A Model-Based Approach for Multi-View Complex Building Description

Z. Kim, A. Huertas, and R. Nevatia
Institute for Robotics and Intelligent Systems, University of Southern California

ABSTRACT: We present an approach to detecting and describing compositions of buildings with flat or complex rooftops by using multiple, overlapping images of the scene. First, 3-D features are generated by using multiple images, and rooftop hypotheses are generated by neighborhood searches on those features. For robust generation of 3-D features, we present a probabilistic approach to address the epipolar alignment problem in line matching. Image-derived unedited elevation data is used to assist feature matching, and to generate rough cues of the presence of 3-D structures. These cues help reduce the search space significantly. Experimental results are shown on some complex buildings.

1 INTRODUCTION

Automatic building modeling is an active research area (Gruen & Nevatia 1998), which can greatly improve the automation of 2-D and 3-D map generation, and be used in many applications including GPS (global positioning system), virtual and augmented reality, military simulations and training, and radio signal reachability tests for wireless communications.

Many of the early building description systems used a single intensity image (Huertas & Nevatia 1988, Irving & McKeown 1989, McGlone & Shufelt 1994, Lin & Nevatia 1998) and were effective for simple buildings only. Recent work has focused on the stereo or multi-view analysis because of the additional information it provides and because such data are widely available. When two or more images are used, 3-D information is typically obtained by matching intensity values or image features. Matching of lines and junctions to generate 3-D descriptions buildings has been common (Roux & McKeown 1994, Noronha & Nevatia 2001, Collins et al. 1998, Fischer et al. 1998). Cord et al. (1999) generated rough building models from high resolution range data and refined using line features. This approach requires accurate range data that may be difficult to obtain. Ameri (2000) extract rough models of complex buildings from LIDAR and refine them assisted by image edges. Baillard & Zisserman (1999) use six or more images to find 3-D matched lines, and find the orientations of half-planes for the 3-D lines by using intensity matching. Refined polygonal meshes are obtained by this method but it does not give explicit building models but just a collection of planar surfaces.

In most building description systems, building models are constructed by extruding polygonal rooftops. The shapes of rooftops vary from simple rectangles to unrestricted polygons; however, as the complexity of rooftops increases, the computation required for rooftop hypotheses generation may grows exponentially. Hence, rectilinear rooftops have been modeled by a collection of rectangular components (Noronha & Nevatia 2001) or by simple blocks with gable rooftops (Noronha 1998) as these simpler models can be derived with a reasonable amount of computation. While collections of rectangular rooftops can represent many rectilinear buildings (Figure 1a) this representation shows a limitation on the detection of even simple rectilinear buildings. An example is shown in Figure 1b. Components a and b of the building are likely to have low evidence support in the image because major parts of the roof boundaries are not

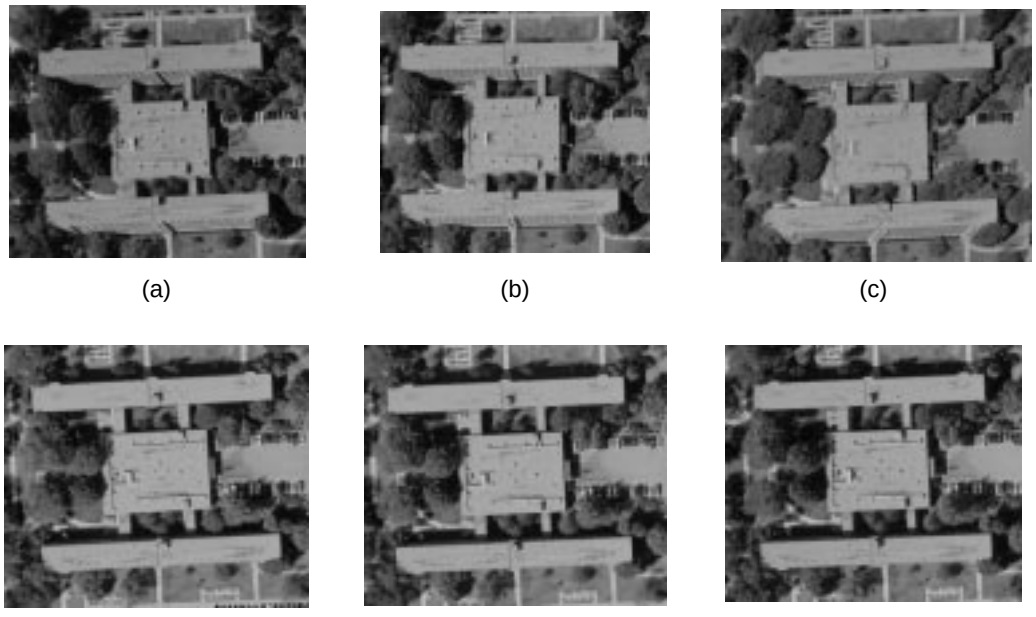
(a) (b)

Figure 1. (a) Rectilinear buildings that can be described as collections of rectangular buildings and (b) a building where collections of rectangles shows a limitation.

present in the image. Component c is unlikely to be even detected because it has even lower evidence support; half of the roof boundary is missing. On the other hand, modeling general rectilinear rooftops imposes large computational demands and/or results in poor detection rates.

We introduce an Automatic Building Extraction and Reconstruction System (ABERS) where the complexity of the building model is increased; ABERS describes buildings with any flat polygonal rooftops or sloping roofs where the outer boundaries of the rooftops (*eaves*) and the *ridges* are parallel to the ground.

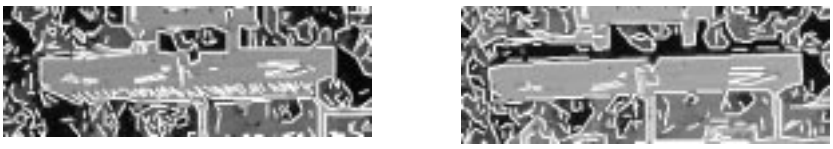
We use multiple (more than 3) images. Example aerial images of a building at 1.0m/pixel resolution are shown in Figure 2. The use of multiple images provides advantages for 3-D object description problems over those of image pairs. They provide alternative views for occluded parts of objects and also help compensate for missing evidence and accidental alignments. An example is illustrated in Figure 3. Figure 3a shows linears extracted from bottom part of Figure 2b. Due to accidental illumination lines are missing for the southern part of the building. Therefore it is hard to describe the building only with the stereo pair of Figures 2a, b. But when multiple views are used, an alternative view (Fig 2c), provides good support for the line missing in two views (Fig 3b). Another instance of "accidental alignment" is *epipolar alignment*. For example, horizontal line segments extracted from Figure 2a, which are important features to detect the building, are aligned with epipolar lines with respect to the images of Figures 2b, c. Therefore, a small position error of a horizontal line segment causes very significant height error when it is matched with a segment from Figures 2b, c. But when it is matched with a line of Figures 2c, d, e the height estimation becomes accurate because the epipolar lines are vertical with respect to those images. Using multiple images also provides robust line matching because we can eliminate accidental wrong line or junction matches by filtering out lines and junctions which have matches in only one or two other views.



(a) (b) (c)

(d)_ (e) (f)

Figure 2. Example aerial images.



(a) (b)

Figure 3. Compensation of missing evidence by using multiple views.

In addition, we use image-derived unedited digital elevation models (DEM's). The DEM's are computed by using correlation methods and, hence, have inherent limitations and produce errors at and near building (and other) depth discontinuities. In contrast with Cord et al. (1999) and Ameri (2000), these DEM's are not accurate enough to retrieve building model directly from them. Research to increase the quality of DEM's using more than 10 images is found in Vestri & Devernay (2000).

2 OVERVIEW

Figure 4 shows a flow diagram of our method. The system uses multiple images and a rough DEM generated by the stereo matching of images. Note that the DEM is used as auxiliary information to provide cues that help reduce the search spaces and validate feature matches. The first step is to extract 2-D line features. Rough building cues generated from the DEM help filter the lines. Then, the lines are grouped into junctions and parallel relationships. Next, we derive 3-D features from groups of matched 2-D features over multiple views to generate rooftop hypotheses. All 2-D features may not be present in all views, therefore, to generate 3-D features, feature matching across pair-wise views is performed first, followed by grouping of the matched pairs. DEM information is used to eliminate some of the false groupings.

The next steps are those of rooftop hypotheses generation, verification and overlap/rooftop analysis. For flat buildings, rough cues for building components generated from the DEM help reduce the search space for the hypotheses generation. The components are arranged in *layers* of various heights. In this case, hypotheses generation, verification, and overlap analysis are repeatedly performed for each layer. In the hypotheses generation step, rooftop boundary hypotheses are generated by neighborhood searches on 3-D features. A level-of-detail technique is applied to reduce the search space. First, coarse-level 3-D features are generated by merging nearby 3-D

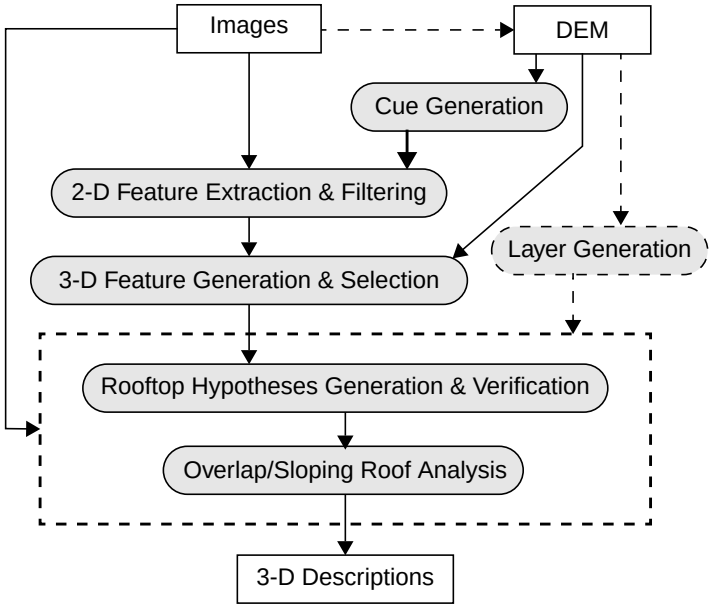


Figure 4. A flow diagram of ABERS.

features; then a neighborhood search is performed on them to generate coarse-level hypotheses. Finally, the coarse hypotheses are refined by a relaxation technique. Evidence support for hypotheses is gathered and used to verify them. *Expandable Bayesian networks* (Kim & Nevatia 2000) are used to combine the diverse evidence. Finally, overlap analysis and sloping roof analysis are performed on the verified hypotheses to give final building descriptions.

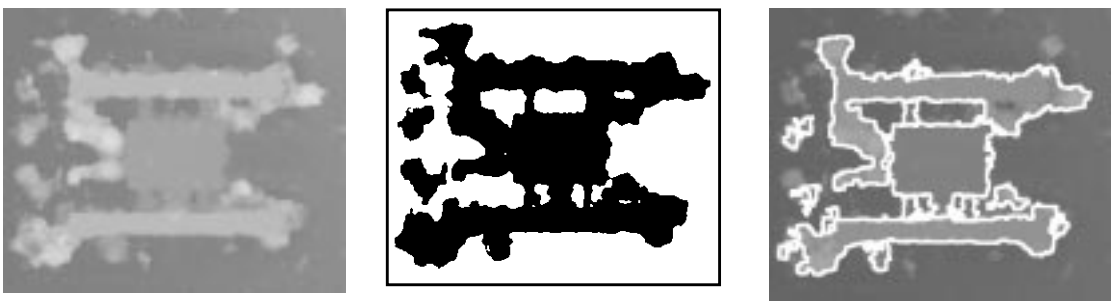
In Section 3, initialization procedures, including the processing of DEM data is described. Section 3.2 describes 2-D feature extraction and filtering. Section 4 deals with the use of multiple images to generate 3-D features. Specifically, we suggest a probabilistic approach to handle the epipolar alignment problem. We also describe the use of DEM information to reduce the matching ambiguity. In Section 5, we describe the rooftop hypotheses generation and verification. Section 6 deals with overlap analysis and the generation of the sloping roof. In Section 7, the time complexity of ABERS is analyzed and the experimental results are shown.

3 PREPROCESSING AND 2-D FEATURE EXTRACTION

3.1 *Preprocessing*

DEM data can be available from stereo image pairs or active sensors such as IFSAR or LIDAR. We focus on the use of image-derived unedited DEM's (of about 1/2 of the image resolution) generated by the commercial "SocetSet" product from BAE Systems Inc. Our approach can be applied to any type of range data but details may need to be changed according to the accuracies of the available DEM's. Figure 5a shows an unedited DEM derived from the images in Figure 2. Although DEM data does not give an explicit building model, it still gives a rough idea of where the buildings are located. We follow the approach of Huertas et al. (1998) to generate rough cues from a DEM image for further processing. The DEM image is first convolved with a Laplacian-of-Gaussian filter to smooth the image and locates the object boundaries. Then the positive-valued regions bounded by the zero-crossings in the convolution output is extracted. Figure 5b shows cues generated from the DEM image of Figure 5a. The detailed usage of these cues is described in Section 3.2. For the buildings of flat rooftops, *layers* are also generated from a DEM image to reduce the search space in hypotheses generation. Layers are defined as flat (parallel to the ground) planar connected surface patches. Layers are generated by segmenting the DEM image. To get layers, adaptive smoothing (Saint-Marc & Medioni 1988) is applied to an image followed by histogram based segmentation. The resulting layers for the DEM image of Figure 5a are shown in Figure 5c.

We need to choose a good combination of five or six images from up to 12 available images since using all the available images results in redundant computation. To compensate for the *epipolar alignment* problem, an image window combination, where the epipolar lines of image pairs are less aligned to each other, should be selected. In ABERS, image windows are automatically selected by such a criteria. First, the epipolar lines at the center of images are calculated for all image pairs. Then, for all the possible image combinations, *epipolar diversities* of the images are calculated. The *epipolar diversity* for the *i*-th image v_i given an image combination, $\boldsymbol{v} = \{v_1, \ldots, v_n\}$ of n images, is determined from the angle differences of the epipolar lines,



(a) (b) (c)

Figure 5. (a) A digital elevation model (DEM) for the example building complex of Figure 2; (b) rough building cues; and (c) *layers*.

which are obtained by sorting epipolar lines by their angles and finding the angle differences between the nearby ones. Once $n-1$ angle differences are obtained, the *epipolar diversity* is determined by the inverse of the standard deviation among them. Finally, the image combination, where the average of the epipolar diversities is biggest, is selected.

The height of the ground is needed to extrude rooftops for building extraction. In ABERS, the ground height is automatically set during the DEM layer generation. The lowest layer with a reasonably large area is considered to be the ground and its height is set to be the ground height.

3.2 *2-D Feature extraction and filtering*

We follow the approach of (Noronha & Nevatia 2001) for 2-D feature extraction. Line segments are extracted from images and grouped into "linear" features by collapsing collinear lines of small gaps. The linear features extracted from Figure 2a are shown in Figure 6a. We see large numbers of distracting linears from trees, roads and other structures nearby. ABERS uses DEM cues (Figure 5) to eliminate many of these distracting linears. First, the linears far from the cues are eliminated. Note that the boundary lines of (polygonal) building rooftops are mostly of two or three dominant angles. In ABERS, a length-weighted histogram of line angles is generated from the linears of each cue and only linears near the *dominant angles* are used for the rooftop boundary hypotheses generation. The *dominant angles* are determined by choosing the two highest peaks in the histogram. The resulting filtered linears are shown in Figure 6b. Figure 6c shows junctions grouped from the filtered linears.

As building designs are largely geometric and have rectilinear boundaries, the junctions among the linears represent strong point features that are well localized. Thus, ABERS uses junctions as an important feature. Junctions are extracted by grouping nearby linears which make angles larger than $60°$. T-junctions are also extracted. A T-junction is considered to be two L-junctions for further processing. Building structures exhibit a great deal of parallelism in their design and construction. We extract parallel relationships of the linears from each view.

(a) (b) (c)

Figure 6. (a) "Linear" features extracted from Figure 2a; (b) linears filtered by locations and orientations;.and (c) 2-D junctions grouped from the linears of (b).

4 3-D FEATURE GROUPING

In ABERS, two types of 3-D features are used; 3-D linears and 3-D junctions. A 3-D linear is a group of matched linears from different views ("member" linears), and a 3-D junction is a group of matched junctions (member junctions). For a 3-D linear, the 3-D height is estimated and a group of 3-D end-points are generated from the 2-D end-points of the member linears. For a 3-D junction, its 3-D position estimate and the branches, which are 3-D linears, are kept. As the rooftop boundary hypotheses generated are restricted to be flat (parallel to the ground), 3-D features are restricted to be parallel to the ground.

3-D features are obtained from pair-wise matches of 2-D features (linears and junctions). We follow the approach of Noronha & Nevatia (2001) for the epipolar matching. A match for a line segment in one view must lie at least partially within a quadrilateral defined by the epipolar

geometry and the 3-D height constraints. A match for a junction must be near to the epipolar lines (within a certain error tolerance), and the corresponding branches need to be matched. The error tolerance for the epipolar matching is automatically determined in the preprocessing stage with epipolar geometry.

4.1 *Combining height estimates of pair-wise matching*

When linears of a match are nearly aligned to the epipolar lines, the estimated height may contain a large error while the error will be smaller when they are perpendicular to the epipolar line. In fact, "accidental epipolar alignments" of important features are not rare in aerial images. In many cases, buildings are aligned with epipolar lines because aerial photos are taken from a flight along a road. The errors in height estimations will depend on the angles of linears and camera parameters. We estimate the height of a 3-D linear from the pair-wise height estimations. However, how to combine these pair-wise estimates of various degrees of errors is not clear. We also face similar problems with junction matches. Although matching junctions does not suffer from the epipolar alignment problem, the accuracy of height estimation depends on the physical distance of the cameras that the images were taken from (baseline length).

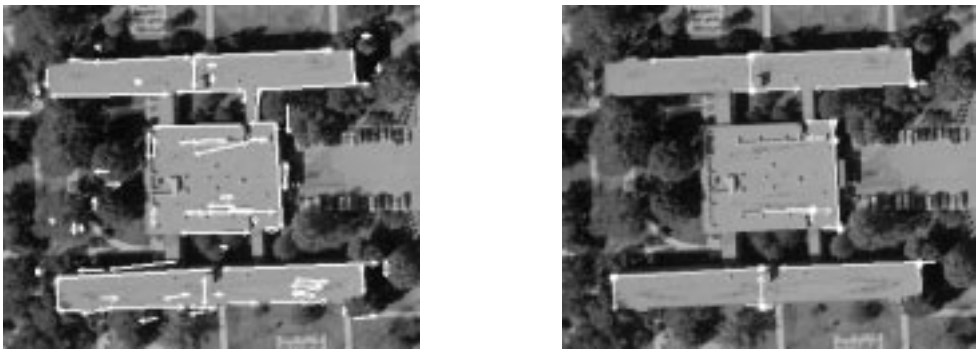
In ABERS, the height estimates are represented as Gaussian random variables. Given a line pair, the height (mean of the Gaussian random variable) is estimated from stereo analysis, and a confidence interval (standard deviation) is obtained assuming possible displacement errors of lines in the image space (by assuming 2.5 pixel location error). Based on this formalism, we estimate the height $\hat{h}$ and confidence interval σ of a 3-D linear from n pair-wise height estimates, $(h_1, \sigma_1), \ldots, (h_n, \sigma_n)$ by the following formula (Kim et al. 2000);

$$\hat{h} = \left(\sum_i \frac{h_i}{\sigma_i^2}\right) \Big/ \left(\sum_i \frac{1}{\sigma_i^2}\right) \text{ and } \sigma^2 = 1 \Big/ \left(\sum_i \frac{1}{\sigma_i^2}\right) \tag{1}$$

We also use probabilistic reasoning for height *compatibility* tests in hypothesis generation. Two height estimates, (h_1, σ_1) and (h_2, σ_2), are considered to come from the same height distribution when $\mu - \sigma \le 0 \le \mu + \sigma$, where $\mu = h_1 - h_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

4.2 *3-D feature generation and selection*

To get a 3-D junction, a matched junction pair of a small height-error-bound is used as a *seed*. Given a seed, junctions are collected from other views which match with at least one of the junctions in the seed and have *compatible heights* (Section 4.1). Similar method is applied to get 3-D linears. We only retain a 3-D linear when it has a member 2-D linear in more than one third of the images. However, still a large number of false groupings exist due to many parallel lines and epipolar alignment. DEM data (Figure 5a) is used to select among the possible groupings. A 3-D linear is projected onto the DEM image. The projected line is used to compute statistics of the DEM values in regions (sampling windows) adjacent and on both sides of the projected line. The height of a sampling window is regarded as a Gaussian random variable with the mean and the standard deviation of the DEM values. A 3-D linear is selected when its height is *compatible* (Section 4.1) with that of a sampling window of either side. Then, a *polarity* is assigned to each 3-D linear. The polarity is defined as positive when its height is the same as that of its left-hand side. When the height of both sides are similar two 3-D linears of opposite polarities are generated. Three-D junctions are selected when both of their branches are selected. The polarity of a 3-D junction is positive when its height is the same as that of its inner-side. Figure 7 shows the 2-D linears and the junctions which are the members of the 3-D selected linears and junctions. Polarities are shown with arrows. The polarity information is used in the rooftop hypothesis generation process described next.



(a) (b)

Figure 7. 2-D features which belong to the selected (a) 3-D linears; and (b) 3-D junctions. Polarities are shown with arrows.

5 HYPOTHESES GENERATION AND VERIFICATION

5.1 *Hypotheses generation*

For flat roof buildings, hypotheses generation, verification and overlap analysis is repeatedly performed for each DEM layer. For each layer, only the 3-D features near the boundary of a given layer with compatible heights (to the layer) are used. Without DEM layers, hypotheses generation, verification and overlap analysis are performed once.

Rooftop boundary hypotheses are generated by a neighborhood search on 3-D features. Even with DEM filtering, we get clouds of 3-D linears per building side which can increase the search complexity exponentially. In ABERS, a level-of-detail technique is applied to reduce these complexity. First, coarse-level 3-D features are generated by grouping nearby 3-D features ("member" 3-D features). 3-D coarse junctions are generated first. A branch of a 3-D coarse junction is defined by the group of branches of the member 3-D junctions. The remaining 3-D linears are grouped with near-by 3-D linears to make 3-D coarse linears.

Then, a neighborhood search is performed on the coarse-level 3-D features to group them into *coarse hypotheses*. Neighborhood relationships are defined by the branch relationships of 3-D coarse junctions and linears. Since polarities are defined for all the 3-D features we only generate counter-clockwise rooftop hypotheses.

The next step is to refine the hypotheses generated from the 3-D coarse features. Given a hypothesis, among the *member* 3-D features of a 3-D coarse feature, the one with the best supporting score is chosen, where the supporting score consists of local scores (intensity contrast of sides for linears, and supporting wall vertical for junctions) and compatibilities with neighborhood features. An iterative relaxation procedure is applied to combine the local scores and the neighborhood compatibilities (Kim 2001).

Grouping only neighbor features does not generate many of the desired hypotheses since some features may be missing in the images. Therefore, we also use parallel relationships; two hypotheses of a compatible height, which have linears that are parallel to each other, are combined once again to make another hypothesis.

The generated and refined hypotheses are either *chains* of 3-D features (usually not closed) or pairs of the chains (when parallel relationships are applied). We need to obtain 3-D rooftop boundaries from them by hypothesizing closed polygons. The 3-D positions of 3-D junctions are used to determine the corner points of the rooftop boundaries. The junctions always have branches (3-D linears) as neighbors. Thus, the chains always end with 3-D linears, and closures are made from them. Figure 8 shows possible alignments of 3-D linears (arrows) and suggested closures. We only consider closures parallel or perpendicular to those 3-D linears. Note that there is more than one pair of end-points for a 3-D linear because it consists of more than one 2-D linear. Closures are generated from all the possible end-points. Selection among these alternatives is based on the support for the closure lines in the 2-D images.

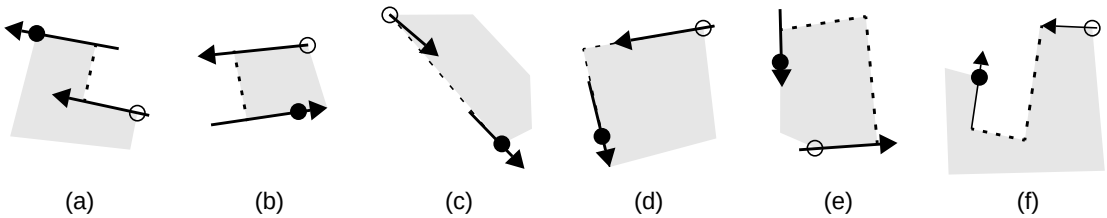


Figure 8. Suggested closures (dashed lines) for (a) parallels with the same polarity; (b) parallels with opposite polarities; (c) collinears; (d) closed L; (e) open L, and; (f) general L. Arrows represent 3-D linears with polarities. Blank circles represent the previous corners and filled circles represent the next corners. Buildings are regarded to lie on the shaded sides.

5.2 *Hypotheses verification*

Once rooftop hypotheses are obtained, supporting evidence is collected for the hypotheses. Evidence support consists of line support, wall vertical line support, darkness of the shadow region, and closeness of a hypothesis to the boundary of a DEM layer. Line support consists of the coverage of supporting line segments (**RP**) and the strength of distracting line segments (**RN**). Wall vertical line support (**WV**) is the coverage of supporting line segments for the possible wall verticals (building corners). We consider only the visible lines; self-occluded lines and lines in the shadow side of a building are not considered. Shadow darkness evidence (**SD**) is the percentage of dark pixels in the possible shadow region. Two evidence variables are generated to measure the closeness of a hypothesis to the corresponding DEM layer. The first one (**DEM1**) is to measure how much the boundary of a rooftop hypothesis is close to that of the DEM layer. ABERS uses a simple estimation which is the maximum of the distances of the rooftop corner points or mid-points of boundary lines to the layer boundary. The second one (**DEM2**) is the area coverage of the hypothesis on the DEM layer. Note that a single DEM layer often times represents more than one building component. **DEM2** may work as negative evidence in such a case. Therefore, **DEM2** is used only for overlap analysis (Section 6); in the overlap analysis, we prefer hypotheses which cover more of the building area.

Note that the computation for some of the evidence, such as the darkness of shadow region (**SD**), is relatively more expensive than others. To reduce this computation, we apply a filtering procedure; in the first stage (*hypotheses selection*) hypotheses with bad roof line support (calculated from **RP** and **RN**) are filtered out, and the rest of the evidence is collected and applied for the remaining hypotheses. Since the number of the image-derived evidence variables varies at runtime according to the number of images, we apply *expandable Bayesian network* (EBN; Kim & Nevatia 2000) to compute P(**Building**|**Evidence**) with varying number of images. A simple expandable Bayesian network (EBN) for the hypotheses selection is shown in Figure 9a and the one using full evidence is shown in Figure 9b. Shaded nodes (boldface) are *repeatable nodes* for the image-derived evidence variables which make their instances at runtime according to the number of images. Additional information, such as the size (area) of the projected rooftop hypotheses on an image (**Size**) and projected length of wall verticals (**WL**), is used. All the continuous evidence variables were discretized into 5 levels, a binary node was used for **Size**, and a ternary node was used for **WL**.

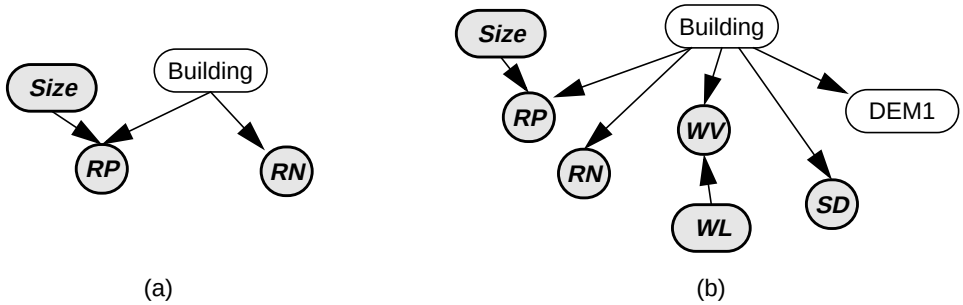
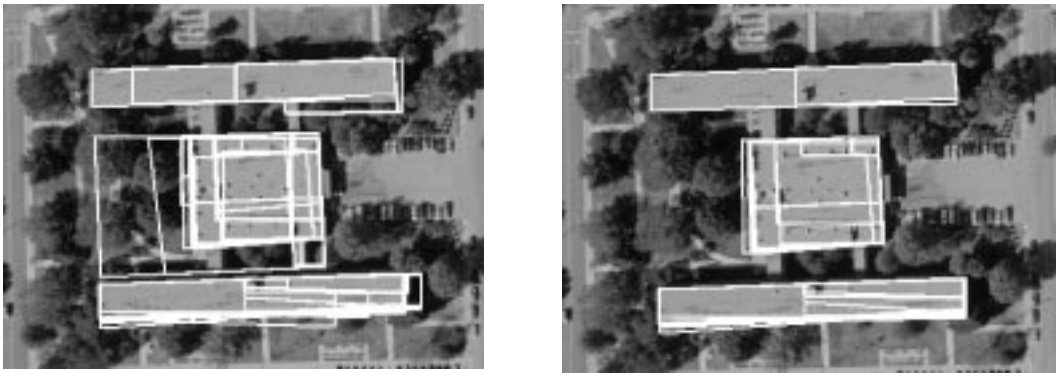


Figure 9. An EBN used for (a) hypotheses selection; and (b) hypotheses verification with full evidence. Shaded nodes (boldface) are *repeatable nodes* for image-derived evidence variables.

The learning dataset for the classifiers was obtained by manually specifying positive and negative building hypotheses. For the hypotheses selection, 283 hypotheses were collected from hypotheses from 6 sites, and, for the verification, 574 were collected from the selected hypotheses of 5 sites. Figure 10a shows all the building hypotheses generated from the image combination of Figure 2, and the verified ones are shown in Figure 10b.



(a) (b)

Figure 10. (a) All the rooftop hypotheses; and (b) the verified ones.

6 OVERLAP AND SLOPING ROOF ANALYSIS

6.1 *Overlap analysis*

It is common that more than one hypothesis is verified for a single building component, where these hypotheses represent parts of an actual building. Therefore, we need to choose the best possible building component. However, comparing two verified hypotheses according to their verification score, $P($**`Building`**$|$**`Evidence`**$)$ of the EBN in Figure 9b, is not an accurate way because that binary classifier is not designed and learned to compare two good building hypotheses but to determine whether a certain hypothesis corresponds to a building or not. Therefore, we need a *comparative classifier,* which takes two sets of evidence variables as input and determines the probability that one hypothesis is better than another.

For *comparative classification*, we use the vector difference between two sets of evidence variables since all the evidence variables are continuous. Another EBN, which takes the vector differences as input, was designed and learned for the overlap analysis. The structure of the *comparative* EBN is similar to the one for the hypothesis verification (Figure 9b) but the use of **`DEM2`**.

The learning dataset for the comparative classifier is obtained by specifying *accurate* and *inaccurate* hypotheses by hand. Then, for all the overlapping pairs of accurate and inaccurate hypotheses, the learning data is obtained by taking the vector difference of their evidence variables. Note that the classification result should be *commutative*; $C(v) = C(-v)$, where v is the vector difference between two sets of evidence comparing, and $C(v)$ is the classification result for v. Therefore, for each evidence pairs, two learning data, v and $-v$, are generated.

An overlap analysis is applied for each layer as well as the rooftop analysis described next. However, overlap analysis is also applied for the final results from all the layers because overlapping hypotheses can also be generated from two different DEM layers.

The final description results for the example building complex is shown in Figure 11. Six images of the 417.2×357.3 average resolution were used. Total time spent was 5 minutes and 48 seconds on a Sun Ultra 2 workstation. The pre-processing time, such as for loading images and generating DEM's, is not included. Although all the rooftops are rectangular, it is hard to detect them correctly without the aid of multiple views because some corners are occluded by trees and some of the important lines are broken by small superstructures or missing because of accidental illumination (see Figure 3).

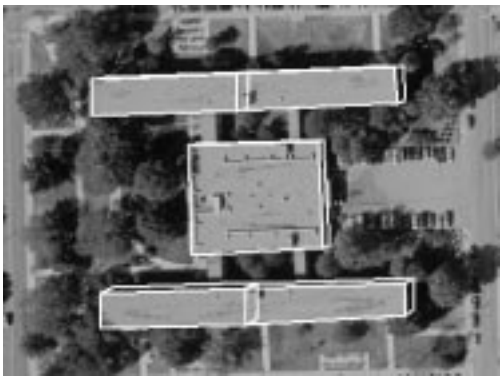
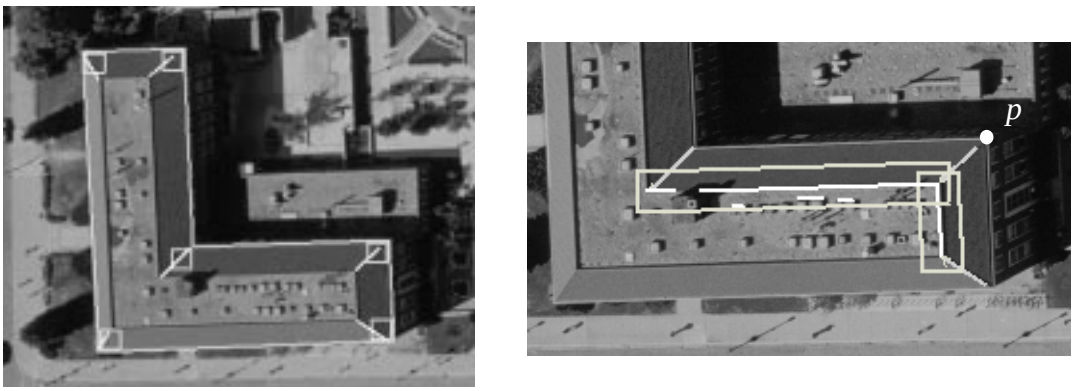


Figure 11. An automatic building detection and description result.

6.2 *Sloping roof analysis*

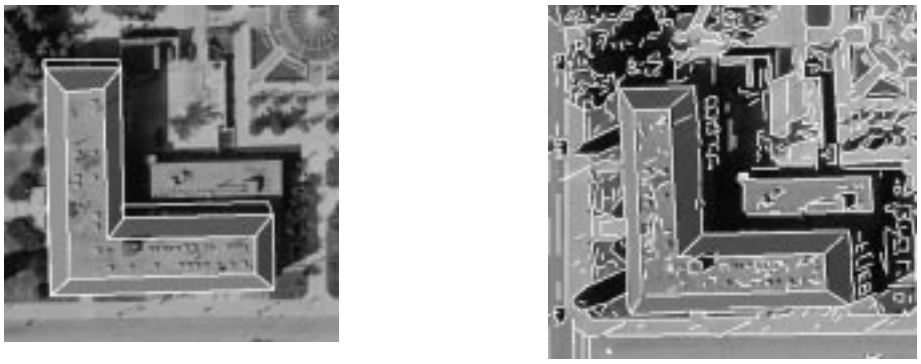
The next step is that of finding interior surfaces of the roof from the detected outer boundaries (eaves). For each corner of the eaves, possible *hips* (diagonal lines from the corners) are gathered. Figure 12a shows the search windows and the hip lines found for an example rooftop boundary hypothesis. Once hip lines are collected from all the images, 3-D linears (*3-D hip linears*) are generated by matching them. When 3-D hip linears are found for most of the corner points, a search is performed to find *ridges* (interior lines parallel to eaves). Example search results for a building corner p are illustrated in Figure 12b. First, 2-D linears parallel to the neighboring building sides are gathered. When 3-D hip linears are available, we can make relatively smaller search windows from the 3-D end-points of the linears as shown in Figure 12b. The ridge candidates are the 3-D linears of plausible heights which have those 2-D linears as their members. The ridge corner point corresponding to the boundary corner point p will be the intersection of a pair of the candidate 3-D linears. For all the candidate pairs, local *line support* (Section 5) and compatibilities with candidate pairs of neighbor corners are calculated, and the best pair for each corner is determined. The ridge corner point is set from these pairs. Note that not all the ridge corner points can be found from the above method. For a missing ridge corner, a corner hypothesis is generated from the neighboring corner points satisfying the parallel relationship. Such a corner hypothesis is verified with the line support. Finally, the rooftop is generated by collecting all the ridges and the hips.

A result on the building with a complex rooftop (Figure 12) is shown in Figure 13a. Five images of the average 406.8×442.8 resolution were used. Total time spent was 4 minutes and 58 seconds on a Sun Ultra 2 workstation. Figure 13b shows the extracted line segments from an image, which shows the complexity of the problem. We see broken boundary lines and large numbers of distracting lines both on the rooftop and the ground.



(a) (b)

Figure 12. Searches for (a) *hips* (bright lines); and (b) ridges. Rectangular search windows are also shown.



(a) (b)

Figure 13. (a) An automatic detection result of a building with a complex rooftop; and (b) the line segments which show the difficulty of the problem.

7 COMPUTATIONAL COMPLEXITY

The time complexity of 3-D feature generation is $O(ln^2)$, where l is the average number of 2-D linears per image, and n is the number of images. The number of 3-D coarse linears is bound by $O(l)$ and the number of coarse hypothesis, before parallel relationships are applied, is $O(lb^k)$, where k is the maximum number of rooftop corners, and b is the *branching factor* of the graphs. Although $b = O(l)$, in the worst case, the actual number of the coarse hypothesis is usually small (smaller than three times the number of coarse junctions, in most cases) because of broken linears and missing junctions. Also, when we limit the size of DEM cues and layers, b^k is bound by $O(1)$. The total number of hypotheses generated are $O(l^2 b^{2k})$ when parallel relationships are applied. However, when we limit the maximum width of a building, it can be reduced to $O(lb^k)$. Hence, the total time complexity of ABERS is $O(ln^2) + O(lb^k)$.

8 CONCLUSIONS AND ACKNOWLEDGEMENT

3-D building detection and description is a difficult problem especially when the target objects have complex shapes. The computational complexity may increase exponentially as the model complexity increases. This paper suggests an approach to combine information from multiple images effectively to control such increasing computation. Experimental results show that the suggested approach is promising. Our system has, so far, been tested on a limited number of examples only. Even though results on theses examples are very good, more extensive testing may reveal limitations and needs for additional mechanisms. More complex roofs may also require addition of other methods. However, we believe that our basic approach can accommodate such extensions.

This research was supported by a subgrant from the Army Research Office MURI grant No. DAAH04-96-1-0444 awarded to Purdue University.

REFERENCES

Ameri, B. 2000. Feature based model verification (FBMV): a new concept for validation in building reconstruction. *Proc. ISPRS Geoinformation for All, 24-35, Amsterdam, the Netherlands, July, 2000.*

Baillard, C. & Zisserman, A. 1999. Automatic reconstruction of piecewise planar models from multiple views. *Proc. IEEE Conf. Computer vision and Pattern Recognition, 2: 559-565*. Los Alamitos, CA: IEEE Computer Society.

Collins, R., Jaynes, C., Cheng, Y.-Q., Wang, X., Stolle, F., Riseman, E. & Hanson, A. 1998. The ascender system: automated site modeling from multiple aerial images. *Computer Vision and Image Understanding* 72(2): 143-162.

Cord, M., Jordan, M, Cocquerez, J.-P., & Paparoditis, N. 1999. Automatic extraction and modelling of urban buildings from high resolution aerial images. *Proc. ISPRS Automatic Extraction of GIS Objects from Digital Imagery, 187-192, Munich, Germany, September, 1999*.

Fischer, A., Kolbe, T., Lang, F., Cremers, A., Förstner, W., Plümer L., & Steinhage, V. 1998. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Computer Vision and Image Understanding* 72(2): 185-203.

Gruen, A. & Nevatia, R. (eds) 1998. *Computer Vision and Image Understanding: Special Issue on Automatic Building Extraction from Aerial Images* 72 (2)

Huertas, A., Kim, Z., & Nevatia, R. 1998. Use of cues from range data for building modeling. *Proc. DARPA Image Understanding Workshop, Monterey, CA, 577-582*.

Huertas, A. & Nevatia, R. 1988. Detecting buildings in aerial images. *Computer Vision, Graphics and Image Processing* 41(2): 131-152.

Irving, R. & McKeown, D. 1989. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Trans. Systems, Man and Cybernetics* 19(6): 1564-1575.

Kim, Z. 2001. *Multi-View 3-D Object Description with Uncertain Reasoning and Machine Learning*. Ph.D. Thesis, Computer Science, University of Southern California.

Kim, Z. & Nevatia, R. 2000. Learning Bayesian networks for diverse and varying numbers of evidence sets. *Proc. Int'l Conf. Machine Learning, 479-486*. San Francisco: Morgan Kaufmann.

Kim, Z., Huertas, A., & Nevatia, R. 2000. Automatic description of complex buildings with multiple images. *Proc. 5th IEEE Workshop on Applications of Computer Vision, 155-162*. Los Alamitos, CA: IEEE Computer Society.

Lin, C. & Nevatia, R. 1998. Building detection and description from a single intensity image. *Computer Vision and Image Understanding* 72(2): 101-121.

McGlone, J.C. & Shufelt, J. 1994. Projective and object space geometry for monocular building extraction. *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 54-61*. Los Alamitos, CA: IEEE Computer Society.

Noronha, S. 1998. *3-D Building Detection and Description from Multiple Intensity Images using Hierarchical Grouping and Matching of Features*. Ph.D. Thesis, Computer Science, University of Southern California.

Noronha, S. & Nevatia, R. 2001. Detection and description of buildings from multiple aerial images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 23 (5): 501-518.

Roux, M. & McKeown, D. 1994. Feature matching for building extraction from multiple views. *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 46-53*. Los Alamitos, CA: IEEE Computer Society.

Saint-Marc, P. & Medioni, G. 1988. Adaptive smoothing for feature extraction. *Proc. DARPA Image Understanding Workshop, 1100-1113*.

Vestri, C. & Devernay, F. 2000. Improving correlation-based DEMs by image warping and facade correlation. *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1: 438-443*. Los Alamitos, CA: IEEE Computer Society.