FUSION OF LIDAR DATA AND AERIAL IMAGERY FOR A MORE COMPLETE SURFACE DESCRIPTION

Toni Schenk
CEEGS Department
The Ohio State University
schenk.2@osu.edu

Bea Csathó
Byrd Polar Research Center
The Ohio State University
csatho.1@osu.edu

Commission III, Working Group 5

KEY WORDS: Fusion, Lidar, Aerial Imagery, Surface Reconstruction, DEM/DTM

ABSTRACT

Photogrammetry is the traditional method of surface reconstruction such as the generation of DTMs. Recently, LIDAR emerged as a new technology for rapidly capturing data on physical surfaces. The high accuracy and automation potential results in a quick delivery of DEMs/DTMs derived from the raw laser data. The two methods deliver complementary surface information. Thus it makes sense to combine data from the two sensors to arrive at a more robust and complete surface reconstruction. This paper describes two aspects of merging aerial imagery and LIDAR data. The establishment of a common reference frame is an absolute prerequisite. We solve this alignment problem by utilizing sensor-invariant features. Such features correspond to the same object space phenomena, for example to breaklines and surface patches. Matched sensor invariant features lend themselves to establishing a common reference frame. Feature-level fusion is performed with sensor specific features that are related to surface characteristics. We show the synergism between these features resulting in a richer and more abstract surface description.

1 Introduction

It has long been recognized that surfaces play an important role in the quest of reconstructing scenes from sensory data such as images. The traditional method of reconstructing surfaces is by photogrammetry. Here, a feature on the ground, say a point or a linear feature, is reconstructed from two or more overlapping aerial images. This requires the identification of the ground feature in the images as well as their exterior orientation. The crucial step in this process is the identification of the same ground feature. Human operators are remarkably adept in finding conjugate (identical) features. DEMs generated by operators on analytical plotters or on softcopy workstations are of high quality but the process is time and cost intensive. Thus, major research efforts have been devoted to make stereopsis an automatic process.

Recently, airborne and spaceborne laser altimetry has emerged as a promising method to capture digital elevation data effectively and accurately. In the following we use LIDAR (LIght Detection And Ranging) as an acronym for the various laser altimetry methods. An ever increasing range of applications takes advantage of the high accuracy potential, dense sampling, and the high degree of automation that results in a quick delivery of products derived from the raw laser data.

Photogrammetry and LIDAR have their unique advantages and drawbacks for reconstructing surfaces. It is interesting to note that some of the shortcomings of one method can be compensated by advantages the other method offers. Hence it makes eminent sense to combine the two methods—we have a classical fusion scenario where the synergism of two sensory input data considerably exceeds the information obtained by the individual sensors.

In Section 2 we elaborate on the strengths and weaknesses of reconstructing surfaces from LIDAR and aerial imagery. We also strongly advocate an explicit surface description that greatly benefits subsequent tasks such as object recognition and image understanding. Useful surface characteristics are only implicitly available in classical DEMs and DSMs. Explicit surface descriptions are also very useful for fusing LIDAR and aerial imagery.

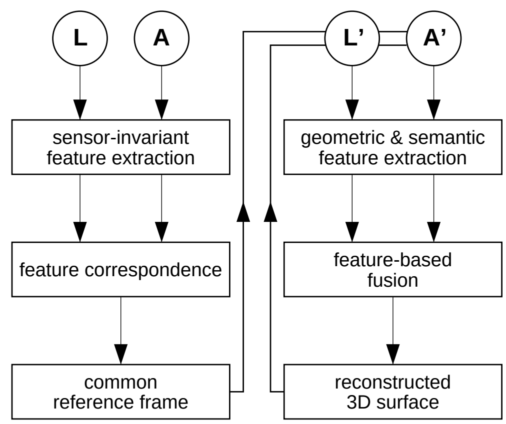

Figure 1: Flow chart of proposed multisensor fusion framework.

Fig. 1 depicts the flowchart of the proposed multisensor fusion framework. Although we consider only LIDAR (L) and aerial imagery (A) in this paper, the framework and the following discussions can be easily adopted for including additional sensors, such as a hyperspectral system. The processes on the left side of Fig. 1 are devoted to the establishment of a common reference frame for the raw sensory input data. The result is a unique transformation between the sensor systems and the reference frame. Section 3 discusses this part of the fusion problem in detail.

The processes on the right side of the figure are aimed at the reconstruction of the 3D surface by feature-based fusion. This task benefits greatly from having the sensory input data (L' and A') aligned. Since the reconstructed surface is described in the common reference frame, it is easy to go back from object space to sensor space in order to extract specific features that might be needed during the inference process to validate hypotheses, for example. Section 4 provides a more detailed discussion.

2 Background

In this section we briefly compare the advantages and disadvantages of the two most prominent methods for surface reconstruction. We also elaborate on how to represent surfaces and emphasize the need for explicit surface descriptions. Combining the advantages of LIDAR and stereo photogrammetry is an interesting fusion problem.

2.1 Surface reconstruction

With surface reconstruction we refer to the process of deriving features in the 3D object space. The traditional method of reconstructing surfaces is by photogrammetry. Here, a feature on the ground, say a point or a linear feature, is reconstructed from two overlapping aerial images—a process known as stereopsis. This requires the identification of the ground feature in both images as well as the exterior orientation of the images. The crucial step in stereopsis is the identification of the same ground feature, also referred to as correspondence problem or image matching. Human operators are remarkably adept in finding conjugate (identical) features. DEMs generated by operators on analytical plotters or on softcopy workstations are of high quality but the process is time and cost intensive. Thus, major research efforts have been devoted to make stereopsis an automatic process.

The success of automatic surface reconstruction from aerial imagery is marginal. Despite considerable research efforts there is no established and widely accepted method that would generate surfaces in more complex settings, say large-scale urban scenes, completely, accurately, and robustly. A human operator needs to be involved, at least on the level of quality control and editing.

On the other hand, LIDAR has been touted as a promising method to capture digital elevation data effectively and accurately. LIDAR can be viewed as a system that samples points of the reflective surface of the earth. The samples are irregularly spaced. We call the original surface measurements point cloud or mass points in this text. Laser points are computed from navigation data (GPS/INS) and range measurements. It is important to realize that there is no inherent redundancy in the computation of a laser point. In general, laser points do not carry semantic information about the scene.

Table 1: Advantages and disadvantages of lidar and aerial imagery for surface reconstruction.

	LIDAR	aerial imagery
advantages	high point density high vertical accuracy waveform analysis	rich in scene information high H + V accuracy redundant information
disadvantages	no scene information occluded areas horizontal accuracy? no inh. redundancy	stereo matching occluded areas degree of automation?

Table 1 lists a number of important advantages and drawbacks of LIDAR and photogrammetry in terms of surface reconstruction. LIDAR has a relatively high point density and a high accuracy potential. However, in order to reach the potential of a vertical accuracy of 1 dm and a horizontal accuracy of a few decimeters, the LIDAR system must be well calibrated. As is obvious from several accuracy studies, actual systems often do not yet reach this potential. Recently, some LIDAR system offer the option to record the entire waveform of the returning laser pulse. Waveform analysis yields additional information about the footprint area, for example roughness and slope information. We have already elaborated on the disadvantages. Laser points are positional, there is no additional scene information directly available from a single point.

In contrast to laser points, surfaces derived from aerial images are potentially rich in scene information. Also, 3D features in object space have a redundancy, r, of $r = 2n - 3$ with n the number of images that show the same feature. The Achilles heel of photogrammetry is the matching, that is, finding corresponding features on n images, where $n \geq 2$. The degree of automation is directly related to the matching problem.

From this brief comparison it is obvious that some of the disadvantages of one method are offset by advantages of the other method. This is precisely the major argument for combining, or fusing if you wish, the two methods.

2.2 Implicit vs. explicit surface description

Common to the techniques of acquiring digital elevation data is a cloud of 3D points on the visible surface. Except for measuring DEMs on analytical plotters, the mass points are irregularly distributed. Consequently, the next step is to interpolate the raw points into a regular grid. Digital Elevation Models (DEM) are immensely popular in many engineering applications. With DEM we refer to a surface representation with bare-earth $z-$values at regularly spaced intervals in $x-$ and $y-$direction. A bare-earth DEM is void of vegetation and man-made features—in contrast to a Digital Surface Model (DSM) that depicts elevations of the top surfaces of features elevated above the bare earth. Examples include buildings, vegetation canopies, power lines, and towers. Finally, the term Digital Terrain Model (DTM) is used as a DEM, augmented with significant topographic features, such as breaklines and characteristic points. A breakline is a linear feature that describes a discontinuity in the surface. Such discontinuities may signal abrupt changes in elevations across the breakline, or may refer to abrupt changes of the surface normal.

Although very useful, it is important to realize that a DEM or DSM does not make surface properties explicit. An explicit description of surface characteristics, such as planar or higher-order surface patches, surface discontinuities and surface roughness, is important for most subsequent tasks. Object recognition and image understanding rely on the knowledge of explicit surface properties and even the generation of orthophotos would greatly benefit from an explicit description of the surface. In this line of thinking we consider DEMs and DSMs as entirely implicit surface descriptions and DTMs as partially explicit (if breaklines and distinct elevation points are present). The challenge of describing the surface explicitly can be met easier if aerial imagery and LIDAR point clouds are fused. As described in *Lee*

(2002), the generation of explicit surface descriptions from irregularly distributed mass points is accomplished by way of segmentation and grouping processes, embedded in the theoretical framework of perceptual organization.

2.3 Motivation for fusing LIDAR and aerial imagery

We already noticed that by combining (fusing) LIDAR data and aerial imagery most of the disadvantages associated with either method are compensated (see, e.g. Table 1). The complementary nature of the two methods is even more evident when we attempt to describe the surface explicitly. Table 2 lists the most important surface properties that make up an explicit description. Surface patches are characterized by an analytical function. For urban scenes with many man-made objects, planar surface patches are very useful. A scene may consist of many planar and second-order surface patches. The exact boundary of surface patches is another important surface property. Boundaries can be explicitly represented by polylines, for example. Equally important is the explicit information about surface discontinuities, such as breaklines. Object recognition and other image understanding tasks greatly benefit from information about the roughness and material properties of surfaces. The interested reader may note that our proposed explicit surface description is conceptually related to the 2.5-D-sketch (*Marr* (1982)).

Table 2: Sources that predominantly determine surface properties.

surface properties	LIDAR	aerial imagery
patches	X	
boundaries		X
discontinuities		X
roughness	X	
material properties		

As Table 2 illustrates, LIDAR makes direct contribution towards surface patches. Boundaries, however, are easier to obtain from aerial imagery, as well as surface discontinuities. If the entire waveform of the returning signal is recorded, then LIDAR data contain information about the surface roughness. From neither sensor, material properties can be easily obtained. This information could come from a hyperspectral sensor, for example.

We propose to combine aerial imagery and LIDAR data on two levels. On the first level we are concerned with establishing a common reference frame for the two sensors. This is an absolute prerequisite for the second step where extracted information from both sensors is fused for a more complete and explicit surface description. Note that this proposed multisensor fusion procedure is resembles the feature-level data fusion as proposed by *Hall* (1992). Conceptually, our approach is also close to the definition of fusion provided by *Wald* (1999), "Fusion aimes at obtaining information of greater quality..."

3 Common reference frame for LIDAR data and aerial imagery

This first step of our fusion approach is also known as *sensor alignment* or *registration*. We prefer the term *referencing* and mean by that the establishment of a common reference frame for the sensor data. It entails a transformation, forward and backward, between sensor data and reference frame. LIDAR points are usually computed in the WGS84 reference frame. Hence it makes sense to reference the aerial images to the same system. Referencing aerial imagery to LIDAR can then be considered an orientation problem. This is attractive because the success of fusion can be quantitatively judged from the adjustment results. Moreover, the features extracted from LIDAR, serving as control features, can be stochastically modeled and the error analysis of the orientation will allow to discover remaining systematic errors in the LIDAR data.

We propose to use sensor invariant features to determine the relevant orientation parameters. Before elaborating further on how to obtain sensor invariant features and their use in orientation, we briefly summarize some aspects of the direct orientation.

3.1 Direct orientation

If both sensors are mounted on the same platform then the navigation system (GPS/INS) provides position and attitude data for the aerial camera and the LIDAR system. Since the two sensors, the GPS antenna, and the IMU unit are physically separated, the success of direct orientation hinges on how well the relative position and attitude of the various system components can be determined and how stable they stay during data acquisition. If the two sensors are flown on separate platforms, obviously each must have its own navigation system. In this case, the sensor alignment is now also affected by potential systematic errors between the two navigation systems.

By and large, the direct orientation and hence the referencing of aerial imagery and LIDAR data works well. However, there is no inherent quality control. If the two sensors are misaligned, for example by an unknown mounting bias, one may only find out much later, if at all. Another problem with direct orientation is related to the camera orientation. It is well-known that the interior and exterior orientation of aerial images are strongly correlated. *Schenk* (1999), shows that most errors in the interior orientation are compensated by exterior orientation parameters. In fact, interior and exterior orientation are a conjugate pair that should be used together for an optimal reconstruction of the object space from images.

3.2 Determining sensor invariant features

The geometric transformation, necessary for establishing a common reference frame, requires features in the object space that can be extracted from both sensory input data. Thus, the condition for aligning sensors in this indirect fashion is that they respond, at least partially, to the same phenomena in object space, referred to as *sensor invariant features* in this paper.

Table 3 lists features that can be extracted in several steps from LIDAR point clouds and aerial imagery. On the level of raw data the two sensors have nothing in common that can be directly used to register them. On the first level one can extract 3D surface patches from laser points. Such patches may be planar or higher order surfaces, depending on the scene. In built-up areas, usually many planar surface patches exist, corresponding to man-made objects. After surface patches have been extracted, a grouping process establishes spatial relationships. This is followed by forming hypotheses as to which patches may belong to the same object. Adjacent patches are then intersected if their surface normals are different enough to guarantee a geometrical meaningful solution. In fact, if the adjacency hypothesis is correct then the intersection is a 3D boundary of an object. *Lee* (2002) treats the steps of extracting and grouping patches as a perceptual organization problem.

Table 3: Multi-stage feature extraction from LIDAR and aerial images.

	LIDAR	**aerial imagery**
raw data	3D point cloud	pixels
feature extraction	patches	2D edges
processing	grouping	matching
results	3D edges	3D edges, patches

Let us now examine the features that can be extracted from images. The first extraction level comprises edges. They correspond to rapid changes in grey levels in the direction across the edges. Most of the time, such changes are the result of sudden changes in the reflection properties of the surface. Examples include shadows and markings. More importantly, boundaries of objects also cause edges in the images because the two faces of a boundary have different reflection properties too. Hence we argue that some of the 2D edges obtained from aerial imagery correspond to 3D edges obtained from laser points. That is, edges are potentially sensor invariant features that are useful for solving the registration problem. Note that the 2D edges in one image can be matched with conjugate edges in other images. It is then possible to obtain 3D features in model space by performing a relative orientation with linear features.

Are 3D surface patches also sensor invariant? They certainly correspond to some physical entitities in object space, for example a roof plane, face of a building, or a parking lot. Surface patches are first order features that can be extracted from laser point clouds relatively easily. However, it is much more difficult to determine them from images. One way to determine planar surfaces from images is to test if spatially related 3D edges are lying in one plane. Surface patches then can also be considered sensor invariant features.

3.3 Referencing aerial images to LIDAR data

From the discussion in the previous section we conclude that 2D edges in images, 3D edges in models, and 3D surface patches are desirable features for referencing aerial images with LIDAR. Table 4 lists three combinations of sensor invariant features that can be used to solve the fusion problem. As pointed out earlier, we consider this first step as the problem of determining the exterior orientation of aerial imagery. Extracted features from LIDAR data serve as control information.

Table 4: Sensor invariant features for fusing aerial imagery with LIDAR.

LIDAR	**aerial imagery**	**Method**
3D edges	2D edges	SPR, AT
3D edges	3D edges	ABSOR, AT
3D patches	3D patches	ABSOR, AT

Orientation based on 2D image edges and 3D LIDAR edges The first entry in Table 4 pairs 2D edges, extracted in individual images, with 3D edges established from LIDAR points. This is the classical problem of block adjustment, except that our fusion problem deals with linear features rather than points. Another distinct difference is the number of control features. In urban areas we can expect many control lines that have been determined from LIDAR data. It is quite conceivable to orient every image individually by the process of single photo resecting (SPR), that is, the problem can be solved without tie features. This offers the advantage that no image matching is necessary—a most desirable situation in view of automating the fusion process.

Several researchers in photogrammetry and computer vision have proposed the use of linear features in form of straight lines for pose estimation. Most solutions are based on the coplanarity model. Here, every point measured on a straight line in image space gives rise to a condition equation in that the point is forced to lie on the plane defined by the perspective center and the control line in object space, see, e.g. *Habib et al.* (2000). The solutions mainly differ in how 3D straight lines are represented.

Although straight lines are likely to be the dominant linear features in our fusion problem, it is desirable to generalize the approach and include free-form curves. *Zalmanson* (2000) presents a solution to this problem for frame cameras. In contrast to the coplanarity model, the author employs a modified collinearity model that is based on a parametric representation of analytical curves. Thus, straight lines and higher-order curves are treated within the same representational framework.

With the recent emergence of digital line cameras it is necessary to solve the pose estimation problem for dynamic sensors. The traditional approach is a combination of direct orientation and interpolation of orientation parameters for every line. This does not solve our fusion problem because no correspondence between extracted features from LIDAR and imagery is used—hence no explicit quality control of the sensor alignment is possible. In *Lee, Y.* (2002) the author presents a solution of estimating the pose for line cameras by using linear features. In this unique approach every sensor line is oriented individually, without the need for navigation data (GPS/INS).

Orientation based on 3D model edges and 3D LIDAR edges We add fusion with 3D model edges more for the purpose of completeness than practical significance. In contrast to the previous method, edges must be matched between images to obtain 3D model edges. In general, image matching, especially in urban areas, is considered difficult. We should bear in mind, however, that in our fusion problem, the surface can be considered to be known. Hence, the problem of geometric distortions of features can be well controlled and matching becomes feasible, assuming that reasonable approximations of the exterior orientation parameters are available. In fact, matching in object space, using iteratively warped images, becomes the method of choice. This matching procedure also offers the opportunity to match multiple images. Now we have the chance for a detailed reconstruction of complex surfaces from multiple images.

Orientation with surface patches Originally, the idea of using surfaces in the form of DEMs for orienting models was suggested by *Ebner and Strunz* (1988). The approach is based on minimizing the $z-$differences between the model points and points in object space found by interpolating the DEM. The differences are minimized by determining the absolute orientation parameters of the model. *Schenk* (1999a) modified the approach by minimizing the distances between corresponding surface elements. We propose the latter method for fusing aerial images with LIDAR.

The advantage of using patches as sensor invariant features is the relatively simple process to extract them from laser points. The fitting error of the laser points to a mathematical surface serves as quality control measure. In contrast to the previous methods, no planimetric features need to be extracted. Patches are more robust than features derived from them, for example 3D edges.

Unfortunately, the situation is quite different for determining surface patches in aerial images. Although theoretically possible by texture segmentation and gradiant analysis, it is is very unlikely that surface information can be extracted from single images. Hence, image matching (fusion) is required. Quite often, surface patches have uniform reflectance properties. Thus, the grey level distribution of conjugate image patches is likely to be uniform too, precluding both, area-based and feature-based matching methods, respectively. The most promising approach is to infer surface patches from surface boundaries (matched edges).

As shown by *Jaw* (1999), the concept of using control surfaces for orienting stereo models can be extended to block adjustment. In analogy to tie points, the author introduces tie surfaces. To connect adjacent models, the only condition is to measure points on the same surface. However, the points do not need to be identical—clearly, a major advantage for automatic aerial triangulation.

Alternative solution with range images A popular way to deal with laser points is to convert them to range images. This is not only advantageous for visualizing 3D laser point clouds but a plethora of image processing algorithms can operate on range images. For example, an edge operator will find edges in a range image, suggesting that the 3D edges used as sensor invariant features be determined from range images. At first sight, this is very appealing since it appears much simpler than the method described in Section 2. Let us take a closer look before making a final judgment, however.

Generating range images entails the interpolation of the irregularly spaced laser points to a grid and the conversion of elevations to grey values. While the conversion is straightforward, the interpolation deserves closer attention. Our goal is to detect edges. Edges in range images correspond to rapid changes of elevations in the direction across the edge. This is precisely where we must expect large interpolation errors. It follows that the localization of edges in range images may not be accurate enough for precise fusion.

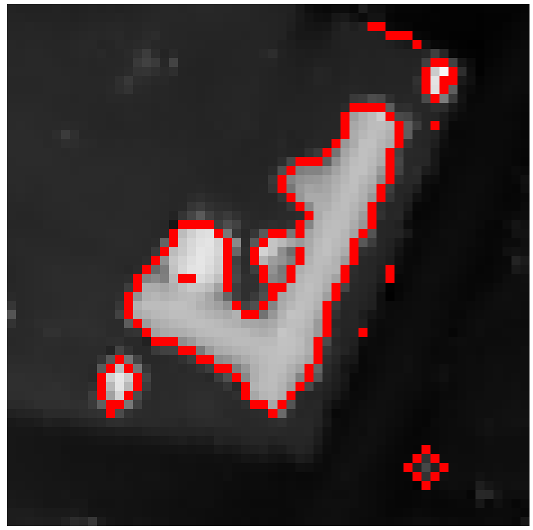

Figure 2: Edge detection performed on a range image. The edges are affected by interpolation errors and usually not suitable for sharp boundary delineation.

Fig. 2 depicts a sub-image with a fairly large building (see also Fig. 3b). The DEM grid size of 1.3 meters (average distance between the irregularly distributed laser points) leads to a relatively blocky appearance of the building and to jagged, fragmented edges. Moreover, edges are predominantly horizontal. It is quite difficult to detect non-horizontal edges (boundaries) in object space from range images. Finally, when comparing the edges obtained from the range image with those determined by intersecting adjacent planar surface patches it becomes clear that their use for fusing aerial images with LIDAR becomes problematic.

4 Fusion of aerial imagery with LIDAR data

After having established a common reference frame for LIDAR and aerial imagery we are now in a position to fuse features extracted from the two sensors to a surface description that is richer in information as would be possible with either sensor alone. We have strongly argued for an explicit description to aid subsequent processes such as object recognition, surface analysis, bare-earth computations, and even the generation of orthophotos. Since these applications may require different surface descriptions, varying in the surface properties (quality and quantity), an important question arises: is there a general description, suitable for applications that may not even be known by the time of surface reconstruction?

Surfaces, that is their explicit descriptions, play an important role in spatial reasoning—a process that occurs to a varying degree in all applications. We consider the surface properties listed in Table 2 essential elements that are, by and large, application dependent. In a demand-driven implementation, additional properties or more detailed information can be obtained from the sensory input data upon request.

Surface patches are obtained by segmenting the laser point cloud. The segmentation process will leave gaps, that is, patches do not contiguously cover the visible surface. A variety of reasons contribute to this situation. For one, occlusions and low reflectance (e.g. water bodies) result in regions with weakly populated laser points. Moreover, certain surfaces, such as the top of canopies, or single trees and shrubs do not lend themselves to a simple analytical surface description. It is conceivable to augment the set of surface patches obtained from LIDAR data by surfaces obtained from aerial imagery. An interesting example is vertical walls, such as building facades. The number of laser points reflected from vertical surfaces is usually below the threshold criterion for segmentation. It is therefore very unlikely that vertical surface patches are extracted. During the analysis of spatial relationships among patches it is possible to deduce the existence of vertical patches. These hypotheses can then be confirmed or rejected by evidence gained from aerial images.

Boundaries: it is assumed that surface patches correspond to physical surfaces in object space. As such, they are only relevant within their boundaries. Thus, the complete boundary description, $\mathcal{B}$, is important. The simplest way to represent the boundary is by a closed sequence of 3D vectors. The convex hull of the laser points of a surface patch serves as a first crude estimation of the patches' boundary. It is refined during the perceptual organization of the surface. However, boundaries inferred from LIDAR data remain fuzzy because laser points carry no direct information about boundaries. A much improved boundary estimate can be expected from aerial imagery. Matching extracted edges in two or more overlapping images is greatly facilitated by the LIDAR surface and by the knowledge where boundaries are to be expected. Thus it stands to reason to replace the somewhat fuzzy boundaries obtained from LIDAR by 3D edges derived from aerial imagery.

Discontinuities are linear features in object space that signal either an abrupt change in the surface normal or an abrupt change in the elevation. Discontinuities constitute very valuable information, not only for automatic scene interpretation but also for mundane tasks such as the generation of orthophotos. Like boundaries, discontinuities are represented as 3D polylines. With a few exceptions, boundaries are, in fact, discontinuities. Whenever patches are adjacent their common boundary must be a discontinuity. Take a saddle roof, for example. If the adjacency of the two roof planes is confirmed then their common boundary (e.g. intersection of roof planes) is a discontinuity. Since discontinuities are richer in information than boundaries, it is desirable to replace boundaries whenever possible by discontinuities.

Discontinuities are derived from aerial images in the same fashion as boundaries. Moreover, some of them can be obtained from LIDAR by intersecting adjacent surface patches. As noted earlier, corresponding 3D edges from images and 3D edges from LIDAR are used for establishing a common reference frame between images and LIDAR.

Roughness is a surface patch attribute that may be useful in certain applications. It can be defined as the fitting error of the surface patch with respect to the laser points. The waveform analysis of returning laser pulses yield additional information about the roughness of the laser footprint and hence the surface patch.

5 Experimental results

In this section we briefly demonstrate the feasibility of the proposed approach to reconstruct surfaces in an urban scene. We use data from the Ocean City test site. As described in *Csathó et al.* (1998b) the data set comprises aerial photography, laser scanning data, and multispectral and hyperspectral data. Fig. 3(a) depicts the southern part of Ocean City, covered by a stereomodel. In the interest of brevity we concentrate on a small sub-area containing a large building with a complex roof structure, surrounded by parking lots, garden, trees and foundation plants that are in close proximity to the building, see Fig. 3(b). The aerial photographs, scale $\approx 1:4,200$, have been digitized with a pixelsize of $15\ \mu$m. The laser point density is ≈ 1.2 points/m^2.

First we oriented the stereopair with respect to the laser point cloud by using sensor invariant features, including straight lines and surface patches. The intersections of adjacent roof planes are examples of straight-line features extracted from the LIDAR data (Fig. 3(d)). In the aerial images, some of these roof lines are detected as edges, see e.g. Fig. 3(e,f)) and consequently used in the orientation process. In order to avoid image matching we oriented the two images first individually by single photo resectioning. For checking the internal model accuracy we performed a relative orientation with the exterior orientation parameters and the laser points as approximations. The average parallax error, obtained from matching several thousand back-projected laser points, was $\pm 2.6\ \mu$m. The error analysis revealed a horizontal accuracy of sensor invariant features of $\pm 2.6\ \mu$m, confirming that the LIDAR data sets (NASA's Airborne Topographic Mapper, ATM) are indeed well calibrated

We now move on to the surface reconstruction of the sub-area, beginning with the LIDAR data. As described in detail in *Lee* (2002), the laser point cloud is subjected to a three-stage perceptual organization process. After having identified suitable seed patches, a region-growing segmentation process starts with the aim to find planar surface patches. In a second step, the spatial relationship and the surface parameters of patches are examined to decide if they can be merged. At the same time, boundaries are determined. Fig. 3(c) shows the result after the first two steps. A total of 19 planar surface patches have been identified. The white areas between some of the patches indicate small gaps that did not satisfy the planar surface patch conditions. We see confirmed that the boundaries of physical surfaces, e.g. roofs, are ill-defined by laser points. The third step of the perceptual organization process involves the intersection of planar surface patches that satisfy adjacency condition. The result of intersecting adjacent planes with distinct different surface normals is depicted in Fig. 3(d). Although extremely useful for spatial reasoning processes, the segmentation results from LIDAR are lacking well defined boundaries. Moreover it is desirable to increase the discrimination between surfaces that may belong to different objects. An interesting example is patch 3 (roof), 19 (tree), and 11 (foundation plant). With LIDAR data only one cannot determine if these three patches belong to the same object.

After having oriented the aerial imagery to the LIDAR point cloud we can fuse features extracted from the images with the segmented surface. Figs. 3(e,f) depict the edges obtained with the Canny operator. We show them here to demonstrate the difficulty of matching edges to reconstruct the object space by stereopsis. With the segmented surface and the exterior orientation parameters available it is possible to constrain the edge detection process to special areas, such as the boundaries of segmented regions, to adapt the parameters of the edge operator, or even choose other operators that may be better suited in a particular case. Figs. 3(g,h) show the effect of using all the knowledge that has been gained about scene before extracting edges. The segmentation of the LIDAR points led to planar surface patches and boundaries. These boundaries are projected back to the images and thus specify image regions where we look for edges. The edges obtained in both images are then projected into the segmented scene, for example by intersecting the planar surface patches with the plane defined by the projection center and the edge. With this procedure we have now boundaries in object space that have been derived either from LIDAR points or from aerial images, or from a combination. Fig. 3(i) shows the final result. The color-coded boundaries reflect the combinations that are also a useful measure to express the confidence and accuracy. For example, the red roof edge was determined from LIDAR and confirmed by edges from both aerial images.

6 Concluding remarks

We have shown in this paper that fusing aerial imagery with LIDAR data results in a more complete surface reconstruction because the two sensors contribute complementary surface information. Moreover, disadvantages of one sensor are partially compensated by advantages of the other sensor. We have approached the solution of the fusion problem in two steps, beginning with establishing a common reference frame, followed by fusing geometric and semantic information for an explicit surface description.

Many higher order vision tasks require information about the surface. Surface information must be represented explicitly (symbolic) to be useful in spatial reasoning processes. Useful surface information comprises surface patches, described by an analytical function, their boundaries, surface discontinuities, and surface roughness. Note that the explicit surface description is continuous, just like the real physical surface. This is in contrast to the better known discrete representations such as DEMs, DSMs, and DTMs. Here surface information is only implicitly available with the notable exception of a DTM that contains breaklines. Unlike explicit descriptions, grid and triangular representations (TIN) have no direct relationships with objects.

The fusion of aerial imagery and LIDAR offers interesting applications. The first step for example establishes an excellent basis for performing a rigorous quality control of the LIDAR data. This is particularly true for estimating the horizontal accuracy of laser points and for discovering systematic errors that may still remain undetected even after careful system calibration. Another interesting application is change detection. Imagine a situation where aerial imagery and LIDAR data of the same site are available but with a time gap between the separate data collection missions. Differences between the two data sets that exceed random error expectations, must have been caused by systematic errors or by changes in the surface.

After having completed the fusion approach as described in this paper, future research will concentrate on applications in order to test the suitability of the explicit surface description in spatial reasoning processes as they pertain to object recognition and other image understanding tasks.

REFERENCES

Csathó, B. and T. Schenk (1998). Multisensor data fusion for automatic scene interpretation. *ISPRS Intern. Archives*, vol. 32, part 3/1, pp. 429–434.

Csathó, B., W. Krabill, J. Lucas and T. Schenk (1998). A multisensor data set of an urban and coastal scene. *ISPRS Intern. Archives*, vol. 32, part 3/2, pp. 15–21.

Ebner, H. and G. Strunz (1988). Combined Point Determination Using Digital Terrain Models as Control Information. In *International Archives of Photogrammetry and Remote Sensing*, **27**(B11/3), 578–587.

Habib, A., A. Asmamaw, D. kelley and M. May (2000). *Linear features in photogrammetry.* Report No. 450, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH 43210.

Hall, D. (1992). *Mathematical Techniques in Multisensor Data Fusion.* Archtech House, New York. 301 p.

Jaw, J.J. (1999). *Control Surface in Aerial Triangulation.* PhD dissertation, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH 43210.

Lee, I. (2002). *Perceptual Oganization of Surfaces*. PhD dissertation, Department of Civil and Environmental Engineering and Geodetic Science, OSU, 157 p.

Lee, Y. (2002). *Pose estimation of line cameras using linear features*. PhD dissertation, Department of Civil and Environmental Engineering and Geodetic Science, OSU, 162 p.

Marr, D. *Vision*. W.H. Freeman, San Francisco,

Schenk, T. (1999a). *Matching Surfaces*. Technical Notes in Photogrammetry, No. 15, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH 43210, 21 pages.

Schenk, T. (1999b). *Digital Photogrammetry*. TerraScience, Laurelville, OH 43135, 428 pages.

Schenk, T. and B. Csatho (2001). Modellierung systematischer Fehler von abtastenden Laseraltimetern. *Zeitschrift für Photogrammetrie, Fernerkundung und GIS*, **5**(2001), 361-373.

Wald, L. (1999). Some Terms of Reference in Data Fusion. *IEEE Trans. on Geoscience and Remote Sensing*, **37**(3), 1190–1193.

Zalmanson, G. (2000). *Hierarchical Recovery of Exterior Orientation from Parametric and Natural 3D Curves.* PhD dissertation, Department of Civil and Environmental Engineering and Geodetic Science, OSU, 121 p.

.

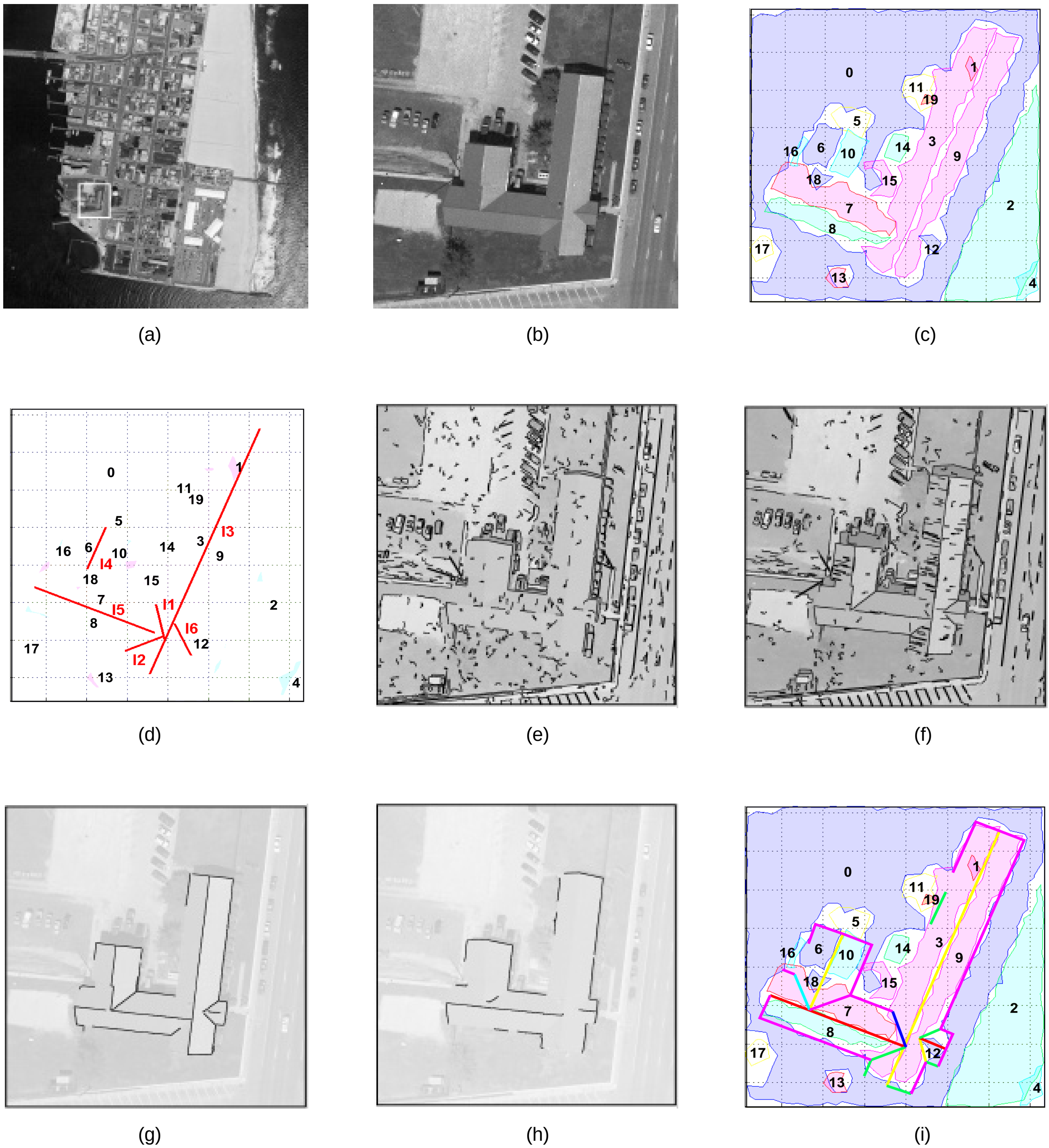

Fig. 3: Results of fusing aerial images with LIDAR data. Fig. (a) shows the test site of Ocean City and (b) depicts the sub-area that was used to demonstrate the detailed surface reconstruction (indicated by white box in a). In (c), the results of segmenting the LIDAR point cloud is shown with a total of 19 planar surface patches. The next step of the perceptually organized LIDAR point cloud is shown in (d) where adjacent planes are intersected, resulting in the six breaklines, $I1 - I6$. Fig. (e,f) contain the edges of the aerial stereopair obtained by the Canny operator, illustrating the dif?culty of image matching for stereopsis. The next ?gures (g,h) show more speci?c edges that are obtained using the current knowledge about the scene. These edges were matched in object space with the segmented LIDAR surface. The ?nal result of the surface reconstruction is shown in (i). The color code for the region boundaries corresponds to: red: LIDAR+aerial+aerial; yellow: LIDAR+aerial; magenta: aerial+aerial; blue: LIDAR; green: aerial.