8-2015

Automatic Registration of Optical Aerial Imagery to a LiDAR Point Cloud for Generation of City Models

Bernard Olushola Abayowa
University of Dayton

Alper Yilmaz
The Ohio State University

Russell C. Hardie
University of Dayton, rhardie1@udayton.edu

Automatic Registration of Optical Aerial Imagery to a LiDAR Point Cloud for Generation of Large Scale City Models

Bernard O. Abayowa[a,c], Alper Yilmaz[b], Russell C. Hardie[a]

[a]*Department of Electrical and Computer Engineering, University of Dayton, Dayton, OH 45469, USA*
[b]*Photogrammetric Computer Vision Laboratory, The Ohio State University, Columbus, OH 43210, USA*
[c]*Sensors Directorate, Air Force Research Lab, WPAFB, OH 45433, USA*

Abstract

This paper presents a novel framework for automatic registration of both the optical and 3D structural information extracted from oblique aerial imagery to a Light Detection and Ranging (LiDAR) point cloud without prior knowledge of an initial alignment. The framework employs a coarse to fine strategy in the estimation of the registration parameters. First, a dense 3D point cloud and the associated relative camera parameters are extracted from the optical aerial imagery using a state-of-the-art 3D reconstruction algorithm. Next, a digital surface model (DSM) is generated from both the LiDAR and the optical imagery-derived point clouds. Coarse registration parameters are then computed using geometric invariants of salient regional features extracted from the LiDAR and optical imagery-derived DSMs. The registration parameters are further refined using the iterative closest point (ICP) algorithm to minimize global error between the registered point clouds. The registration framework is tested on a simulated scene and aerial datasets acquired in real urban environments. Results demonstrates the robustness of the framework for registering optical and 3D structural information extracted from aerial imagery to a LiDAR point cloud, when co-existing initial registration parameters are unavailable.

Keywords: large scale scene modeling, aerial imagery, sensor fusion, 3D registration

1. Introduction

Textured models of large-scale urban environments are desired in many applications involving scene visualization, analysis and understanding. These applications include, but are not limited to, urban development planning, 3D scene classification, obstacle avoidance for low flying autonomous vehicles, and generation of 3D city maps, such as Google Earth and Microsoft Virtual Earth.

There are several methods for modeling urban environments including the use of aerial or terrestrial Light Detection and Ranging (LiDAR), 3D reconstruction from image sequences, airborne Synthetic Aperture Radar (SAR) Interferometry, and a combination of these methods. The combination of aerial LiDAR and aerial optical imagery is becoming increasingly popular for modeling urban environments.

Aerial LiDAR provides highly accurate height measurement and direct geo-registration of large regions in a fast manner. In a conventional LiDAR data collection, a downward-looking LiDAR sensor is mounted on a airborne platform or satellite. The LiDAR sensor sends short pulses that travel down to the scene surface, some of which are reflected and detected by the sensor. The elevation at each surface point is then determined from the total travel time of the pulses.

The LiDAR point cloud (LPC) generation process provides very little to no data on vertical surfaces such as building facades. Moreover, the LiDAR sensor does not provide color information for the scene. Although, there are systems that are capable of co-registering images to the LPC at the time of collection, these systems have several limitations because they require fixing the relative position and orientation of the sensors [1].

Detailed structural and optical information are necessary for generating photo-realistic 3D models of urban environments. High resolution oblique aerial images efficiently provide optical and structural information complementary to aerial LiDAR with high planimetric accuracy. An optical imagery-derived point cloud (OIPC), representing the structure contained in the optical aerial imagery, can be obtained through a 3D reconstruction

process.

Image-based 3D reconstruction is well developed in computer vision and photogrammetry fields and has been an active research area for decades. Innovations such as better image quality, improved radiometry, advances in multiple-view geometry and emergence of graphics processing unit (GPU) makes image-based modeling practical for large scale scenes. [2–5].

Although, OIPCs can provide information on the vertical details of a scene, they are usually noisy. Moreover, surfaces with weak texture are often removed from the reconstruction output as they are difficult to model using image based 3D reconstruction methods. Radiometric errors, such as sensor sensitivity, illumination changes, and atmospheric effects may also lead to missing regions and errors in the final 3D model.

Since the information from optical aerial imagery and aerial LiDAR are complimentary, both sensor types can be combined to acquire more details than what is obtainable by either sensors when used alone. The aerial LiDAR and optical aerial imagery have to be transformed to a common reference frame before a model can be generated from both data. Earlier methods require coexisting initial registration parameters such as Global Positioning System (GPS)/Inertial Measurement Unit (IMU) metadata for texture mapping the optical information extracted from aerial imagery onto the LiDAR data. Moreover, the majority of the existing algorithms are directed towards texture mapping the optical information extracted from aerial imagery onto the LiDAR data. In addition to optical information, 3D structure can be extracted from optical aerial imagery that may be complementary to the LPC.

In this paper, we present a novel framework for automatic registration of both the optical and 3D structural information extracted from oblique aerial imagery to a LPC without prior knowledge of an initial alignment. The framework employs a coarse to fine strategy in the estimation of the registration parameters.

In the coarse step, the framework takes advantage of the advances in image-based 3D reconstruction to extract a dense OIPC, and the associated relative camera parameters, from the optical aerial imagery. The extraction of the OIPC reduces the mapping between the optical aerial imagery and the LPC from a perspective projection to a 3D similarity transformation. Next, a LiDAR digital surface model (LDSM) and an optical imagery-derived digital surface model (OIDSM) is generated from the LPC and OIPC, respectively. The digital surface model (DSM) representation enables the extraction of salient regional features in the scene. Coarse 3D registration parameters are computed using geometric invariants of the salient regional features extracted from the LDSM and OIDSM. In the fine step, the coarse registration parameters are refined using the iterative closest point algorithm (ICP) to obtain fine estimates of the registration parameters.

An overview of the proposed registration framework is shown in Fig. 1. The input to the proposed framework is a sequence of optical aerial images of the region of interest to be modeled, and the LPC containing the region of interest covered by the optical aerial imagery. The output of the framework are fine estimates of the 3D transformation and the absolute camera poses relating the OIPC and the optical aerial imagery to the LPC, respectively.

The remainder of this paper is organized as follows. We begin with a review of relevant literature in the remaining part of Section 1.1. The coarse registration parameters estimation process is described in Section 2. This includes the extraction of the OIPC and relative camera parameters from the optical aerial imagery, LDSM and OIDSM generation, salient regional feature extraction, and coarse 3D transformation estimation. Section 3 discusses the refinement of the registration parameters. Finally, results and conclusions follow in Section 4 and Section 5, respectively.

1.1. Related Work

The problem of registering images to LiDAR have received considerable attention in the literature. In [6, 7], an approach was presented for texturing terrestrial LiDAR with ground images using features on building facades. They extract parallel line segments in the images using vanishing points. They also extract parallel line segments in the LiDAR data. The intrinsic and extrinsic parameters of the camera are then estimated using groups of rectangular structures obtained from the parallel lines in both sensor types. The camera parameters are further refined in [8] through sparse point clouds obtained by applying structure from motion techniques on the images and in [1, 9] by utilizing low level linear features. These approaches provides photo-realistic models of urban environments. However, they are limited to modeling interior and exteriors of a few buildings. In [10, 11], the terrestrial LiDAR based modeling approach is extended to larger scenes.

In [12] Ding et al. presented a texture mapping approach that is also based on geometric features extracted from optical images and a LDSM. First they generate an initial estimate of the camera orientation and position by utilizing vanishing point information and GPS/IMU measurements. Next, they refine the rough camera parameter estimates by extracting and matching features

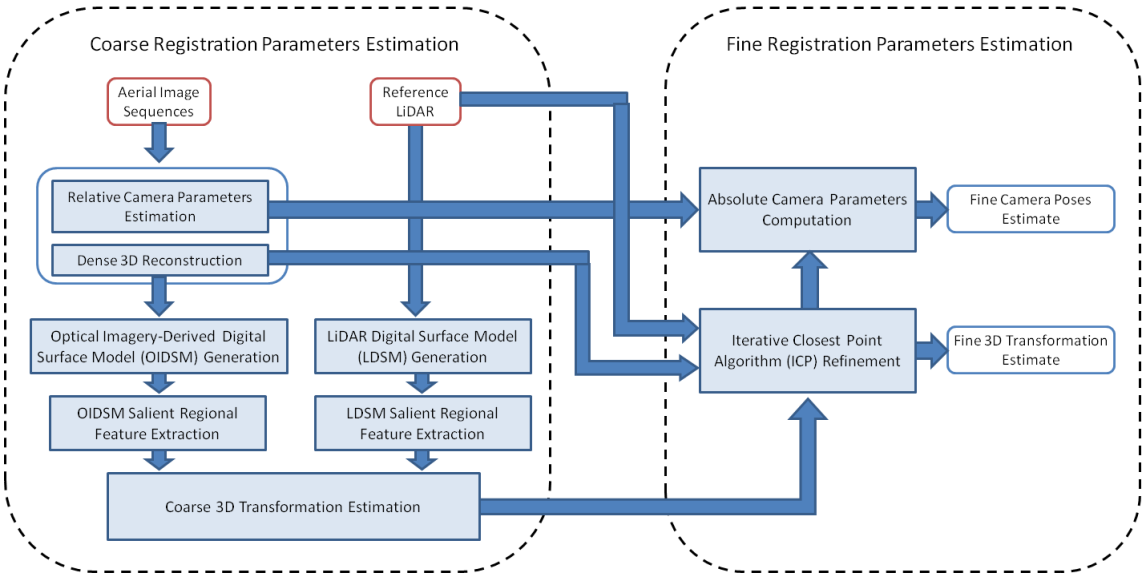

Figure 1: An overview of the registration framework.

corresponding to orthogonal structural corners in the LDSM and optical images. With this approach many outliers are generated and sometimes sufficient corner feature cannot be extracted.

A similar approach is presented by Wang and Neumann that does not require the computation of vanishing points and allow for large errors in initial GPS/IMU registration [13]. They achieve a higher level of accuracy in the camera parameters than [12] by using more robust features computed from connected line segments in the LDSM and the optical images.

Another approach for registering optical images onto aerial LIDAR is to exploit the statistical dependence in these multi-sensor data. Mastin et al. [14] introduced a registration method that estimates the camera parameters between the LPC and optical images by maximizing the mutual information between the images, a projection of the LPC onto the image plane, and an image representing the number of photon returns for the LPC.

The aforementioned approaches for registering optical aerial imagery and LiDAR focuses on adding optical detail to the LiDAR model using features common to both sensor types. However, the scene structure obtainable from the imagery are not integrated with the LiDAR. The facade details from these methods are often inferred from range information obtained from the LPC, using surface reconstruction algorithms such as [15, 16]. These modeling methods do not accurately represent structures elevated above the ground or hidden from overhead view such as bridges, elevated roads, building facades and roof overhangs

In [17], Zhao et al. developed an alignment approach for registering structure obtained from oblique aerial images to a nadir-view LPC. The camera position and orientation parameters for the image sequences are recovered by applying the method described in [18] to matching features in the sequences. A dense model of the scene is then reconstructed by combining the camera pose and focal length data with a plane-plus-parallax framework presented in [19, 20]. They use GPS/IMU measurements for initial alignment of the point cloud from the two sensor types. The similarity transformation between the point clouds obtained is then refined using the Iterative Closest Point (ICP) algorithm [21–23]. Although this work provides an efficient method for aligning an OIPC to an LPC, the alignment framework employed requires knowledge of initial registration parameters. The GPS/IMU measurements required for initialization may not be available, or they may be very erroneous.

Registration of the OIPC to the LPC is a 3D-3D registration problem with vast literature. The ICP algorithm is the gold standard for solving this registration problem. However, the algorithm requires a rough registration estimate close to the true solution in order to converge. Therefore, the point clouds have to be roughly aligned before ICP can be used to refine the registration of the point clouds.

Several automatic algorithms have been proposed for coarse registration of 3D point clouds. However, most

algorithms are designed for registration of objects with high density surfaces where local descriptors based on surface normals, surface geometry and principal curvatures can be reliably computed [24–27]. The more robust approaches, applicable to large-scale urban scenes, require the point cloud to be recovered from a similar sensor such as LiDAR where the 3D points are directly recorded with minimal noise such that the structure of the point cloud is not lost by subsampling [28, 29].

A surface registration algorithm based on the invariance of intersecting coplanar line segments was introduced in [28]. However, the algorithm is designed for point clouds that are directly obtained from a similar 3D sensor. Also, the algorithm requires the point clouds to be of the same scale.

Existing algorithms for registering optical aerial imagery and LiDAR rely on manual correspondence or GPS/IMU metadata for initial alignment. Manual correspondence is a tedious process and may be impossible for large scenes. GPS/IMU measurements are not always available and its accuracy is easily affected by interference such as dropouts. They are also prone to drift and biases because the accelerometer and gyroscopes of the IMU are sensitive to temperature fluctuations due to changes in aircraft altitude. These systematic errors accumulate overtime and measurements may become unreliable for long video sequences. Moreover, highly accurate GPS/IMU systems are very expensive and may not be available for oblique images collected on an unmanned aerial vehicle(UAV). GPS/IMU information for the oblique images and/or the LIDAR data may also be missing entirely.

2. Coarse Registration Parameters Estimation

This section describes the coarse estimation of the parameters relating the optical imagery and the OIPC to the LPC. The mapping between the optical aerial imagery and the LPC is reduced from a perspective projection to a 3D similarity transformation through the recovery of the OIPC. A coarse estimate of the 3D transformation between the OIPC and LPC is then computed using geometric invariants of salient regional features extracted from the OIDSM and LDSM.

2.1. Relative Camera Parameters and OIPC Recovery

The recovery of the relative camera parameters and the OIPC is the first step in the registration framework. Errors in this step will cause the registration framework to fail. Hence, a robust and accurate system for recovering relative camera parameters and OIPC is necessary for this step.

The relative camera parameters describe the mapping of the 3D points in the OIPC to the corresponding points in an optical image. This mapping is given by

$$\mathbf{x} = \mathbf{P}_O \mathbf{X}_O, \tag{1}$$

where $\mathbf{X}_O \in \Re^{4xn}$ is the OIPC in homogenous coordinates [30], $\mathbf{x} \in \Re^{3xn}$ is the corresponding optical image points in homogenous coordinates, and $\mathbf{P}_O \in \Re^{3x4}$ is the relative camera projection matrix. n is the cardinality of the OIPC. The relative projection matrix can be decomposed as

$$\mathbf{P}_O = \mathbf{K}\mathbf{R}_O \left[\mathbf{I}| - \mathbf{t}_O\right], \tag{2}$$

where $\mathbf{K} \in \Re^{3x3}$ is the calibration matrix. $\mathbf{R}_O \in \Re^{3x3}$ contains the relative camera orientation, $\mathbf{t}_O \in \Re^{3x1}$ represents the relative camera position, and $\mathbf{I} \in \Re^{3x3}$ is an identity matrix. The calibration matrix is given by

$$\mathbf{K} = \begin{bmatrix} s_x & \kappa & u_0 \\ 0 & s_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where s_x and s_y are the scale factors in the $x-$ and $y-$ *coordinate* directions, $[u_0, v_0]^T$ are the coordinates of the principal point, and κ is the skew factor.

In this registration framework, the relative camera parameters and the OIPC are recovered using Visual Structure from Motion (VisualSFM), an end to end system for large scale 3D reconstruction from unordered image sequences. VisualSFM is selected for recovering the relative camera parameters and the OIPC because of its accuracy and robustness for large scale scenes. The system does not require user intervention after setup and it is fast and efficient.

In VisualSFM, features and correspondence are first computed using a GPU implementation of scale invariant feature transform (SIFT) [31]. A multicore bundle adjustment technique [32] is then used to estimate the relative camera parameters for each image, as well as a sparse point cloud recovered from the images.

Bundle adjustment is a collective nonlinear refinement of the relative camera parameters as well as the 3D locations of matching image points present in a scene. This process can take a long time for large scenes. Multicore bundle adjustment significantly reduces the processing time for large scenes by exploiting CPU and GPU parallelism. The camera parameters computed for each image are the focal length, a single radial distortion parameter, the camera orientation, and the position. Fig. 2 shows camera positions, orientations and sparse point cloud recovered by Visual SFM through SiftGPU and multicore bundle adjustment.

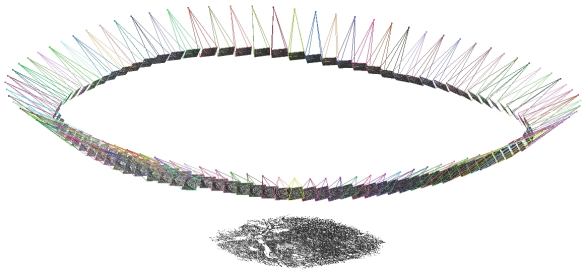

Figure 2: Camera positions, orientations and sparse point cloud recovered by Visual SFM through SiftGPU and multicore bundle adjustment.

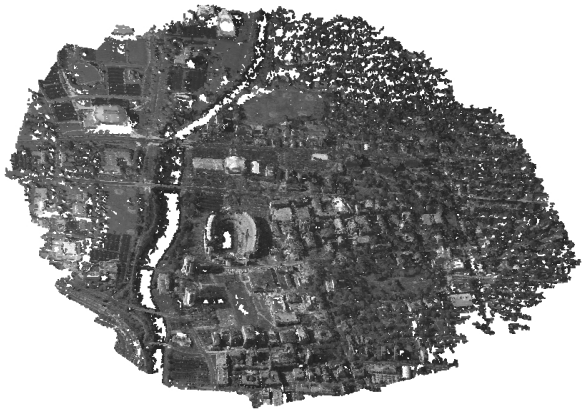

Figure 3: Textured OIPC recovered by VisualSFM through PMVS.

VisualSFM generates the OIPC from the optical images using Patch-Based Multi-View Stereo (PMVS) [5]. PMVS follows the stereo triangulation approach to dense reconstruction. The process starts with sets of corresponding points which are developed into dense patches through a match, expand and filter technique. The sparse points are repeatedly expanded to neighboring points with correspondences, and visibility constraints are used to filter out false correspondences.

PMVS reconstructs a dense OIPC of static objects visible in the images. Non-rigid and moving objects in the scene are ignored in the reconstruction process. In our experiments, only image patches that are visible in at least three views are reconstructed. This reduces the chances of poor reconstruction in weakly textured regions. The input to PMVS are the image sequences and the computed camera parameters corresponding to each image. Fig. 3 shows textured OIPC recovered by VisualSFM through PMVS.

2.2. LDSM and OIDSM Generation

The LDSM and OIDSM are image representations of the earth surface where each location carries a value based on the relative height of the location with respect to the ground level. The generation of the LDSM and the OIDSM enables the extraction of regional features in the LPC and OIPC.

2.2.1. Statistical Filtering of Outlier Points

The LPC and OIPC contains noise and outliers that needs to be removed before the LDSM and the OIDSM can be generated. This filtering step is especially important for the OIPC as the 3D reconstruction process usually generate more noisy point cloud compared to the LPC. A statistical outlier filter [33] is used to reduce the noise and outliers in the LPC and the OIPC. Let d_i be the average distance from each point $\mathbf{X}_i$ to its nearest neighbors. If we assume the resulting d_i forms a gaussian distribution, the filtered point cloud is defined as

$$\mathbf{X}_F = \{\mathbf{X}_i \in \mathbf{X} \mid d_i < \mu + \alpha\sigma\}. \quad (4)$$

where $\mathbf{X}$ is a set representing the original point cloud. The parameters μ and σ are the mean and standard deviation of the vector formed by average distances d_i, respectively. The parameter α is the multiplier that determines what distances from μ is considered in determining what points are inliers. The average distance for each point is computed using using 50 nearest neighbors, and α is set to 1 in our experiments. An example of inliers and outliers after filtering of the LPC and OIPC is shown in Fig. 4 and Fig. 5, respectively. The inliers are shown as blue dots and outliers as red dots.

As mentioned earlier, aerial LiDAR does not capture the detail on facade of buildings in the scene. Due to the sparse nature of these facade points, they are removed as outliers in the filtering process along with true outliers such as the region of the scene containing water bodies. The facade information is usually captured in more detail in the OIPC. However, the OIPC contains more outliers compared to a LPC. Most of these outliers are in regions between buildings in the scene, resulting from errors in stereo correspondence. The removal of the outliers from the point clouds help reduce the chances of erroneous height values in the LDSM and OIDSM generation process, which leads to better detection of true regional features in the scene.

2.2.2. Nadir-view Rotation of the OIPC

The point clouds have to be in nadir-view before the LDSM and OIDSM can be generated. The LDSM can be directly generated from the LPC since the LPC is collected in nadir-view. The orientation of the OIPC is arbitrary; therefore, the point cloud needs to be rotated

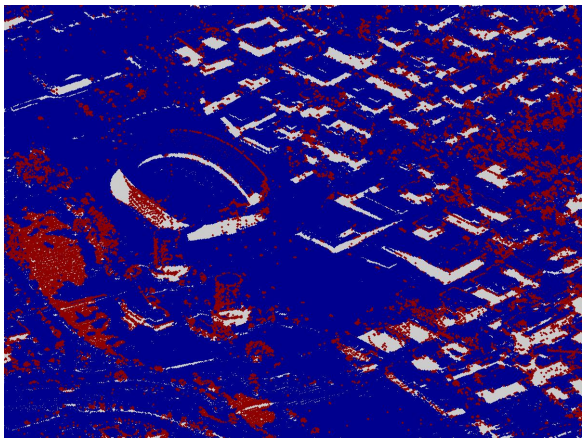

Figure 4: Inliers and outliers in the LPC after statistical outlier filtering. The inliers are shown as blue dots and outliers as red dots.

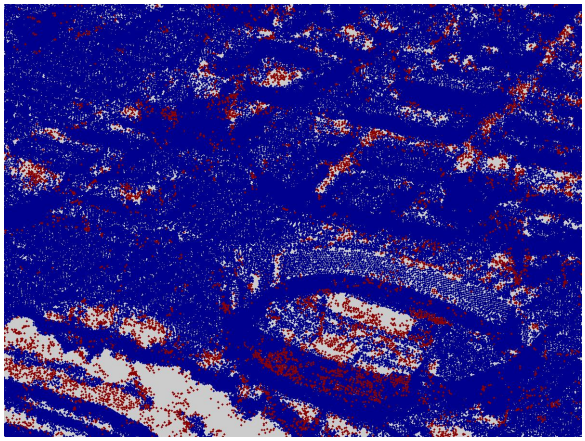

Figure 5: Inliers and outliers in the OIPC after statistical outlier filtering. The inliers are shown as blue dots and outliers as red dots.

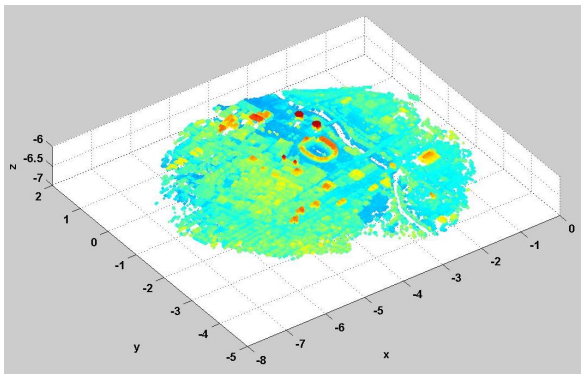

Figure 6: Nadir-view transformation of the OIPC using the eigenvectors of the dominant plane inliers.

so that the orientation of its principal axes corresponds to nadir-view.

Most scenes have three principal directions. The two major axes are parallel to the ground plane, and the minor axis is perpendicular to the ground plane. Assuming the region of interest represented by the OIPC is large enough such that the ground plane have minimal variance in the point cloud data, the principal axes can be extracted using Principal Component Analysis (PCA) [34].

In some urban environments, the terrain may not be relatively flat. As a result, direct application of PCA to the data may not yield principal axes corresponding to nadir-view. To ensure the selected dominant plane is representative of the ground plane, an efficient RAndom Sample Consensus (RANSAC) plane extraction approach is used to search for the dominant plane in the scene [35]. Following the detection of the dominant plane, the eigenvectors of the major axes are extracted from the inliers of the dominant plane using PCA. The OIPC is then transformed to nadir-view using a rotation only transformation matrix containing the eigenvectors of the dominant plane. The transformation is given by

$$\mathbf{H}_N = \begin{bmatrix} \mathbf{V} & \mathbf{0}^T \\ \mathbf{0} & 1 \end{bmatrix}, \tag{5}$$

where $\mathbf{V} \in \Re^{3x3}$ be a matrix whose rows are the eigenvectors computed on the dominant plane inliers. An example of a nadir-view transformation of the OIPC using the eigenvectors of the dominant plane inliers is shown in Fig. 6.

2.2.3. LDSM and OIDSM Resampling

The LPC and the OIPC may have different point spacings and point densities. Therefore, direct generation of

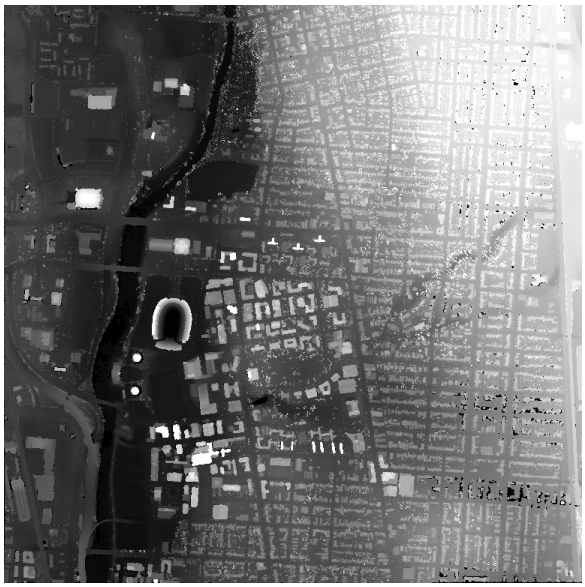

Figure 7: A LDSM generated from the LPC.

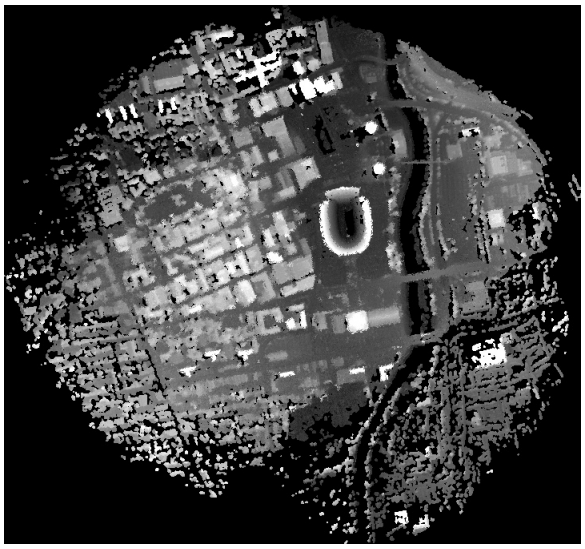

Figure 8: A OIDSM generated from the OIPC.

the LDSM and OIDSM my result in large resolution differences. To reduce the number of candidate sets in the feature matching process, the LDSM and OIDSM needs to have close resolutions. Since the LPC and the OIPC are obtained from the same scene, the variance along the minor axis of both point clouds should be similar. The ratio between the LPC and OIPC minor axis variance will depend on the level of overlap between the optical aerial imagery and the LPC coverage.

Let λ_l and λ_o represent the average distance between each point and its nearest neighbors in the LPC and OIPC, respectively. Also, let σ_l and σ_o represent the standard deviations along the normal direction to the ground plane for the LPC and OIPC.

The LDSM can be resampled to a new spatial resolution

$$\hat{\lambda}_l = \tau\sigma_l. \tag{6}$$

Similarly, the OIDSM can be resampled to a new spatial resolution

$$\hat{\lambda}_o = \tau\sigma_o, \tag{7}$$

where $\tau = \max(\frac{\lambda_l}{\sigma_l}, \frac{\lambda_o}{\sigma_o})$.

An example of a LDSM generated from the LPC and a OIDSM generated from the OIPC are shown in Fig. 7 and Fig. 8, respectively.

2.3. Salient Regional Feature Extraction

Urban scenes usually contain maximal structures elevated from the terrain level, such as buildings. These structures are salient regional features which can be extracted from the LDSM and OIDSM. A morphological reconstruction transformation technique is used to extract the salient regional features. Consider a grayscale DSM whose pixel values correspond to the height at the pixel locations. The goal here is to extract regions in the scene that are distinct from neighboring regions based on their height. Hence, we define a salient regional feature as connected components of pixels whose external boundary pixels all have a lower height value, and whose centroid is consistent over a specified range of morphological reconstruction thresholds.

2.3.1. Regional Feature Extraction by Morphological Reconstruction

Morphological reconstruction have been proven to be an effective technique for extracting regional features from different image types including DSMs [36–39]. One of the advantages of morphological reconstruction is that unlike classical morphology, the extracted features are not dependent on the choice of structuring element. Instead, the process makes use of a flat structuring element that works for different types of shapes of interest in an image. The flat structuring element commonly used assumes a 8-neighborhood connectivity. i.e. a 3x3 matrix of ones. The structuring element determines what neighboring pixels are considered in the morphological filtering of each pixel.

Morphological reconstruction iteratively transforms a marker image by geodetic dilation using the flat structuring element until stability is reached. The transformation is constrained by another image called the mask image. The marker image is derived from the mask image by subtracting a threshold from the mask image. The threshold determines what regional features in the image are detected. The range between the minimum and the maximum pixel values of a regional feature extracted at a specified threshold is less than the threshold. This allows for detection of different regional feature sets at different thresholds of the mask image.

Furthermore, the same threshold can be applied to extract regional features of similar heights regardless of their location in the image. For example, if building A is on a hill and building B is on a valley. If both building A and B are of the same height from their surrounding terrain, they will both be detected as regional features at the same threshold.

Given a grayscale mask image I and a reconstruction threshold h. The marker image is given by

$$J = I - h. \tag{8}$$

An elementary geodesic dilation of marker image J constrained by mask image I using a flat structuring element S is given by

$$R_I^{(1)}(J) = (J \oplus S) \wedge I \tag{9}$$

where $\oplus$ is the dilation operator and $\wedge$ represents the point-wise minimum between the dilated marker image and the mask image. The morphological reconstruction $R_I^{(n)}(J)$ of the marker image J is obtained by n successive elementary geodesic dilations of the marker image as follows

$$R_I^{(n)}(J) = \underbrace{R_I^{(1)}(J) \circ R_I^{(1)}(J) \circ \ldots R_I^{(1)}(J)}_{n \text{ times}} \tag{10}$$

where n is the iteration number at which stability is reached. i.e successive elementary geodesic dilations does not change the image any longer.

An Illustration of the morphological reconstruction of a one dimensional (1D) signal as described in [39] is shown in Fig. 9. The marker signal J is obtained from mask signal I by subtracting threshold h from mask signal I (top-left). Successive elementary geodesic dilations is then applied to marker image J until the pixel values does not change any longer (top-right). The regional features signal at threshold h is generated by subtracting the transformed marker image from the mask image (bottom).

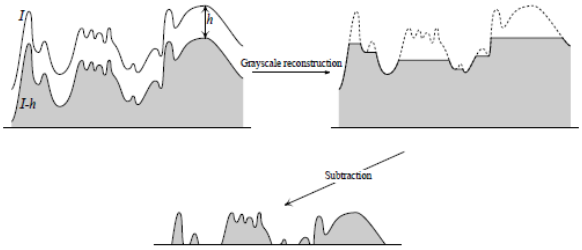

Figure 9: An Illustration of the morphological reconstruction of a one dimensional (1D) signal.

For large images such as large scale DSMs, successive elementary geodesic dilations can quickly become intractable in reasonable time before stability is reached. To overcome this problem, efficient morphological reconstruction techniques that only consider pixels whose values will be modified during the geodesic transformation are often used in practice. The fast hybrid grayscale reconstruction algorithm proposed in [39] is used for geodesic dilation of the LDSM and the OIDSM.

An illustration of the regional features images extracted from the LDSM at different thresholds is shown in Fig. 10. Fig. 10(a) is the original LDSM image Figures 10(b), 10(c) and 10(d) are the regional features binary images generated at thresholds 5, 25 and 50, respectively. At each threshold, the reconstructed image is subtracted from the LDSM. A binary image is generated from the resulting image by setting every pixel greater than zero to one. The regional features of the binary image are then detected using connected component analysis [40]. The white region in Fig. 10(d) is due to the presence of trees close to the buildings in the region. The boundary of the region changes with successive morphological reconstruction thresholds and therefore the regional features is unstable. Unstable regional features are eliminated in the salient regional feature extraction process. The regional features of the OIDSM are detected in a similar manner. Similar to Fig. 11, Fig. 11 shows the original OIDSM and regional features extracted at various thresholds.

2.3.2. *Salient Regional Features*

The morphological reconstruction transformation generates numerous regional features, most of which are not representative of interest objects in the scene. Moreover, other scene objects such as incomplete models in the OIDSM and trees in both the LDSM and OIDSM are detected as regional features in the scene. Most of the false regional features are eliminated from the OIDSM and LDSM based on their saliency.

The average height of the regional features varies

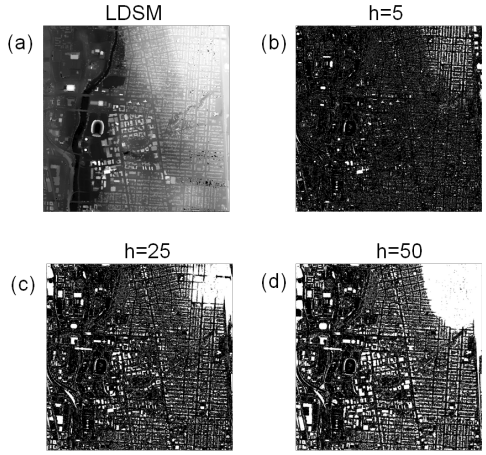

Figure 10: An Illustration of the regional features images extracted from the LDSM at different thresholds: (a) is the original LDSM image (b), (c) and (d) are the regional features binary images generated at thresholds 5, 25 and 50, respectively.

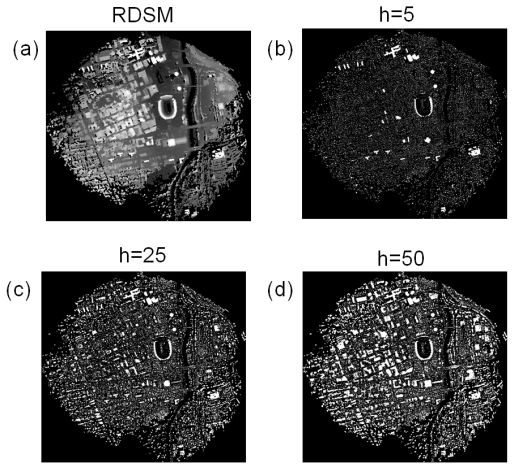

Figure 11: An Illustration of the regional features images extracted from the OIDSM at different thresholds: (a) is the original OIDSM image (b), (c) and (d) are the regional features binary images generated at thresholds 5, 25 and 50, respectively.

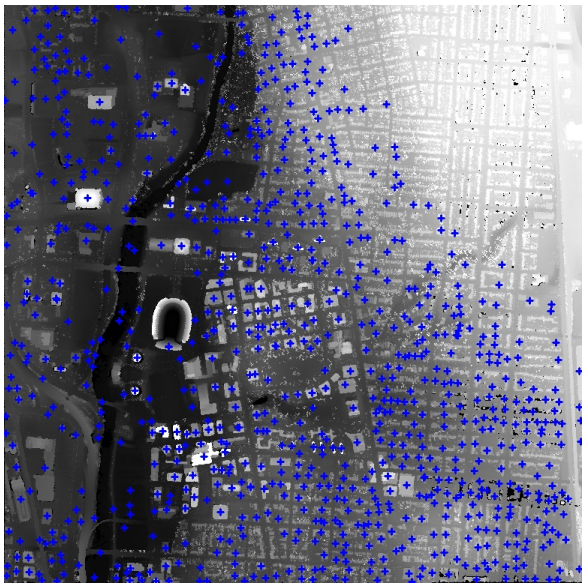

Figure 12: LDSM salient regional features

among different urban environments. In order to capture true regional features at different heights, regional features images are generated for the LDSM and OIDSM at different morphological reconstruction thresholds. At each threshold feature points are extracted from the each detected regional features by computing the centroid of the regional features across all thresholds. The centroid of each regional feature is the average coordinate of all the connected components pixel coordinates belonging to the regional feature.

It is expected that all true regional features of the scene will have a consistent centroid regardless of the morphological reconstruction thresholds at which they are detected. The range of the morphological reconstruction thresholds need to be set such that it is higher than the range between the minimum and maximum grey values of the true regional features in the scene. In our experiments, the morphological reconstruction thresholds ranges from 5 to 100 in steps of 5. These thresholds ensures that different shapes of interest objects in the scene are detected. Regional features whose centroid do not change across a minimum of 10 morphological reconstruction thresholds are considered salient regional features of the scene.

Fig. 12 and Fig. 13 shows the salient regional features extracted from the LDSM and OIDSM, respectively. The salient regional features detected contains interest objects such as buildings in the scene.

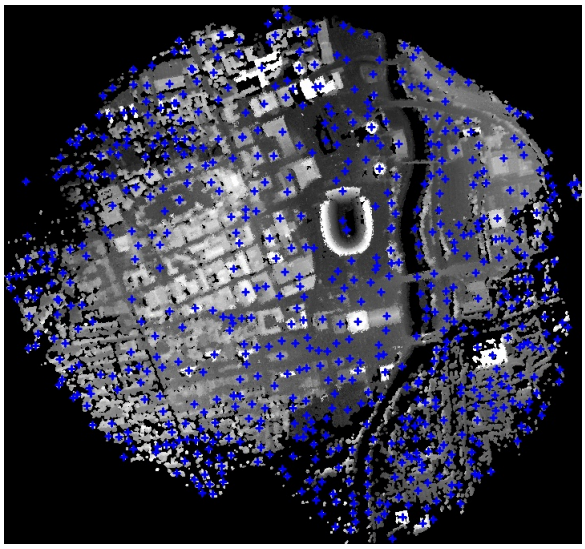

Figure 13: OIDSM salient regional features

2.4. Coarse 3D Transformation Estimation

The salient regional features extracted from the OIDSM and LDSM is used to obtain a coarse similarity transformation that registers the OIPC to the LPC. This process provides an initial registration similar to what is obtained through GPS metadata or manual point correspondences. However, the transformation parameters are automatically computed.

2.4.1. Geometric Invariants of Salient Regional Features

In order to register the OIPC to the LPC, we need at least three point correspondences. Let m and n represent the cardinality of the 3D points corresponding to the LDSM and OIDSM salient regional features centroids, respectively. Trying all possible pair of three point correspondences will require $O(n^3m^3)$ combinations, which is intractable for large scale urban environments. To overcome this problem, a subset of the salient regional features centroids that satisfy certain geometric invariant constraints are used in the correspondence search process.

A set of geometric invariant properties are defined on intersecting line segments computed from the OIDSM salient feature points. The invariant descriptors of the intersecting line segments are then used to find corresponding segments in the LDSM. While there are many intersecting line segments that satisfies these constraints, the true matching line segments are verified based on the similarity of overlapping regions in the

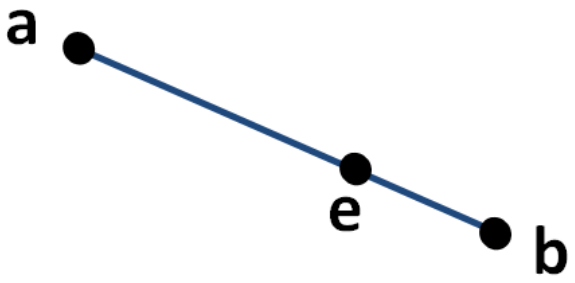

Figure 14: Consider a line segment $\overline{ab}$ and a point e on $\overline{ab}$. The ratios $\frac{|a-e|}{|a-b|}$ and $\frac{|e-b|}{|a-b|}$ are preserved under affine transformations.

OIDSM and LDSM generated by the line segment's feature points.

The ratio of lengths on collinear or parallel line segments in a plane is invariant under affine transformation [30]. Consider a line segment $\overline{ab}$ and a point e on $\overline{ab}$ as shown in Fig. 14. The ratios $\frac{|a-e|}{|a-b|}$ and $\frac{|e-b|}{|a-b|}$ are preserved under affine transformations.

The coarse 3D registration algorithm makes use of salient regional features in urban environments to reduce the number of candidate points used for registration. Moreover, the nadir-view representation of the LDSM and OIDSM is used to resolve the possible large scale difference between the LPC and OIPC. Furthermore, the angle between two line segments and the ratio of intersecting line segments, which are both invariant under similarity transformation, are used as additional constraints in the correspondence search. The feature extraction and constrained correspondence search enables the registration of OIPC of large scale scenes to a LPC without assumption about initial alignment.

The geometric constraints are used to find a minimum of four pairs of corresponding salient feature points in the LDSM and OIDSM that yields maximum Normalized Cross Correlation (NCC) score between overlapping region patches of the DSMs. This is possible because the relationship between the LDSM and the OIDSM images is a similarity transformation, which is a special case of affine transformation consisting of rotation, translation and an isotropic scaling. Although the DSMs is a 2.5D image, the z dimension is ignored in this initial transformation. The planar transformation of the surface models roughly aligns the z axis of the LDSM and OIDSM.

Let O_p be the set of OIDSM salient regional features centroids and L_p the set of LDSM salient regional features centroids. Given two sets $A = \{a, b, c, d\}$, $A \subseteq O_p$ and $B = \{p_i, q_i, p_j, q_j\}$, $B \subseteq L_p$, $i \neq j$, both consisting of four salient regional features centroids, not all collinear

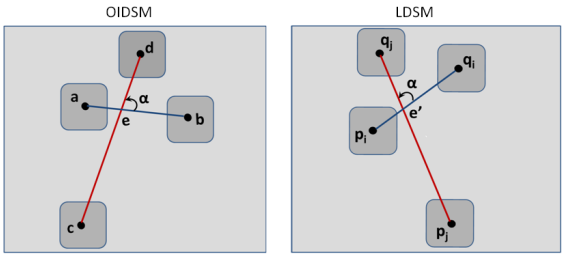

Figure 15: Geometric invariants of salient regional features

as shown in Fig. 15. Let e be the intersection point of line segments $\overline{ab}$ and $\overline{cd}$. The ratios $r_1 = \frac{|a-e|}{|a-b|}$, $r_2 = \frac{|c-e|}{|c-d|}$, $\kappa = \frac{|c-d|}{|a-b|}$, and the signed angle α defined on set A are invariant under similarity transformation. If line segment $\overline{p_iq_i}$ corresponds to $\overline{ab}$, then a line segment $\overline{p_jq_j}$ corresponding to $\overline{cd}$ must satisfy the following constraints:

i. share a common intersection point e' such that $e' = p_i + r_1(q_i - p_i)$ and $e' = p_j + r_2(q_j - p_j)$;
ii. given θ, the signed angle from x-axis to $\overrightarrow{p_iq_i}$, β, the signed angle from x-axis to $\overrightarrow{p_jq_j}$ can be expressed as $\beta = \theta + \alpha$;
iii. the ratio of lengths between $\overline{p_iq_i}$ and $\overline{p_jq_j}$ can be expressed as $\|p_jq_j\| = \kappa\|p_iq_i\|$.

2.4.2. *OIDSM Salient Regional Feature Selection*

The salient regional feature extraction process is able to eliminate most of the spurious regional features in the LDSM. However, several spurious regional features are detected as salient in the OIDSM due to the noise and outliers in the OIDSM that cannot be captured by statistical outlier filtering process. A minimum of four corresponding pair of regional features centroids in the LDSM and OIDSM are required for computing the registration parameters using geometric invariants. A small subset of the OIDSM salient features are used in searching for the 3D registration parameters. The subset is selected from a rank ordered list of the OIDSM salient regional features.

The first ranking criterion is based on the area of the OIDSM salient regional features. The salient regional features are ranked in descending order according to the actual number of pixels in the largest regional feature detected for each feature centroid across all morphological reconstruction thresholds. This ranking criteria ensures large regional features in the scene that corresponds to interest regions such as buildings are highly ranked.

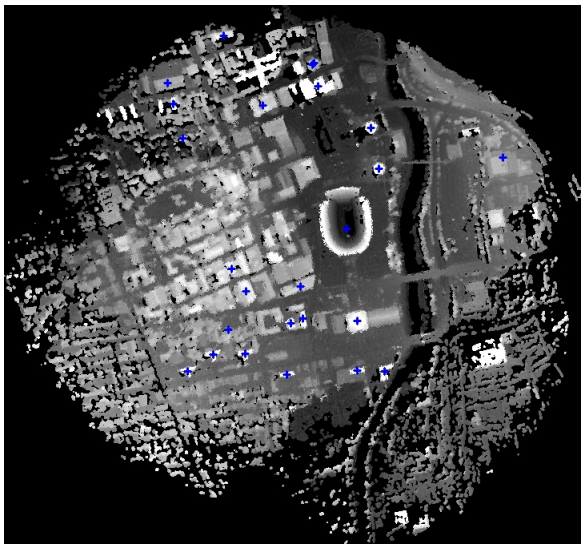

Figure 16: OIDSM top salient features

The second ranking criterion is based on the local height of the salient regional features largest areas. A $3x3$ window is moved over the boundary pixels of the largest regional feature detected for each feature centroid. At the boundary locations the minimum and maximum values inside the moving window are computed. The local height is then defined as the difference between the median of the maximum and the minimum boundary values. This ranking criteria ensures regional features that are much higher than surrounding regions are highly ranked.

The feature point areas and local heights are weighted equally in the computation of the rank values. The combination of the ranking criteria gives higher ranking values to regional features that have large areas and are much higher than surrounding regions. Fig. 16 shows the top salient features extracted from an OIDSM.

2.4.3. *Salient Feature Point Set Matching*

Given the top salient OIDSM feature points $O_{pT} \subseteq O_p$ and salient LDSM feature points L_p, the best pair of four point correspondences in the OIDSM and LDSM are computed iteratively. At each iteration a set of four salient feature points are selected from O_{pT} and candidate sets from L_p satisfying the geometric invariant ratios and angle are extracted. The closest match for each O_{pT} subset of four feature points is selected based on the average NCC score of overlapping patches generated by transforming the OIDSM to the LDSM coordinates using L_p candidate subsets. The point set pairs that yields the best average NCC score is selected as the

best matching salient feature set pairs. The algorithm terminates when the minimum average NCC score has been achieved or the maximum number of iterations is reached. The pseudocode for selection of best matching salient feature point sets is summarized in Algorithm 1.

There are four possible intersection point locations on each line segment $\overline{pq}$ formed by $\{p, q\} \subseteq L_p$ for each $\{a, b, c, d\} \subseteq O_{pT}$. The possible intersection point locations are:

- $\overline{pq} = \overline{ab} \Rightarrow e' = p + r_1(q - p)$
- $\overline{pq} = \overline{ba} \Rightarrow e' = q + (1 - r_1)(p - q)$
- $\overline{pq} = \overline{cd} \Rightarrow e' = p + r_2(q - p)$
- $\overline{pq} = \overline{dc} \Rightarrow e' = q + (1 - r_2)(p - q)$

We only need two equations containing ratios r_1 and r_2 to obtain all the possible intersection locations.

A tolerance value ε is used to limit the number of candidate line segments generated from feature points in L_p based on their lengths compared to the query line segments generated from the feature points in R_{PT}. In our experiments we use $\varepsilon = 2$. This assumes the variation of points along the normal direction to the LPC ground plane is less than twice the variation along the OIPC ground plane or vice versa. In situations where their are large variations in the terrain and scene content of the optical aerial imagery compared to that of the LiDAR coverage area, a larger ε will be required.

Once the possible intersection locations are known, B candidates that approximately correspond to query set A are extracted based on geometric invariant constraints in Section 2.4.1, using a kd-tree nearest neighbor search. The OIDSM is then transformed to the LDSM coordinates using each of the B candidates with A to create overlapping regions needed for NCC.

The $2D$ similarity transformation matrix $\mathbf{H}_B = [\mathbf{u}|\mathbf{v}]$ that transforms the OIDSM to the LDSM coordinates is computed for each B by minimizing the objective function

$$F(\mathbf{H}_B) = \sum_{k=1}^{4} \|B_k - \mathbf{u}A_k^T - \mathbf{v}\|^2, \tag{11}$$

where $[\mathbf{u}|\mathbf{v}]$ is of the form $\begin{bmatrix} u_1 & u_2 & v_1 \\ -u_2 & u_1 & v_2 \end{bmatrix}$, a 2D rotation, translation and isotropic scaling.

The average NCC similarity score between the LDSM and each transformed optical imagery-derived digital surface model (TOIDSM) is computed from the dot product of four normalized overlapping patches as follows

$$\rho_B = \frac{1}{4} \sum_{i=1}^{4} \left\langle \frac{\hat{L}_i}{\|\hat{L}_i\|}, \frac{\hat{T}_i}{\|\hat{T}_i\|} \right\rangle, \tag{12}$$

where L and T are the LDSM and TOIDSM respective overlapping patches centered at pi, qi, pj and qj. $\hat{L} = L - \overline{L}$, and $\hat{T} = T - \overline{T}$.

The most likely B corresponding to each A is selected as the B that generates the maximum NCC score ρ_B. The matching pair of A and B that approximately satisfies the geometric invariant constraints and yields that maximum ρ_B among all $\{A, B\}$ pairs is used for the initial 3D registration of the OIPC to the LPC.

Fig. 17 shows overlapping NCC patches generated from best matching point set pairs in the LDSM and TOIDSM. Fig. 18 shows OIDSM and LDSM matching line segments generated from the matching salient feature point sets.

The $x-$ and $y-coordinates$ of the 3D locations corre-

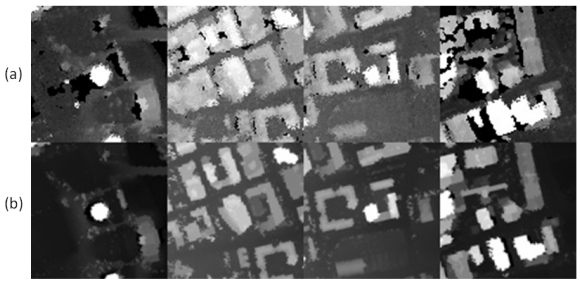

Figure 17: Overlapping NCC patches generated from best matching point set pairs in the LDSM and TOIDSM. (a) TOIDSM patches. (b) LDSM patches.

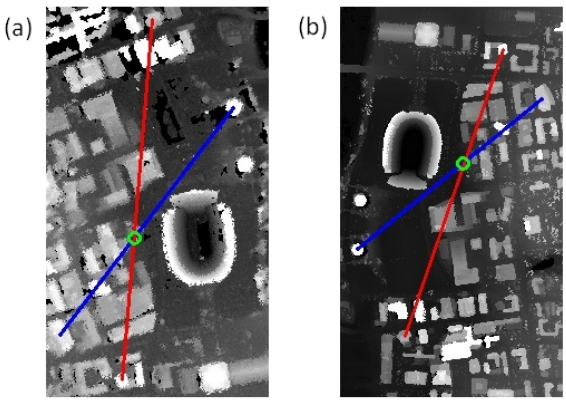

Figure 18: OIDSM and LDSM matching line segments generated from matching salient feature point sets. (a) OIDSM line segments. (b) LDSM line segments.

Algorithm 1 Salient feature point set matching

while $\{Iter < MaxIter\}$ & $\{\rho < \rho_{min}\}$ **do**
  for each $A = \{a, b, c, d\} \subseteq O_{pT}$ **do**
    Compute ratios r_1, $1 - r_1$, r_2, $1 - r_2$, κ, and angle α
    for each $\{p_j, q_j\} \subseteq L_p$, $\frac{\|cd\|}{\varepsilon} \leq \|p_j q_j\| \leq \varepsilon\|cd\|$ **do**
      Compute possible intersection locations e' based on ratios r_2 and $1 - r_2$
      Compute β_i, the signed angle from x-axis to $\overrightarrow{p_j q_j}$
      Compute $\|p_j q_j\|$
    for each $\{p_i, q_i\} \subseteq L_p$, $\frac{\|ab\|}{\varepsilon} \leq \|p_i q_i\| \leq \varepsilon\|ab\|$ **do**
      Compute possible intersection locations e' based on ratios r_1 and $1 - r_1$
      Compute θ_i, the signed angle from x-axis to $\overrightarrow{p_i q_i}$
      Compute $\|p_i q_i\|$
      Find closest $\{p_j, q_j\}$ that approximately satisfies invariant constraints in 2.4.1 using a k-d tree search
    for each candidate feature point set $B = \{p_i, q_i, p_j, q_j\}$ **do**
      Compute the 2D similarity transformation matrix $\mathbf{H}_B$
      Transform the OIDSM to LDSM coordinates using $\mathbf{H}_B$
      Extract four overlapping patches in the LDSM and TOIDSM, centered at p_i, q_i, p_j and q_j
      Compute the overlapping patches average NCC ratio ρ_B
    Extract B corresponding to A as $\underset{B}{\operatorname{argmax}}(\rho_B)$
  Return best matching salient feature set pairs as $\underset{A,B}{\operatorname{argmax}}(\rho_B)$

sponding to the best matching $\{A, B\}$ can be directly extracted from the 3D points which are used to determine pixel locations of the salient feature points in the LDSM and OIDSM. However, there are no direct mapping between the height values of the salient feature points in the DSMs and their corresponding values in the point clouds.

The $z-$ *coordinate* of the feature points in 3D are estimated as the median of the $z-$ coordinate of all points inside the largest regional features boundary that is used to generate the respective salient features in the LDSM and OIDSM. The 3D similarity transformation $\mathbf{H}_{NC} \in \Re^{4x4}$ that coarsely map the nadir-view OIPC to the LPC coordinates is then estimated from the matching 3D points, using the closed-form least-squares solution approach in [41]. The 3D similarity transformation that coarsely registers the original OIPC to the LPC can then be computed from $\mathbf{H}_N$ and $\mathbf{H}_{NC}$ as

$$\mathbf{H}_C = \mathbf{H}_{NC}\mathbf{H}_N. \tag{13}$$

3. Fine Registration Parameters Estimation

This section discuses the refinement of the coarse 3D registration parameters and the computation of the absolute camera poses that aligns the optical images to the LPC. The coarse 3D transformation parameters are refined using the ICP. The absolute camera poses for all the images are jointly computed through the refined 3D transformation parameters.

3.1. 3D Transformation Refinement

Ideally, the 3D similarity transformation $\mathbf{H}_C$ should be sufficient for registration of the OIPC to the LPC. However, due to errors in the nadir-view rotation of the OIPC and in the estimation of 3D coordinates of the matching salient feature points, an additional refinement process is necessary to obtain an absolute registration in LiDAR coordinates. The ICP algorithm is used to complete this registration process.

Given two point clouds, the ICP algorithm attempts to optimally align an observed point cloud to a reference point cloud. The ICP algorithm involves three main step that are iteratively repeated until the change in the global error between the point cloud falls below a given threshold. First each point in the observed point cloud is associated with its nearest neighbor in the reference point cloud. Next, the orientation and translation parameters that best align each observed point to its nearest neighbor is estimated. Finally the observed point cloud is transformed based on the estimated parameters.

Let $\mathbf{X}_L$ be the LPC and $\mathbf{X}_C = \mathbf{H}_C\mathbf{X}_O$ be the coarsely

registered OIPC, the Mean Squared Error (MSE) objective function, $F(\mathbf{H}_R)$, minimized by the ICP algorithm is given by

$$F(\mathbf{H}_R) = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{H}_R\mathbf{X}_C - \mathbf{X}_L\|^2, \tag{14}$$

where $\mathbf{H}_R$ is the refinement transformation matrix that represents the difference between the coarsely aligned OIPC and absolute registration to the LPC.

Hence, the absolute transformation matrix $\mathbf{H}$ that transforms the OIPC to the LPC coordinates is given by

$$\mathbf{H} = \mathbf{H}_R\mathbf{H}_C. \tag{15}$$

The absolute transformation $\mathbf{H}$ can be decomposed as

$$\mathbf{H} = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}, \tag{16}$$

where $\mathbf{R} \in \Re^{3x3}$ is the orientation matrix, $\mathbf{t} \in \Re^{3x1}$ is the translation matrix, s is an isotropic scale, and $\mathbf{0} \in \Re^{1x3}$ is a null vector.

The ICP algorithm can be heavily influenced by outliers. To reduce the effect of outliers on registration refinement process, the algorithm is implemented with rejection strategies. The point clouds are winsorized by rejecting 10% of the point pairs with the largest point to point distances. The OIPC point matches to the LiDAR edge vertices are also rejected to avoid incorporation of poor reconstruction at the edges in MSE estimation.

3.2. Absolute Camera Poses Computation

Once the absolute similarity transformation matrix $\mathbf{H}$ is known, the absolute camera pose for registering the optical aerial imagery to the LPC can be computed by

$$\mathbf{P} = \mathbf{P}_O\mathbf{H}^{-1}. \tag{17}$$

4. Experimental Results

The registration algorithm is tested on three experimental data sets. These include one simulated scene as a proof of concept, and two real world environment data sets. The simulated data set represents optical aerial imagery collected by an aircraft flying a circular path around a region of interest, and the corresponding LiDAR data sampled from the scene. The second data set is obtained from the Ohio State University (OSU) Campus and neighboring area [42, 43]. The third data set covers the downtown region of Providence, Rhode Island. [44, 45].

In all data sets, the input optical aerial imagery and the LPC are acquired at different resolutions. However, the presence of salient regional features in the scene enables the selection of features which are used in the registration of the data sets. The registration of the optical information and OIPC from the aerial imagery to the LiDAR produces a scene model that draws the complimentary characteristics from both sensor types.

The registration results for the three data sets are shown in Figures 19–21. The results for each data set show (a) a single frame of optical aerial imagery, (b) the LPC, (c) the textured OIPC, (d) the textured LPC, (e) the textured registered model, (f) the zoomed in image of the blue box region of the textured OIPC, (g) the zoomed in image of the blue box region of the textured LPC, and (h) the zoomed in image of the blue box region of the textured registered model.

The density of the OIPC varies significantly across the scene, especially in the real world data sets as shown in Fig. 20(c) and Fig. 21(c). This is because of the unrepresented locations in the OIPC reconstruction process. The OSU scene contains trees which are difficult to reconstruct from images. Poor photo-consistency and imagery matching errors also produce missing reconstructions. The LPC provides height information for these missing regions in the OIPC as can been seen in the registered models shown in Fig. 19(e), Fig. 20(e) and Fig. 21(e).

When a scene is texture rich and the images acquired from the scene have little radiometric errors and good photo-consistency, the 3D structure of the scene can be recovered in very good detail such as the OIPC model from simulated data in Fig. 19(c). However, this is not the case in real world scenarios. Nonetheless, the OIPC provides optical augmentation to the LPC and structural addition in regions not captured by LPC.

As shown in Fig. 19(d), Fig. 20(d) and Fig. 21(d), the LPC does not capture the facade details of the structures in the scene. Although these vertical details can be inferred from roof boundaries, the true structure will not be correctly represented. In Fig. 19(e), 20(e), and 21(e), the missing facade details are filled in with the OIPC. The optical information are also projected on the LPC using the computed registration parameters.

Since the LPC only measures the highest point for each $x-$ and $y-$ *coordinate* locations, the accuracy of the registration framework is evaluated based on the displacement error between the LiDAR points and the corresponding points from the registered OIPC used to generate the OIDSM.

The coarse registration process provides registration parameters, which is sufficient for the ICP refinement

initialization. The total coarse registration displacement Root Mean Square Error (RMSE) for the simulated scene, the OSU region and Providence downtown area are 2.4078 meters, 5.7951 meters, and 7.7635 meters, respectively. The RMSEs are below the 25 meters displacement error required for ICP based on the experiments in [17]. The RMSE is computed from nearest neighbor correspondences in the point clouds.

A natural neighbor interpolation scheme is also used to get an estimate of the RMSE between the registered point cloud and the OIPC taking into account the resolution differences between the point clouds. The interpolant is created from the LPC as it has better distribution of points in each scene. The OIPC is then resampled to a new grid based on the interpolation function generated from the LPC.

The RMSE after coarse registration, final refinement, and interpolation for the simulated scene, the OSU region, and the Providence downtown area are shown in Tables 1–3, respectively. An histogram of the distribution of errors for the respective scenes are also shown in Figures 22–24.

5. Conclusions

A framework for registering oblique aerial imagery to a LPC is presented. The framework simultaneously registers both the optical and structural information extracted from oblique imagery to a LPC without requiring prior knowledge of an initial alignment. Moreover, the ability of the registration framework to register aerial imagery and LPC at different resolutions has been demonstrated. Results on a simulated scene and real world environments shows the robustness of the registration approach. The final textured model generated from the registration process shows the complimentary characteristics of optical aerial imagery and LPC.

In future works it would be interesting to look at methods for significantly reducing the number of outlier features in the OIDSM and LDSM as this will reduce the number of candidate point sets used to search for coarse registration parameters. Also, since the search for candidate matches for each OIDSM query point set are independent of other OIDSM query point sets, a GPU implementation would take advantage of the highly parallel nature of the correspondence search process. Another interesting direction would be to extend the algorithm for registration of large-scale multi-modal aerial imagery and aerial imagery of the same modality collected at different times.

References

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

[2] F. Leberl, A. Irschara, T. Pock, P. Meixner, M. Gruber, S. Scholz, A. Wiechert, Point clouds: Lidar versus 3d vision, PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING 76 (10) (2010) 1123–1134.

[3] M. Pollefeys, D. Nistér, J. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. Kim, P. Merrell, et al., Detailed real-time urban 3d reconstruction from video, International Journal of Computer Vision 78 (2) (2008) 143–167.

[4] A. Akbarzadeh, J. Frahm, P. Mordohai, B. Clipp, C. Engels, D. Gallup, P. Merrell, M. Phelps, S. Sinha, B. Talton, et al., Towards urban 3d reconstruction from video, in: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), IEEE Computer Society, 2006, pp. 1–8.

[5] Y. Furukawa, J. Ponce, Accurate, dense, and robust multiview stereopsis, IEEE transactions on pattern analysis and machine intelligence (2009) 1362–1376.

[6] I. Stamos, P. Allen, Geometry and texture recovery of scenes of large scale, Computer Vision and Image Understanding 88 (2) (2002) 94–118.

[7] L. Liu, I. Stamos, Automatic 3d to 2d registration for the photorealistic rendering of urban scenes, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 2, IEEE, 2005, pp. 137–143.

[8] L. Liu, G. Yu, G. Wolberg, S. Zokai, Multiview geometry for texture mapping 2d images onto 3d range data, in: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol. 2, IEEE, 2006, pp. 2293–2300.

[9] I. Stamos, L. Liu, C. Chen, G. Wolberg, G. Yu, S. Zokai, Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes, International Journal of Computer Vision 78 (2) (2008) 237–260.

[10] C. Frueh, A. Zakhor, Constructing 3d city models by merging ground-based and airborne views, in: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, Vol. 2, IEEE, 2003, pp. II–562.

[11] C. Früh, A. Zakhor, An automated method for large-scale, ground-based city model acquisition, International Journal of Computer Vision 60 (1) (2004) 5–24.

[12] M. Ding, K. Lyngbaek, A. Zakhor, Automatic registration of aerial imagery with untextured 3d lidar models, in: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, 2008, pp. 1–8.

[13] L. Wang, U. Neumann, A robust approach for automatic registration of aerial images with untextured aerial lidar data, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 2623–2630.

[14] A. Mastin, J. Kepner, J. Fisher, Automatic registration of lidar and optical images of urban scenes, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 2639–2646.

[15] F. Lafarge, C. Mallet, Creating large-scale city models from 3d-point clouds: A robust approach with hybrid representation, International journal of computer vision (2012) 1–17.

[16] Q. Zhou, U. Neumann, 2.5 d dual contouring: A robust approach to creating building models from aerial lidar point clouds, Computer Vision–ECCV 2010 (2010) 115–128.

[17] W. Zhao, D. Nister, S. Hsu, Alignment of continuous video onto 3d point clouds, IEEE transactions on pattern analysis and machine intelligence (2005) 1305–1318.

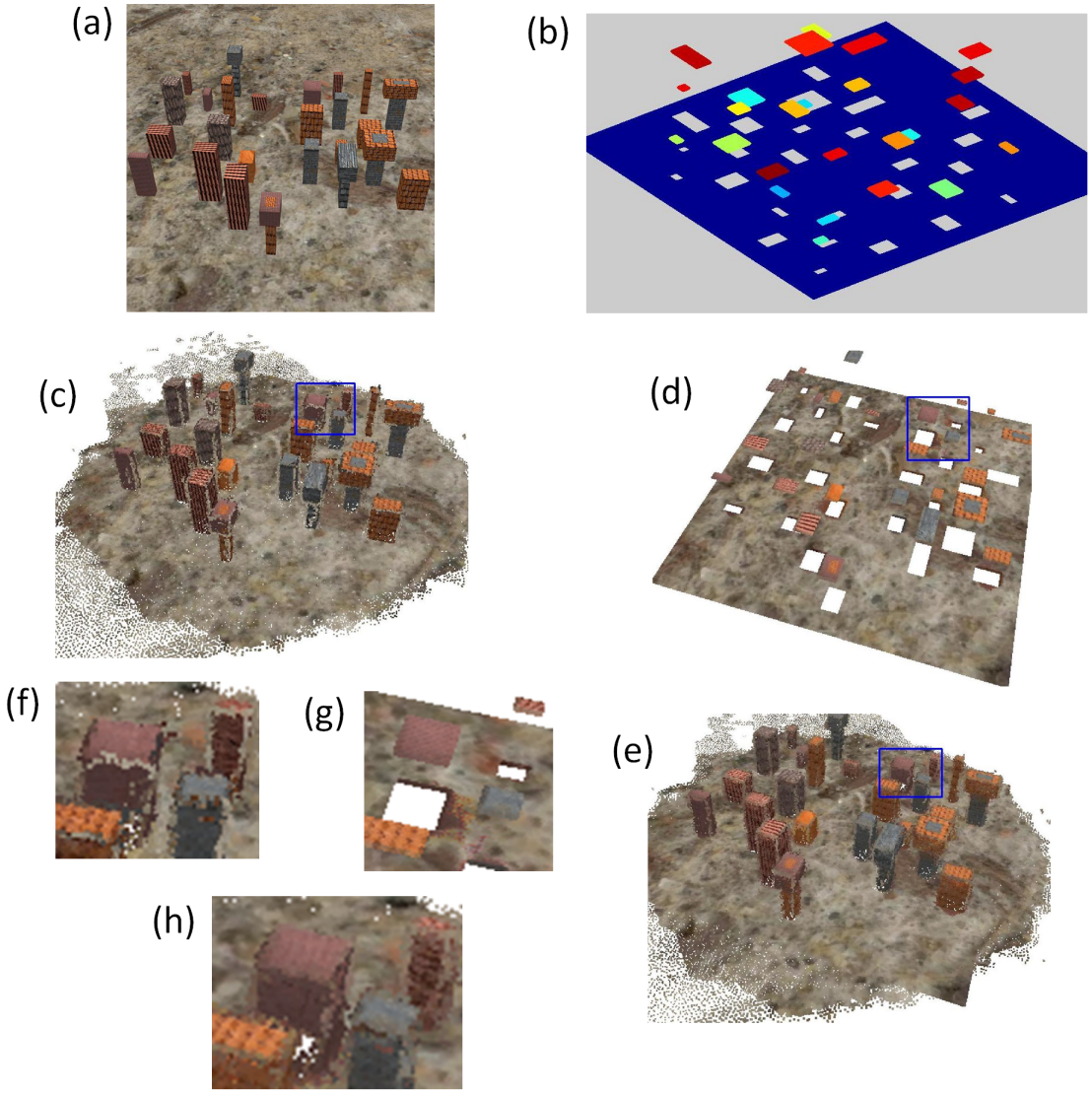

Figure 19: Registration of optical aerial imagery and LPC of the simulated scene. (a) single frame of optical aerial imagery, (b) LPC, (c) textured OIPC, (d) textured LPC, (e) textured registered model, (f) zoomed in region of the OIPC shows varying point cloud densities due to missing reconstructions, (g) zoomed in region of the LPC shows missing vertical details, (h) zoomed in registered model with optical and structural details from the oblique aerial imagery and LPC

Table 1: Simulated scene displacement RMSE (meters)

	x	y	z	Euclidean distance
Coarse	0.3097	0.3108	2.3675	2.4078
Refined	0.2884	0.2883	0.4634	0.6173
Interp	-	-	0.5079	0.5079

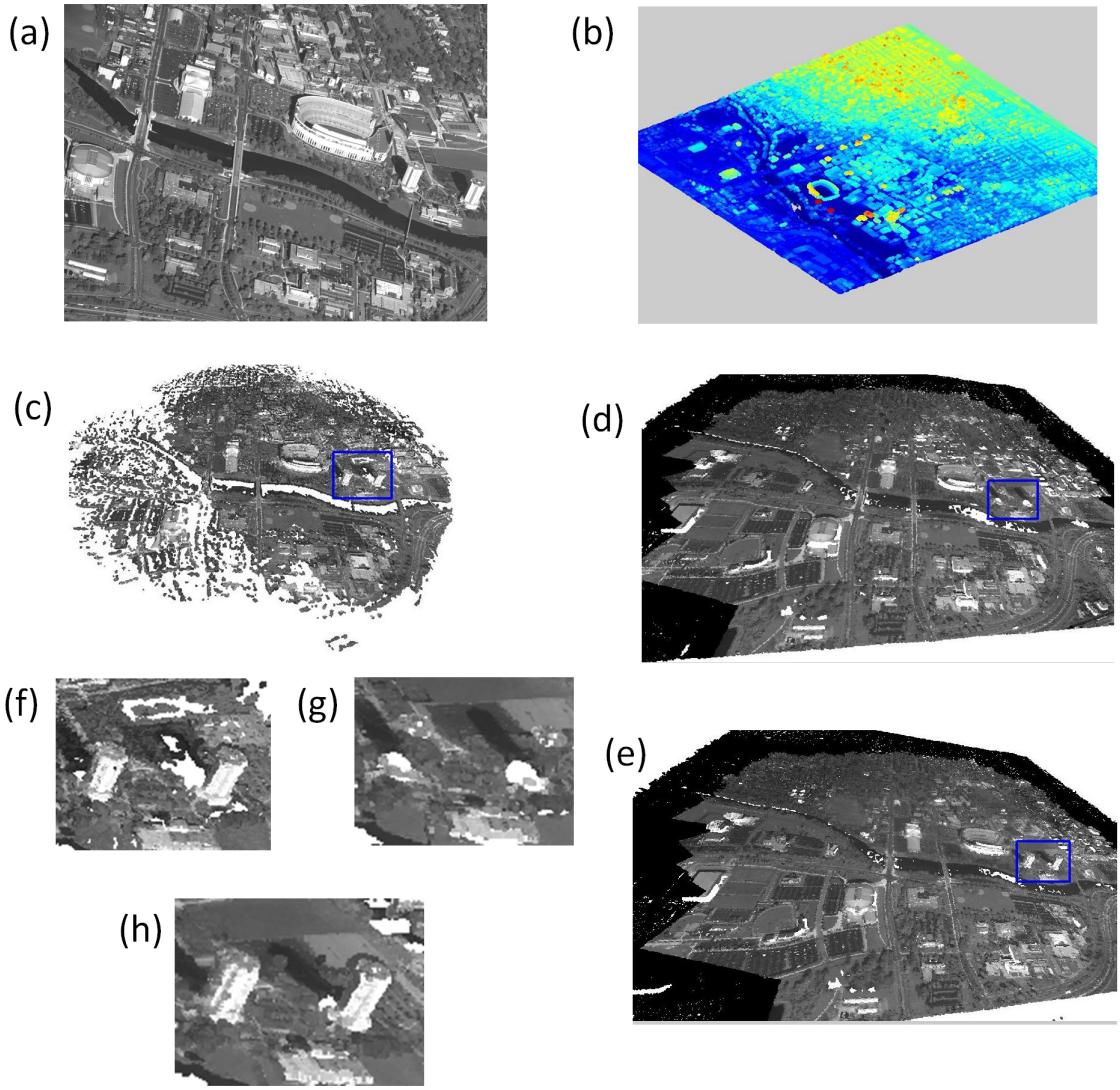

Figure 20: Registration of optical aerial imagery and LPC of the OSU region. (a) single frame of optical aerial imagery, (b) LPC, (c) textured OIPC, (d) textured LPC, (e) textured registered model, (f) zoomed in region of the OIPC shows varying point cloud densities due to missing reconstructions, (g) zoomed in region of the LPC shows missing vertical details, (h) zoomed in registered model with optical and structural details from the oblique aerial imagery and LPC

Table 2: OSU region displacement RMSE (meters)

	x	y	z	Euclidean distance
Coarse	3.3922	3.1805	3.4584	5.7951
Refined	0.8325	0.7344	1.1302	1.5843
Interp	-	-	1.3393	1.3393

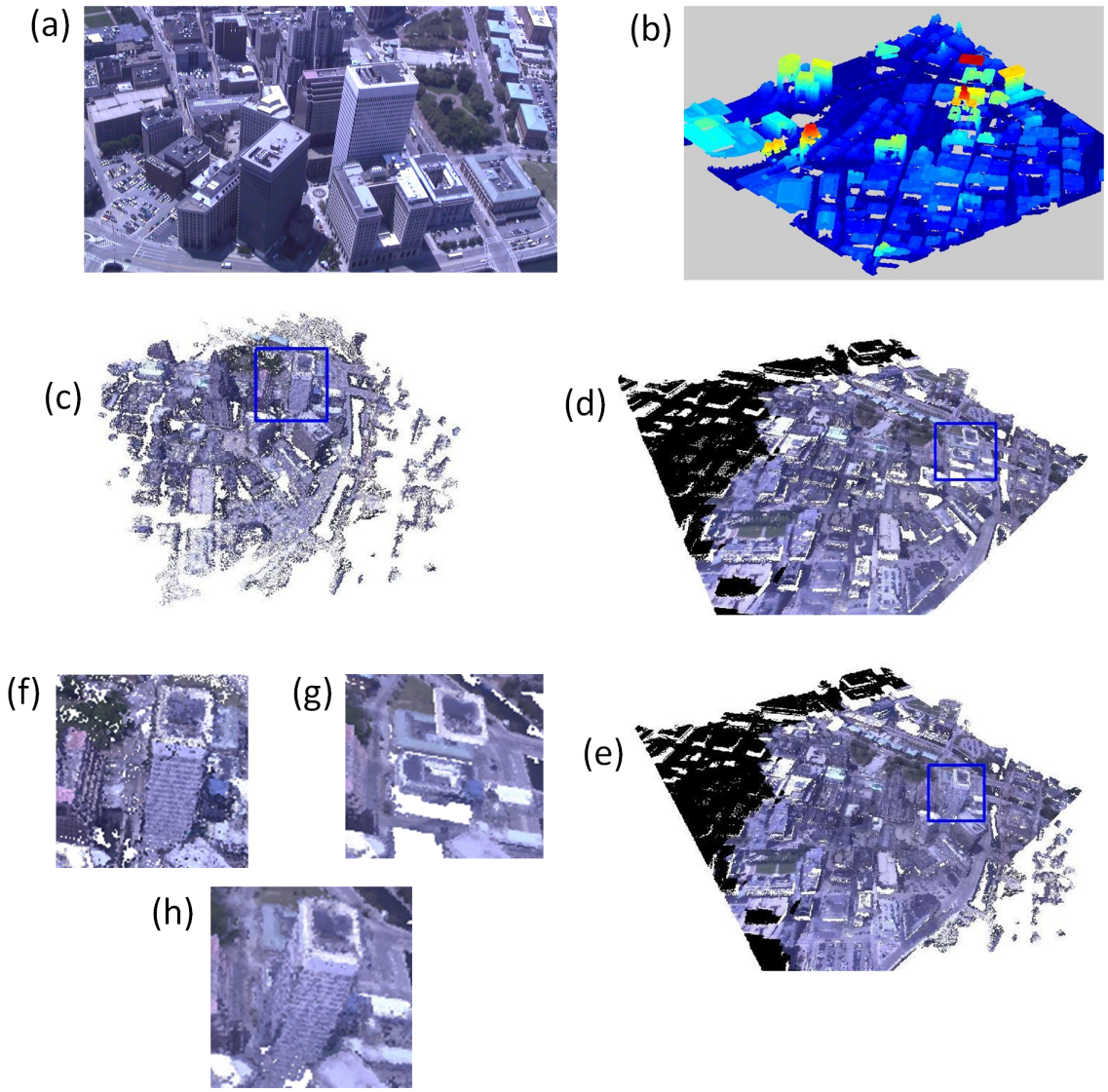

Figure 21: Registration of optical aerial imagery and LPC of the Providence downtown region. (a) single frame of optical aerial imagery, (b) LPC, (c) textured OIPC, (d) textured LPC, (e) textured registered model, (f) zoomed in region of the textured OIPC shows varying point cloud densities due to missing reconstructions, (g) zoomed in region of the textured LPC shows missing vertical details, (h) zoomed in textured registered model with optical and structural details from the oblique aerial imagery and LPC

Table 3: Providence downtown displacement RMSE (meters)

	x	y	z	Euclidean distance
Coarse	3.7646	2.9488	6.3679	7.9635
Refined	0.4539	0.4773	0.8998	1.1151
Interp	-	-	0.9539	0.9539

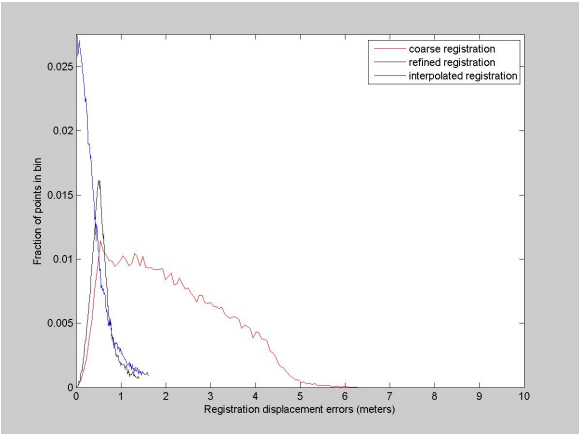

Figure 22: Simulated scene displacement error distribution

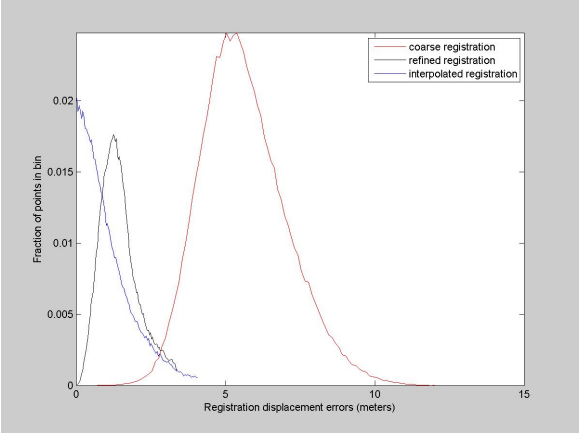

Figure 23: OSU region data displacement error distribution

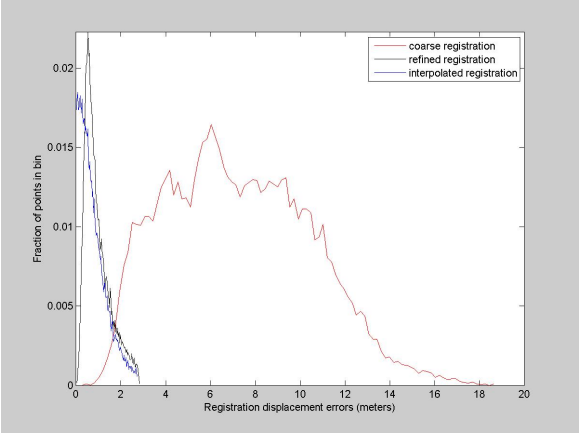

Figure 24: Providence downtown data displacement error distribution

[18] D. Nistér, An efficient solution to the five-point relative pose problem, Pattern Analysis and Machine Intelligence, IEEE Transactions on 26 (6) (2004) 756–770.

[19] R. Kumar, P. Anandan, K. Hanna, Direct recovery of shape from multiple views: A parallax based approach, in: Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on, Vol. 1, IEEE, 1994, pp. 685–688.

[20] M. Irani, P. Anandan, Parallax geometry of pairs of points for 3d scene analysis, Computer VisionECCV'96 (1996) 17–30.

[21] S. Rusinkiewicz, M. Levoy, Efficient variants of the icp algorithm, in: 3dim, Published by the IEEE Computer Society, 2001, p. 145.

[22] P. Besl, N. McKay, A method for registration of 3-d shapes, IEEE Transactions on pattern analysis and machine intelligence (1992) 239–256.

[23] K. Pulli, Multiview registration for large data sets, in: 3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on, IEEE, 1999, pp. 160–168.

[24] X. Li, I. Guskov, Multi-scale features for approximate alignment of point-based surfaces, in: Proceedings of the third Eurographics symposium on Geometry processing, Eurographics Association, 2005, pp. 217–es.

[25] N. Gelfand, N. Mitra, L. Guibas, H. Pottmann, Robust global registration, in: Proceedings of the third Eurographics symposium on Geometry processing, Eurographics Association, 2005, pp. 197–es.

[26] R. Rusu, N. Blodow, M. Beetz, Fast point feature histograms (fpfh) for 3d registration, in: Robotics and Automation, 2009. ICRA'09. IEEE International Conference on, IEEE, 2009, pp. 3212–3217.

[27] Q. Mérigot, M. Ovsjanikov, L. Guibas, Robust voronoi-based curvature and feature estimation, in: 2009 SIAM/ACM Joint Conference on Geometric and Physical Modeling, ACM, 2009, pp. 1–12.

[28] D. Aiger, N. Mitra, D. Cohen-Or, 4-points congruent sets for robust pairwise surface registration, in: ACM SIGGRAPH 2008 papers, ACM, 2008, pp. 1–10.

[29] Y. Meng, H. Zhang, Registration of point clouds using sample-sphere and adaptive distance restriction, The Visual Computer (2011) 1–11.

[30] R. Hartley, A. Zisserman, Multiple view geometry, Vol. 642, Cambridge university press Cambridge, UK, 2000.

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

[32] C. Wu, S. Agarwal, B. Curless, S. Seitz, Multicore bundle adjustment, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 3057–3064.

[33] R. Rusu, Z. Marton, N. Blodow, M. Dolha, M. Beetz, Towards 3d point cloud based object maps for household environments, Robotics and Autonomous Systems 56 (11) (2008) 927–941.

[34] I. Jolliffe, Principal component analysis, Wiley Online Library, 2005.

[35] R. Schnabel, R. Wahl, R. Klein, Efficient ransac for point-cloud shape detection, in: Computer Graphics Forum, Vol. 26, Wiley Online Library, 2007, pp. 214–226.

[36] R. C. Gonzalez, R. E. Woods, Digital image processing using MATLAB, PrenticeHall, Upper Saddle River, NJ, 2008.

[37] H. Arefi, M. Hahn, A morphological reconstruction algorithm for separating off-terrain points from terrain points in laser scanning data, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (2005) 3.

[38] P. Soille, Morphological image analysis: principles and applica-

tions, Springer-Verlag New York, Inc., 2003.
[39] L. Vincent, Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms, Image Processing, IEEE Transactions on 2 (2) (1993) 176–201.
[40] M. B. Dillencourt, H. Samet, M. Tamminen, A general approach to connected-component labeling for arbitrary image representations, Journal of the ACM (JACM) 39 (2) (1992) 253–280.
[41] B. Horn, Closed-form solution of absolute orientation using unit quaternions, JOSA A 4 (4) (1987) 629–642.
[42] S. D. M. System, Columbus large image format (clif) 2006.
URL https://www.sdms.afrl.af.mil/index.php?collection=clif2006
[43] O. S. I. Program, 2006- 2010 osip i.
URL http://gis3.oit.ohio.gov/geodata/
[44] M. I. Restrepo, B. A. Mayer, A. O. Ulusoy, J. L. Mundy, Characterization of 3-d volumetric probabilistic scenes for object recognition, Selected Topics in Signal Processing, IEEE Journal of 6 (5) (2012) 522–537.
[45] T. U. of Rhode Island, Spring 2011 rhode island statewide lidar data.
URL http://www.edc.uri.edu/rigis/data/download/lidar/2011USGS/