Information Fusion in Cartographic Feature Extraction from Aerial Imagery

David M. McKeown
Frederic P. Perlant
Jeffrey Shufelt

Digital Mapping Laboratory
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

The extraction of buildings from aerial imagery is a complex problem for automated computer vision. It requires locating regions in a scene that possess properties distinguishing them as man-made objects and opposed to naturally occurring terrain features. The building extraction process requires techniques that exploit knowledge about the structure of man-made objects. Techniques do exist that take advantage of this knowledge; various methods use edge-line analysis, shadow analysis, and stereo imagery analysis to produce building hypotheses. It is reasonable, however, to assume that no single detection method will correctly delineate or verify buildings in every scene. As an example, imagine a feature extraction system that relied on analysis of cast shadows to predict building locations in cases where the sun was directly above the scene.

It seems clear that a cooperative-methods paradigm is useful in approaching the building extraction problem. Using this paradigm, each extraction technique provides information which can then be added or assimilated into an overall interpretation of the scene. Thus, our research focus is to explore the development of a computer vision system that integrates the results of various scene analysis techniques into an accurate and robust interpretation of the underlying three-dimensional scene.

This paper briefly describes research results in two areas. First, we describe the problem of building hypothesis fusion using only monocular cues in aerial imagery. Several building extraction techniques are briefly surveyed, including four building extraction, verification, and clustering systems that form the basis for the work described here. A method for fusing the symbolic data generated by these systems is described, and applied to monocular image and stereo image data sets. Evaluation methods for the fusion results are described, and the fusion results are analyzed using these methods.

The second research area examines how estimates of three-dimensional scene structure, as encoded in a scene disparity map, can be improved by the analysis of the original monocular imagery. In some sense this procedure is counter-intuitive. Since we have already used the imagery to perform stereo matching, what information could be available in either of the single images that would improve on the stereo analysis? We describe the utilization of surface illumination information provided by the segmentation of the monocular image into fine surface patches of nearly homogeneous intensity to remove mismatches generated during stereo matching. Such patches are used to guide a statistical analysis of the disparity map based on the assumption that such patches correspond closely with physical surfaces in the scene. This technique is quite independent of whether the initial disparity map was generated by automated area-based or feature-based stereo matching.

1 Introduction

The extraction of significant man-made structures such as buildings and roads from aerial imagery is a complex problem that must be addressed in order to produce a fully automated cartographic feature extraction system. We focus on the building extraction process since buildings are present in almost all sites of cartographic interest and their robust detection and delineation requires techniques that exploit knowledge about man-made structures. There exist a multitude of techniques that take advantage of such knowledge; various methods use edge-line analysis, shadow analysis, stereo disparity analysis, and structural analysis to generate building hypotheses[1, 2, 3, 4, 5, 6, 7, 8].

It is reasonable, however, to assume that no single building extraction technique will perfectly delineate man-made structures in every scene. Consider the use of an edge-analysis method on an image where the ground intensity is similar to the intensity of the roofs of the buildings in the scene. As another example, consider the use of a shadow analysis method on an image in which the sun was directly above the scene.

Clearly, a cooperative-methods paradigm is useful in approaching the building extraction problem. In this paradigm, it is assumed that no single method can provide a completely accurate or complete set of building hypotheses for a scene; each method can, however, provide a subset of the information necessary to produce an improved interpretation of building structure in the scene. For instance, a shadow-based method can provide useful information in situations where ground and roof intensity are similar; an edge-line analysis method can provide disambiguating information in cases where shadows were weak or nonexistent, or in situations where structures were sufficiently short that disparity analysis would not provide useful information. The implicit assumption of this paradigm is that the information produced by each detection technique can be integrated into a more meaningful collection of building hypotheses.

Stereo matching provides a direct measurement of building location and height. In complex urban scenes, stereo matching based upon feature matching, i.e., edges, lines, and contours, appears to provide more accurate and robust matching than area-based techniques. This is primarily due to the ability of feature-based approaches to detect large depth discontinuities found in urban scenes. However, feature-based techniques generally provide only a sparse set of match points from which a three-dimensional surface is usually interpolated. In Section 3 we describe a method to integrate monocular surface intensity information with the stereo disparity map to refine the height estimates and reduce the effect of stereo matching errors.

TIB Hannover

2 Fusion in Monocular Scene Analysis Systems

Our work in man-made feature extraction from monocular views of a scene has used both edge-line intensity based techniques as well as shadow-analysis based techniques. BABE utilizes intensity edges to form corners, which then undergo structural analysis in order to generate plausible building hypotheses. These hypotheses are then evaluated in terms of size and line intensity constraints[3, 4]. Figure 1 shows a typical BABE result for a suburban area in Washington, D.C.

SHADE is a building detection system based on a shadow analysis technique. SHADE utilizes a shadow intensity estimate generated by BABE to produce shadow regions, which are analyzed to locate shadow/building edges. These noisy edges are then smoothed and broken at corners by using an imperfect sequence finder[9]. The line segments that form nearly right-angled corners are joined, and the corners that are concave with respect to the sun are extended into parallelograms. Figure 2 shows a typical SHADE result.

SHAVE is a system for verification of building hypotheses that examines the relationships between these hypotheses and the shadow regions in an image to rank the quality of these building hypotheses. SHAVE determines which segments of each building could cast shadows. Intensity walks are then performed for each pixel of these segments to delineate the cast shadows. Each segment is then scored based on the variance of shadow length along each segment. These scores can then be used to estimate the likelihood that a building hypothesis corresponds to a building, based on the extent to which it casts shadows. Figure 3 shows a typical SHAVE result.

GROUPER is a system that utilizes shadows to cluster fragmented building hypotheses. GROUPER extends the shadow/building edges produced by SHADE along the sun illumination angle to form closed regions of interest in which man-made structures might occur. GROUPER intersects each building hypothesis with these regions of interest. Hypotheses that have sufficient areas of overlap with regions of interest are grouped together to form a composite building cluster.

Figure 1: DC38008 industrial scene (smoothed)

Figure 2: Histogram-splitting segmentation

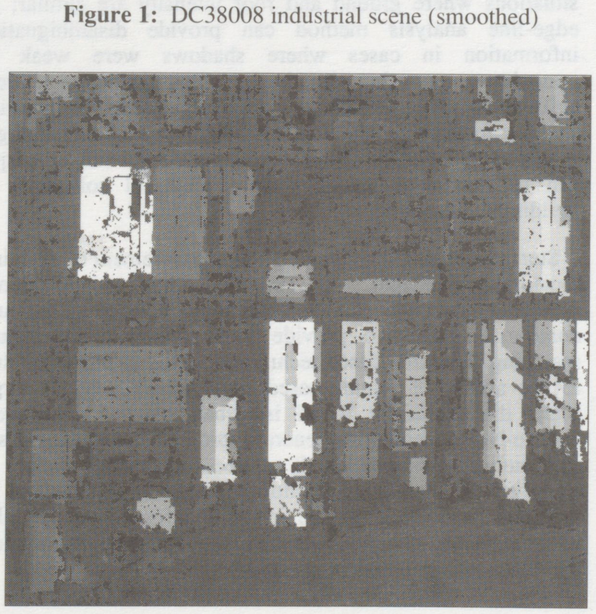

Figure 3: Refined S2 disparity map

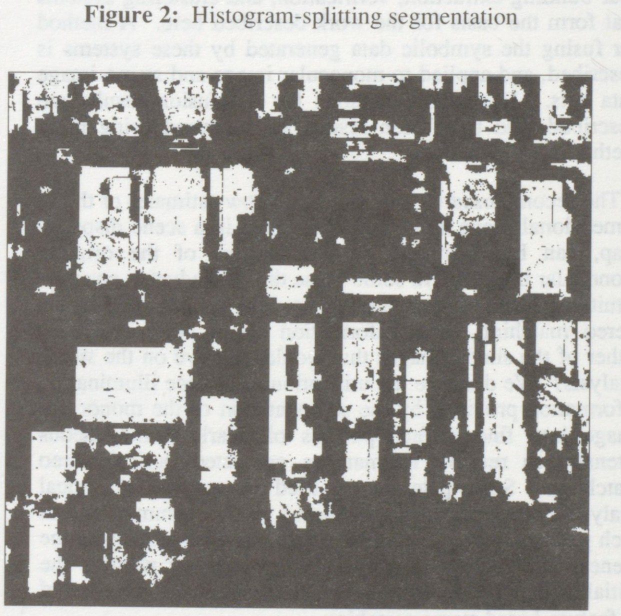

Figure 4: Extracted building regions

TIB Hannover

Recent work has addressed the problem of fusing the information provided by analysis systems such as BABE, SHADE, SHAVE, and GROUPER into more complete and accurate sets of building hypotheses. A monocular fusion system has been developed that merges the geometric building boundaries into composite boundaries. Figure 4 shows the fusion result for the Washington, D.C. scene. Performance analysis indicates that the percentage of building pixels identified correctly in this scene for SHADE, SHAVE and GROUPER is 37.5%, 47.2%, and 48.7%, respectively. The results of monocular fusion, however, improve the overall building pixel classification rate to 77.7%. We have performed fusion analysis over a test database of 8 images and have observed similar performance improvements[10, 11].

2.1 Further Issues in Monocular Fusion

The problems of building structure hypothesis integration are amenable to the techniques of information fusion. Recent work in this area has shown that simple image processing techniques can integrate complementary sources of building hypothesis data to provide improved detection of man-made objects in aerial imagery[11]. These building fusion methods provide a simple and effective means for increasing the building detection rate for a scene. There remain, however, other issues to be addressed in the fusion of monocular scene analysis systems.

The fusion techniques produce qualitatively accurate building delineations, in the sense that few buildings are overlooked in the extraction process; quantitative delineations, however, are typically poor due to the accumulation of delineation errors in the data produced by the monocular analysis subsystems. While applications such as flight simulation are not necessarily adversely affected by this problem, cartographic applications may require very accurate delineations of man-made structures. We would like to address this issue by examining the interactions between data during the information fusion process. For example, during the fusion process, an estimate of "building density" can be produced; that is, an estimate of the likelihood of building structure for each pixel in an image. These likelihood estimates could be used to refine the fused building delineations. In addition, the fusion process provides a composite set of building boundaries. Various subsets of these boundaries could be combined and evaluated with respect to image gradient, shadow casting, and disparity measures to produce improved boundaries. Another approach would use these boundaries as intermediate results to be analyzed by each of the component monocular analysis systems. These systems could refine their initial building estimates in accordance with the fused boundaries, and then these refined estimates could themselves be integrated to produce improved fusions. This idea suggests an iterative process in which building fusions are used to refine their components until some quality measure is exceeded.

If multiple views of a scene are available, either from different vantage points, or from sensors with different spectral characteristics, then more information is available for an information fusion process. Under the cooperative-methods paradigm, monocular analysis techniques could be applied to multiple images to produce initial building detection data, which could then be integrated by information fusion techniques to produce improved building detection rates for the scene. Although multiple sets of data may not always be available, their presence requires the development and/or extension of fusion techniques to take advantage of the additional information such data provides.

3 Improvement and Interpretation of 3D Scene Analysis

The main goal of many vision systems, and in particular automated cartographic feature extraction systems, is to recover the three-dimensional structure of a scene. Stereo analysis is a common technique used for this problem, and much work has been done in this area. Nevertheless, the difficulty of the task (in terms of registration and stereo matching) limits the success of any single technique. To obtain complete and accurate height estimates for a scene, it seems clear that we will need to utilize more knowledge than that used by any stereo matching process. Some work is being done on the interpolation of sparse disparity information to generate reliable and dense height information, and work is being done on post-processing of stereo matching results to ensure consistency of the derived height estimates.

Many three-dimensional estimation improvement approaches use external knowledge about the scene, in the form of models that constrain certain aspects of scene structure. An interesting approach would be the use of information from multiple sources to achieve consistent disparity results, by applying a simple information fusion model. Such an approach allows for the integration of incomplete and inconsistent information to refine height estimates.

Initial work under this paradigm used a region-based interpretation model for the information fusion and refinement process[7]. The assumption of this model is that uniform image radiometry is produced by planar surfaces, of specific orientations and materials. Under this assumption, the segmentation of the monocular images into fine surface patches of nearly homogeneous intensity will ideally result in a segmentation delineating planar surfaces in the scene.

Having obtained segmentations of the images in which regions of nearly homogeneous intensity are delineated, it becomes possible to refine initial height estimates for the scene. Since each region is assumed to correspond to a planar surface in the scene, we can assign disparity values to each region to produce an initial refinement of the height estimates. This is done by histogramming the disparity values of each region and selecting the most representative value for the region. Figure 5 shows a smoothed image of an industrial area in Washington, D.C. Figure 6 shows a segmentation obtained by a recursive histogram splitting technique[12], and Figure 7 shows a refined S2 disparity map[13].

Experiments have shown that such region-based representations of a scene provide useful models for the refinement of height estimates. Using the fusion approach, it becomes possible to segment the scene into building regions based solely on disparity information. (Figure 8 shows the building regions extracted by such an approach). Other sources of information could be utilized at the refinement stage to further enhance disparity data. Left/right consistency constraints could augment the fusion model, as well as more sophisticated models of image radiometry and surface structure. The region-based refinement approach could also be used to refine scene segmentations (such as those produced by feature extraction systems). In summary, the information fusion and refinement paradigm provides a framework for the enhancement of height estimates and allows for the incorporation of possibly inaccurate or inconsistent information in a robust manner.

TIB
TIB Hannover

4 Summary

We have described an information fusion approach to cartographic feature extraction. We have discussed two problems in aerial image interpretation, the fusion of monocular scene analysis data and the improvement of three-dimensional scene interpretations. We have briefly shown the application of information fusion techniques to these problems, as well as results for typical suburban aerial imagery. We have also discussed potential avenues for further research utilizing the information fusion approaches described here.

Cartographic feature extraction from aerial imagery is a difficult problem, which requires analysis techniques that utilize image domain cues as well as *a priori* or contextual information. The analyses produced by such techniques can be complementary; they can also be redundant, or inaccurate. Information fusion methods provide a means for the integration of such data into a more accurate and comprehensive interpretation of the imagery.

5 Bibliography

1. Fua, P., Hanson, A. J., "Resegmentation Using Generic Shape: Locating General Cultural Objects", Tech. report, Artificial Intelligence Center, SRI International, May 1986.

2. Mohan, R., Nevatia, R., "Using Perceptual Organization to Extract 3-D Structures", *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 11, No. 11, November 1989, pp. 1121-1139.

3. Huertas, A. and Nevatia, R., "Detecting Buildings in Aerial Images", *Computer Vision, Graphics, and Image Processing*, Vol. 41, April 1988, pp. 131-152.

4. Nicolin, B., and Gabler, R., "A Knowledge-Based System for the Analysis of Aerial Images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-25, No. 3, May 1987, pp. 317-329.

Figure 5: Suburban BABE results

Figure 6: Suburban SHADE results

Figure 7: Suburban SHAVE results

Figure 8: Suburban fusion results

TIB Hannover

5. R. B. Irvin and D. M. McKeown, ''Methods for exploiting the relationship between buildings and their shadows in aerial imagery'', *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, No. 6, November 1989, pp. 1564-1575.

6. Aviad, Z., McKeown, D. M., Hsieh, Y., ''The Generation of Building Hypotheses From Monocular Views'', Tech. report, Carnegie-Mellon University, 1990, to appear

7. Perlant, F., McKeown, D. M., ''Improved Disparity Map Analysis Through the Fusion of Monocular Image Segmentations'', *Proceedings: IAPR Workshop on Multisource Data Integration in Remote Sensing*, June 1990.

8. Hsieh, Y., Perlant, F., and McKeown, D. M., ''Recovering 3D Information from Complex Aerial Imagery'', *Proceedings: 10th International Conference on Pattern Recognition, Atlantic City, New Jersey*, June 1990, pp. 136-146.

9. Aviad, Z., ''Locating Corners in Noisy Curves by Delineating Imperfect Sequences'', Tech. report CMU-CS-88-199, Carnegie-Mellon University, December 1988.

10. Shufelt, J., McKeown, D. M., ''Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery'', *Proceedings: IAPR Workshop on Multisource Data Integration in Remote Sensing*, June 1990.

11. Shufelt, J., McKeown, D. M., ''Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery'', *Proceedings: DARPA IUS Workshop*, September 1990.

12. Ohlander, R., Price, K., and Reddy, D. R., ''Picture Segmentation Using a Recursive Region Splitting Method'', *Computer Graphics and Image Processing*, Vol. 8, 1978, pp. 313-333.

13. Hsieh, Y., Perlant, F., and McKeown, D. M., ''Recovering 3D Information from Complex Aerial Imagery'', *Proceedings: DARPA IUS Workshop*, September 1990.

TIB

TIB Hannover