Conference Paper

Use of DTMs/DSMs and orthoimages to support building extraction

Author(s):
Baltsavias, Emmanuel P.; Mason, Scott; Stallmann, Dirk

Publication Date:
1995

Permanent Link:
https://doi.org/10.3929/ethz-a-004333550

Rights / License:
In Copyright - Non-Commercial Use Permitted
Use of DTMs/DSMs and Orthoimages to Support Building Extraction

Emmanuel Baltsavias, Scott Mason and Dirk Stallmann
Institute of Geodesy and Photogrammetry
Swiss Federal Institute of Technology (ETH)
8093 Zurich, Switzerland
{manos, mason, dirk}@p.igp.ethz.ch

1. Introduction
The acquisition of 3D models of buildings and other man-made objects is currently an issue of high importance to many users of geoinformation, including planners, geographers, architects, etc. Aerial imagery has proven to be a valuable data source for these models. The project AMOBE, a joint research effort between the photogrammetry and image sciences groups at ETH, aims firstly at developing practical algorithms to support the semi-automatic reconstruction of man-made objects from aerial imagery and secondly, at developing improved techniques for automatic digital terrain and surface model generation. In the latter case it is important to differentiate between terrain models (DTMs), which model the terrain, and surface models (DSMs), which model all 3D objects. In this paper, we explore the roles of both DTMs and DSMs and their derived products orthoimages in supporting the extraction of buildings from aerial imagery. In Section 3 the quality of commercially available automatic DTM/DSM generation software is investigated. In Section 4 a number of methods for automatically detecting buildings in DSMs are presented and evaluated. In Section 5, the use of DSMs in deriving coarse building models is described. Applications of orthoimages are discussed in Section 6. First, however, the fundamental assumptions and strategy employed in AMOBE is outlined. Note finally that the techniques and tests described in this paper constitute preliminary investigations. Promising directions, as noted, are the subject of ongoing research.

2. Project AMOBE
Primary attention in AMOBE is focussed on the extraction of buildings as one of the more predominant and frequently occurring 3D man-made structures. The employed imagery is assumed to be digitised photogrammetric photography of the type typically used for topographic mapping. In this nadir imagery primarily building roofs, and not walls,

1. Support for this research by ETH under project 13-1993-4 is gratefully acknowledged. The project is entitled “Automation of Digital Terrain Model Generation and Man-Made Object Extraction from Aerial Images” (AMOBE).
can be extracted. Given assumptions, e.g. from the context, or estimates of the overhang from terrestrial field measurements, the full volume structure of a building can be inferred from the reconstructed roof. The main features of the strategy we are following are illustrated in Figure 1 and include:

- Automatic generation of a DSM exploiting all overlapping images.
- The DSM and the extracted 2D features should be exploited to automatically derive a coarse building model that can be employed in the following steps of feature matching and model-based building reconstruction.
- As 3D object reconstruction can best proceed using 3D features, 3D information should be derived as early as possible by multi-image matching of 2D image features, e.g. straight lines, exploiting their attributes and graph relations obtained from low level segmentation (cf. Henricsson, 1995). The DSM is exploited along with the epipolar constraint for the determination of an approximate search space.
- The matching step produces a “cloud” of 3D linear features. Generic object models are then used to structure the 3D linear features into roof structures. The form of these models is currently being investigated. The analysis of the DSM blob for each building (see Section 5) may provide for automatic generic model selection. Information derived from 2D grouping (cf. Henricsson, 1995) will be exploited in the matching process.
- Due to the high complexity and range of building (roof) shapes, generic models will be needed to support the reconstruction process. Our concept, as shown in Figure 2, is to decompose a building’s roof into 3D surface primitives, e.g. planar entities such as rectangles, trapezoids, and triangles. Models of these primitives are then used to automatically extract the roof components which, when combined, constitute the reconstructed roof. For simple buildings, parametric volume models may be applicable.
- It is presumed that operator interaction will be required to support automatic extraction procedures. The type and level of this interaction will depend on the quality of the results produced by the automated steps.

Figure 1: Strategy employed in AMOBE for building extraction from aerial imagery.
Figure 1 clearly shows that DSMs play an important role at all stages of the building extraction strategy. In the following sections, this role is described and demonstrated. The results are derived from the “Avenches” dataset¹ (Mason et al., 1994). This dataset consists of a residential and industrial scene and fulfills the aforementioned imagery assumptions: large image scale, color imagery, four-way image overlap, precise orientation data and high resolution scanning (15µm).

3. DSM Generation

The purpose in DSM generation is the derivation of a complete 3D model of the visible surface with the highest possible accuracy and taking into account terrain discontinuities. State-of-the-art automatic algorithms based on stereo imagery are not able to distinguish between the terrain surface and objects on and above this surface, e.g. man-made structures such as buildings or trees. Their output is consequently a 2.5D DSM as opposed to the generally desired DTM. For building extraction from aerial imagery, however, DSMs are well-suited offering a number of possible uses:

- Approximate detection of buildings. These results can be used to guide 2D image feature extraction, thereby speeding up processing time, support 2D feature grouping, and reduce the number of candidates in feature matching.
- Separation of building from other non-DTM objects, e.g. trees.
- Approximations for feature matching.
- Selection of a model or reduction of hypotheses in model-based building reconstruction (see Section 5), support of 3D grouping.
- Orthoimage generation (see Section 6).
- Derivation of coarse building models (e.g. building outline and maximum height) as a final product.

---

¹. This dataset is available via ftp by contacting the authors.
3.1. DSM Generation by Stereo Correlation

In this section we report on the quality of DSM generation using the commercial state-of-the-art approach supported by the Helava DPW 770. This software employs hierarchical, stereo-based correlation. A number of strategies are offered amounting to parameter selections appropriate to different topographic classes and accuracy requirements. These selections are experience-based. We investigated different strategies as well as the sensitivity of the results with respect to the choice of the Ground Sample Distance (GSD), the sampling density of the DSM in object space. For the residential scene the strategy “steep_1” for high accuracy extraction under steep terrain conditions (here, the buildings and trees) offered the best results.

Figure 3 shows that, even though it will not guarantee close modelling of a building, the choice of a small GSD is a requirement; a coarse GSD can, in isolated cases, lead to the building being poorly modelled or not at all.

![Figure 3: DSM results for varying ground sample distances: (a) image of the building; (b) 0.1m; (c) 0.25m; (d) 1.0m; (e) and (f) illustrate the result of merging multiple DSMs derived from pairwise matching of six images (GSD 0.25m).](image)

For the dataset, a GSD of 0.25m was determined to be a good comprise between modelling accuracy and processing requirements (cf. Table 1). Figure 3 illustrates the “worst
case” result achieved by the Helava DPW 770, i.e. with a GSD of 1m the 12 x 10 x 6 m house was not modelled. All other buildings in the dataset were modelled for GSD = 1m.

Table 1: DSM generation by stereo correlation

<table>
<thead>
<tr>
<th>Test</th>
<th>GSD (m)</th>
<th>Numbera</th>
<th>Timeb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>49216</td>
<td>23.2</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>7925</td>
<td>3.3</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>505</td>
<td>1.0</td>
</tr>
</tbody>
</table>

a. Number of acceptably matched points
b. Processing time relative to Test 3

3.2. Alternatives

The results above demonstrate some of the limitations of image matching procedures for DSM generation including:

• The measurement density may not be sufficient to model discontinuities such as buildings, buildings close to each other etc. Buildings may even totally disappear.
• Discontinuities are smoothed.
• Many matching errors occur around buildings due to occlusions, trees and shadows.

We are currently developing a multi-image approach based on the Multiphoto Geometrically Constrained matching (MPGC) (Baltsavias, 1991). In this approach, the points to be matched are selected in one image by an interest operator, basically a direction sensitive edge extractor, using a variable-size window. We believe that the inclusion of points on the discontinuities in the matching will lead to better modelling. Results will be reported in due course. Sensors such airborne laser range scanners and, to a lesser extent, interferometric SAR are also expected to provide viable alternatives to and/or support for DSM generation by image-based techniques.

3.3. Merging DSMs from Multiple Stereopairs

Inherent occlusions associated with 3D objects in stereo matching may, to some degree, be reduced by a multi-image approach. Moreover, it may be anticipated that the accuracy of the resultant DSM would be improved by this approach. An alternative to simultaneous multi-image approaches, such as in MPGC, is to merge the results of multiple correlations¹. For example, if four-way stereo overlap is available, a total of six separate DSMs can be produced. We conducted an experiment in which six such DSMs were automatically generated using the Helava DPW 770 for the residential scene in the “Avenches dataset”. The DSMs were then merged on the basis of a “figure of merit” (FOM) priority. The FOM is an estimate of the quality of each correlated point and therefore of the resultant height. Figure 3e and Figure 3f illustrates the results of merging the

1. This approach is clearly suboptimal in terms of efficiency and the statistical testing afforded by the simultaneous approach.
six DSMs. Clearly the modelling of the house remains unsatisfactory, because the FOM is not a reliable measure of the matching accuracy.

4. Use of DSMs in Building Detection

Because a DSM ideally models the man-made objects as well as the terrain, it may be applied to automatically detect, for example, buildings in the scene and thus, by projecting the positions of the detected buildings into the images, provide initial windows for building extraction schemes. A number of difficulties immediately arise: (1) buildings will be at best modelled as smooth “lumps” by matching approaches and thus will appear similar to other 3D objects such as trees, small mounds, etc.; (2) closely situated 3D objects will often be “melted” together; (3) in residential areas, trees are often found close to buildings; (4) in some cases, the matching process will fail to model a building (see Figure 3). In particular due to (4), the results of detection should not be relied upon but used as a starting point.

Three schemes for building detection from DSMs are presented in Figure 4. The middle scheme is based on applying the morphological operation “opening” to smooth the DSM. Subtracting the processed DSM from the original DSM results in the extraction of the smoothed peaks. By thresholding this new “image” according to the expected building height range for the scene, the individual peaks can be simply segmented, e.g., using a component labelling algorithm. Shape descriptors, e.g. the compactness, area, etc., can be applied to judge whether or not an individual component is hypothesised as a building. Limits for these descriptors can be obtained from the context, e.g. residential buildings have an area greater than 50 m². There are clearly a number of weaknesses in this approach. First, it is sensitive to the choice of structuring element size, particularly in dense urban areas. Second, the opening operation shows no discretion when a building is closely situated to other DSM “lumps”, e.g. steep terrain changes such as retaining walls.
might be extracted with the building leading to gross errors in the shape descriptors. Figure 5 illustrates some of these problems.

Figure 5: Results of morphological building detection: (a) DSM generated by DPW 770 for the residential dataset; (b) and (c) illustrate the effect of varying the structuring element size from 3 m to 5 m on the hypothesised building components.

A more promising approach to building detection is to segment the DSM with a 3D edge detector (left-hand scheme in Figure 4). Clearly, shape descriptors will still need to be used to form hypotheses but the problematic choice of the structuring element is avoided. Neighbouring “lumps” in the DSM will still be discernible, as illustrated in Figure 6a. This approach is being investigated further.

If a DTM is available, and for many developed countries this is the case, a subtraction of the DTM from the DSM (right-hand scheme in Figure 4) can provide a useful clue in building detection. The quality of the detection results using this method (see Figure 6d) may be a little optimistic, as in this case the DTM was derived to a higher accuracy and at a higher density than that of the Swiss\(^1\) national 25m DTM (DHM25).

Figure 6: (a) segmentation of DSM with a range image edge operator (Boulanger, 1994); (b) DTM; (c) DTM subtracted from the DSM; (d) building detection based on DSM - DTM.

An alternative to the above building blob detection methods is the use of multiple height bins (MHB). In the MHB method the DSM heights are grouped into consecutive bins (height ranges) of a certain size. This results in segmentation of the DSM in relatively few regions that are always closed and are easy to extract (see Figure 7). The method is

---

\(^1\) The DTM for the dataset was measured as a 10m grid to an accuracy of about 20 cm. The Swiss DHM25 has a 25m grid spacing and an accuracy of 2-4 meters for this region.
applied hierarchically whereby the large bins are used to detect possible buildings, and the small ones to verify and refine the coarse detection, find the approximate building model and separate buildings that are close to each other (see Section 5). The MHB method is simple and fast and can be applied globally or locally. The maximum and minimum bin sizes are determined from the known height accuracy of the DSM (e.g. 0.5 - 1 m) and the estimated minimum building height in the image (e.g. 3 - 4 m). The method can be also combined with the 3D edge extractor (see Figure 6a) to help bridge contour gaps and provide information in the interior of the contours, i.e. building roofs.

4.1. Blob classification: seeing the buildings from the trees (and other objects)

The blobs detected in a DSM can be attributed to many different 3D objects. To reduce the false alarm rate in building detection we need to be able to separate those belonging to buildings and those belonging to other objects, e.g. trees. Here we assume that context knowledge, such as shape bounds (height, size, elongation, etc.) on buildings in a given area, has already been applied to preclassify the blobs and rely now on information in the images. A simple procedure for distinguishing buildings from trees is illustrated in Figure 8. A histogram of the gradient orientations in the range [0°, 180°] at all pixels belonging to strong gradients in an image region associated with a DSM blob is computed.
Histograms belonging to buildings will contain significant peaks 90° apart (for regularly shaped buildings), with additional peaks (usually one or two) being detected for more angular buildings. On the contrary, histograms belonging to trees are flat. Additional classification cues that can be used to distinguish buildings from trees and other 3D objects include the number and length of long straight lines, texture, colour including infrared for vegetation segmentation, motion and context.

An important product of classifying the DSM blobs is the opportunity to generate a DTM. Those blobs which belong to non-terrain features can be subtracted from the DSM, and in the blob regions heights can be interpolated from the surrounding terrain or in case of forests the DSM heights can be reduced by an estimation of the tree heights, to derive a DTM.

5. Estimation of Approximate Building Model

An analysis of the blobs detected in a DSM can, when the DSM is dense and relatively accurate, provide in combination with the 2D extracted features an approximation of the building model. As an example, small elongated regions in the middle of a DSM blob (see top right in Figure 7) that include long straight lines in the image form a very strong hypothesis that there is a gable on the roof. The approximate model can then be used to (i) support grouping of 2D features, and (ii) automatically select the appropriate model or restrict the possible hypotheses in the 3D reconstruction stage. The aims of the analysis for approximate building detection may be to:

- Determine approximate building size (height, length, width).
• Distinguish between flat and peaked roofs. In the latter case, distinguish whether the roof plane(s) intersect at a highest point or at a gable.
• Distinguish between rectangular, T, L, U and + shaped buildings.
• Determine the number and form of the major roof planes.
• Detect buildings that are close together (see top right in Figure 7).
• Exclude small disturbances on roofs (skylights, chimneys).
• Detect smaller buildings on building roofs and group them.

The above analysis makes use of certain properties (e.g. symmetry) and rules (e.g. a gable is the intersection of either two roof planes or one roof plane and a wall).

In certain applications, e.g. telecommunications, flight simulators etc., only a coarse building model, e.g. building boundary and maximum height, is required. This information can be provided by the above analysis. Another attractive alternative is to automatically derive the outline of the buildings from digitised maps (cf. Carosio and Nebiker, 1995) and the height from the DSM. The latter is accomplished by estimating the maximum building height using the nearest DSM blob to the projected building outline, as shown in Figure 9. This procedure can lead to rapid establishment of 3D building databases, a very important practical application.

Figure 9: Coarse building modelling. Building outline digitised from a 1: 25000 map overlaid on an orthoimage from a DTM (left) and a DSM (right).

6. Orthoimages and Orthorectified Stereo Images

Orthoimages are potentially useful in building extraction:

• Geocoded information such as DTMs, DSMs, roads or utility lines from, e.g. a GIS or a map, can be directly combined with the orthoimage.

• 3D measurements can be made by detecting corresponding points in orthorectified stereo images. These 3D measurements are correct even if the DTM or DSM used for the orthoimage generation is totally wrong (Baltsavias, 1993). Matching by use of multiple orthoimages and geometric constraints is possible. Matching in orthoimages is easier due to good approximations and small geometric distortions.

There are, however, a number of problems with their use. First, buildings in an orthoimage produced using a DSM will be deformed due to DSM errors. These deformations
cause problems in the extraction of straight linear features and matching between orthorectified stereo images (see Figure 10). In orthoimages produced from a DTM the buildings lie on the ground and are partly out of the detected DSM blobs. However, straight lines remain straight and can be extracted and matched. Orthorectified stereo images generated from a DTM can also be used for building reconstruction by detecting corresponding characteristic roof points in a semi-automatic approach.

Figure 10: Orthophotos generated using (a) DTM, (b) DTM and 3D building model, and (c) DSM.

Ideally, 3D non-terrain objects could also be detected by differencing orthorectified stereo images produced by a DTM. However, there will be no differences in the overlapping areas of the non-DTM objects, and there will be differences due to radiometric differences of the images and geometric misregistration (see Figure 11). The latter problem can be avoided by subtracting images in a higher pyramid level. More useful is the subtraction of orthorectified stereo images produced by a DSM (see Figure 12b) for quality control of the DSM and manual or automated corrections. In regions of large differences either DSM errors or radiometric differences will exist. This is a very simple and practical way of controlling the DSM quality in the absence of ground truth.

Figure 11: Left: differencing orthorectified stereo images. Not all differences can be attributed to non-DTM objects. Right: differences due to a hair (noise) on one of the scanned films.
7. CONCLUSIONS

DSMs provide a very useful source of information in extracting buildings from aerial imagery. They can be employed to automatically detect buildings as blobs and to provide approximations in matching of features in overlapping images. The detected buildings blobs and the extracted 2D features, in turn, can be analysed to hypothesize a coarse model of the building. This model can be used in 2D and 3D feature grouping and 3D roof reconstruction. Alternatively, for certain applications the coarse building may be the final product. Clearly these methods require that the DSM models the buildings to a sufficient extent. To this end it has been shown that in automatically generating DSMs state-of-the-art matching procedures require a dense sampling rate. It has also been shown that the quality of a DSM can be assessed by subtracting orthorectified stereo images produced with it. Finally, by classifying and removing non-terrain objects from a DSM, a DTM may be generated.

8. REFERENCES


Figure 12: Differences of orthorectified stereo images: (a) using DTM, (b) using DSM