FRAMEWORK TO AUTOMATICALLY CHARACTERIZE REAL PROPERTY USING HIGH RESOLUTION AERIAL IMAGES

Philipp Meixner, Franz Leberl
Institute for Computer Graphics and Vision
Graz University of Technology 2009
{meixner, leberl}@icg.tugraz.at

ABSTRACT

Accurate and realistic 3-dimensional models of urban environments are increasingly important for applications like virtual tourism, city planning, internet search and many emerging opportunities in the context of "ambient intelligence". Applications like Bing-Maps or Google Earth are offering virtual models of many major urban areas worldwide. While initially, these data sets support visualization they are inherently capable of addressing a broader purpose. On the horizon are urban models that consist of semantically interpreted objects; an urban 3D visualization will be computer generated, with a fundamental advantage: the urban models can be searched based on object classes.

This paper presents a framework which specifies the processing steps that are necessary for a reasonable semantic interpretation and evaluation of real property using high resolution aerial images.

We first describe the different source data which have to be brought into a common coordinate system. In this process we build an integrated geometry and semantic object data set that can be analyzed for various purposes. Our focus is on characterizing individual properties and to determine the size of buildings, their number of floors, status of vegetation, roof shapes with chimneys and sky lights etc.

We start out by merging the aerial imagery with the cadastral information to define each property as a separate entity for further analysis. The cadastral data may also contain preliminary information about a building footprint.

In the next step the building footprints get refined vis-à-vis the mere cadastral prediction, using an image classification and the definition of roof lines. 3-D façade coordinates are computed from aerial image segments, the cadastral information and the DTM. This helps to determine the number of floors and the window locations to see if there are attics and basement windows.

The paper uses an experimental data set of the City of Graz (Austria) with 216 buildings representing 321 separate properties. We present the results of the characterization with as much information as possible being extracted about each property and related to manually collected ground truth.

Key words: Aerial Photogrammetry, Digital Aerial Images, Semantic image interpretation, window detection, floor detection, real property characterization

INTRODUCTION

For the last couple of years more and more data about the earth has been collected to create 2- dimensional maps and also a 3-dimensional urban model of the earth. The driving force for this rapid development was the internet-search. "Internet maps" in this context consist of the street-maps used for car navigation, augmented by addresses, furthermore the terrain shape in the form of the Bald Earth and all this being accompanied by photographic texture from orthophotos. The so called location-aware Internet (Leberl, 2007) has evolved since about 2005 and companies like Microsoft (www.bing.com/maps) and Google (maps.google.com) as well as many regional websites are offering their services.

3 dimensional "urban models" have been a topic of academic research since the early 1990's (Gruber, 1997). This has evolved into a massive and systematic effort to map buildings in 3D to support a certain location-awareness in Internet-searches. In November 2006 Microsoft announced the availability of a 3D Internet mapping program (Leberl, 2007). The vertical man-made buildings are modeled as triangulated point clouds and get visually embellished by photographic texture. Since April 2008, vegetation is being classified and identified, and computer-generated vegetation is being placed on top of the Bald Earth. Figure 1 shows an example of the 3D mapping program by the Microsoft-web site Bing/Maps.

Figure 1. Typical 3D content in support of an Internet search. Capitol in Denver (Microsoft's Bing-Maps).

While Internet-search may be the most visible and also initial driving application, there of course are others like city planning, virtual tourism, disaster preparedness, military or police training and decision making or car navigation.

The 3D-data representing the so-called location awareness of the Internet serve to please the user's eye – one could speak of "eye candy" -- but cannot be used as part of the search itself. This is unlike the 2D content with its street map and address codes that can be searched. An interpreted urban 3D model would support searches in the geometry data, not just in the alphanumeric data. One may be interested in questions involving intelligent geometry data. Questions might address the number of buildings with more than 5 floors in a certain district, or properties with a built-up floor area larger than 200 m², with impervious areas in excess of 30% of the land area, or with a window surface in excess of a certain minimum.

Such requirements lead towards the interpretation of the image contents and represent a challenge for computer vision (Kluckner, Bischof, 2009). While currently driven by "search", applications like Bing-Maps or Google Earth have a deeper justification in light of the emerging opportunities created by the Internet-of-Things and Ambient Intelligence. These have a need for location awareness (O'Reilly & Batelle, 2008).

PROCESSING FRAMEWORK

We start with geometric data from essentially two sources: the aerial imagery and the cadastral information. Figure 3 shows an example for a 400 m x 400 m urban test area in the city of Graz (Austria). We merge these two data sources to define each property as a separate entity for further analysis. The cadastral data may also contain preliminary information about a 2D building footprint.

In a next step we produce dense 3D point clouds associated with the aerial photography and extracted from it by means of a so-called dense matcher applied to the triangulated aerial photographs (Klaus, 2007) and we extract the data per building. This gives us the areas occupied by a building as well as its height. The building footprints get refined vis-à-vis the cadastral prediction using image segmentation and classification to define roof lines.

From the building one proceeds to the facades: building footprints become façade baselines. For each façade we must find its intersection with the ground, thus its footprint. This footprint is the basis for an extraction of the façade in 3D by intersecting it with the elevation data. We compute the corner points of each façade. These can then be projected into the block of overlapping aerial photography. We can search in all aerial images for the best representation of the façade details; typically this will be the image with the maximum number of pixels for a given façade. Since there are multiple façade images, we prepare for a multi-view process.

What follows is a search for rows and columns of windows in the redundant photographic imagery. First of all, this serves to determine the number of floors. Second, we also are interested in the window locations themselves. And finally, we want to take a look at attics and basement windows to understand whether there is an attic or basement. Figure 3 summarizes the workflow towards a property characterization and represents the framework in which the effort is executed.

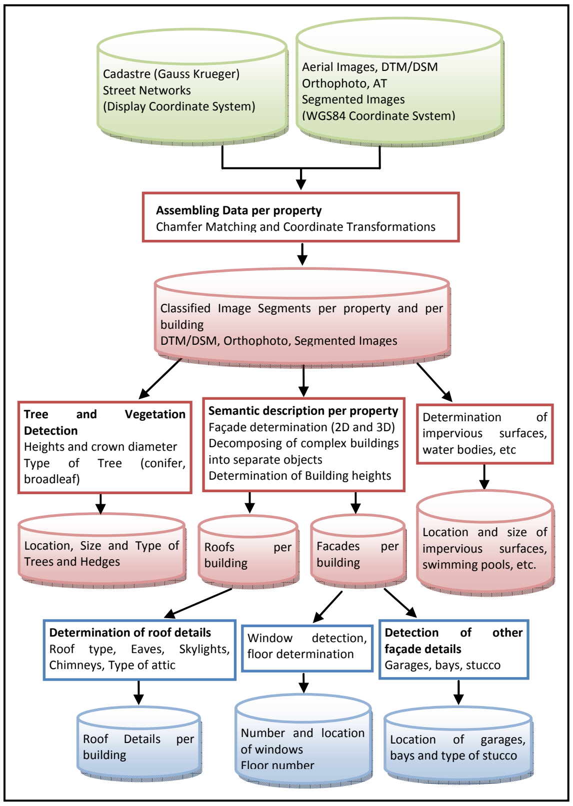

Figure 2. Diagram of the proposed work flow to characterize real properties from aerial images and associated cadastral data.

SOURCE DATA

Figure 3 is an orthophoto of a segment of the City of Graz and covering 400 m x 400 m with a pixel size of 10 cm and image overlaps in the range of 80% forward and 60% sideward (Scholz and Gruber, 2009). A point on the ground will thus be imaged 10 times and the orthophoto will not have to have any occluded regions. Both a traditional orthophoto with relief displacements of vertical objects such as buildings and trees is a common product, and increasingly the true orthophoto is as well since the ability to avoid occlusions is essential in this case, and the novel high overlaps ensure that such occlusions get eliminated. However, in order to produce a true orthophoto at a good quality, one needs a Digital Surface Model DSM with well-defined building roof lines to avoid "ragged" building edges. A high-quality DSM requires a 3D capability at an accuracy level that is not needed for traditional orthophotos.

Associated data are computed from the aerial images. They consist firstly of the results of the aerial triangulation with their pose information per image.

Secondly, we have available the dense point cloud of the DSM and its filtered Bald Earth DTM. It may be remarkable that the DSM is computed at an interval of the elevation postings at only 2 pixels. Traditional photogrammetry had postulated a distance between elevation postings as a multiple of the height accuracy. That horizontal spacing was recommended

Figure 3. A 400 m x 400 m segment of an orthophoto of the urban core of the city of Graz (Austria). The pixel size is at 10 cm. The orthophoto is of the type "true"; therefore the facades are not visible.

to be in the range of perhaps 20 times the elevation error. If one were to assume an elevation error of ± 1 pixel, then the postings were to be spaced 20 pixels apart. However, these recommendations were based on 2-image stereo. This is now changing to a 10-image multi-view geometry (Hartley, Zisserman, 2000), and thus to a concept of "super-resolution", as if the pixel sizes were in effect much smaller than they actually are. The result is a much denser DSM than was ever computed previously (Klaus, 2007). This leads to well-defined horizontal edge information such as building roof lines.

The third type of derived information is the image classification into roofs, gras, vegetation, water bodies and circulation spaces such as roads, parking spaces, driveways and other impervious surfaces.

DATA PER PROPERTY

The image classification result is in the same coordinate system as the orthophoto. Therefore the cadastral map can be used directly to cut a classification map into data per property. Figure 4 illustrates the result.

Zebedin et al. (2006) deliver an accuracy of 90%. This is consistent with the current effort's conclusion. A source for discrepancies between cadastre and image is seen where the cadastral boundary line coincides with a building façade. One observes the existence of façade details such as balconies, or roof extensions in the form of eaves. Having the cadastre available offers one the option of changing the segmentation and classification.

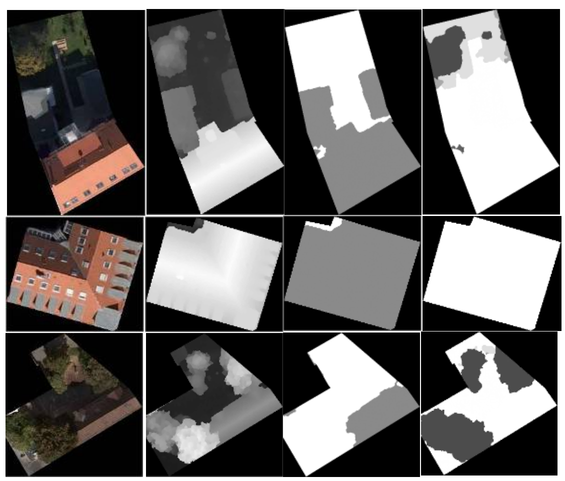

Figure 4. Three separate sample properties and the source data per property.

DESCRIPTION OF A PROPERTY

Several descriptions have become available as a byproduct of conflating the 2D cadastral data with the 2D imagery. We have not only defined the properties, but also the areas used up by the various object classes such as building, vegetation, water bodies or impervious surfaces. These measurements of surface area have previously been determined to be available at an accuracy of 90%.

However, we have yet to introduce into the work the 3rd dimension in the form of the dense point cloud. This will add the most relevant information

These considerations create the need for methods to automatically improve the alignment of the cadastral line work and the segmentation boundaries. Until such algorithms get developed and implemented, we perform such improvements by hand. Figure 5 illustrates the discrepancies and their removal.

The overriding role is associated with the buildings, and these are in the initial focus of the effort. All the work being applied is per property.

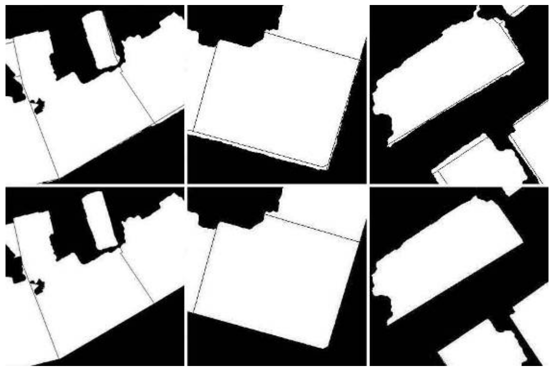

Figure 5. Overlay of segmented image and cadastre for areas. Above is with the discrepancies due to roof eaves and façade detail, below is a manually cleaned-up version.

Facades Footprint 2D

Vectorizing the Building Contour: The building objects obtained from the image classification are an approximation of the intersection of a façade with the ground. One needs to isolate the contour of each building object in a given property. Initially, this contour is in the form of pixels in need of a vectorization. This is a well developed capability, one therefore has a choice of approaches. The Douglas-Peucker algorithm (Douglas, Peucker, 1973) is being used here. The goal is to replace the contour pixels by straight lines, each line defining a façade.

Vectorizing the Points along the Vertical Elements in the DSM: Separately, the 3D point cloud found for a building object also is a source for façades. Passing over the X-rows and Y-columns of the point cloud, one finds the building outline from the first derivative of the z-values – they represent the tangent to the point cloud and where this is vertical, a façade is present.

Reconciling the Segmentation Contour with the DSM Facade Points: The façade footprints from the image classification are based on color and texture and need to be reconciled with the footprint based on the 3D point cloud. One approach is to define the mean between the two largely independent measures.

A Property Boundary Cutting though two Connected Buildings: In the special case where a property boundary cuts through a building or a pair of connected buildings, one does not have a façade. Such cases need to be recognized. An approach is the use of the 3rd dimension, as shown below. The output of this step is a series of straight line segments representing multiple facades.

Decomposing a Building into Separate Building Objects: The option exists to fit into the pattern of façade footprints a series of predefined shapes of (rectangular) building footprints. In the process one hopes to develop a set of separate non-overlapping basic building objects. The 3rd dimension is being considered via roof shapes. Having more than one local maximum in the roof height is an indication that the single building should be segmented into multiple building objects.

Façades in the 3rd Dimension

Along the footprints of the façade one finds elevation values in the DSM. These do attach to the façade a 3rd dimension. Depending on the shape of the roof, a façade could have a complex shape as well. However, for use as a descriptor one might be satisfied with a single elevation value for each façade. We have now defined a vertical rectangle for each façade footprint.

A refinement would consist of a consideration of the change of elevations along the façade footprint. This could be indicative of a sloping ground, or of a varying roof line, or a combination of both. The slope of the ground is known from the DTM. The variations of the roof line are read off the difference between the DSM and the DTM.

The issue of connected buildings along a property line exists. One needs to identify such façade footprints since they are virtual only. Such facades can be identified via a look at the dense point cloud. The elevation values above the Bald Earth along a façade footprint will be zero at one side of the footprint. If they are not, then buildings are connected and this façade is only virtual.

Image Texture per Façade

The definition of the façade quadrilaterals produces 4 façade corner points in 3D object coordinates. These must be projected into each of the aerial images to associate image content to each façade. Typically, many aerial images will show the texture of each façade. Figure 6 is an example for one of the separate facades one building. The projection is based on the pose values of each image from the aerial triangulation.

Figure 6. Of one single façade of one building will obtain multiple aerial image segments. These have been rectified into a façade coordinate system. From an aerial image block showing for each object point typically 10 images, not all will contain useful data for a specific vertical façade. Selected here are the 4 best, where "best" is defined as the largest area of a façade quadrilateral in the projection into an image.

Floors

From the building's appearance, floors get defined by windows. In turn, windows form a defining structure in describing a façade's detail. A procedure for finding a floor count has been developed using the following steps.

For each façade i of a building j, repeat:

Import all n image segments showing this façade i.

- For each image segment repeat:
- Transform the segment into the façade coordinate system.
- Apply a contrast enhancement.
- Apply the Prewitt edge detection horizontally.
- Apply the Prewitt edge detection vertically.
- Convert the maximum horizontal and vertical edge values into a binary format.
- Create for each image row, and image column, a summation of all pixel values, resulting in a vertical and horizontal edge profile.
- From the summation, remove outliers, normalize the values and remove low values as "noise".
- Determine the number of maxima of the sums of vertical gradients and use this as the number of floors.
- Perform a verification by eliminating floors that do not have the proper vertical spacing (minimum distance between floors); and removal values from along the edges of the image texture inside the façade quadrilateral.

This approach will result in data as illustrated in Figure 7.

Figure 7. Binary Prewitt edges in (a) are vertical, in (b) horizontal. The sums of edge values are shown in (c) as a count of the number of floors.

A floor count can be applied to each of a set of overlapping façade images. If there were a discrepancy in the result, some logic would have to get applied to resolve the ambiguity.

Windows

Window detection has been of some interest in recent years. Algorithms like "boosting" have been applied by Nguyen et al. (2007) to detect cars and windows in aerial images. Czech and Sara (2007) have developed a window detection based on a library of window shapes. Lee and Nevatia (2004) have based their approach on edge images.

These approaches have been subjected to only limited experimental analysis, but are generally reported to find windows in a rather robust manner.

Given our floor counts, we are reusing the intermediate Prewitt edges to also find the windows. An approach that simply "intersects" the locations along the pixel rows and columns with the maximum edge sums will work if all windows are regularly arranged. While this is often the case, it is not always true. Therefore Lee and Nevatia (2004) have proposed a variation of the approach.

To refine the locations of the windows a one dimensional search for the four sides of a window is performed. For every line of a window hypothesized lines are generated by moving the lines to its perpendicular direction. The refined positions of the windows are determined where the hypothesized line has the best score for the window boundary. For a more detailed description of the used algorithm read Lee and Nevatia (2004).

The big advantage of this method is that one can also use images with lower resolution, and that not only rectangular windows but almost all window designs can be automatically detected rather quickly without training the program in advance.

The window count is applicable in each image segment of a given façade, separately. Or one might want to merge the edge data sets and apply a single window detection to the sum of all edges. Initial tests have shown that the window count is a rather robust method that delivers no discrepancies between the separate images of one façade in the examples chosen thus far.

A comparison of the various different methods for window detection should be performed and will be the subject of ongoing work.

Multiple Facades per Building

The redundancy not only applies to the image coverage per façade from the high overlaps of aerial photography. We also find that we have multiple measures for the number of floors from multiple facades. These must be consistent with one another. It is possible that a building has different floor counts on a sloping terrain. Since the "bald Earth" as well as the slope of a building footprint are known, they must enter into the floor count.

Figure 8 illustrates the floor counts and detected windows in each façade of that one building. As one can easily determine, the automated floor count and the count of the windows is consistent with a visual inspection.

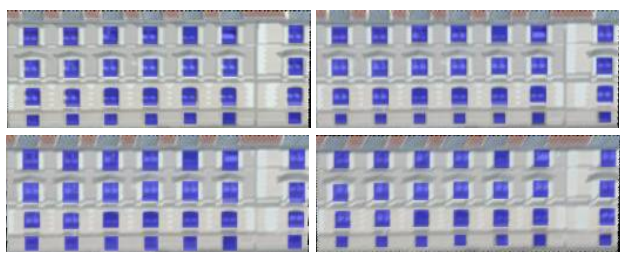

Figure 8. Four facades of one building lead to independent floor counts and window counts. It is to noted that the floor counts and the number of windows coincide with the visual inspection.

We have extended this exercise to a selection of 150 properties in the Graz demo data set. In those properties we have identified 102 buildings with a total of 225 facades. The total number of floors was 387, the number of all windows was 2646. Running the approach through this data set results in the following:

Building detection error: 100%, all 102 building were found

Error of the floor count: 90% of the 387 floor were correctly counted

Error of the window count: 87.1% of the 2646 windows were correctly counted

CONCLUSION

The search for a description of individual buildings per property is but an element in a larger effort. The development of as detailed a description of real properties will have the buildings as the most important element, but other features of a property are also in need of a description. One will want to consider the land, the vegetation, the impervious surfaces, even the interaction between properties casting shadows or affecting privacy. And one will also be interested in the traffic, distances to businesses or public transportation etc. A full system for property descriptions will involve business addresses, traffic information, street network information, as well as sun angles.

In the current contribution we have focused on basic descriptions of buildings. This involves the definition of a building on a property, even if two buildings are connected along a property line. It deals with complex buildings having many facades and a complex roof-scape. From the outside, thus from aerial imagery, one can count the floors and windows, and identify the window areas on a façade for further analysis. At this stage of research we are beginning with the experimental evaluation of the various approaches. We will have to cope with occlusions from vegetation, with ambiguities regarding garages and sheds, the difficulties arising from an inability of matching parcel maps with aerial imagery, and with ambiguities from basement and attic windows.

Initial results are encouraging. Using 150 properties with 102 buildings having 387 facades and 2646 windows, 90% of all floors and 87.1% of all windows were found automatically. The result addresses, however, a specific situation in a mature core area of Graz (Austria). Reasons for misclassifications regarding floors and windows result from inaccuracies of the DTM, occlusions from vegetation and other buildings, partial shadows on the facades, very complex facades and steep camera angles.

The results will need to be confirmed by increasing the sample data in one city, and then by looking at vastly different environments such as a coastal resort environments, historical small towns, alpine terrains and industrial zones.

REFERENCES

Douglas, D., T. Peucker, 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *The Canadian Cartographer,* pp. 112-122.

Čech J., R. Šára, 2007. Windowpane detection based on maximum a-posteriori labeling, Technical Report TR-CMP-2007-10, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic.

Hartley, R, A. Zisserman, 2000. *Multiple View Geometry for Computer Vision*, Cambridge University Press, 1st Edition.

Klaus, A., 2007. Object Reconstruction from Image Sequences, Dissertation, Graz Univ. of Technology, 1997.

Kluckner S., H. Bischof, 2009. Semantic Classification by Covariance Descriptors within a Randomized Forest, *Proceedings of the IEEE International Conference on Computer Vision, Workshop on 3D Representation for Recognition,* (3dRR-09).

Leberl, F., 2007. Die automatische Photogrammtrie für das Microsoft Virtual Earth, *Internationale Geodätische Woche Obergurgl.* Chesi/Weinold(Hrsg.), Wichmann-Heidelberg- Publishers, pp. 200-208.

Lee, S.C., R. Nevatia, 2004. Extraction and Integration of Window in a 3D Building Model from Ground View Images, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* CVP'04.

Nguyen, T., H. Grabner, B. Gruber, H. Bischof, 2007. On-line Boosting for Car Detection from Aerial Images, *Proceedings of the IEEE International Conference on Research, Innovation and Vision for the Future,* (RIVF'07), pages 87-95.

O'Reilly, T., J. Batelle, 2009. *Web Squared: Web 2.0 Five Years On*, O'Reilly Media Inc., Available from www.web2summit.com.

Scholz, S., M. Gruber, 2009. Radiometric and Geometric Quality Aspects of the Large Format Aerial Camera UltraCam Xp., *Proceedings of the ISPRS*, Hannover *Workshop 2009 on High-Resolution Earth Imaging for Geospatial Information*, XXXVIII-1-4-7/W5, ISSN 1682-1777.

Zebedin, L., A. Klaus, B. Gruber-Geymayer, K. Karner, 2006. Towards 3D map generation from digital aerial images, *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 60, Issue 6, September 2006, Pages 413-427.