

Multisource Data Processing for Semi-Automated Radiometrically-Correct Scene Simulation

Stephen R. Lach

United States Air Force/Rochester Inst. of Technology
Chester F. Carlson Center for Imaging Science
Rochester, New York 14623 USA
sfl1194@cis.rit.edu

John P. Kerekes

Rochester Institute of Technology
Chester F. Carlson Center for Imaging Science
Rochester, New York 14623 USA
kerekes@cis.rit.edu

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 initiated a research effort that aims to reduce the man-in-the-loop requirements for several aspects of synthetic hyperspectral scene construction. Through a fusion of 3D lidar data with passive imagery, we are working to semi-automate several of the required tasks in the creation of high-resolution urban DIRSIG scenes. This paper reports on the progress made thus far in achieving these objectives.

I. INTRODUCTION

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. Similarly, infrared data provide information related to surface temperatures, while reflective hyperspectral imagery can characterize the surface material type and condition. With data from all these multiple sources, a comprehensive characterization of the scene becomes possible.

The Digital Imaging and Remote Sensing Image Generation (DIRSIG) tool is a first-principles based model that produces synthetic spectral imagery of pre-defined scenes. It 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 broad band 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 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, takes 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.

In recent years, lidar has proven to be an invaluable tool in extracting geometric properties of a scene. In creating a digital model of the terrain, the three-dimensional first-return point cloud is filtered in order to separate the ground points from the non-ground points. Additionally, once building regions have been defined, an estimate of the building outline and an accurate description of the planar facets may often be reconstructed directly from this data. Processing multiple returns of the lidar signal yields even greater insight into the scene content, as this provides one means of distinguishing trees from buildings. However, the scene geometry is not completely specified by the lidar information alone. High resolution passive imagery complements lidar data well, as edges and other two dimensional spatial details often are represented more accurately in image form than they are in point cloud data sets. Through the use of co-registered oblique images, structure along the sides of building may often be more accurately reproduced as well.

Multi-spectral and hyperspectral classification techniques excel at differentiating vegetation from most man-made objects and therefore also provide a complimentary data set for scene reconstruction. Through the use of simple spectral clustering techniques applied to the building/tree pixels, one may determine to which class each object belongs. When fused with spatial-based discrimination approaches, this clustering becomes even more accurate. Hyperspectral imagery may also be processed to define each surface's spectral signatures and thermal characteristics. This may be done either by using the sensor as a field spectrometer, or by letting the received signal drive selection of a spectrum from a previously-collected database.

Work to date has focused on the spatial registration of multiple high-resolution data sources, terrain extraction, object

2007 Urban Remote Sensing Joint Event

identification and reconstruction from dense (5 points/m²) lidar data, 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 proposed multi-source processing approach, along with some experimental results.

II. 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 [1]. 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 [2], and it was recently expanded by Blevins [3] 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 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.

Two representative synthetic images produced by the DIRSIG model are included in Fig. 1. The first of these depicts a near-infrared image of the Microscene1 test range on the RIT campus, and the second is a topographic lidar product of the same area (with the vehicle removed and from a slightly different perspective).

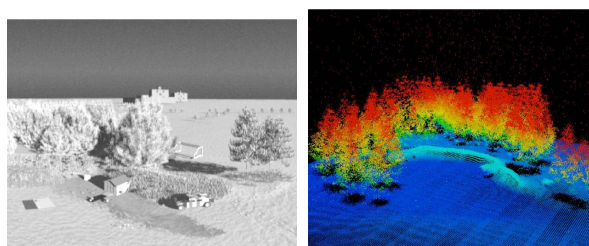


Figure 1. A near-infrared simulation of MicroScene1 produced by DIRSIG (left) and simulated topographic lidar image (right)

III. PROPOSED PROCESS FOR SCENE CONSTRUCTION

The current 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 lower-precision USGS products and are re-sampled to the TIN format required by DIRSIG. Object geometries are produced using CAD-based tools; trees are typically generated in Onyx Tree Professional [4], while man-made structures are defined using more standard tools such as AutoCAD or Rhinoceros [5]. Spectral assignment is done manually on a per-facet basis using the Bulldozer in-house software utility, or via detailed material maps. Spectral texture may be added through additional mapping images. The detailed current process is further described in [6], and it relies on expertise with many different software tools. This process typically requires several months to complete.

Although the current 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 is being implemented where several of the scene design tasks have been replaced by semi-automated methods. The basic concepts for automating much of the DIRSIG scene generation process are depicted below in Fig. 2, and will be explained in greater detail throughout the remainder of this section.

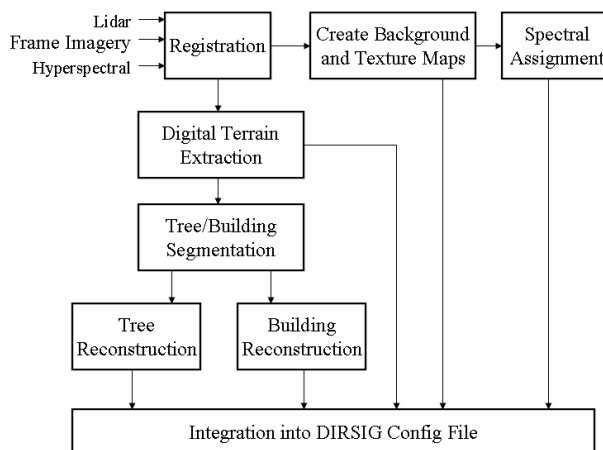


Figure 2. Proposed process for scene construction

2007 Urban Remote Sensing Joint Event

A. Image Registration

A critical step 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 be achieved before data may be combined effectively. Images are usually preprocessed to extract desired features (such as notable points, lines, edges, or surface patches), which not only serves to identify potential matching features, but also significantly reduces the size of the space used during the registration. Once the appropriate features have been identified, correspondences must be established so that a feature depicted in one image may be matched to its equivalent representation in the other images. However, difficulties arise when the images are obtained using sensors of differing modalities and record content using differing spatial structures. In order to deal with this, we present our approach for registering 3D lidar data with single 2D frame-array images and linescanned hyperspectral imagery.

1) Registering 3D Lidar Data to 2D Frame-Array Imagery

If the 2D frame-array data may be processed to yield 3D features, a registration with a lidar point cloud may be carried out in 3D object space. In [7], the authors present two methods for implementing this approach. However, when only a single frame image is to be registered to 3D data, an alternative approach must be employed. A 3D to 3D registration may only be used in these cases when special single image photogrammetric techniques are feasible. The reader is referred to [8] and [9] for additional information on such approaches, which rely on known parallel and perpendicular features or shadow content to be effective.

However, in many cases where individual 2D images must be combined with 3D datasets, it is simpler to carry out the registration in 2D image space. In these instances, the geometrical representation of a feature extracted from the lidar data is linked to a corresponding feature in 2D image space.

In order to perform this registration, the lidar data is first processed to extract a CAD representation of objects. The frame image and lidar-derived CAD models are then converted to a common 2D baseline reference frame. If the 3D point data is reasonably accurate and the approximate exterior orientations of the frame-array cameras are known, the 3D CAD representation in object space may be transformed to a comparable 2D image using an appropriate projective transformation. Both datasets then appear as similar perspective views of the object-space scene. Homologous points in each of these representations are then selected, and the 2D perspective image of the lidar data is re-mapped to the frame image via a second transformation.

Once the frame-image matches the transformed lidar model to satisfaction, relevant features from the CAD model are selected. These features are then identified in a processed frame-image, in which edges have been extracted. Although many techniques are available for performing this edge extraction, we have achieved good results using a Canny edge detector applied to both the saturation and value channels of color imagery. At this point, relevant features in 2D image

space now have an unambiguous specification both in space and in 3D object space.

This process is illustrated for the Carlson Building on RIT's campus in Fig. 3. A commercially obtained oblique image is given in Fig. 3a, and the CAD model derived from the lidar data is given in Fig. 3b. As an example, note that the left edge of the roof has been selected. This edge is then highlighted in a perspective view of the original lidar data in Fig. 3c, and the corresponding frame-array feature (as determined by edge detection) is identified in Fig. 3d.

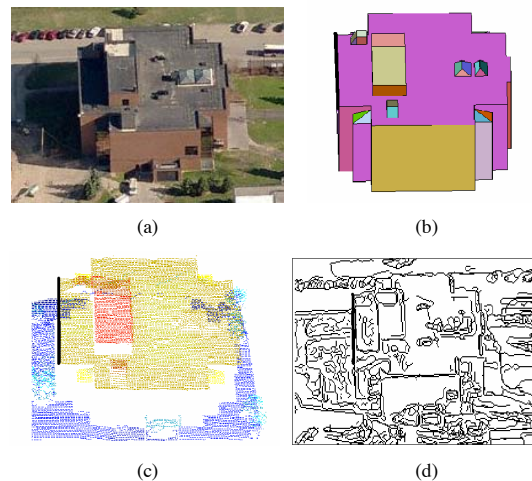


Figure 3. (a) Frame-array image, (b) extracted edges, (c) 2D projection of lidar data, and (d) projected CAD model

2) Registrations Involving Spectral Linescanned Data

While most airborne single-band and color imagery is collected using frame-array architectures, multi-spectral, hyperspectral and lidar datasets are often collected with linescanning sensors. The non-instantaneous image collection inherent in this design introduces additional difficulties when registering data collected by these imagers. To circumvent these difficulties, when matching features to linescanned data, an orthographic (geo-referenced) projection space is preferred to the baseline frame array image space.

The registration of frame array and linescanner imagery has been well documented in the literature. Although non-parametric (image warping) techniques have been proposed (see [10] and [11] for example), it is generally agreed upon that they produce less than satisfying results in many instances [12]. Furthermore, in the cases where nonparametric techniques are employed, a large number of ground control points are typically required, and the image-to-image mapping usually requires manual intervention [12].

As such, we have opted to use parametric techniques to bring our linescanned data into a common frame with other image sources. This approach models the 3D geometry between the aircraft, the scanning instrument, and the terrain, and effectively performs a ray-tracing solution that maps image intensities to points on the terrain model. Next, a re-gridding is

2007 Urban Remote Sensing Joint Event

performed to convert these projected intensities into a rasterized digital ortho-image. Finally, a polynomial transform using control points is used to fine-tune the result when such control points are available.

Once all of the images (frame-array, linescanned multi- and hyperspectral and lidar) have been remapped as 2D orthographic perspective projections, a final least squares polynomial mapping using homologous features between all image sources is performed to remove most of the remaining geometrical discrepancies. Although this final transformation may be obtained and applied locally to sub-regions of the imagery, it has been found that when the initial ortho-rectification has been completed accurately, the final polynomial mapping may be applied globally with no noticeable degradation.

In automated systems, a final step is to evaluate the quality of the registration given the geometric model, and then use this evaluation to fine-tune the model parameters. Such an evaluation requires use of a metric function, and in standard image processing applications, maximization of the cross-correlation of the registered images is often used. However, when the images are taken with sensors recording differing phenomenologies, cross-correlation often produces sub-par results, as corresponding features may have differing radiometric properties across the bandwidths of the sensors being used. In these instances, metrics such as maximization of mutual information (MMI) or phase correlation are often preferable. Additional information regarding the use of this metric may be found in [13].

B. Extraction of the Digital Terrain Model

In creating the digital terrain model (DTM), the irregularly-sampled lidar point cloud may be processed directly, or the data may first be interpolated onto a regular grid to form a range image. In either case, the fundamental concept is to separate the ground points from the non-ground points, then to remove the non-ground points (and potentially interpolate across them). In [14], the author describes several techniques that may be applied directly to the 3D point data, but we found that satisfactory results could be obtained easily by applying simple image processing techniques 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



Figure 4. Illustration of the Modified Median Filter at a single location; median height value is calculated in smaller window, points 3m above median in larger window are flagged as non-ground points

employed directly on the point cloud, if desired. Fig. 4 illustrates this concept for a single location of the filter's sliding windows.

For regions with very high building or tree density, an alternative technique must be employed. An erosion-type approach has been shown to be effective, whereby the lowest data value in each voxel 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 very similar to the approach presented in [15].

Once the points that are significantly higher than their surroundings have been flagged as being non-ground locations, a second filtering operation must 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 on the Laplacian of the form

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \implies \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

may be used for this purpose. We have also noted 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 data set, and an initial terrain model is obtained by interpolating across the removed points. If desired, 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 DTM.

2007 Urban Remote Sensing Joint Event

C. Segmenting Buildings from Trees

Through the methods proposed in the previous section, we extracted a set of points that were significantly higher than their neighbors (the initial set of "non-ground" points). In many cases, it may be assumed that these points represent either large man-made objects (collectively termed 'buildings' for the purpose of this paper) or trees, and 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.

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. Through the visible light region, vegetation has a relatively low reflectance on the order of 5% up through approximately 680nm. However, there is an abrupt increase of reflectance in the near-infrared, 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.

In general, use of spectral information is the most robust method that we have explored for separating building and tree regions, although in many cases other methods must be employed as well. In the event that spectral information is not available or the spatial resolution of the spectral imagery is not satisfactory to perform a vegetation/building classification completely, spatial-based approaches may also be used.

To this end, we have implemented a spatial segmentation approach as well. This method is a lidar range-image based binary morphological technique in which each region of the flagged non-ground point image is first identified through a connected-components analysis as described [16]. The image is then opened by a pre-determined structuring element. By sizing this kernel to preserve just a few pixels from the smallest building in the scene, tree regions are effectively removed. This holds true even for large areas containing many trees, as several ground pixels are present in almost all tree canopy regions. Additional ground points may be ensured in these regions by including secondary return data, but this is not required in most cases. Objects in the connected-components image that contained pixels in common with the opened image are then specified as being buildings, and the remaining objects are classified as trees.

D. Geometrical Reconstruction of Trees

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 [4], 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 [17], 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. 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 [18].

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.

E. Geometrical Reconstruction of Buildings

This section focuses on the proposed techniques to derive CAD building models based on the 3D lidar point data. As in many building reconstruction approaches, we assume that buildings are strictly composed of planar faces, and that their dominant internal and external edges are linear. Although in general building edges may be aligned in arbitrary directions, most structures have dominant orientations of their vertical surfaces. As such, we identify the orientation of dominant boundaries, and force them to lie in 45 degree increments relative to each other. Additionally, it is assumed that the lidar data was collected from a predominantly downward-looking sensor, and that few point samples are available from any

2007 Urban Remote Sensing Joint Event

building component other than roof structures. As such, we use the roof boundary to describe the location and orientation of the exterior vertical walls, and no effort is made to identify wall structure from actual wall data points.

In general, buildings may be composed of multiple roof levels, and each roof level may be composed of groupings of smaller objects (each potentially composed of multiple planes). In order to deal with this assumed building model, we have developed a data-based approach using an intersection-of-planes methodology to find vertex points. In this approach, we use the following steps:

1. Generate a histogram of the data point heights and then segment these heights so that each roof layer may be processed individually.
2. Within a given height layer, points are then clustered into groups that define similar objects.
3. Each of these objects is then processed individually in order to determine the object's outer boundary
4. Internal vertices are found through an intersection of planes approach

Once all vertices are extracted, the CAD model is assumed to be complete. It should also be noted that throughout the process, the lidar point values are used directly without resampling to a rasterized grid.

The first processing task (after building points have been isolated from the surrounding vegetation) is to decompose the points into groups based on height. To this end, a histogram of the elevation data is created, and the resultant function is grouped into regions of similar height. In order to illustrate this process, actual lidar data collected over Building 7 at RIT will be used. A 3D scatter plot of this data is given in Fig. 5.

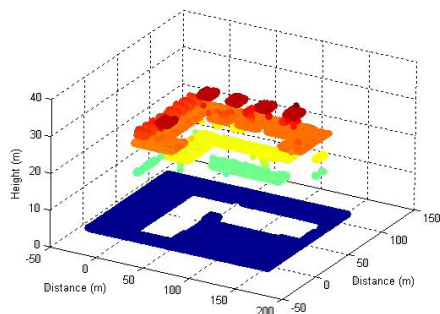


Figure 5. 3D scatter plot of the building data

In order to segment this height histogram function, many techniques are available. We have opted to use the comparatively simple method discussed in [19]. This approach first smoothes the histogram with a gaussian filter then takes the gradient of the result. The zero-crossings of this gradient represent local maxima in the histogram, and by comparing these to adjacent minima, threshold regions may be defined to complete the segmentation. The output of this processing is shown in Fig. 6.

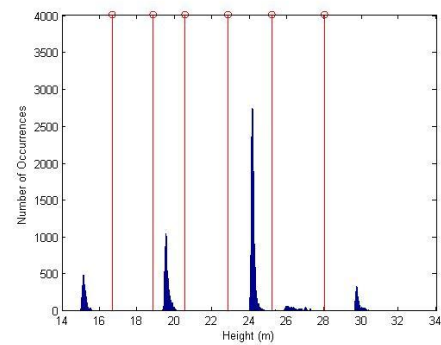


Figure 6. Height histogram segmentation

Once the data have been segmented based on height values, each height group must be segmented into groups of points corresponding to building features (sub-objects) at that particular level. This is done via a segmentation routine using the data (x,y,z) coordinates as feature vectors. Unfortunately, many common clustering algorithms such as k -means and ISODATA fail when applied to data of this modality, as the low-dimensionality of the feature space does not permit the algorithms to adapt to poor initial conditions.

As such, an alternative segmentation routine is required, and we have selected the mean-shift technique [20]. The mean-shift is a non-parametric mode-seeking algorithm that uses a kernel for density gradient estimation. Comanicu discusses applying the mean-shift clustering to joint spatial-range data, but we have found that this technique also works well with spatial-only data in many cases. Since a detailed description of how the mean-shift approach may be used for spatial segmentation is beyond the scope of this paper, the reader is directed to [21] for information regarding our implementation.

Once the individual objects on a given layer have been identified, each object is then reconstructed individually using an intersection of planes approach. However, since many lidar datasets were 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. However, in many cases it may be assumed that exterior walls are oriented directly under the outer boundary of a building object. Therefore, in modeling the geometry of a given object on a given building layer, the first step is to determine the exterior boundary of that object.

Several published approaches use the exterior triangles from a Delaunay triangulation to determine this outer boundary (see [22], for example). However, the use of this methodology assumes that the object perimeter is convex, and the technique fails for more complex geometries. In Fig. 7, we are able to see that the third layer of Building 7 is approximately 'L-shaped,' and the Delaunay-based convex hull fails to highlight the interior angle. An alternative approach is to use a polygon that minimizes the enclosed area. In many cases this provides a better solution than the triangulation approaches, but as can be seen in Fig. 7 (right), it too may provide substandard results.

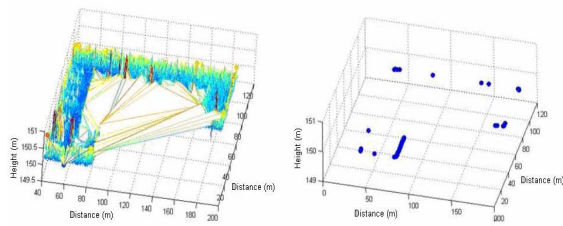


Figure 7. Determination of a group's exterior boundary. Convex hull points obtained from Delaunay triangulation (left) and solution points for polygon of least area approach (right)

To this end, we have opted to use *alpha shapes* for the determination of our exterior object 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. Additionally, for a given point set, the resultant alpha shape is a function of a pre-determined structuring element. Therefore, the final 'shape' is actually an entire family of shapes.

As described in [23], we may think of alpha shapes in the following manner. Assume we have a volume of chocolate chip ice cream, where the chips represent a set of points. If we then used a spherical scoop to scoop away all the ice cream possible without bumping into the chips, and are somehow even able to scoop out holes in the same manner on the inside of the volume, we would eventually end up with a shape bounded by arcs and points. If we then replace these round facets with triangles by using line segments connecting the points, the result would be a representation of the alpha shape of the points. In this example, the parameter alpha is related to the radius of the scoop that was used.

As can be seen from the above description, a very small value of alpha will permit us to scoop away all of the ice cream between the chips, just leaving us with the point locations in the final shape. Similarly, a large scoop will be unable to scoop away concave regions between points, and the resultant shape will be identical to the convex hull of the points. For intermediate alpha values, there is a set of shapes that decrease and eventually develop cavities as alpha decreases. Fig. 8 depicts the convex hull and one of the alpha shapes for a given set of 2D data points.

To use alpha shapes for boundary extraction of lidar data, we need to augment the above technique with a few additional steps. First, although the lidar data is in 3D, we will omit the height information, and simply perform the shape extraction in R^2 . Additionally, when using linescanning sensors, one dimension of the data may have a higher sampling density

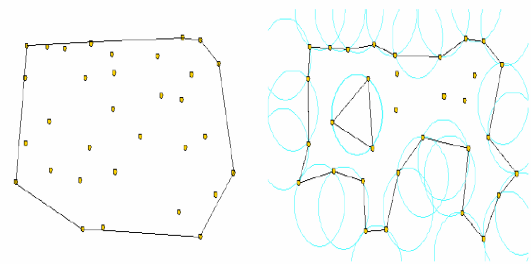


Figure 8. 'Shape' of a collection of 2D points. Convex hull (left) and alpha shapes (right)

than the orthogonal dimension. To compensate, we may opt to use rotated ellipses as our structuring element. However, a simpler approach is to stretch the data along either the in-track or cross-track dimension, whichever is sampled at a greater frequency. Once this stretching has occurred, the sampling density will be approximately uniform, and circular structuring elements may be used. A traditional alpha shapes analysis should then be performed, with the structuring element sized to be slightly larger than the inter-data spacing. Boundary points are then ordered, and a curvature algorithm is used to determine which of these boundary points represent significant changes in the boundary direction. These points are then isolated, and the resultant groups of points between the points of high curvature are modeled with a best-fitting line using a Deming orthogonal regression. If desired, these lines may be constrained to fall along the dominant 45 degree lines of the building, which produces more regular results in many cases. The resultant boundary lines are then intersected to obtain vertices, and the connectivity of these vertices defines the outer boundary of the object being modeled. This process is illustrated for the main roof layer of Building 7 in Fig. 9.

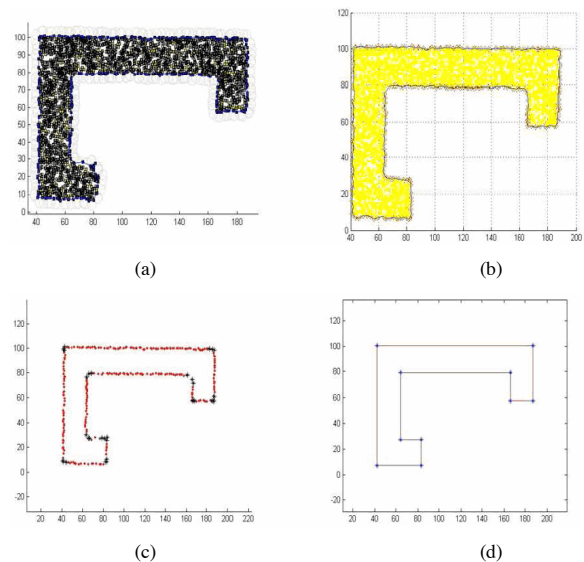


Figure 9. Determination of an object's exterior boundary: (a) Identification of exterior points using 2D alpha shapes (b) Ordering of the exterior points (c) Identification of boundary points with high curvature (d) Best-fit lines and resultant vertices (all units in meters)

2007 Urban Remote Sensing Joint Event

Once the outer boundary has been determined, interior breaklines and the resultant vertices still need to be defined. This is accomplished through a methodology similar to that presented in [14]. First, each data point is assigned a normal vector according to the plane best fitting the data in a voxel centered on the point of interest. This plane is determined through a 3D Deming regression. The mean shift algorithm is then used to segment the points into several groups, using the normal vectors and point locations as the defined features. Points lying between dominant classes are then removed, and the remaining data are fit with planes, again using the Deming methodology. The planes are intersected with the other internal planes as well as vertical planes through the boundary and the resultant lines are also intersected to produce vertices. This internal feature extraction was performed on a simulated 3-sided gabled roof, and the results are presented in Fig. 10.

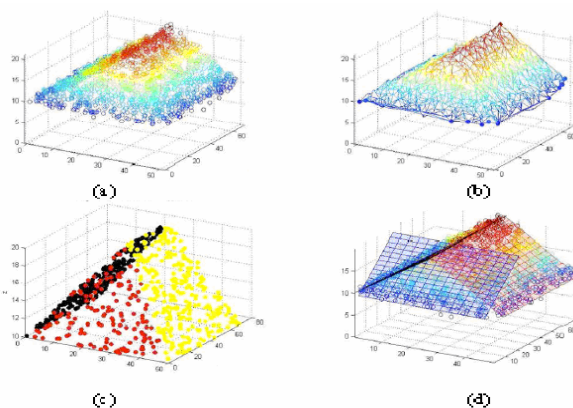


Figure 10. Internal Plane Analysis: (a) Data Points, (b) Extraction of the External Boundary, (c) Classification of Points Based on Orientation of their Local Normal Vectors, (d) Resultant Object Model

F. Background Maps

A variety of background characterization methods have been added to the DIRSIG model in the form of a collection of map layers. Each layer assigns various attributes to surfaces within a scene, and in general, we extract these map layers from spectral data that is frequently sharpened by higher resolution passive imagery. In most cases, this spectral data will be used to geometrically segment the image, and the desired properties will then be assigned via a look-up table that accesses previously created attributes. The mapping images should be available at spatial resolutions significantly higher than the resolution of the sensor to be simulated, so that DIRSIG can sufficiently oversample the scene to allow for modeling of the detector point spread function.

As a case in point, road maps in DIRSIG have traditionally been created in a drawing utility such as Adobe Photoshop. An improvement may be realized by segmenting the scene using traditional spectral classification techniques, then refining the road regions through use of high-resolution panchromatic imagery (see [24] for an example of performing this refinement via edge extraction). Although this often works well without additional processing, we have found that industrial roof-tops are frequently categorized in the same class as roads.

Therefore, augmenting the spectrally-derived classification map with the lidar-derived building locations provides a means of improving the final roadmap. Similar techniques are used to create material, texture, mixture, and reflectance maps.

G. Spectral Assignment

DIRSIG assigns spectra to spatial regions in the scene using two complimentary methodologies. The first of these is on a per-facet basis, where material types and corresponding spectra are directly assigned to object facets, while the second is through the use of spectral material and texture maps. In order to specify spectra for building models, a spectral classification routine is used to assign a material identification index to each horizontal and sloped roof 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.

Vertical walls and most materials on the terrain use a material map in lieu of assigning materials on a per facet basis. This material map is typically a classification map obtained from spectral imagery that has been draped over the appropriate surface. Using this technique, a single facet may have multiple material types associated with it. As before, final spectral curve selection is driven by a texture map image, which is usually obtained by frame-array imagery.

IV. PRELIMINARY RESULTS

In this section, we briefly demonstrate the feasibility of the proposed approach by applying these techniques to a portion of the RIT campus. The imagery used in this analysis came from the following sources:

- Lidar data was supplied by an outside vendor flying a commercial Optech sensor. The data contained approximately 6 points/m², roughly uniform in both the in- and cross-track dimensions. First-return range and intensity data was provided.
- Hyperspectral imagery came from RIT’s Modular Imaging Spectrometer Instrument (MISI), a 70-band VNIR lincanning spectrometer with a 3m ground resolution.
- Multi-spectral imagery was provided the Wildfire Airborne Sensor Program (WASP) 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.

We chose to apply our technique to the portion of the RIT campus highlighted in Fig. 11. This area contains several buildings, a region with densely populated trees, numerous isolated trees, roads, grass and dirt. There is also a 10m change in ground elevation throughout 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.

2007 Urban Remote Sensing Joint Event



Figure 11. Area of interest for scene simulation

We began the process by registering the lidar data to the WASP frame-array imagery. This registration was successful to within 4 pixels throughout most of the image, although it should be noted that after the initial 2D projection of 3D points was performed, the selection of homologous points for the second (refining) transform was done manually.

We next used the projection approach to geo-rectify the hyperspectral imagery collected from MISI. The result of this rectification is shown in Fig. 12. The left image depicts the raw hyperspectral image as collected, and the image on the right shows the result of projection and resampling one of the data bands. Although the geo-rectified image shows a marked improvement, a detailed inspection revealed spatial artifacts in excess of 10m were present throughout much of the image. It is believed that these artifacts are the result of GPS/INS errors, and a means to correct these distortions using additional imagery is being investigated. We were able to demonstrate improved spectral clustering for producing material maps using MISI imagery relative to using multi-spectral datasets. However, the spatial inaccuracies precluded the use of this hyperspectral data for the scene generated for this paper.

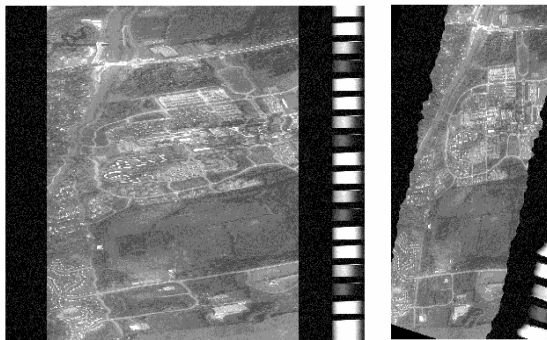


Figure 12. Un-registered MISI image (*left*) and geo-rectified image (*right*)

Once the lidar/WASP registration was completed, we processed the lidar data as specified above to extract the terrain model. Both the modified median filter and high-pass filter techniques were used to flag points for removal, and a simple linear interpolation was used to reconstruct the full terrain image. A 5x5 averaging filter was applied after interpolation to

smooth out some of the high-frequency content before facetization.

One problem area we noticed when analyzing the final terrain model was that holes adjacent to buildings were not handled well. There is a basement-accessible loading dock on the eastern side of Building 7, and once the building points were removed, this hole caused local problems with the subsequent interpolation. Future research will focus on a method to address this issue.

We recreated two of the buildings in the scene using the intersection of planes approach to determine model vertices. As specified earlier, building layers were first binned based on height, then individual objects were identified through spatial clustering. The parameters used for the height histogram segmentation were the same for each building, but height bins that did not have relevant features were manually ignored, and additional intervention was required to combine spatial classes related to a single object in some cases. Fig. 13 shows a detailed CAD model of Building 7 using the proposed extraction technique. Although a few details were misrepresented in the final models, the overall building geometries match truth data quite well.

In the future, we aim to refine the registration of the MISI imagery, and then use this data to accurately attribute building materials. However, in the interim, a more rudimentary technique was employed. For this work, building spectra were assigned on a per-facet basis based on facet orientation. Near horizontal facets were classified as roof material, while vertical surfaces were specified as brick. A small number of facets were also manually classified as glass.

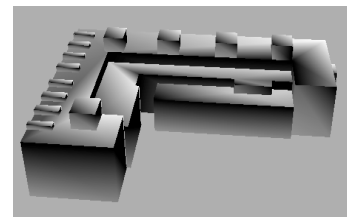


Figure 13. Extracted model of Building 7

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 produce a more uniform result. Detailed regions were then sharpened manually using Photoshop. A grayscale version of the same WASP image was used as the texture map.

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 14. 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.



Figure 14. Simulated color image using the extracted DIRSIG scene

V. 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, and the generation of background maps. 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 the RIT campus. The DTM extraction, identification of buildings and trees, and low-level geometric reconstruction of buildings were shown to be successful.

The overall scope of this research is quite large, and the work presented here represents the initial steps of a much larger effort. Future research will focus on improved methods of data registration, as well as additional ways to integrate traditional photogrammetric techniques into the geometrical object reconstruction. Additionally, the identification and specification of spectral signatures from registered hyperspectral data will be explored further.

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.

ACKNOWLEDGMENT

This work has been supported in part by the U.S. Government under University Research Initiative HM1582-05-1-2005. The authors also wish to acknowledge the significant contributions of many members of the Digital Imaging and Remote Sensing group at RIT in preparing this document.

REFERENCES

- [1] Schott, J.R., Brown, S. D., Raqueno, R. V., Gross, H. N., and Robinson, G., "An Advanced Synthetic Image Generation Model and Its Application to Multi/Hyperspectral Algorithm Development", *Canadian Journal of Remote Sensing*, Vol. 25, No. 2, June 1999.
- [2] Burton, R., "Elastic LADAR Modeling for Synthetic Imaging Applications", Ph.D. thesis, RIT, NY, 2002.
- [3] Blevins, D., "Modeling Scattering and Absorption for a Differential Absorption LIDAR System", Ph.D. thesis, RIT, Rochester, NY, 2005.
- [4] "Tree Professional" software, information available online at <http://www.onyxtree.com>
- [5] "Rhinceros: NURBS Modeling for Windows" software, information available online at <http://www.rhino3D.com>
- [6] Ientilucci, E. J., Ewald, K., Marcin, J., Spivey, A., "Guide to Building Large Scale DIRSIG Scenes", Digital Imaging and Remote Sensing Laboratory, RIT, 2003.
- [7] Habib, Ayman, Mwafag Ghanma, Michael Morgan, and Rami Al-Ruzouq. "Photogrammetric and Lidar Data Registration Using Linear Features." *Photogrammetric Engineering and Remote Sensing*, 71 (June 2005): 699–707.
- [8] Williamson, James R. and Michael H. Brill. *Dimensional Analysis Through Perspective: A Reference Manual*. ISBN: 0-8403-5673-0. Dubuque, Iowa: Kendall Hunt through ASPRS, 1990.
- [9] Shufelt, Jeffrey A. "Performance Evaluation and Analysis of Monocular Building Extraction From Aerial Imagery." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21 (April 1999): 311–326.
- [10] Wiemker, Rafael. "Registration of Airborne Scanner Imagery Using AKIMA Local Quintic Polynomial Interpolation." San Francisco, CA: , June 1996.
- [11] Brown, Lisa Gottesfeld. "A survey of image registration techniques." *ACM Comput. Surv.* 24 (1992): 325–376.
- [12] Breuer, Michael and Jorg Albertz. "Geometric Correction of Airborne Line-Scanner Imagery." *ISPRS Commission III, Working Group I, Vienna XXXI (1996): 19–23.*
- [13] Fan, Xiaofeng, Harvey Rhody, and Eli Saber. "Automatic Registration of Multi-Sensor Airborne Imagery." *International Conference on Computer Vision Theory and Applications*, 2007.
- [14] Ma, Ruijin. *Building Model Reconstruction from Lidar Data and Aerial Photographs*. PhD thesis, The Ohio State University, 2004.
- [15] Sithole, G. "Segmentation and Classification of Airborne Laser Scanner Data" *Publications on Geodesy* 59, May 2005.
- [16] Gonzalez, Rafael C. and Richard E. Woods. *Digital Image Processing*. Second edition. New Jersey: Prentice-Hall Inc., 2002.
- [17] Gray, R., Brown, S., Schott, J., "Scene construction methodologies and techniques for simulating forest areas", 11th Annual Ground Targets Modeling and Validation Conference, August 2000.
- [18] Solberg, Svein, Erik Naesset, and Ole Martin Bollandsas. "Single Tree Segmentation Using Airborne Laser Scanner Data in a Structurally Heterogeneous Spruce Forest." *PE&RS* 72, Dec 2006.
- [19] Li, Feng and Stephen Lach. 2006 "Geometric Scene Reconstruction Using Lidar Data." Final Report for Adv DIP Course, RIT. 2006.
- [20] Comaniciu, Dorin I. *Nonparametric Robust Methods for Computer Vision*. PhD thesis, Rutgers, 2000.
- [21] Lach, Stephen. "Semi-Automated DIRSIG Scene Construction Using Multi-Modal Imagery," Proposal for Doctoral Research, Rochester Inst. Of Tech. March 2007.
- [22] Morgan M. and Habib A., "Interpolation of Lidar Data and Automatic Building Extraction," *ACSM-ASPRS Annual Conference Proceedings*, 2002.
- [23] Edelsbrunner, H. and E. P. Mucke. 1992 "Three-dimensional Alpha Shapes." Dept. Comp Sci, U. of Illinois Urbana-Champaign.
- [24] Wang, R., Zhang, Y., "Extraction of Urban Road Network Using Quickbird Pan-Sharpned Multispectral and Panchromatic Imagery by Performing Edge-Aided Post-Classification", *ISPRS Joint Workshop on Spatial Temporal and Multi-dimensional Data Modeling and Analysis*, Quebec, Canada, October, 2003.