

Fusion of multiple image types for the creation of radiometrically-accurate synthetic scenes

Stephen R. Lach,^{a,b} John P. Kerekes,^a and Xiaofeng Fan^a

^a Digital Imaging and Remote Sensing Laboratory,
54 Lomb Memorial Drive, Rochester, NY 14623,
Chester F. Carlson Center for Imaging Science,
Rochester Institute of Technology
sfl1194@cis.rit.edu, kerekes@cis.rit.edu, xxf3764@cis.rit.edu

^b Air Force Institute of Technology,
2950 Hobson Way,
WPAFB, OH 45433,
United States Air Force

Abstract. The Digital Imaging and Remote Sensing Image Generation (DIRSIG) model is an established, first-principles based scene simulation tool that produces synthetic multi-spectral and hyperspectral images from the visible to long wave infrared (0.4 to 20 microns). Over the last few years, significant enhancements such as spectral polarimetric and active Light Detection and Ranging (lidar) models have also been incorporated into the software, providing an extremely powerful tool for algorithm testing and sensor evaluation. However, the extensive time required to create large-scale scenes has limited DIRSIG's ability to generate scenes "on demand." To date, scene generation has been a laborious, time-intensive process, as the terrain model, CAD objects and background maps have to be created and attributed manually. To shorten the time required for this process, we have developed a comprehensive workflow aimed at reducing the man-in-the-loop requirements for many aspects of synthetic hyperspectral scene construction. Through a fusion of 3D lidar data with passive imagery, we have been able to partially-automate many of the required tasks in the creation of high-resolution urban DIRSIG scenes. This paper presents a description of these techniques.

Keywords: lidar, hyperspectral, fusion, DIRSIG, building reconstruction, synthetic scene

1 INTRODUCTION

Over the past twenty years, there has been a significantly increasing demand for accurate three dimensional scene models. Although until recently most of these models were primarily created through manual methods, they still proved to be cost-effective tools in such diverse applications as town planning [1], computer graphics [2], and military training [3]. As the technical and commercial advantages to scene simulation became more widespread, additional applications such as climate and environmental research, noise propagation, navigation, and large-scale urban planning created the need to generate such scenes in less time and with even greater accuracy [4]. To this end, there has been a concerted research effort over the last 10 years focusing on the automatic or semi-automatic retrieval of accurate scene geometries from remotely-sensed data. Initially this research was dominated by traditionally photogrammetric approaches, but the increased availability of laser scanning systems over the past few years has made the direct processing of lidar-produced point clouds a viable

alternative [5]. In addition to the exploitation of single modality images, several authors have also proposed techniques of fusing these two approaches to produce improved results. However, in nearly all cases, images created from the extracted models are required only to *appear* realistic to a human observer, and the radiometric accuracy of the real scene is not preserved in either the extracted scene model or simulated imagery based on such models.

In contrast to the applications utilizing appearance-based models, the Digital Imaging and Remote Sensing Image Generation (DIRSIG) tool [6] is a first-principles based model that produces *physically-accurate* synthetic spectral imagery of pre-defined scenes. This synthetic spectral imagery is then used in a host of applications, including spectral sensor development, algorithm test and evaluation, and image analyst training. DIRSIG uses detailed computer-aided-design drawings for man-made and natural objects, along with material maps and associated characteristics as first level input parameters. Standard atmospheric propagation codes are then used to predict the at-sensor radiance from the scene, as would be seen by a broadband or spectral imaging sensor. Detailed models for the sensor are then applied to the received radiance, producing a radiometrically-correct simulated digital image.

DIRSIG has been used to create simulated imagery of complex synthetic urban scenes on the order of several square kilometers in area at meter-scale resolution. However, the definition and construction of these scenes is quite labor intensive and, in many cases, has taken several months to complete. This work aims to describe our efforts in semi-automating the construction of DIRSIG scenes by processing data from multiple sources.

With the continued development and deployment of new remote sensing technologies, it is increasingly common to have data from multiple sources over a common area. These various remote sensing techniques, with their distinctive phenomenology, measure what are often complementary characteristics of a scene. For example, high resolution optical imagery can provide fine two-dimensional spatial details, while lidar can provide accurate 3D position information [7]. Similarly, infrared data provide information related to surface temperatures, while reflective hyperspectral imagery can characterize the surface material type and condition [8]. With data from all these multiple sources, a comprehensive characterization of the scene becomes possible.

In this research effort, we have fused a few previously-published techniques with several novel approaches for accomplishing the many diverse tasks required for physically-accurate scene models extracted from remotely-sensed data. Work to date has focused on the spatial registration of multiple high-resolution data sources, terrain extraction, object identification and reconstruction from dense (6 points/m²) lidar data and frame array imagery, and surface characterization using hyperspectral imagery. The outputs from the various data processing steps are then combined in the proper format for their use as inputs to DIRSIG. The following sections will describe in greater detail the multi-source processing approach, along with selected experimental results.

2 THE DIRSIG MODEL

The DIRSIG model is a complex synthetic image generation utility that has been developed at the Digital Imaging and Remote Sensing (DIRS) Laboratory at the Rochester Institute of Technology (RIT) over the last 20 years. The tool was originally designed to model the thermal infrared region of the electromagnetic spectrum but was expanded several years ago to cover the full visible to long-wave infrared (0.4 to 20 micron) range [6]. It effectively models broadband, multi-spectral and hyperspectral imagery using a suite of first-principles-based radiation propagation modules. These modules perform specific tasks such as predicting bi-directional reflectance functions (BRDF), computing time and material dependent surface temperature values, and computing the dynamic viewing geometries of scanning instruments on a moving platform. In addition to these DIRSIG-specific modules,

the tool leverages several utilities (such as MODTRAN and FASCODE) that have been used extensively by the remote sensing community.

In 2002, a first-principles-based lidar model was incorporated into the passive radiometry framework enabling the model to calculate arbitrary, time-gated photon counts at the sensor for atmospheric, topographic, and volumetric backscattered returns. The DIRSIG lidar model was first developed by Burton [9], and it was recently expanded by Blevins [10] to handle a wide variety of complicated scene geometries, diverse surface and participating media optical characteristics, and a variety of sensor models.

Although the passive and active models used by DIRSIG are valuable in their own right, in many applications the ability to integrate passive electro-optical (EO) and lidar simulations using a common scene (and potentially common viewing geometries) is of even greater utility. The multi-modal capabilities of the DIRSIG model allow designers to evaluate both passive and active approaches to solving specific imaging problems, as well as potential fusion techniques using both image modalities. Additionally, the upper performance limit for a given approach may be more easily determined when using a common scene.

3 SEMI-AUTOMATED SCENE CONSTRUCTION PROCESS

The baseline method of creating a scene for use in DIRSIG simulations is straightforward, but unfortunately it is also quite labor and time intensive. Terrain models are produced manually or extracted from 'small scale' USGS Digital Elevation Models [11], which are available in 1x1 degree blocks for most of the United States. These array-based elevation models are subsequently re-sampled to the triangular irregular network (TIN) format required by DIRSIG. Object geometries are produced using CAD-based tools; trees are typically generated in Onyx Tree Professional [12], while man-made structures are defined using tools such as Rhinoceros [13]. Spectral material assignment is done manually on a per-facet basis using the Bulldozer in-house software utility, or via detailed, manually-derived, material maps. If multiple spectral curves of a given material are available, additional mapping images may be used to drive in-material spatial-spectral variations. To achieve this, relatively-high spatial resolution aerial panchromatic or spectral imagery is used to define the particular spectral curve used for each location in the final scene. This selection is based on a look-up table relating pixel values of the texture image to spectral signatures. The detailed process for the baseline scene generation technique is further described in [14] and [15], and it relies on expertise with many different software tools. This process typically requires several months to complete a large scale scene.

Although the baseline process for creating DIRSIG scenes has proven to yield excellent resultant images, a streamlining of the construction of large scenes would be beneficial to many users. To this end, a new process has been implemented where many of the scene design tasks have been replaced by semi-automated methods. The basic approach is to initially extract enough feature information from the lidar data to perform a registration among the various datasets, then to refine and add to these features using all available imagery. These steps are depicted below in Fig. 1, and they will be explained in greater detail throughout the remainder of this section.

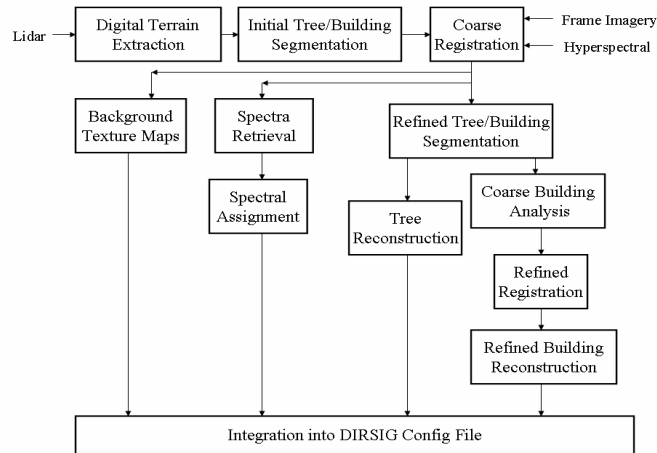


Fig. 1. Semi-automated process for DIRSIG scene construction.

3.1 Digital Terrain Model Extraction

In order to explicitly define scene information from the lidar data, we must first extract an accurate geometric model of the bare earth, where data points derived from buildings, trees and other non-ground objects have been removed. In creating this digital terrain model (DTM), the irregularly-sampled lidar point cloud may be processed directly, but we typically interpolate the data onto a regular grid to form a (rasterized) range image to decrease the computational requirements. In either case, the fundamental concept is to separate the ground points from the non-ground points, then to remove the non-ground points.

Due to the many features that distinguish ground points from adjacent objects, a multitude of DTM extraction techniques have been proposed in the literature. Using the assumption that object points are higher than the adjacent ground, morphology filters have been applied to distinguish terrain points from other features in the lidar data [16], [17], [18]. These techniques define a window that is larger than the largest building, and this window is used to specify a neighborhood of pixels used at each point during the processing. For each pixel in the DEM, the minimum value in its neighborhood (with the window centered on the pixel of interest) is found and assigned to that pixel. After this erosion operation is completed, a dilation is performed, whereby each pixel in the processed image is examined relative to its neighbors, and the maximum pixel value in the neighborhood is assigned to the pixel of interest. This method of terrain extraction has proven to be successful in many applications, but it has been found to be highly susceptible to noise in the DEM. Although a median filter may be applied to mitigate some of the noise effects, such an approach is unable to compensate for larger patches of noisy data [19]. In [20], the author describes a modification to the erosion/dilation technique, where he proposes using a slope-based filter to identify object points. In this approach, a point is classified as being non-ground if the maximal slope of the vectors connecting a point to its neighbors is above a pre-defined threshold. This technique was further modified in [21] and [22], where additional heuristics are added to increase accuracy within building outlines and local terrain features are accounted for. The approach proposed in [23] uses active contours to define the ground model. This methodology starts with a surface that is initially defined to be lower than the lidar data. The surface is then iteratively moved upwards while undergoing continual changes in shape. When it finally is in close contact with the ground, a parametric model of the terrain is available, as are the original data points which are in contact with this model.

Although these techniques work well, for many terrain regions we found that satisfactory results could also be obtained by applying a simple sliding-window filtering technique to the

range image. In cases where neither the tree nor building density is high, a variation of the spatial sliding window median filter may be used to identify points that are significantly higher than their neighbors. Although in certain cases a standard median filter could be used, care must be taken to ensure that the kernel is large enough to span the roof structures of the largest buildings in the scene. If it is not, the central points of large objects may not be flagged as being non-ground, and more complicated processing would be required. However, when the kernel is large, such a technique may fail in regions where there is a low ratio of ground to non-ground points. We avoid these issues by computing the median value for a small region (typically 5m x 5m), and then flagging points in a larger region that are significantly higher than this median value. This modified median filter also has the advantage of being much more computationally efficient, and a similar technique may be employed directly on the point cloud, if desired.

For regions with very high building or tree density, an alternative technique must be employed. A minimum-filter (or morphological erosion-type) approach has been shown to be effective, whereby the lowest data value in each grid is retained to form an initial ground model. Data that is a certain threshold above this initial model is then flagged as being non-ground. This is conceptually similar to the approach presented in [19]. However, this approach has the disadvantage of erroneously categorizing highly sloped terrain as being non-ground.

If the median filter approach is used, once the points that are significantly higher than their surroundings have been flagged as being non-ground locations, a second filtering operation may be performed in order to remove additional objects from the terrain. These include large tree canopies, isolated small trees, vehicles, and other man-made objects, which would all be ignored by the initial processing. In order to effectively remove these points, the lidar range image is high-pass filtered, thereby highlighting regions with rapid changes in elevation. Bright pixels in this second filtered image (those with maximum rate of change) are also flagged as being non-ground. A simple sliding window filter based of the Laplacian may be used for this purpose. We have also found that first-derivative based approaches also work well in most applications for determining transition regions.

After the non-ground pixels have been identified, they are removed from the dataset, and an initial rasterized terrain model is obtained by interpolating across the removed points. If necessary, a smoothing filter may be applied to remove remaining high-frequency content in the resultant data, and the output is then faceted to produce the final DIRSIG DTM. Fig. 2 shows the result of this processing for an area containing buildings and trees on slightly undulating terrain.

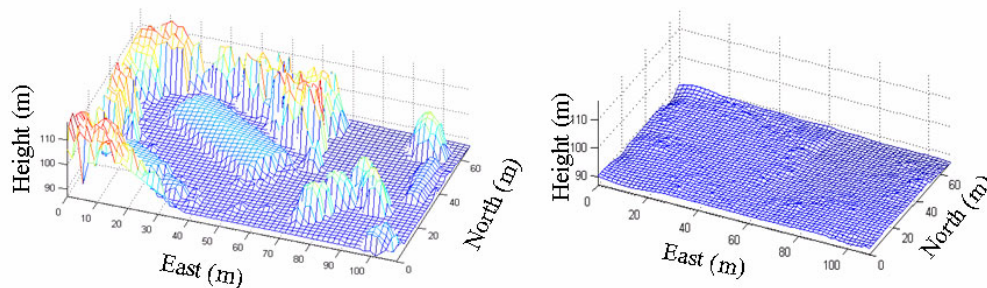


Fig. 2. Original range image (left) and DTM (right); Data down-sampled to 5m post ground spacing for display, points are relative to scene origin.

3.2 Initial Segmentation of Trees and Buildings

Once the DTM has been produced, it is a simple matter to isolate points that are a given threshold higher than the surrounding terrain. By subtracting the terrain from the original data (or a rasterized version of the original data) we obtain the normalized digital elevation model (NDEM). In many cases, we may then isolate the combined set of large man-made objects (collectively termed 'buildings' for the purpose of this paper) and trees by retaining only the data from regions greater than a given threshold above the NDEM. We have empirically found that a 2m threshold works well for most applications, but in certain cases thresholds as low as 0.5m may be preferable. Once we obtain this combined tree/building set, the next logical task is to identify to which class each of these points belongs. With finely registered multi-modal imagery, many features are available for addressing this problem. However, before the initial registration of the lidar and other data is performed, we must complete this segmentation using spatial features from the lidar point cloud alone.

To this end, we have implemented a spatial segmentation approach based on lidar range image entropy. In this technique, each region of the flagged non-ground point image is first identified through a connected-components analysis as described [24]. Regions that exhibit rapid height variations within a small window are then labeled as trees, while regions with lower height-valued entropy are classified as building structures. It should be noted that buildings that are heavily occluded by trees will not be identified as buildings at this step. However, these objects will usually be recovered during the fine building/tree segmentation step when other data sources are also used.

3.3 Coarse Image Registration

A critical requirement when working with multi-modal imagery is the proper registration of all data sources, as the relative spatial mapping of the multi-modal imagery must usually be achieved in order for the data to be combined effectively. Once we have extracted building and tree locations from the lidar data, we are able to perform an initial registration of all data sources.

The registration structure we have chosen to implement is composed of two stages. An initial coarse registration of the lidar data with optical imagery is completed to facilitate the segmentation and initial processing steps described in Fig. 1. Once this registration has been performed, initial geometric models may be extracted such as those describing the terrain, trees, and buildings. A more precise registration is then performed between the lidar data and the 2D frame-array imagery in order to permit refinement of individual object models. This refined registration will be discussed later in Section 3.5.

Traditional registration techniques rely on the ability to identify matching features (typically points or lines) in two or more images [25]. Once such features are identified, a regression is performed to determine an appropriate transformation that brings the feature set from the image to be warped in line with the corresponding features from the baseline image. However, such an approach is often difficult to implement in autonomous systems, due to the complexity of generating feature correspondences [25].

To this end, most autonomous registration techniques rely on a slightly different approach [25]. First, an initial parameter set describing a geometric transform is selected, and that transformation is applied to the image to be warped. This warped image is then compared to the baseline image through a registration quality metric. The transformation parameters are then updated, the new transform is applied to the (pre-warped) image, and the result is again compared to the baseline image. This process continues until a parameter set that leads to a global minimum metric value is obtained. In image processing applications where the images were taken using similar sensors, cross-correlation has proven to be an effective metric. However, difficulties arise when the images are obtained using sensors of differing modalities.

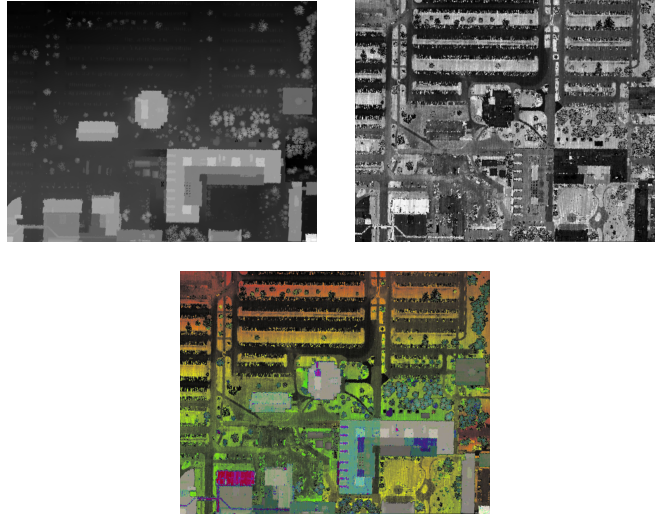


Fig. 3. Range image (top left), intensity image (top right), and CLRI (bottom).

For the coarse lidar to 2D frame-image registration, we construct a composite rasterized lidar image (CRLI) from the lidar height and intensity data. This image is created by using the intensity image for locations within 2m of the DTM, and a scaled range image elsewhere. Rasterized range, intensity and CLRI images from a portion of the RIT campus are shown in Fig. 3.

Once the CRLI is obtained, we may then register it to overhead imagery via a robust multi-modal registration technique. Unfortunately, since the CLRI and frame array images recorded differing phenomenologies, using cross-correlation as the quality metric often produces sub-par results. In these instances, a metric such as maximization of mutual information (MMI) [26] has been found to be preferable.

If we define the entropy, $H(X)$, of the random variable X with probability distribution $P(X)$ as

$$H(X) = -E_x[\log(P(X))] = -\sum_{x_i \in \Omega_x} \log(P(X = x_i))P(X = x_i), \quad (1)$$

Then the mutual information between pixels located at X and Y is specified by

$$\begin{aligned} I(A(X), B(Y)) &= H(B(Y)) - H(B(Y) | A(X)) \\ &= H(A(X)) + H(B(Y)) - H(A(X), (B(Y))) \end{aligned} \quad (2)$$

The mutual information may be viewed as a measure of similarity of the two pixel value distributions. As noted in [26], Y may be interpreted as X after a transformation by α , so that

$$I(A(X), B(Y)) = H(A(X)) + H(B(\alpha X)) - H(A(X), (B(\alpha X))) \quad (3)$$

Thus, the objective of the initial registration algorithm is to find the set of transformational parameters α that maximizes Equation (3).

We have found that this approach may be improved by first extracting edge and corner features in the imagery, then performing the MMI only on these featured pixels. Additionally, instead of using the actual pixel values in the algorithm, non-feature pixels are replaced with the value 0, edge pixels are assigned a value of 1, and corners are given a pixel value of 2. This modified MMI technique is termed Feature-Enhanced MMI (FE-MMI), and it has proven to be successful in registering lidar, visible and long-wave infrared frame array

imagery, linescanned imagery and maps with each other. A detailed description of the FE-MMI technique may be found in [27], and a typical registration result is shown below in Fig. 4.

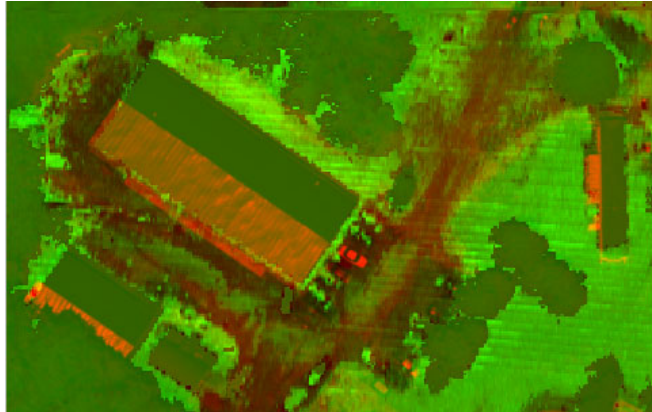


Fig. 4. Coarse registration of lidar and frame array imagery for a simple scene.

3.4 Refined Tree Segmentation and Reconstruction

Although an initial segmentation of the tree and building regions was possible using only the lidar data, in regions where buildings are heavily occluded by trees we may use co-registered spectral imagery to refine the segmentation. In general, use of spectral information is the most robust method that we have explored for separating building and tree regions.

The spectral reflectance of plants has a distinct signature, and therefore multi-spectral and hyperspectral classification techniques excel at differentiating vegetation from most man-made objects [8]. Through the visible light region, vegetation has a relatively low reflectance on the order of 5% up through approximately 650nm. However, there is an abrupt increase of reflectance in the near-infrared (NIR), where a peak of 50% is often recorded at about 730nm. This rapid variation in reflectance is commonly termed called the red edge, and this distinctive signature may be used to easily distinguish trees from most man-made objects. As such, through the use of simple spectral clustering techniques applied to the building/tree pixels, one may usually determine to which class each object belongs.

Once all of the lidar points depicting trees have been effectively identified, further processing of this data is required to determine the geometrical properties of each individual tree. Once these geometries have been defined, CAD objects representing these trees may be produced in one of two ways. The first is to use the extracted geometrical properties to define input parameters for software such as TreeProfessional [12], and then to create each tree separately from scratch using these parameters. The second approach is to use the calculated tree geometry to select the best-matching member from a pre-defined object library, and then scale this object within pre-determined bounds. Although the first approach has been explored to a limited extent, the primary method employed in the preparation of this paper was the latter.

Traditionally, tree parameters have been extracted from standard optical images. As an example, in [28], the authors describe an approach to streamline the construction of large forested scenes using high-resolution digital photographs. They assert that most coniferous tree crowns may be identified and characterized using radially-symmetric correlation methods. By first blurring the imagery then using circle functions at various scales, features with high radial symmetry may be uncovered. A similar approach is also used on processed images where tree regions are first extracted through an analysis of the NIR/red spectral ratio. Although the authors' primary intent was merely to recreate similar spatial aspects for the forested regions in the synthesized scene, in many cases the results were an accurate

representation of the real-world objects. However, in dense tree regions, individual tree locations and sizes were often in error, and it should be stated that this work also notes poorer results when attempting to isolate deciduous species. Additionally, such methods only provide information regarding tree location and canopy diameter, while tree height must be inferred.

A recent alternative approach has been to extract similar parameters from lidar imagery [29], [30]. Through methods outlined earlier, lidar permits an accurate localization of tree data without the need for spectral ratios, and with the fusion of multiple image modalities even better results may often be obtained. By blurring the lidar data from the tree regions in order to remove much of high frequency content, likely tree centers may be found by identifying local maxima in the resultant image. A radial analysis (using heights in place of grey-values) may then be applied to determine individual tree canopy sizes, although alternate approaches such as watershed analysis and other local minima extraction techniques have also been employed [31].

For the initial results presented here, we used passive imagery to define the tree boundaries while using lidar data to extract tree center locations and height. In future analyses, we will be using hyperspectral data to determine tree type (coniferous or deciduous), and will augment this with basic shape parameters if the lidar data has a sufficient sampling density. Fig. 5 presents an image depicting extracted tree centers for a region consisting of dense forest as well as a few isolate trees.

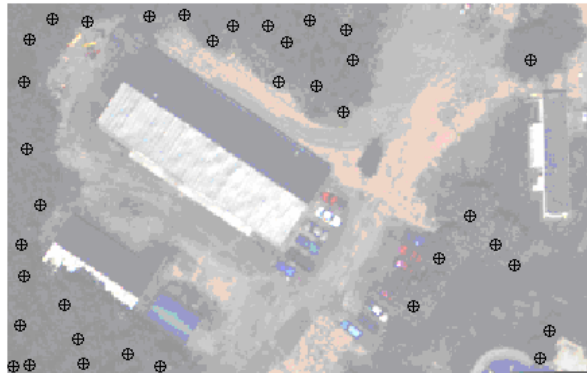


Fig. 5. Autonomously extracted tree centers.

3.5 Building Analysis and Modeling

Once the initial building regions have been identified, they are analyzed one at a time in order to produce geometric model of the building structure. In this research, the coarse building model estimates are initially derived using only lidar data. These coarse models are then used to register the lidar data with other image modalities which may then be used to refine the coarse model estimates.

Most of the recently proposed techniques for constructing building models directly from lidar point data may be categorized into one of two general classes. The first of these is termed model-based, and these techniques attempt to fit previously defined parametric models to the data by optimizing the model parameters. Detailed descriptions of this approach may be found in [32], [33], and [34]. Although this methodology often produces excellent results when the assumed model is correct, it tends to break down when applied to most complex building geometries [19].

The second class of techniques is data-based, which means they extract composite geometric features directly from the data and combine these features to complete the model.

Typically, these features are the dominant planes in the building's roof structure (see [19] and [35] for example), although recent work has also shown that other geometries may be used as well [36]. In general, these techniques are more robust when handling complex geometries, but the features of interest must be adequately sampled in order for them to be recovered.

In this research, we use a data-based approach for the initial model hypothesis, and attempt to fit parametric models to regions where the initial building model does not fit the original data well. Dominant roof planes are identified by segmenting the original points using local point properties as feature vectors, and adjacent planes are then analyzed to determine the inner roof-segment boundaries. Since vertical surfaces of the building structure are not directly represented in the data, a separate methodology must be used to determine the outer roof boundary. Once the initial building model is completed, the geometry is refined through both a comparison of the model with the original point data and the introduction of geometric constraints. To this end, we have implemented the following steps in our approach, which is more fully detailed in [37]:

1. Determine initial exterior boundary estimate from lidar data using alpha shapes [38] and line-fitting.
2. Segment the lidar range image such that each segment represents a planar face.
3. Determine internal boundaries through an intersection-of-planes approach
4. Determine vertices through intersections of inner and/or outer boundaries.
5. Refine the building model by introducing geometric constraints
6. Refine the lidar-derived model through a verification process using the original point data

Since many lidar datasets are obtained from near-nadir orientations, very few data points are available that lie on vertical surfaces. As such, it is often difficult to determine the planes corresponding to exterior walls from data points on these walls. This problem may be partially alleviated if we make the simplifying assumption that exterior walls are oriented directly under the outer boundary of a building object, a condition that is true in many building types. Therefore, in modeling the geometry of a given object on a given building layer, the first step is to determine the exterior roof boundary of that object.

Due to potential concavity in this boundary, simple shape descriptions such as the convex hull do not provide an adequate description of the outer roof shape. To this end, we have opted to use *alpha shapes* [38] for the determination of our exterior roof boundaries. Like the convex hull, alpha shapes are simply another approach to formally describe the 'shape' of a set of spatial point data. Unlike the convex hull, however, alpha shapes are not limited to convex geometries, and may even represent holes inside the geometry.

As described in [38], we may think of alpha shapes as a family of shapes for a given point dataset, where each shape is defined as the intersection of all closed discs with radius $1/\alpha$. In practice, α is set such that 2α is set to be 25% larger than the largest point-point spacing in the sampled lidar data. Fig. 6 depicts the convex hull and one of the alpha shapes for a given set of 2D data points.

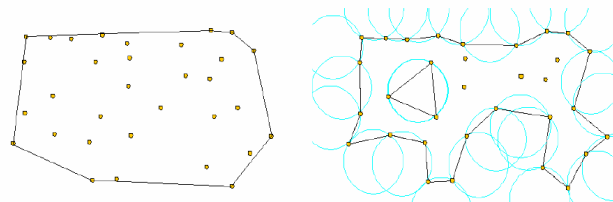


Fig. 6. 'Shape' of a collection of 2D points. Convex hull (left) and alpha shapes (right).

Once the outer boundary has been determined, inner plane edges and corresponding vertices still need to be defined. This is accomplished through a methodology similar to that presented in [35]. First, each data point is assigned a normal vector according to the plane best fitting the data in a one m^3 voxel centered on the point of interest. This plane is determined through a 3D Deming regression. In a manner similar to that presented in [19], the mean shift algorithm [39] is then used to segment the points into several groups, using the normal vectors and point locations as the defining features. Coplanar adjacent regions are then merged, and planes are fit to each data region. Planes from adjacent regions are then intersected with each other to determine candidate facet boundaries. Where the candidate boundaries match the actual data region boundaries well, the candidate boundary is used to define the inner roof edge. In places where the candidate boundary does not match the actual data region boundaries, a breakline is assumed, and a piecewise linear boundary is fit to the data region boundary. Vertical walls are projected down from both breaklines and outer edge surfaces until they intersect another roof plane or the previously extracted terrain model.

After the initial roof structure has been coarsely modeled, we often refine this model by introducing geometric constraints. Although we typically only require the outer roof edges to lie along lines that are oriented at increments of 45 degrees relative to each other, additional constraints are possible. These include ensuring certain edges lie at a constant height, or that specific inner edges meet exactly at outer boundary corner points.

A second refinement stage may also be performed, where the distance is calculated between each original data point and the resultant model. In regions where the errors are large, we search for additional (and often under-sampled) features such as window dormers that were not captured by the original segmentation. A reconstruction of these additional features is then attempted both through an intersection of planes approach and through parametric modeling of common roof objects. When this additional feature extraction step fails and errors between the model and the original data remain high in a localized region, we then create facets directly from the point data using a Delaunay triangulation.

Once the building model has been extracted and refined using the lidar data, enough information is available to refine the lidar/frame-array registration to the point where features in the optical imagery may be used to improve the building model geometry. This registration basically seeks to determine the transform that relates the 3-dimensional world coordinates of the building model to the 2-dimensional coordinates in the frame-array image. This is implicitly the camera projection matrix P described in [40], a transform matrix that is based on the interior and exterior orientation parameters of the frame-array imagers. If the internal parameters (such as focal length, sensor element geometries, principal point offsets and distortion coefficients) are known, we may derive the exterior orientation, and hence P , from a combination of 3 (or more) image to CAD model feature correspondences. Although a similar approach using standard photogrammetric equation has been presented in [19], we have opted to use the methods described in [40] in this research.

Once the refined registration has been completed, the building outer edges and internal breaklines are refined using the frame array-imagery. Such features have large errors when derived from lidar data alone, but these boundaries are usually directly visible in the optical imagery. Rather than using a Canny edge detector as presented in [19], we do a spatial segmentation of the RGB imagery using a proprietary gradient-based algorithm, and then fit lines to the transition regions. These lines are then back-projected onto the CAD model using the derived camera model.

Fig. 7 shows a detailed CAD model of both a simple building as well as a more complicated structure taken from a collect over a residential neighborhood. Both models were produced autonomously using the described techniques.

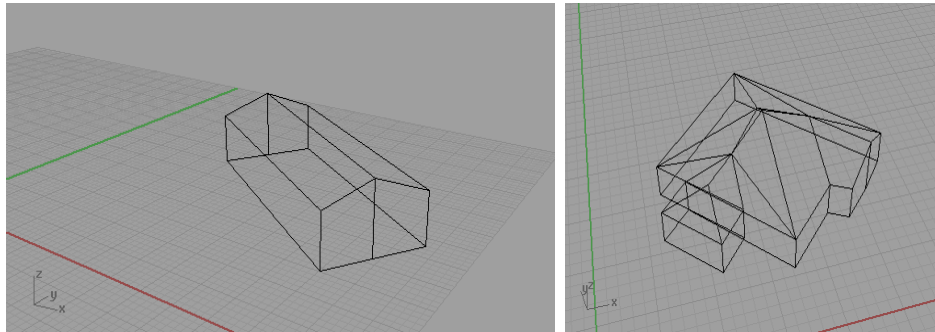


Fig. 7. Extracted model of simple scene building (left) and more complicated structure (right).

3.6 Spectra Retrieval and Spectral Assignment

DIRSIG assigns spectra to spatial regions in the scene using two complimentary methodologies. The first of these is on a single material per-facet basis, where material types and corresponding reflectance spectra are directly assigned to object facets. Although this technique has been used heavily in manually-defined scenes, the direct facet approach is not used in our current process.

The preferred method of assigning spectra is through the use of spectral material and texture maps. In order to specify spectra using this approach, a spectral classification routine is used to assign a material identification index to each location of a chosen facet. A texture image is then used to select an appropriate spectral curve from a library of spectra of that particular material type for each location on the facet. As such, two adjacent roof shingles may both be classified as "type 1 asphalt roof shingles", but their assigned spectra may be different due to differing gray levels in the texture image.

Typically, we generate material maps for the terrain via a supervised classification of a nearly orthonormal RGB or spectral image. An unprocessed image of the same type is often used for the terrain texture image. For building surfaces, oblique views of RGB images are reprojected such that the facet of interest matches the geometry of the CAD model, as is shown in Fig. 8. Currently, the features used to determine the appropriate projective transform are derived manually, but research to automate this process is ongoing.



Fig. 8. Oblique image of building, and texture map image for the facet representing the building side Image obtained via [42].

The reflectivity curves used in the spectral assignment are obtained in two ways. When quality hyperspectral data is not available, or the spatial resolution is such that the materials of

interest are significantly sub-pixel, it is usually preferable to use the available image data to select the best reflectance spectra from a previously collected library. However, a description of the technique used to determine the appropriate spectra is beyond the scope of this paper.

When high SNR hyperspectral imagery is available with sufficient resolution to have fully-resolved pixels on each material type, the sensor may be used to determine the spectral curve directly. However, it should be emphasized that the spectral radiance received by the sensor is not solely based on the reflectivity of the material being imaged. Rather the received radiance is dependent on the incident radiation which is then reflected and modified by the atmosphere, an effect that must be compensated for before reflectance properties may be isolated.

There are many methods of performing this atmospheric compensation in the literature. The Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) [41] algorithm uses a radiative transfer model to retrieve surface reflectance values from the directly measured radiances. The University of Colorado's Atmospheric Removal (ATREM) [42] uses a similar approach.

As noted in [8], if large (approximately 3 times the ground instantaneous field of view), near-Lambertian ground panels of known reflectance are available, one of the most attractive techniques for recovering reflectance curves from the sensed spectral radiance is through the Empirical Line Method (ELM). This is the method used for reflectance retrieval in this work.

Per [8], we assume a received radiance model of

$$L_{sensor} = \left(\frac{E'_s \cos \theta \tau_1}{\pi} + F \cdot L_d \right) \tau_2 r + L_u, \quad (4)$$

where E'_s is the direct exo-atmospheric solar irradiance, θ is the angle at which direct solar irradiance is incident upon the target, L_d is the downwelled radiance, F is a shape factor between 0 and 1 used to scale L_d , τ_2 is the target-sensor path transmission, r is the target reflectance and L_u is the upwelled radiance. The calibration may therefore be achieved by noting this model is linear with respect to reflectance,

$$L_{sensor} = m \cdot r + b, \quad (5)$$

and the slope (m) and intercept (b) may be determined through a regression with known reflectance values for each band. In most cases, b is assumed to be constant throughout the scene. However, when lidar data or other three-dimensional information is available, improved results may be obtained by using proper angular values at each point being considered. By modifying the method of [8], the slope effects may be compensated for if the ratio

$$l = \frac{L_d}{\left(E'_s \cos \theta \tau_1 \pi^{-1} + L_d \right)} \quad (6)$$

is known by using the form

$$L_{sensor} = \left((m - l \cdot m) \frac{\cos \theta_t}{\cos \theta_c} + \frac{F_t}{F_c} \cdot l \cdot m \right) r + b, \quad (7)$$

where the subscripts t and c represent the values at target of interest and at the calibration panels, respectively.

In practice, I may be obtained through a field measurement at the time of the collection. If such a measurement is not feasible, this ratio may be estimated either through in-scene techniques such as that presented in [43] or through atmospheric propagation models such as Moderate Resolution Atmospheric Transmission (MODTRAN) [44].

In cases where BRDF effects are non-negligible, the uncompensated ELM solution should be used to solve for a reflectance curve on each roof facet, and the solar angle, imaging angle and roof orientation for each facet should also be recorded. When subsequent synthetic imagery is then to be produced from the scene model, the reflectance curve for each facet should be chosen based on the best match between the synthetic imaging geometry and the original imaging geometry. Additional details regarding this technique may be found in [37].

4 PROCESS VERIFICATION: CAMP EASTMAN

In this section, we briefly demonstrate the feasibility of the proposed approach by applying these techniques to a region within Camp Eastman, a public park in Irondequoit, NY. The imagery used in this analysis came from the following sources:

- Lidar data was supplied by Leica Geosystems flying a commercial Optech sensor. The data contained approximately 6 points/m², roughly uniform in both the in- and cross-track dimensions. Multiple-return range and intensity data were provided.
- Spectral imagery came from the COMPact Airborne Spectral Sensor (COMPASS), a hyperspectral imager that captures incident radiation from 400nm to 2350nm on a single focal plane. This sensor was developed at the Army Night Vision and Electronic Sensors Directorate (NVESD). The average bandwidth of this sensor is approximately 8nm over the instrument's spectral range, and the spatial resolution was on the order of 1m. A hyperspectral materials library of ground truth was also created by measuring many materials in the scene with an Analytical Spectral Devices hand-held spectrometer.
- Near-nadir multi-spectral imagery was provided by RIT's Wildfire Airborne Sensor Program (WASP) sensor, a high-resolution (half meter) RGB frame-array sensor co-mounted with three lower resolution (3m) IR cameras operating in the short-wave, mid-wave, and long-wave regions.
- Oblique airborne imagery was taken from Microsoft's Maps-Live Web Application [45].

We chose to apply our technique to the portion of Camp Eastman depicted in Fig. 9. This area contains several buildings, a region with densely populated trees, numerous isolated trees, a road, grass and dirt. There is also a subtle change in ground elevation within the region, and a significant amount of truth data for this locale has been collected. As such, this has proven to be an excellent test scene for the proposed techniques.

We began the process by coarsely registering the lidar data to the WASP RGB frame-array imagery using the FE-MMI approach. This registration was successful to within approximately 2m (horizontal accuracy) throughout the image, and may be seen in Fig. 4. We next used a manual tie-point registration approach to rectify the hyperspectral imagery collected from COMPASS with the lidar data. In the future, this registration will be automated by first projecting the hyperspectral pixels to the lidar-derived height map, thereby producing a true orthoimage. However, the sensor position and orientation parameters were not available at the time of this work, so this projection operation was not possible.



Fig. 9. Area of interest for scene simulation (image taken by WASP RGB sensor).

Once the data registrations were complete, we processed the lidar data as specified in Section 3.1 to extract the terrain model. Both the modified median filter and high-pass filter techniques were used to initially flag points for removal, and a simple linear interpolation was used to reconstruct the full terrain image. However, due to the limited number of ground points in the forested region, this method was subsequently replaced with a variation of Vosselman's slope-based filter [19]. In order to speed the implementation, this filter was applied to the rasterized range image rather than to the irregularly-sampled original point cloud, and the result was shown earlier in Fig. 2. It should be noted that when labeling points as either ground or non-ground, the goal was not to minimize the total classification error. Since the labeling of object points as ground points (Type II error) is significantly more detrimental to the DTM generation process than mislabeling points actually on the terrain (Type I error), we typically chose our filter such that it drives the Type II error rate to below 1%. For the small scene at hand, with a filter slope of 20 degrees we were able to eliminate all Type II error points, while keeping the Type I error rate to less than 4%. However, this error rate is highly scene dependent, and significantly different results are seen when applying the technique to other locales. We also compared the semi-autonomously produced DTM to a USGS small-scale terrain model [11] for a slightly larger, hillier region in the same general area. This analysis found the average discrepancy between the models to be 2cm with a 0.6m variance.

We recreated the buildings in the scene using the intersection of planes approach to determine initial model vertices. Although this building has a very simple structure, it provided an adequate proof of concept for the end-to-end scene creation process. For the main shed roof structure, all vertices were accurate to within 0.5m, and the roof angles were correct to within one-half of a degree. Significantly more complex building geometries have also been successfully recreated using the approach described here (see [37], for example), but quality hyperspectral imagery was not available for the other locations in which lidar data were collected.

Trees were identified and reconstructed as described above with a slight change. Instead of using multiple tree models that were selected based on height and width estimates. This model was then scaled in both height and width to fit the actual tree parameter estimates. All tree regions were correctly classified using the described approach. However, within the tree regions, only 36 tree center locations were identified while a ground survey of the area indicated that the actual tree count in this area to be significantly higher (~47). This large under-estimation of the tree-count is due to the extremely high tree density in this region; in many cases trees were as close as 2m together. Manual identification of tree centers from

aerial imagery of the region provided an identification rate similar to that of the semi-automated process, as an average of 30 trees were identified using the available data.

Spectral assignment for features on the terrain was accomplished through the use of a material map overlaid on the terrain model. This map was produced by doing a supervised minimum-distance classification to the WASP RGB imagery, then performing spatial filtering to produce a more uniform result. Material maps for the vertical building surfaces were created by performing a perspective projection on each building face extracted from an oblique airborne image, as shown in Fig. 8. This vertical map surface was then segmented and materials were manually assigned to each class. Reflectance spectra for these materials were obtained from applying the FLAASH algorithm [41] to COMPASS imagery. Fig. 10. presents a single reflectance spectrum of the green roof material as determined by FLAASH and compares it to a lab-measured spectrum of a similar (but slightly darker) material. Note that despite the relative brightness difference, the two curves share a fairly similar spectral signature.

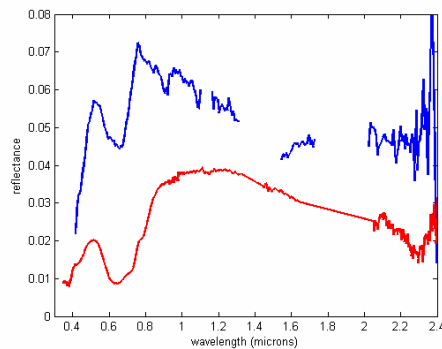


Fig. 10. FLAASH-derived reflectance spectrum for green roof material in the scene (blue) and lab-measured spectrum for a similar sample (red).

A complete DIRSIG scene using the extracted features, maps and objects was then created. A simulated aerial view of this scene is given in Fig. 11. It should be noted when viewing this scene that trees and buildings have been re-created in true 3D, but smaller objects such as vehicles have not. The reason that some of these smaller objects seem to appear in the DIRSIG image is that they were present in the texture image used for spectral assignment, and the spectral variations give the appearance of additional objects.



Fig. 11. Simulated color image using the semi-autonomously generated DIRSIG scene.

5 SUMMARY

This work presented an approach to reducing man-in-the-loop requirements for several aspects of synthetic hyperspectral scene construction. Through a fusion of 3D lidar data and passive imagery, we were able to partially automate several of the required tasks in the DIRSIG scene generation process. These included extraction of a bare-earth digital terrain model, identification of buildings and trees, object reconstruction, the generation of material and texture maps, and the generation of reflectance curves. Through the proper application of these techniques, we aim to enable the creation of synthetic scenes where truth data is not available, as well as to significantly reduce the time required by current scene-building methods.

After describing the proposed process for scene construction, we demonstrated the feasibility of the techniques by applying them to a portion of Camp Eastman. The DTM extraction, identification of buildings and trees, and low-level geometric reconstruction of buildings were shown to be successful.

Future research will focus on improved methods of mapping material and texture information onto occluded surfaces, as well as additional ways to integrate traditional photogrammetric techniques into the geometrical object reconstruction.

Acknowledgments

This work has been supported in part by the U.S. Government under University Research Initiative HM1582-05-1-2005. The authors also wish thank the Leica Corporation for providing lidar data used in this work, Merrick for assistance in obtaining their MARS lidar processing/viewing software and the many members of the Digital Imaging and Remote Sensing group at RIT for their continued assistance with this project.

Disclaimer

The views expressed in this article are those of the authors and do not reflect the official policy or position of the U.S. Air Force, Department of Defense, or U.S. Government.

References

- [1] W. Förstner, "3D-city models: automatic and semiautomatic acquisition methods," *Photogrammetric Week '99*, 291-303 (1999).
- [2] D. Hearn and M. Baker, *Computer Graphics with OpenGLz: 3rd Edition*, Prentice Hall, Saddle River, NJ (2003).
- [3] T. Mason, "Rapid mapping of the 3D urban environment," *Proc. 20th ISPRS Congress* **35-3B** (2004).
- [4] C. Brenner, "Interactive modeling tools for 3D building reconstruction," *Photogrammetric Week '99*, 23-34 (1999).
- [5] C. Brenner, "Building reconstruction from images and laser scanning," *Int. J. Appl. Earth Observation Geoinform.* **6**(3-4), 187-198 (2005) [doi: 10.1016/j.jag.2004.10.006].
- [6] J. R. Schott, S. D. Brown, R. V. Raqueno, H. N. Gross, and G. Robinson, "An Advanced synthetic image generation model and its application to multi/hyperspectral algorithm development," *Can. J. Rem. Sens.* **25**(2), 99-111 (1999).
- [7] T. Schenk and B. Csatho, "Fusion of LIDAR and aerial imagery for a more complete surface description," *Int. Archives Photogram. Rem. Sens.* **34**(3A), 310-317 (2002).
- [8] J. Schott, *Remote Sensing: The Image Chain Approach: 2nd Edition*, Oxford University Press, New York (2007).

- [9] R. Burton, "Elastic LADAR modeling for synthetic imaging applications," Ph.D. Thesis, Rochester Institute of Technology, NY(2002).
- [10] D. Blevins, "Modeling scattering and absorption for a differential absorption LIDAR system," Ph.D. thesis, Rochester Institute of Technology, NY (2005).
- [11] United States Geological Survey, "The National Map Seamless Server – online resource," <http://seamless.usgs.gov> (2008).
- [12] Onyx Computing, "Tree Professional software webpage," <http://www.onyxtree.com> (2008).
- [13] McNeel North America, "Rhinceros: NURBS Modeling for Windows webpage," <http://www.rhino3D.com> (2008).
- [14] E. J. Ientilucci, K. Ewald, J. Marcin, and A. Spivey, "Guide to building large scale DIRSIG scenes," *Internal publication: Digital Imaging and Remote Sensing Laboratory*, Rochester Institute of Technology, NY (2003).
- [15] S. D. Brown, "DIRSIG Homepage," <http://dirsig.cis.rit.edu/> (2008).
- [16] J. Lindenberger, "Laser-Profilmessungen zur topographischen Gelandaufnahme," Ph.D. thesis, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften (1993).
- [17] H. Masaharu and K. Ohtsubo, "A filtering method of airborne laser scanner data for complex terrain," *Int. Archives Photogram. Rem. Sens.* **34**(3B), 165-169 (2002).
- [18] K. Zhang, S. Chen, D. Whitman, M. Shyu, J. Yan, and C. Zhang, "A Progressive Morphological Filter for Removing Nonground Measurements from Airborne LIDAR Data," *IEEE Trans. Geosci. Rem. Sens.* **41**(4), 872-882 (2003) [doi: 10.1109/TGRS.2003.810682].
- [19] R. Ma, "Building model reconstruction from Lidar data and aerial photographs," PhD thesis, Ohio State University (2004).
- [20] G. Vosselman, "Slope-based filtering of laser altimetry data," *Int. Archives Photogram. Rem. Sens.* **33**(B3), 935-942 (2000).
- [21] G. Vosselman and H. Maas, "Adjustment and filtering of raw laser altimetry data," *Proc. OEEPE Work. Airborne Laser Scanning Interferometric SAR DEMs*, Paper 5, Royal Inst. Tech., Stockholm(2001).
- [22] G. Sithole, "Filtering of laser altimetry data using a slope adaptive filter," *Int. Archives Photogram. Rem. Sens.* **34**(3/W4), 203-210 (2001).
- [23] M. Elmqvist, "Ground surface estimation from airborne laser scanner data using active shape models," *Int. Archives Photogram. Rem. Sens.* **34**(3A), 114-118 (2002).
- [24] R. C. Gonzalez and R. E. Woods, *Digital Image Processing 2nd ed.* Prentice-Hall, Saddle River, NJ(2002).
- [25] L. Brown, "A survey of image registration techniques," *ACM Computer Survey*, **24**, 325–376 (1992) [doi: 10.1145/146370.146374].
- [26] X. Fan, H. Rhody, and E. Saber, "A spatial-feature enhanced MMI algorithm for multi-modal airborne image registration," submitted to *IEEE Trans. Geosci. Rem. Sens.* (2008).
- [27] X. Fan, H. Rhody, and E. Saber, "Automatic registration of multi-sensor airborne imagery," *Appl. Imagery Pattern Recognition '05*, 81-86 (2005) [doi: 10.1109/AIPR.2005.21].
- [28] R. Gray, S. Brown, and J. Schott, "Scene construction methodologies and techniques for simulating forest areas," *11th Annual Ground Targets Modeling Validation Conf.* (2000).
- [29] J. Hyypä and M. Inkinen, "Detecting and estimating attributes for single trees using laser scanner," *Photogram. J. Finland* **16**, 27-42 (1999).
- [30] A. Persson and U. Soderman, "Detecting and measuring individual trees using an airborne laser scanner," *Photogramm. Eng. Rem. Sens.* **68**(9), 925-932(2002).

- [31] F. Morsdorf, E. Meier, B. Allgower, and D. Nuesch, "Clustering in airborne laser scanning raw data for segmentation of single trees," *Proc. ISPRS Working group III/3 Worksh. 3-D Reconstruction from Airborne Laserscanner InSAR Data* (2003).
- [32] U. Weidner and W. Förstner, "Towards automatic building extraction from high-resolution digital elevation models," *ISPRS J. Photogramm. Rem. Sens.* (1995).
- [33] H. G. Maas and G. Vosselman, "Two algorithms for extracting building models from raw laser altimetry data," *ISPRS J. Photogramm. Rem. Sens.* **54**(2), 153-163 (1999) [doi:10.1016/S0924-2716(99)00004-0].
- [34] N. Haala, C. Brenner, and K. H. Anders, "3D urban GIS from laser altimeter and 2D map data," *Int. Archives Photogram. Rem. Sens.* **32**, 339-346 (1998).
- [35] F. Rottensteiner and C. Briese, "Automatic generation of building models from Lidar data and the integration of aerial images," *Int. Archives Photogram. Rem. Sens.* **34**, (2003).
- [36] P. Gurram and S. Lach, "3D scene reconstruction through a fusion of passive video and lidar imagery," *Applied Imagery Pattern Recognition*, Washington D.C. (2007).
- [37] S. Lach, "Semi-automated DIRSIG scene modeling from 3D lidar and passive imaging sources," Ph.D. thesis, Rochester Institute of Technology, (2008).
- [38] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, "On the shapes of a set of points in the plane," *IEEE Trans. Inform. Theor.* **29**(4) (1983) [doi: 10.1109/TIT.1983.1056714].
- [39] D. Comaniciu, "Nonparametric Robust Methods for Computer Vision," PhD thesis, Rutgers University (2000).
- [40] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press (2004).
- [41] M. Adler-Golden, Bernstein, Levine, Berk, Richtsmeier, Acharya, Anderson, Felde, Gardner, Hike, Jeong, Pukall, Mello, Ratkowski, and Burke, "Atmospheric correction for shortwave spectral imagery based on MODTRAN4," *Proc. SPIE* **3753**, 61-69 (1999) [doi:10.1117/12.366315].
- [42] Center for the Study of Earth from Space (CSES), *Atmosphere Removal Program (ATREM), Version 3.1, Users Guide*. University of Colorado, Boulder (1999).
- [43] K. R. Piech, and J. E. Walker, "Interpretation of Soils," *Photogramm. Eng. Rem. Sens.* **40**, 87-94 (1974).
- [44] A. Berk, L. S. Bernstein, G. P. Anderson, P. K. Acharya, D. C. Robertson, J. H. Chetwynd, and S. M. Adler-Golden, S.M., "MODTRAN cloud and multiple scattering upgrades with application to AVIRIS," *Rem. Sens. Environ.* **65**, 367-375 (1998) [doi: 10.1016/S0034-4257(98)00045-5].
- [45] Microsoft: Maps-Live Web-based application, <http://maps.live.com/> (2008).

Author Biographies

Stephen R. Lach received the B.S. and M.S. degrees in electrical engineering from Villanova University in 1996 and 1998 and the Ph.D. degree in imaging science from the Rochester Institute of Technology in 2008. His research interests include adaptive array antenna design, signal processing for spread spectrum systems, and three-dimensional scene analysis and reconstruction from fused imagery.

John P. Kerekes received the B.S., M.S., and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1983, 1986 and 1989. From 1983 to 1984, he was a Member of the Technical Staff with the Space and Communications Group, Hughes Aircraft

Co., El Segundo, CA, where he performed circuit design for communications satellites. From 1986 to 1989, he was a Graduate Research Assistant, working with both the School of Electrical Engineering and the Laboratory for Applications of Remote Sensing at Purdue University. From 1989 to 2004, he was a Technical Staff Member with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington. In 2004, he became an Associate Professor in the Center for Imaging Science, Rochester Institute of Technology, Rochester, NY. His research interests include the modeling and analysis of remote sensing system performance in pattern recognition and geophysical parameter retrieval applications. Dr. Kerekes is a Senior Member of the IEEE and a member of Tau Beta Phi, Eta Kappa Nu, the American Geophysical Society, the American Meteorological Society, the American Society for Photogrammetry and Remote Sensing and SPIE.

Fan Xiaofeng received his B.S. and M.S. degrees in Electrical Engineering from the University of Science and Technology of China in 1997 and 2001. He has been pursuing his Ph.D. in imaging science at the Rochester Institute of Technology in Rochester, NY since 2003. His primary research interests include autonomous and semi-autonomous registration of multi-modal image types and feature detection in aerial imagery.