Improved reconstruction and rendering of cities and terrains based on multispectral digital aerial images

Konrad KARNER, Gerd HESINA, Stefan MAIERHOFER, Robert F. TOBLER

(VRVis Research Center, Donau-City Str. 1/3, A-1220, Wien)

1. ABSTRACT

In this paper we describe the 3D modeling and interactive visualization of 3D city models generated from multispectral digital aerial images. The proposed 3d modeling approach integrates spectral classification to obtain a semantic description of the reconstructed scene. The workflow consists of several consecutive steps, namely an initial land use classification (LUC), the aerial triangulation (AT), a dense matching process, a refined LUC using the DSM to fuse redundant information from all input images and a texture extraction step.

Based on the available, reconstructed geometry data optimized rendering methods are employed in the subsequent visualization stage. Due to the reconstruction process, most data is available in a 2½ D format, i.e. height values are available for a grid of locations. In order to render this type of data, a chunked level-of-detail approach is employed, based on the work Ulrich. By extending Ulrich' level of detail method to run multiple instances of the algorithm in parallel we make it possible to visualize arbitrary ranges and layers of data at arbitrary precision.

Together, the proposed reconstruction and visualization methods result in a very fast, and economical pipeline for visualizing large parts of a city or landscape within a very limited timeframe. The resulting real-time visualization can be used in a number of urban applications that need up-to-date information on changes in the vegetation or building suituation of a city.

2. INTRODUCTION

The traditional workflow for reconstructing urban geometry is a multi-step process that is costly both in time and money. Speeding up the process and automating the workflow will therefore automatically result in a cheaper more viable method for generating urban models. In order to achieve this goal, we use images from an aerial camera to perform an improved reconstruction of urban geometry by using specialized Land Use Classification methods (LUC).

For rendering terrain data a number of algorithms have been published, the most recent ones being the one by Lasasso and Hoppe [2004], Asirvatham and Hoppe [2005] and Ulrich [2005]. These algorithms are based on chunked levels of detail, i.e. the geometry for levels of detail is not switched on a per-triangle basis, but based on large chunks of triangles. The speed of current graphics hardware makes this approach significantly more effective than previous more fine grained approaches. In order to render the reconstructed geometry in an optimized manner, we extended the chunked-level-of detail approach to handle large areas of data and added additional layers to cope with data that is not arranged in a 2½ D format.

The following sections describe our improved methods for reconstruction and interactive visualization.

3. 3D CITY MODELING FROM DIGITAL AERIAL IMAGES

Digital aerial cameras change the way of 3D city modeling completely. The availability of high redundant aerial images at almost no additional cost results in a paradigm shift where a transition from minimizing the number of film photos due to human operator intensive processing to maximizing the robustness of automation due to this high redundant image information takes place.

In our fully digital aerial workflow we use images from UltraCam-D cameras which deliver 16 bit pan-sharpend RGB-NIR images with a size of 11500 x 7500 pixels. The camera is able to deliver images at intervals down to 1 sec.

Our workflow includes the following steps: a classification of all images, the aerial triangulation (AT) using area and feature based point of interests, a dense matching to generate a dense digital surface model (DSM), a 'true' orthophoto production, a refined classification using the DSM, and the estimation of roof and façade polygons.

3.1 Land Use Classification (LUC)

The initial land use classification is a supervised classification performed on each of the overlapping color images with 4 color channels RGB and NIR. Our classification approach is based on support vector machines (SVM). Figure 1 shows the RGB image and the corresponding LUC for a small area on the ground.

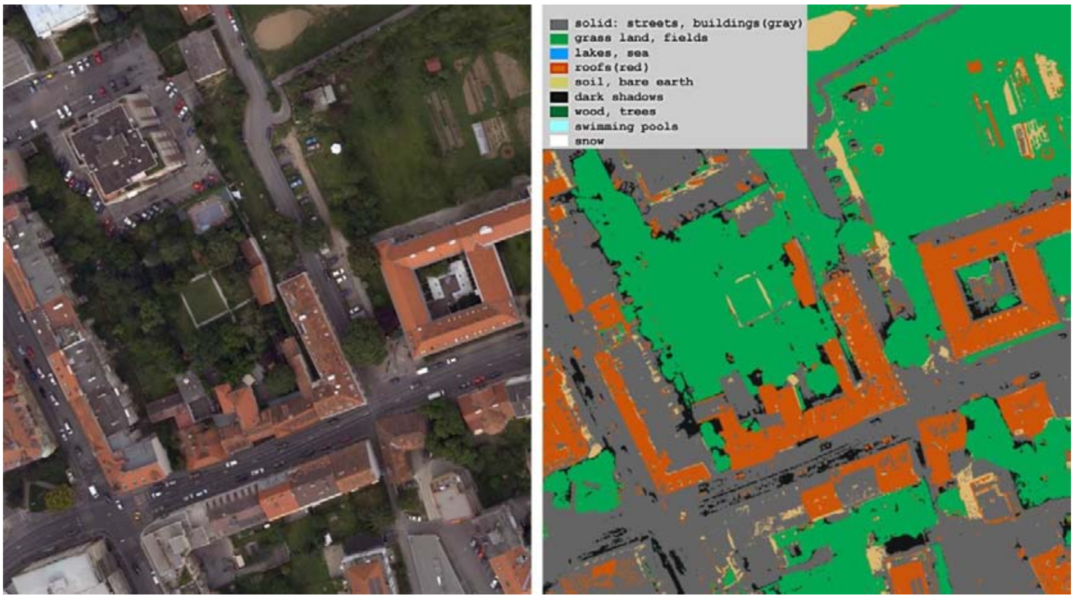

Figure 1: The RGB image (left) and the corresponding initial LUC (right) for a small area within the city of Graz.

3.2 Automatic Aerial Triangulation

Our aerial triangulation starts with a POI (point of interest) extraction which is based on Harris points and POIs from line intersections (Bauer et al. 2004). The POIs from line intersections which we call 'zwickels' are very suitable for urban areas. Zwickels are sections defined by two intersecting line segments, dividing the neighborhood around the intersection point into two sectors. After the POIs extraction in each image we calculate feature vectors in the close neighborhood. Feature vectors are used to find 1 to n correspondences between POIs in two images. Using affine invariant area based matching the number of candidates is further reduced. For all remaining candidates we iteratively apply an affine transformation to maximize the crosscorrelation score. As a result we get a list of corresponding points. In order to fulfill the nonambiguous criteria, only matches with a high distinctive score are retained. The robustness of the matching process is enhanced by processing a back-matching as well. Figure 2 illustrates corresponding POIs from line intersections.

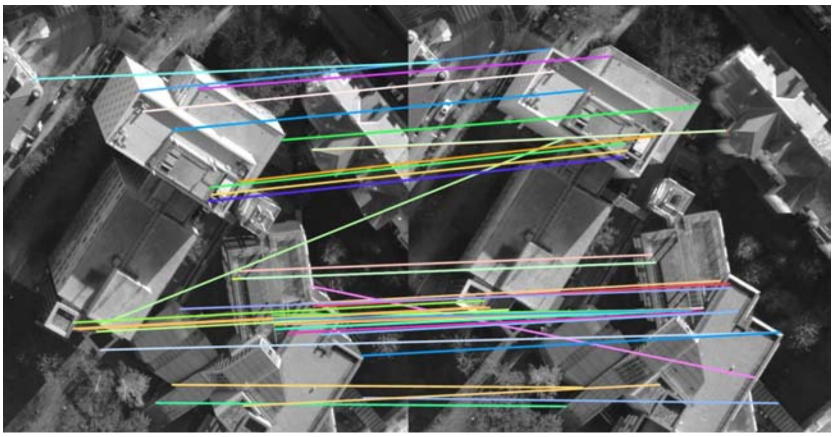

Figure 2: Line intersections are used as POIs in the aerial triangulation. These POIs have a high location accuracy and low outlier rate (there are only two outliers within the best 25 matches) even before the epipolar constraint is enforced. Corresponding POIs are connected by lines.

Another restriction is enforced by the epipolar geometry. Therefore the RANSAC method is applied to the well known five point algorithm (Nister 2003). As a result we obtain inlier correspondences as well as the essential matrix. By decomposition of the essential matrix the relative orientation of the current image pair can be calculated.

This step is accomplished for all consecutive image pairs. In order to get the orientation of the whole set, the scale factor for additional image pairs has to be determined. This is done using corresponding POIs available in at least three images. A block bundle adjustment refines the relative orientation of the whole set and integrates other data like GPS, DGPS, IMU or ground control information. Figure 3 shows an oriented block of 7 x 31 aerial images together with the used 3D tie points on the ground. The whole block of images was processed without any human interaction.

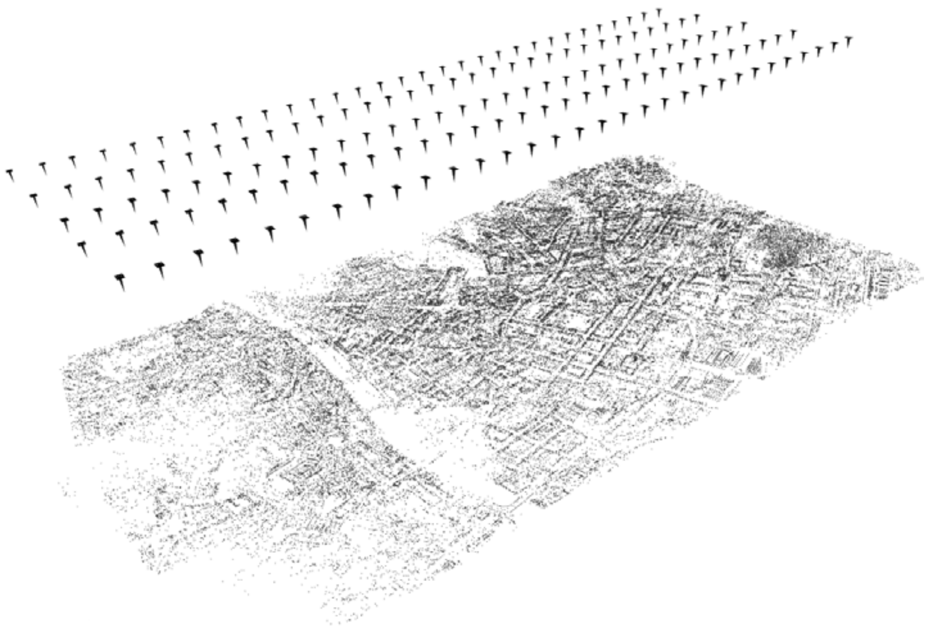

Figure 3: 5 strips of about 31 images each denoted by small arrows are oriented to each other using about 70.000 tie points on the ground which are shown as black dots.

3.3 Dense Matching

Once the AT is finished we perform a dense area based matching to produce a dense DSM (digital surface model). Recent years saw more new dense matching algorithms were introduced. A good comparison of stereo matching algorithms is given in a paper by Scharstein et al.(2002). Recently, a PDE based multi-view matching method was introduced by Strecha et al. (2003). In our approach we focus on an iterative and hierarchical method based on homographies to find dense corresponding points. Figure 4 shows a dense matching results using our approach for the inner city of Graz.

Figure 4. A dense height image with a GSD (ground sampling distance) of 30cm for the inner city of Graz is calculated fully automatically from oriented aerial images.

3.4 True Orthophoto

A 'true' orthophoto is obtained by orthoprojection of the DSM. The color information of the orthophoto is calculated using all available aerial images and is based on view-dependent texture mapping described in (Bornik et al. 2001).

3.5 Refined Classification

Refined classification performs data fusion in a way that the classification results are less scattered, see Figure 5: the initial classification – top middle image – has the roof correctly classified but the chimney and a small roof over a window are classified as solid. The height data – top right image – as well as the height gradients – bottom left image - and the building blocks – bottom middle image - cause a classification of the whole roof as one block, see bottom right image.

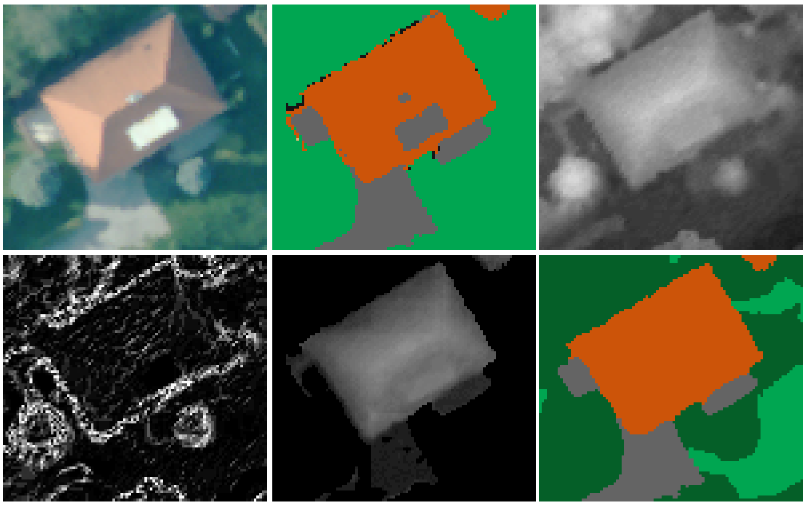

Figure 5: Different stages within our modeling workflow. Top row: one input image (left); initial LUC (middle); height image produced by dense matching (right). Bottom row: height gradients (left); fusion of LUC and height image; refined LUC (right).

3.6 Modeling of Roof and Façade Planes

The information from the height image (Figure 4) together with the refined LUC (Figure 5 bottom right) is used to extract building blocks. Local planes are fitted inside these building blocks. Neighboring planes are intersected to obtain roof polygons. The height gradients at the border of the building blocks are used to estimate façade planes which are refined in a constrained plane sweeping approach and again intersected with the roof polygons and the ground planes. In a last step the best input images, incorporating the visibility information, are searched to texture the façade planes. Figure 6 shows a small area within Graz.

Figure 6: 3D city model of a small area in Graz. Additionally to the geometry and texture information, the semantic of single objects (building block, tree, street, water region, …) is known as well.

4. RENDERING AND VISUALIZATION OF THE RESULTING GEOMETRY

The available reconstructed geometry consists of two types of data: tha actual terrain, which is available in a a 2½ D format, i.e. height values are available for a grid of locations, and the classified, reconstructed data such as facades and roofs.

For rendering the terrain data we use an algorithm based on Ulrich’s chunked level-of-detail method for fast rendering of terrains [Ulrich]. This method is optimized for visualizing large terrains, using a qaud-tree based approach for loading the geometry and associated textures into memory as needed. However, in its original form, Ulrich’s algorithm is limited in its precision: due to the use of IEEE 32-bit floats of current rendering hardware, very large terrains will result in a loss of precision the further one moves out from the origin.

In order to overcome this problem we extended Ulrich’s method by running multiple instances in parallel. Each instance uses its own, optimized coordinate system and the instances are rendered with the users camera subsequently mapped into the coordinate system of each instance. The computation of the camera coordinate system with respect to the coordinate system of each instance can then be performed in double precision (IEEE 64-bit floats), thereby retaining the available precision of the original data.

Using this technique it is possible to stitch the renderings of multiple instances of Ulrichs algorithm and thereby render huge terrain models without loss of precision. The actual instances of the algorithm are started and discarded as the user moves through the large model.

Figure 7: Screenshots from an interactive walkthrough/flyover application.

The reoncstucted façade and roof geometry cannot be rendered using Ulrich's algorithm, since the data is not representable as a 2½ D Grid, but consists of arbitrary geometry. Thus this geometry is handled in a separate layer that is only displayed for all tiles in Ulrich's algorithm that are close enough to be visible in detail. This is sufficient to provide the user with a nearly seamless walkthrough or flyover of the reconstructed terrain and urban model (see figure 7).

5. CONCLUSION AND FUTURE WORK

We demonstrated a significantly improved method and worflow for fast reconstruction of urban geometry based on aerial photography. Due to its fast turnaround time the new method allows new applications of aerial reconstruction such as quick verification of available data or repeated reconstruction for detecting changes. Combined with a fast and interactive visualization method especially optimized for viewiing the resulting data, a very short turnaround time from acquiring the data to visual analysis of the data is achieved.

The current algorithms are based only based on the data obtained from the aerial images. By extending the matching algorithm to incorporate data from additional sources such as municipal land use databases, additional information for improving the quality of the resulting urban model is possible.

The land use classification algorithms can also be extended to provide the locations of additional object types as an input for the visulazation engine: vegetation can be classified as individual trees with height and categorized into different types, vehicles can be detected and classified into different types as well. The resulting information can then be used in the visualization step to render detailed models of the detected tree and vehicle instances.

6. REFERENCES

A. Asirvatham, H. Hoppe, 2005: Terrain rendering using GPU-based geometry clipmaps. GPU-Gems 2 (M. Pharr and R Fernando eds.), Addison-Wsley.

J. Bauer, H. Bischof, A. Klaus, K. Karner, 2004, Robust and fully automated Image Registration using Invariant Features, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXV, Istanbul, Turkey, ISSN 1682-1777.

A. Bornik, K. Karner, J. Bauer, F. Leberl, H. Mayer, 2001, High-quality texture reconstruction from multiple views, *The Journal of Visualization and Computer Animation*, Volume 12, Issue 5, 2001, Online ISSN: 1099-1778, Print ISSN: 1049-8907, John Wiley & Sons, pp. 263-276.

C.-.W. Hsu, C.-C. Chang, and C.-J. Lin, 2003, *A Practical Guide to Support Vector Classification*, Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan.

C. Huang , L.S. Davis, J.R.G. Townshend, 2002, An assessment of support vector machines for land cover classification*, Int. J. Remote Sensing*, Vol. 23, No. 4, pp. 725-749.

A. Klaus, J. Bauer, K. Karner, K. Schindler, 2002, MetropoGIS: A Semi-Automatic City Documentation System, *Photogrammetric Computer Vision* 2002 (PCV'02), ISPRS - Commission III, Symposium 2002, September 9 - 13, 2002, Graz, Austria.

F. Leberl, J. Thurgood., 2004, The Promise of Softcopy Photogrammetry Revisited. ISPRS 2004, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXV, Istanbul, Turkey, ISSN 1682-1777.

F. Leberl, R. Perko, M. Gruber, 2002, Color in photogrammetric remote sensing, *Proceedings of the ISPRS Commission VII Symposium*, Hyderabad, India, Vol. 34, pp. 59-64.

F. Losasso, H. Hoppe, 2004: Geometry Clipmaps: Terrain rendering using nested regular grids. ACM Siggraph 2004, 769-776.

D. Nister, 2003, An efficient solution to the five-point relative pose problem, *CVPR 2003*, pp. 195–202.

F. Roli, J. Kittler, 2002, Fusion of multiple classifiers, *Information Fusion*, Vol. 3, pp. 243.

D. Scharstein and R. Szeliski, 2002, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1/2/3), pp. 7-42.

M. Sormann, A. Klaus, J. Bauer, K. Karner, 2004, VR Modeler: From Image Sequences to 3D Models, *SCCG* (Spring Conference on Computer Graphics) 2004, ISBN 80-223-1918-X, pp. 152-160.

C. Strecha, T. Tuytelaars, L. Van Gool, 2003, Dense Matching of Multiple Wide-baseline Views, *ICCV* 2003, Vol 2, pp. 1194-1201.

J. Thurgood, M. Gruber, K. Karner, 2004, Multi-Ray Matching for Automated 3D Object Modeling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXV, Istanbul, Turkey, ISSN 1682-1777.

T. Ulrich: Chunked Level of Detail Homepage. http://tulrich.com/geekstuff/chunklod.html

C. Zach, A. Klaus and K. Karner, 2003, Accurate Dense Stereo Reconstruction using 3D Graphics Hardware, *Eurographics* 2003, Short Presentations, pp. 227-234.