INCLUSION OF MULTISPECTRAL DATA INTO OBJECT RECOGNITION

Bea Csathó [1], Toni Schenk, Dong-Cheon Lee and Sagi Filin [2]
[1] Byrd Polar Research Center, OSU, 1090 Carmack Rd., Columbus, OH 43210, email: csatho.1@osu.edu, phone: 1-614-292-6641
[2] Department of Civil Engineering, OSU, 2070 Neil Ave., Columbus, OH 43210 email: schenk.2@osu.edu, phone: 1-614-292-7126

KEYWORDS: Data fusion, multisensor, classification, urban mapping, surface reconstruction.

ABSTRACT
In this paper, we describe how object recognition benefits from exploiting multispectral and multisensor datasets. After a brief introduction we summarize the most important principles of object recognition and multisensor fusion. This serves as the basis for the proposed architecture of a multisensor object recognition system. It is characterized by multistage fusion, where the different sensory input data are processed individually and only merged at appropriate levels. The remaining sections describe the major fusion processes. Rather than providing detailed descriptions, a few examples, obtained from the Ocean City test-data site, have been chosen to illustrate the processing of the major data streams. The test site comprises of multispectral and aerial imagery, and laser scanning data.

1. INTRODUCTION

The ultimate goal of digital photogrammetry is the automation of map making. This entails understanding aerial imagery and recognizing objects - both hard problems. Despite of the increased research activities and the remarkable progress that has been achieved, systems are still far from being operational and the far-reaching goal of an automatic map making system remains a dream.

Before an object, e.g. a building, can be measured, it must first be identified as such. Fully automated systems have been developed for recognizing certain objects, such as buildings and roads on monocular aerial imageries, but their performance largely depends on the complexity of the scene and other factors (Shufelt, 1999). However, the utilization of multiple sensory input data, or other ancillary data, such as DEMs or GIS layers, opens new avenues to approach the problem. By combining sensors that use different physical principles and record different properties of the object space, complementary and redundant information becomes available. If merged properly, multisensor data may lead to a more stable and consistent scene description. Active research topics in object recognition include multi-image techniques using 3D feature extraction, DEM analysis or range images from laser scanning, map- or GIS-based extraction, color or multispectral analysis, and/or a combination of these techniques.

Now the cardinal question is how to exploit the potential these different data sources offer to tackle object recognition more effectively. Ideally, proven concepts and methods in remote sensing, digital photogrammetry and computer vision should be combined in a synergistic fashion. The combination may be possible through the use of multisensor data fusion, or distributed sensing. Data fusion is concerned with the problem of how to combine data from multiple sensors to perform inferences that may not be possible from a single sensor alone (Hall, 1992). In this paper, we propose a unified framework for object recognition and multisensor data fusion.

We start out with a brief description of the object recognition paradigm, followed by the discussion of different architectures for data fusion. We then propose a multisensor object recognition system. The remaining sections describe the major fusion processes. Rather than providing detailed descriptions, a few examples, obtained from the Ocean City test-data site, have been chosen to illustrate the processing of the major data streams. Csathó and Schenk (1998) reported on earlier tests using the same dataset. The paper ends with conclusions and an outline of future research.

2. BACKGROUND

2.1. Object recognition paradigm

At the heart of the paradigm is the recognition that it is impossible to bridge the gap between sensory input data and the desired output. Consider a gray level image as input and a GIS as the result of object recognition. The computer does not see an object, e.g., a building. All it has available at the outset is an array of numbers. On the output side, however, we have an abstract description of the object, for example, the coordinates of its boundary. There is no direct mapping between the two sets of numbers.

A commonly used paradigm begins with preprocessing the raw sensory input data, followed by feature extraction and segmentation. Features and regions are perceptually organized until an object, or parts of an object, emerge from the data. This data model is then compared with a model of the physical object. If there is sufficient agreement, the data model is labeled accordingly. In a first step, the sensor data usually require some pre-processing. For example, images may be radiometrically adjusted, oriented and perhaps normalized. Similarly, raw laser altimeter data are processed to 3-D points in object space.

The motivation for feature extraction is to capture information from the processed sensory data that is somehow related to the objects to be recognized. Edges are a typical example.

Edges are discontinuities in the gray levels of an image. Except for noise or systematic sensor errors, edges are caused by events in the object space. Examples of such events include physical boundaries of objects, shadows, and variations in the reflectance of material. It follows that edges are useful features, as they often convey information about objects in one way or another.

Segmentation is another useful step in extracting information about objects. Segmentation entails grouping pixels that share similar characteristics. Unfortunately, this is a quite vague definition and not surprisingly often defined by the application.

The output of the first stage is already a bit more abstract than the sensory input data. We see a transition from signals to symbols, however primitive they may still be. These primitive symbols are now subject of a grouping process that attempts to perceptually organize them. Organization is one of the first steps in perception. The goal of grouping is to find and combine those symbols that relate to the same object. Again, the governing grouping principles may be application dependent.

The next step in model-based object recognition consists of comparing the extracted and grouped features (data model) with a model of the real object (object model), a process called matching. If there is sufficient agreement, then the data model is labeled with the object and undergoes a validation procedure. Crucial in the matching step is the object model and the representation compatibility between the data and object model. It is fruitless to describe an object by properties that cannot be extracted from the sensor data. Take color, for example, and the case of a roof. If only monochromatic imagery is available then we cannot use ‘red’ in the roof description.

The sequential way on how the paradigm is presented is often called bottom-up or data driven. A model driven or top-down approach follows the opposite direction. Here, domain specific knowledge would trigger expectations, where objects may occur in the data. In practice, both approaches are combined.

2.2. Multisensor fusion

Multisensor integration means the synergistic use of the information provided by multiple sensory devices to assist the accomplishment of a task by a system. The literature on multisensor integration in computer vision and machine intelligence is substantial. For an extensive review, we refer the interested reader to Abidi and Gonzalez (1992), or Hall (1992).

At the heart of multisensor integration lies multisensor fusion. Multisensor fusion refers to any stage of the integration process where information from different sensors is combined (fused) into one representation form. Hence, multisensor fusion can take place at the signal, pixel, feature, or symbol level of representation. Most sensors typically used in practice provide data that can be fused at one or more of these levels. Signal-level fusion refers to the combination of signals from different sensors with the objective of providing a new signal that is usually of the same form but of better quality. In pixel-level fusion, a new image is formed through the combination of multiple images to increase the information content associated with each pixel. Feature-level fusion helps making feature extraction more robust and creating composite features from different signals and images. Symbol-level fusion allows the information from multiple sensors to be used together at the highest level of abstraction.

Like in object recognition, identity fusion begins with the preprocessing of the raw sensory data, followed by feature extraction. Having extracted the features or feature vectors, identity declaration is performed by statistical pattern recognition techniques, or geometric models. The identity declarations must be partitioned into groups that represent observations belonging to the same observed entity. This partitioning - known as association - is analogous to the process of matching data models with object models in model based object recognition. Finally, identity fusion algorithms, such as feature-based inference techniques, cognitive-based models, or physical modeling are used to obtain a joint declaration of identity. Alternatively, fusion can occur at the raw data level or at the feature level. Examples for the different fusion types include pixel labeling from raw data vectors (fusion at data or pixel level), segmenting surfaces from fused edges extracted from aerial imagery and combined with laser measurements (feature level fusion), and recognizing buildings by using ‘building candidate’ objects from different sensory data (decision level fusion).

Pixel level fusion is only recommended for images with similar exterior orientation, similar spatial, spectral and temporal resolution, and capturing the same or similar physical phenomena. Often, these requirements are not satisfied. Such is the case when images record information from very different regions of the EM spectrum (e.g., visible and thermal), or if they were collected from different platforms, or else have significantly different sensor geometry and associated error models. In these instances, preference should be given to the individual segmentation of images, with feature or decision level fusion. Yet another consideration for fusion is related to the physical phenomena in object space. Depending on the level of grouping, extracted features convey information that can be related to physical phenomena in the object space. Obviously, features extracted from different sensors should be fused when they have been caused by the same physical property. Generally, the further the spectral bands are apart, the lesser the features extracted from them are caused by the same physical phenomena. On the other hand, as the level of abstraction increases, more and

more different phenomena are described and need to be explained.

2.3. Dataset

To illustrate the main steps of the proposed approach, we use a dataset collected over Ocean City, Maryland on April 25 and 30, 1997 (http://polestar.mps.ohio-state.edu/~csatho/wg35.htm). Ocean City is located along a narrow peninsula by sandy beaches on the east, and harbors and docks on the west coast. High-rise buildings on the east and residential areas on the west side flank the main road. The dataset comprises of aerial photography, multispectral scanner imagery, and scanning laser data. As a part of the laser altimetry data, precise GPS positions and INS attitude have also been recorded. Csathó et al. (1998) describe the dataset in more detail.

Digital elevation data were acquired by the Airborne Topographic Mapper (ATM) laser system. The ATM is a conical scanning laser altimeter developed by NASA to measure the surface elevation of ice sheets and other natural surfaces, such as beaches, with ca. 10 cm accuracy (Krabill et al., 1995). The multispectral data were collected by the Daedalus AADS-1260 airborne multispectral scanner from the National Geodetic Survey (NGS). The AADS-1260 is a multispectral line scanner with eleven spectral bands in the visible, near infrared and thermal infrared providing 1.8-2.5 m pixels on the ground. Aerial photography was acquired with an RC20 camera, also from NGS. The laser scanner and the multispectral scanner were mounted on NASA's P-3B aircraft. The aerial camera was operated independently by NGS, but on the same day.

The original data have been preprocessed. For example, the GPS, INS, and laser range data were converted into surface elevations of laser footprints. The aerial photographs were scanned with a pixel size of 28 microns and an aerial triangulation with control points established by GPS provided the exterior orientation parameters. However, due to the lack of instrument calibration data, we could not transform the multispectral dataset into ground reflectance to establish a match with library spectra.

3. PROPOSED SYSTEM FOR MULTSENSOR OBJECT RECOGNITION

Figure 1 depicts a schematic diagram of a data fusion architecture that is tailored for combining aerial and multispectral imagery, and laser scanning data for object recognition in urban scenes. It includes modules that are considered important in model based object recognition.

The overall goal is to extract and group those features from the sensory input data that lead to a rich data model. This, in turn, permits modeling the objects with additional attributes and relationships so that the differences between data and object model become smaller, resulting in a more stable and robust recognition.

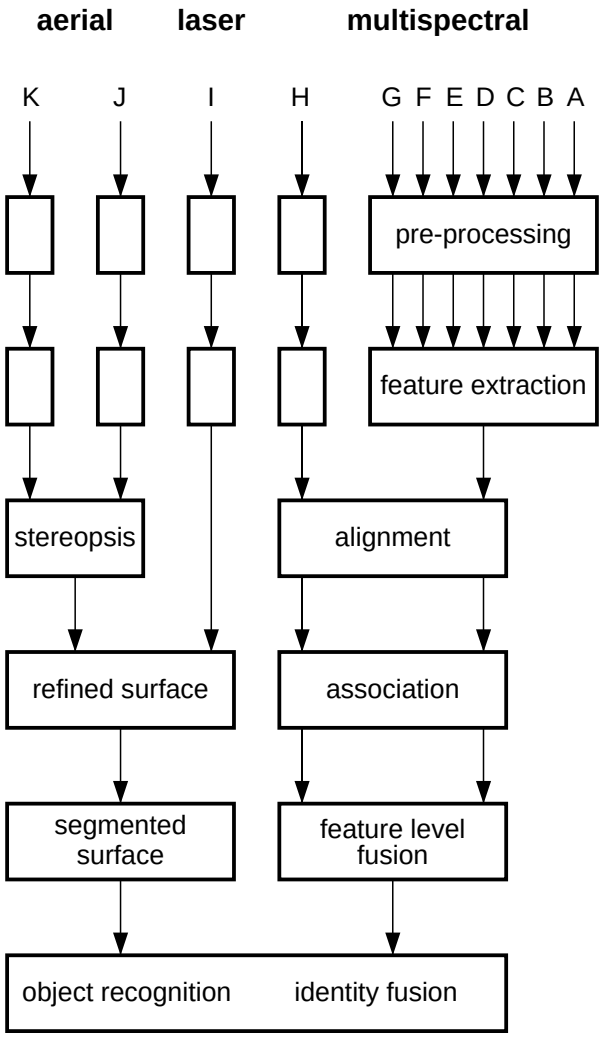

Fig. 1. Combination of sensor fusion and object recognition.

A common approach to fuse multispectral imagery and laser scanning data is to convert the latter into a range image, to add it as a separate channel to the multispectral data and to classify this combined set of data. This elegant approach is based on the assumption that the classes correspond to objects. Our approach is radically different, as we are quite skeptical on whether object recognition can be satisfactorily solved by classification. Rather, we advocate extracting features from the individual sensors and fuse them at appropriate levels, following the general rules described in the previous section.

It has long been realized that surfaces play an important role in object recognition. However, to be useful, surface information must be explicit and suitably represented. This entails extracting surface features from laser scanning data and aerial imagery followed by fusing them, leading to the

refined surface in Figure 1. However, a more abstract representation is needed, called segmented surface. Here, the topographic surface is separated from objects that have a certain vertical dimension, which, in turn, are approximated by planar surface patches or higher order polynomials.

Object recognition and identity fusion is performed in the 3-D object space. This last step is basically a spatial reasoning process, taking into account the grouped features, including the visible surface, and based on knowledge on how the features are related to each other. Abductive inference (Josephson and Josephson, 1994) provides a suitable framework by generating and evaluating hypotheses formed by the features.

Performing fusion in object space requires that sensor data or extracted features are registered to the object space. Preprocessed laser scanning points are already in object space. To establish a relationship between aerial imagery and object space, an aerial triangulation must be performed that will provide the exterior orientation parameters. Instead of generating an orthophoto for the multispectral imagery, we register it to the aerial imagery. This offers the advantage of performing the classification and feature extraction on the original image, thus preserving the radiometric fidelity. The results are then transformed to object space through the aerial images.

We have not implemented the design into a complete system, but performed experiments with the purpose of testing the fusion stages. The following sections report some of these experimental results.

4. MAJOR FUSION PROCESSES

4.1. Multispectral imagery

Multi- (and hyperspectral) systems are capturing images in a number of spectral bands in the visible and infrared region. In the visible-NIR part of the spectra, the dominant energy source is the solar radiation, and features in the images are mostly related to changes in surface reflectance, or in the orientation of the surface elements, or in both. Owing to the complex relationship between the spectral curves of the different materials, objects may look quite different in different spectral domains. For example, note the differences between the gray level images of the same area in visible and NIR frequencies (Figure 2a and b, right parts of images). Different combination of these bands, such as the false color composites in Figure 3a, can facilitate visual interpretation. The non-turbid, deep, clear water of the channels almost completely absorbs the energy resulting in a black color. The different man-made materials have more or less uniform reflectance throughout the visible-NIR domain creating a characteristic gray hue with an intensity that depends on the total brightness of the material. The bright red areas are associated with live green vegetation, which scatters most part of the solar radiation in the NIR. There is almost no energy reflected back from areas in deep shadow along the northern part of the houses.

In thermal infrared sensing the emitted EM radiation is imaged (Figure 2c). The measured radiant temperature of the objects depends on their kinetic or ‘true’ temperature and their emissivity. The temperature of the different objects changes differently throughout the day. For example, trees and water bodies are generally cooler than their surroundings during the day and warmer during the night. Fortunately, not all the objects exhibit this complex temporal behavior. For example, paved roads and parking lots are relatively warm both during day and night. Similarly to the visible images, daytime thermal imagery contains shadows in areas shaded from the direct sunlight. The energy captured by thermal IR sensing is also a function of the emissivity of the objects. In contrast to most natural surfaces, which have very similar emissivities, some man-made materials possess very distinct emissivities. For example, unpainted metal roofs have a very low emissivity (0.1-0.2), causing extremely low gray values in the thermal images. Hence, they provide excellent clues for locating metal surfaces.

Two different approaches were selected and tested for automatic interpretation of multispectral data. In the ‘multispectral edges’ method, edges extracted from selected individual spectral images were fused in the image space. In the more traditional approach, first the visible-NIR bands were segmented in image space by using unsupervised classification. Since visible-NIR and thermal images are based on different physical principles, the thermal imagery was not included in this step. Then, the boundaries between the different classes were extracted. Finally, these boundaries were fused with the ones extracted from the thermal imagery.

Multispectral-edges method. Edges obtained from different portions of the spectrum form a family - not unlike the scale space family - that add a new dimension to the grouping, segmentation, and object recognition processes. For example, Githuku, (1998) analyzed the relationship of 'colored' edges and exploited the uniqueness for matching overlapping images.

Edges extracted from visible, NIR, and thermal images can be strikingly different (2 a-c, left part of images). By extracting the edges from the individual bands and then analyzing and merging them, composite features can be created. Edges extracted from a visible (blue), a NIR (green) and a thermal band (red), are combined in a color composite image in Figure 4. The color of an edge on this image tells us which band had the strongest discontinuity in the location. All man-made objects are bounded by edges. Fortuitously, no edges were extracted along the fuzzy boundaries of some natural surfaces, such as the transition between bare soil, sparse and vigorous vegetation. Note, that the edges of man-made

objects, such as buildings, do not coincide on the different images (e.g., parallel green, yellow and red edges in the upper row of buildings). This suggests that the Instantaneous Field of View of the sensor had a significant spectral dependence.

Unsupervised classification. In solving remote-sensing problems, classification - sometimes combined with contextual information - is usually expected to provide the final answer. To increase the reliability and robustness of classification, many researchers favor supervised techniques. In our object recognition scheme, classification is just one of the early vision processes that provides only partial, incomplete information for object recognition. Since the number of classes is usually not known a priori and no training data is available, we employ unsupervised classification methods.

Unsupervised classification explores the inherent cluster structure of the feature vectors in the multidimensional feature space. Clustering usually results in a grouping, where the variance within a cluster is minimized, while maximizing it between the clusters. Clusters are not intrinsic properties of the set of features under consideration. There is a risk that, instead of finding a natural data structure, we would be imposing an arbitrary or artificial structure, for example, by selecting an unreasonable number of clusters. Therefore, it is inevitable to analyze the distribution of the classes and their separability in feature space.

In this test, we merged the visible-NIR bands (3-10) of the multispectral scanner data by using the well-known ISODATA methods. At the heart of the ISODATA scheme is an updating loop that, using a distance measure, reassigns points to the nearest cluster center, each time the center is moved (Nadler and Smith, 1993). Since the number of different cover-types is scene dependent and usually not known a priori, the dataset was classified several times with increasing the number of classes each time. Because some of the spectral bands are highly correlated, different band combinations were additionally tested. Each classification was compared with the ground truth. Additionally, the separability of classes was analyzed. Different separability measures are described in the literature. To find the best definition is not a trivial task (Schowengerdt, 1997). For our clusterings, the different separability measures (Mahalanobis, divergence, Jeffries-Matusita, etc.) provided very similar results.

We obtained the best clustering results, when using the complete 8-band dataset (see Fig. 3b, 3c). However, when using only 4 bands, selected from different spectral positions, still acceptable results were obtained. Six major cover types were distinguished in the scene (Figure 3b), namely water and roof (black, 1), roof (dark green, 2), vegetation (red, 3 and 4), and roof and bare soil (light gray and white, 5 and 6). Using more classes, for example ten (Figure 3c), some of the classes were split, giving rise to new classes with relatively low separability. Comparing the cluster maps with the aerial photographs reveals that despite the confusion between water and roof pixels, and bare soil and roof pixels, the boundary between man-made surfaces (buildings, walkways, driveways, roads) and vegetated natural surfaces is always recognizable. Note that other boundaries, such as the ones between bare soil and grass, and between vigorous and sparse vegetation, are also present, even though these boundaries are not related to any objects of interest. It is very important to emphasize that no building or roof spectra exist, as it is well known from previous studies. For example, the 6-class clustering classified roof pixels into four different classes with distinctly different spectra throughout the entire range.

To include information about the quality of the clustering in the visual representation, we introduce the concept of weak and strong boundaries. Weak boundaries are located between pixels belonging to classes with low separability; they are of secondary importance. In the 6-class clustering, all boundaries are strong. However, the 10-band clustering rendered 3 weak boundaries, from a total of 45. The use of weak and strong boundaries helps considerably in organizing and simplifying edges.

Fig. 2 a. Visible image and detected edges; **b.** NIR image and detected edges; **c.**Thermal image and detected edges.

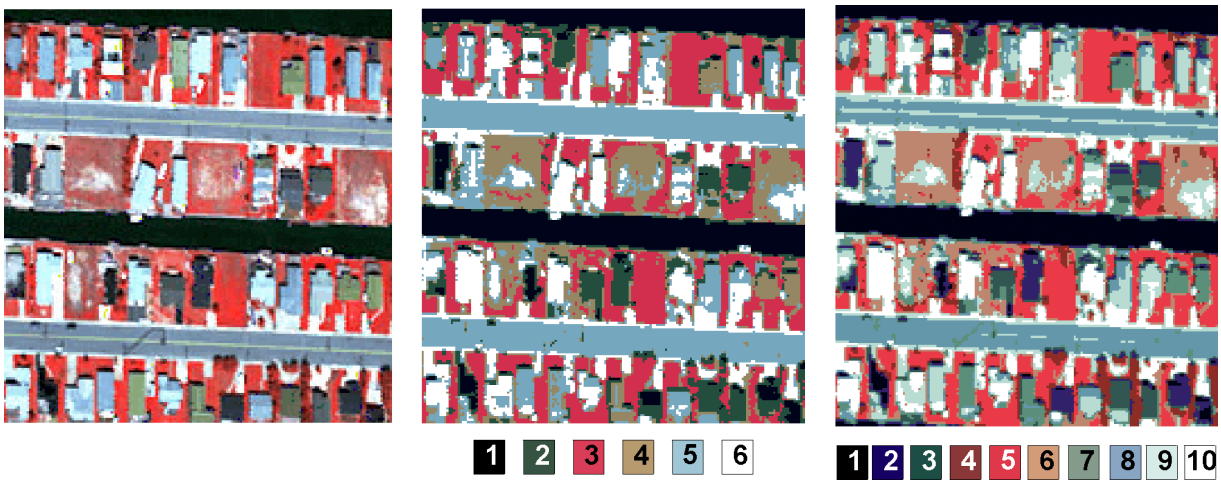

Fig. 3 a. False color composite of NIR (red), red (green) and green (blue) bands; **b**. ISODATA clustering with 6 classes; **c.** ISODATA clustering with 10 classes (weak boundaries between classes 4 and 5, 6 and 7, and 8 and 9). For explanation of the classes see text.

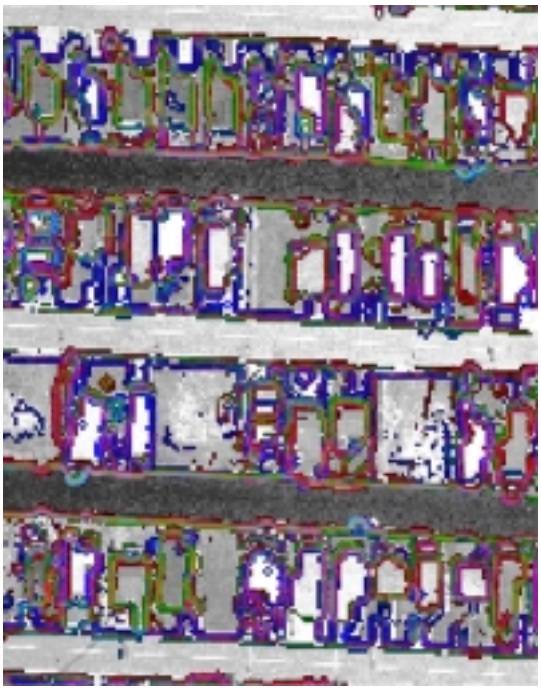

Fig. 4. Edges extracted from the visible (blue), NIR (green) and thermal (red) multispectral images were combined into a color composite and then superimposed on the aerial photograph. The colors indicate which spectral image has the strongest discontinuity along the edge.

4.2. Laser scanning data and aerial imagery

Laser scanning data. Laser scanning systems are increasingly being used in photogrammetry, mainly for generating DEMs. Applications are as diverse as determining the topographic surface and the canopy of forested areas or establishing city models. Usually, the laser system is the only sensor used on the platform. This limits the range of problems that can be solved. More complex applications require several sensors to be used in concert. As briefly described in section 3, the test site in Ocean City includes laser scanner data, aerial, and multipectral imagery.

Laser scanning systems provide a fairly dense set of points on the surface. The accuracy in elevation is about 1 dm and footprint sizes are 1 m or less. The platform orientation system determines the positional accuracy. The critical component is the attitude. While the errors resulting from GPS and ranging are virtually independent of the flying height, the attitude error propagates linearly and thus restricts the flying height. Current airborne laser systems hardly exceed flying heights of 2000 m.

Refined and segmented surface. In our attempt to recognize objects from multisensor data, the information the laser system provides is used for surface reconstruction and generation of hypotheses of man-made objects, such as buildings. It has long been realized that surfaces play an important role in object recognition. The laser points are not directly suitable for representing the surface. For the purpose of object recognition, we need an explicit description of surface properties such as breaklines and surface patches that can be analytically described. We distinguish between the raw, refined, and the segmented surface (Schenk, 1995). The irregularly distributed laser points describe the raw surface. The refined surface includes surface information obtained from aerial imagery. It is the result of fusing features from the two sensors. The fusion process also includes the resolution of conflicts that may exist between the laser surface and the visible surface. The next step is concerned with segmenting the refined surface, resulting in an explicit description that is much more amenable for object recognition than the raw surface, where important surface properties are only implicit.

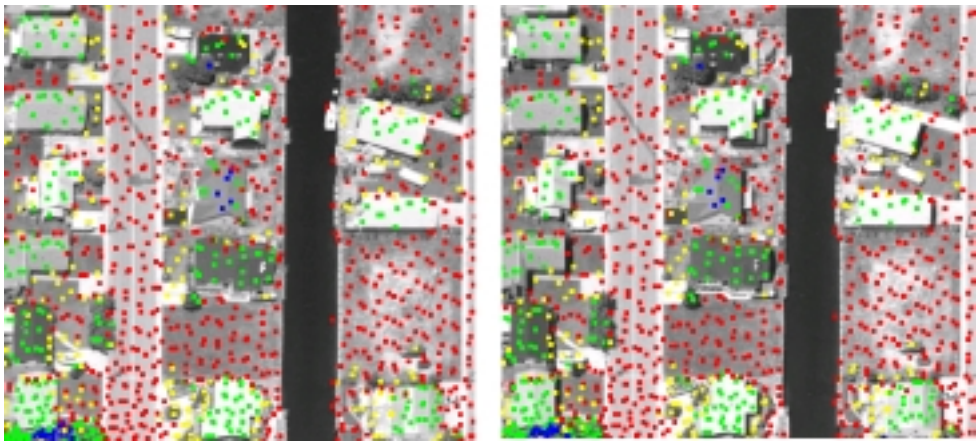

Fig. 5. This stereo pair shows a small image patch from the low altitude flight. Laser points from the same region are projected back to the two images with their exterior orientation parameters. Viewed under a stereoscope, a vivid 3-D scene appears with the laser points on the top of the surface that is obtained by fusing the stereopair. The colored dots indicate different elevations, with red the lowest and blue the highest points. Note that blue points are on top of buildings.

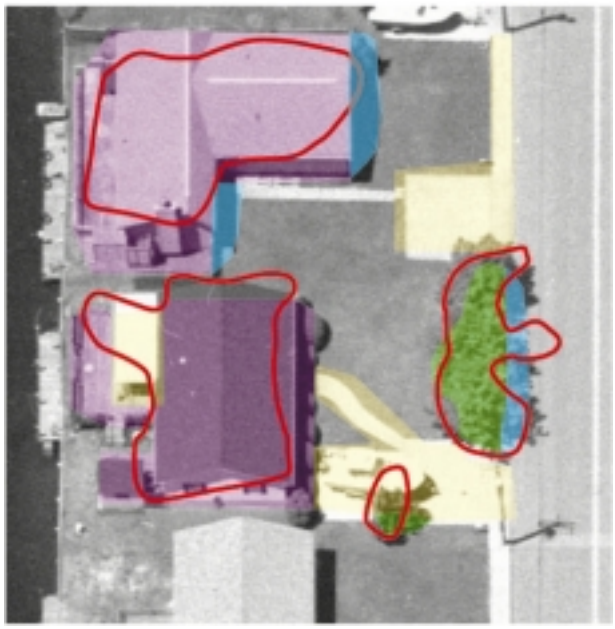

Fig. 6. Superimposed on the aerial image are the results from classifying the multispectral imagery and from segmenting the laser surface. Pink areas indicate dark, non-vegetated regions, and yellow are bright, non-vegetated regions. Green refers to woody vegetation. Finally, blue are shaded areas. The red contours are derived from laser data and they indicate humps. The combination green (from multispectral) and hump (from laser) triggers the hypothesis for a tree, for example. A hump with planar surface patches and a non-vegetated region is used for a building hypothesis.

The fusion of surface information from aerial imagery and laser scanning systems ought to take into account the strengths and weaknesses of the two sensors. The major advantage of laser measurements is the high density and high quality. However, breaklines and formlines must be extracted from irregularly spaced samples. If feature-based matching is employed in aerial stereopairs, we may obtain surface discontinuities directly. This is because edges in aerial images may have been caused by breaklines in object space. Not all edges correspond to breaklines, but there is hardly a breakline that is not manifest as an edge in the image. Fusing surface features is actually a two step process. First, the images covering the same scene are processed using multiple image matching (Krupnik, 1996). Next, the surface obtained during image matching is fused with the laser surface. Obviously, the aerial imagery must be registered to the same object space the laser points are represented in. In turn, this requires an aerial triangulation. To achieve the best fit between the visible and laser surface, the aerial triangulation should be performed by incorporating the laser data (Jaw, 1999).

Figure 5 illustrates the registration of the visible and the laser surface. Here, a small image patch from two overlapping photographs is shown, together with laser points that have been projected back from object space to the images based on their exterior orientation. Viewed under a stereoscope, one gets a vivid 3-D impression of the surface. The figure also demonstrates the distribution of laser points, which is rather random with respect to surface features. For example, features smaller than the (irregular) sampling of laser point may not be captured. Moreover, Figure 6 clearly supports the claim that breaklines, such as roof outlines, should be determined from the aerial imagery.

The refined surface, obtained in a two step fusion process, is now analyzed for humps in an attempt to separate the topographic surface. Then, the difference between the refined and the topographic surface would result in what we call hump-objects of a certain vertical dimension. The prime motivation is to partition the object space, such that the subsequent surface segmentation is only performed in areas identified as humps. The segmentation is a multi-stage grouping process aimed at determining a hierarchy of edges and surface patches. As an example, breaklines are segmented

into straight line segments and arcs (curvilinear segmentation, obtained with the ?-s method), straight lines are grouped to polylines, and polylines with 90 degree vertices are singled out as they play an important role in many man-made objects. Similarly, surface patches within a hump, that can be analytically approximated, are found with various methods. A particularly efficient method is the Hough transform for finding planar surface patches. Here, the parameters of planes defined by the triangles that have been determined by a Delauney triangulation, enter the accumulator array. A cluster analysis of the accumulator array indicates triangles that lie in the same plane. As an optional step, all points found to be in a plane are used in a least-squares adjustment for determining the parameters more precisely. Again, the motivation for this step is related to the observation that man-made objects consist of planes that have certain preferred orientations.

5. CONCLUDING REMARKS

We have described a conceptual multisensor fusion system for the purpose of recognizing objects of urban scenes. The system is characterized by a multi-stage fusion approach that includes laser scanning data, aerial and multispectral imagery. Experimental results confirm the proposed architecture (Figure 6). The experiments are performed with the test dataset of Ocean City. The test site is a rather complex urban area, including residential and commercial areas with buildings of different size and shape, live and dead vegetation, and trees right next to buildings.

In light of an automatic system, we performed unsupervised classification of the multispectral data with the number of classes and the band selection/combination as parameters. The classification results are remarkably robust. That is, the parameter selection does not appear to be very crucial. The results also clearly demonstrate the usefulness of multispectral data for object recognition. Moreover, they confirm that object recognition in complex urban scenes cannot be reliably solved with one sensor only.

Most fusion processes are carried out in the 3-D object space. This requires that for every sensor a relationship between the sensor space and object space is established. Another interesting challenge for fusion is related to the problem that sensors have different resolutions. For example, aerial imagery and multispectral imagery of the test site differ in resolution by a factor of ten. We solve this problem with a scale space approach.

Future research will address the problem of object modeling, taking into account the features that can be extracted from different sensors. By the same token, more research will be devoted to inference processes and identity fusion.

REFERENCES

Abidi, M. A., and R. C. Gonzalez, 1992. *Data Fusion in Robotics and Machine Intelligence*. Academic Press, Inc., San Diego, CA, 546 p.

Csathó, B.M., and T. Schenk, 1998. A multisensor data set of an urban and coastal scene. *Int'l Arch. of Photogr. and Rem. Sensing*, Vol. 32, Part 3/1, pp. 588-592.

Csathó, B.M., W.B. Krabill, J. Lucas, T. Schenk, 1998. A multisensor data set of an urban and coastal scene. *Int'l Arch. of Photogr. and Rem. Sensing*, Vol. 32, Part 3/2, pp. 588-592.

Githuku, A., 1998. A conceptual framework for object recognition and surface reconstruction from high resolution multispectral and hyperspectral imagery. *Int'l Arch. of Photogr. and Rem. Sensing*, Vol. 32, Part 3/1, pp. 452-463.

Hall, D., 1992. *Mathematical Techniques in Multisensor Data Fusion*. Artech House, Boston, London, 301 p.

Jaw, J., 1999. Control Surfaces in Aerial Triangulation. Ph.D. dissertation, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University.

Josephson, J., and S. Josephson, 1994. *Abductive Inference.* Cambridge University Press, 306 p.

Krabill, W. B., R. Thomas, K. Kuivinen, and S. Manizade, 1995. Greenland ice thickness changes measured by laser altimetry. *Geophysical Research Letters*, Vol. 22, No. 17, pp. 2341-2344.

Krupnik, A., 1996. Using theoretical intensity values as unknowns in multiple-patch least-squares matching. Photogrammetric Engineering and Remote Sensing, 62(10):1151-1155.

Nadler, M., and E. Smith, 1993. *Pattern recognition engineering.* John Wiley & Sons, New York, pp. 299-302.

Schenk, T., 1995. A layered abduction model of building recognition. In: Grün A., Kuebler O., Agouris P. (Eds.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, pp. 117-123.

Schowengerdt, R. A., 1997. *Remote Sensing, Models and Methods for Image Porcessing*. Academic Press, San Diego, 522 p.

Shufelt, J. A., 1999. Performance evaluation and analysis of monocular building extraction from aerial imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4), pp. 311-326.