikhail, E. M., Bethel, J. S., & McGlone, J. C. (2001). Introduction to modern photogrammetry (Vol. 1). John Wiley & Sons Inc, New York. 479 pp.

Sattler, T., Leibe, B., & Kobbelt, L. (2011). *Fast image-based localization using direct 2D-to-3D matching.* Paper presented at the Computer Vision (ICCV), 2011 IEEE International Conference on.

Stamos, I., & Alien, P. (2001). *Automatic registration of 2-D with 3-D imagery in urban environments.* Paper presented at the Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on.

Zhao, W., Nister, D., & Hsu, S. (2005). Alignment of continuous video onto 3D point clouds. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27*(8), 1305-1318.