MetropoGIS: A City Modeling System

DI Dr. Konrad KARNER, DI Andreas KLAUS, DI Joachim BAUER, DI Christopher ZACH

VRVis Research Center for Virtual Reality and Visualization, Virtual Habitat, Inffeldgasse 16/2, 8010 Graz, karner@vrvis.at

1 ABSTRACT

We report on a new system to generate photo-realistic 3D models of real cities. So far, we focus on the accurate geometric and radiometric modeling of man-made structures. The modeling process starts with a 3D block model obtained from aerial images, aerial laser scans or GIS data. This block model is further augmented with geo-referenced terrestrial images of the facades. The terrestrial images are either captured using a mobile multi-sensor-platform or an off-the-shelf digital camera. The relative orientation of terrestrial images is calculated automatically and fitted towards the 3D block model with minimized human input using vanishing points. 2D image features like lines are extracted from the images and used to model 3D lines using line matching techniques. The extraction of dense 3D point clouds on the facades is based on a hierarchical point matching strategy over multiple oriented images. Facades are found by robust fitting of plane hypotheses through these 3D line and point clouds. The introduced city modeling system delivers a fully 3D geographic information data set and is called MetropoGIS.

2 INTRODUCTION

The usability of virtual city models strongly depends on the correct modeling of geometric, radiometric and topological structures of the real city. During the whole design and implementation process we kept the various applications such as the simulation of wind, flooding, traffic, and radio transmission as well as city planning, surveying, virtual tourism, cultural heritage documentation, and emergency training in mind. Thus, important requirements are a fully three dimensional representation of the model with a high geometric accuracy. Furthermore, high resolution textures of all objects for photo-realistic visualization and a topological description of building features like windows and doors are necessary. Such virtual city models are very attractive for different uses and pay off in a short time.

3 RELATED WORK

The modeling process of virtual cities starts with building a block model using aerial images which is well known in the photogrammetry and remote sensing community. Nowadays aerial laser scanning is becoming popular and is an alternative approach for the generation of a block model. There exist some commercial products using photogrammetry to build city block models. Such commercial products are the CyberCity-Modeler from CyberCity AG (http://www.cybercity.tv/), inJECT from Inpho (http://www.inpho.de/), or IMAGIS from Supresoft (http://www.supresoft.com/english/). All these products need a human operator who is supported more or less with automated image matching techniques.

Facades are not or only hardly visible from aerial images. Thus, the block model has to be augmented by terrestrial data. A working system for terrestrial image acquisition in urban areas is described by Teller. The system works with uncalibrated images and provides calibrated, geo-referenced images as output. The usage of a GPS-based position estimation allows a fully automatic processing. The sensor provides omni-directional images for better pose estimation. Detailed information about this approach can be found in Antone and Bosse. Our approach is quite similar but differs in some important aspects. One is that we can handle occlusions in a much easier way by evaluating multiple adjacent images simultaneous as described in Bornik et al. In addition our input sensor a digital camera (still or video) can be used in stop and go as well as in continuous mode. A GPS-based positioning system is not necessary and of limited use in narrow streets.

Our 3D modeling of the facades is inspired by work done by Zisserman et al. where an emphasis is put on the automatic extraction of planes from architectural images.

4 WORK FLOW

Our work flow comprises six consecutive steps where the first step consists of the generation of a city block model which uses data from aerial sensors. The steps 2 to 4 deal with terrestrial captured data, whereas in step 5 the transformation of the terrestrial data to the block model is addressed. Step 6 deals with the modeling of the facades. These six steps are described further in the following subsections.

4.1 Block Model Generation

In our pilot project in Graz we use GIS data provided by the surveying office (Stadtvermessungsamt Graz) to obtain the 3D block model (see Figure 1). The GIS data consists of manual measured 3D roof lines which are triangulated and projected onto a 20m grid DEM. The DEM itself is augmented with measured 3D break lines along sharp discontinuities in the terrain. The aerial images are used to improve the accuracy of the measured 3D points and an image consistent triangulation is applied as described by Klaus et al. A result of our approach can be seen in Figure 1.

Figure 1: Overview of the virtual block model of the inner city of Graz

4.2 Feature Extraction

Line extraction starts with an edge detection and linking process and yields contour chains with sub-pixel accuracy. For all contour chains of sufficient size a RANSAC (RANdom Sample Consensus) based line fitting method is applied. The vanishing point detection is based on the method proposed by Rother.

The known position of vanishing points in the image is used to extract more lines pointing to these vanishing points. The extraction is based on a sweep line approach. In a preprocessing step edgels are extracted using a lower threshold than before. The amount of edgels to be processed is reduced by removing edgels with an orientation differing too much from the orientation to the vanishing point. The sweep line starts at the vanishing point and goes through the image plane. All edgels within some perpendicular distance to the sweep line are considered as inliers. Line segments are constructed by least squares fitting each densely chained subset of inliers. Overlapping parallel segments are merged after the sweep. In a post processing step intersections of line pairs from different vanishing points are computed. These intersections serve as points of interest (POI) for the computation of the relative orientation. Figure 2 shows the result of the advanced line extraction process.

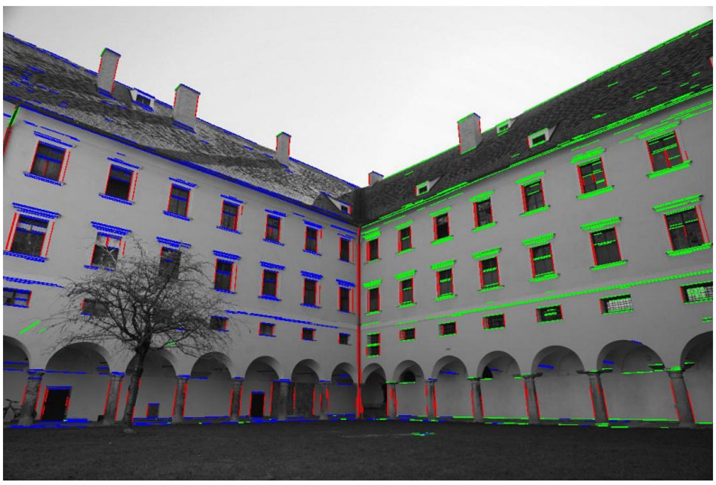

Figure 2: Detected line segments by the advanced line extraction where lines with the same color belong to the same vanishing point.

4.3 Relative Orientation of Image Pairs

In this approach we do not need time consuming point to point correlation in image pairs to find corresponding points. Instead we use vanishing points and POIs from line intersections to estimate the relative orientation. POIs are classified into 8 categories depending on the gradient information of lines and the position of the intersection relative to the lines (left or right, upper or lower corner). These POIs are accurate and invariant to perspective transformations.

Corresponding POIs in different images are found by testing a qualified set of possible POI pairs. For each potential pair the support is measured and the pair with the highest support is selected. The rotation of the left camera which is known from vanishing points is used to determine a plane, on which the POIs of the left image are projected as shown in Figure 3. A correct point pair is found by sampling through all possible combinations. If one potential point pair is assigned, the position of the second camera can only be shifted on the ray that goes through the corresponding 3D point on the plane, the camera center and the corresponding point in the right image plane. In Figure 3 this ray is indicated by the black continuous line.

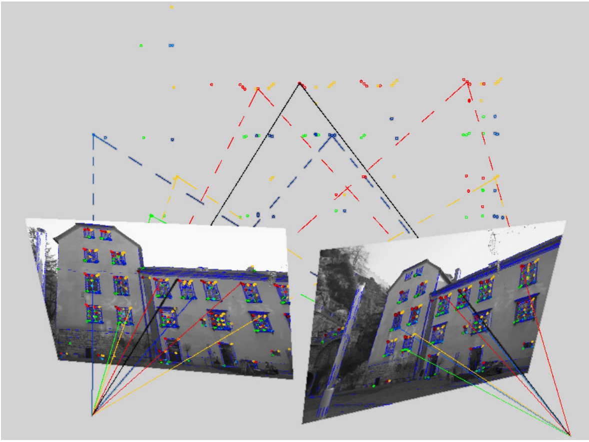

Figure 3: Illustration how to determine relative orientation of an image pair using POI´s.

If the right image is shifted along the corresponding ray the 3D points on the determined plane are projected onto different positions in the right image plane (see Figure 4). The final position is found by shifting the camera along the corresponding ray until the support is maximized.

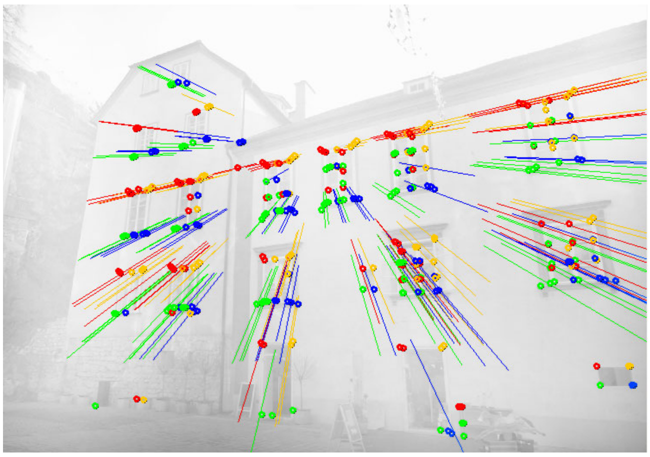

Figure 4: If one corresponding point is given the second camera can be shifted on the given ray only.

4.4 Relative Orientation of Image Sequences

So far we have determined the orientation of image pairs and corresponding points in adjacent images. In order to calculate the orientation of a continuous sequence we start with one image pair and calculate 3D points assuming a fixed baseline (without loss of generality) and the rotation to the third image using the vanishing points. The position is found by selecting a 3D to 2D POI correspondence and shifting the third image along the obtained ray to maximize the support of reprojected 3D points and 2D POIs in the third image. The corresponding points of the new image are used to calculate new and to improve old 3D points. This approach is repeated until all images are oriented to each other.

4.5 Geo-referenced Orientation of Image Sequences

So far we have only obtained a relative oriented sequence, where the position and orientation in geo-referenced coordinates as well as the scale are not yet known. Due to the fact that the rotation of an image is known from vanishing points, only two control points are necessary to transform the image into a geo-referenced coordinate system. These two control points are taken from the known eave lines. In our approach only two images of a continuous sequence have to be fitted in such a semi-automatic way which minimizes the manual work. Figure 5 shows a sequence of images which were transformed semi-automatically.

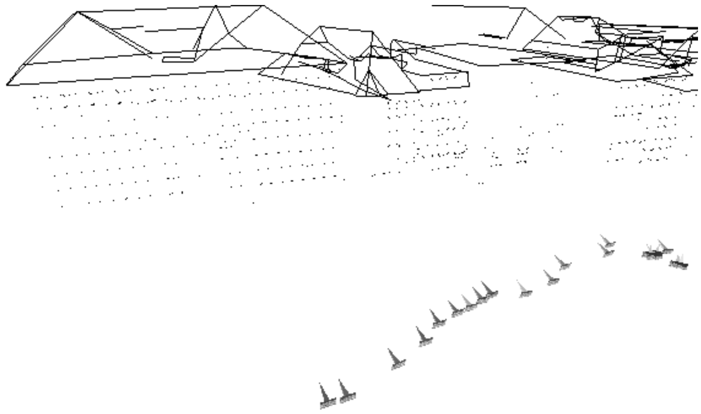

Figure 5: Geo-referenced sequence of the right side of the Grazer Hauptplatz. The measured roof lines are superimposed by cameras represented by a small plane and a direction arrow and POIs indicated by dots.

For each building block a bundle adjustment is carried out to optimally fit the images into the existing block model. The known control points in some images help to fix the transformation to the world coordinate frame and to stabilise the block, particularly in long sections of translational camera movement. Since the vertical direction is known from the detected vanishing points in the terrestrial images, a constraint is added to keep the mean vertical direction of the facades unchanged.

4.6 City Modeling

Our modeling step is divided into a line matching and a dense point matching approach. The following subsections give some overview and shows some results.

4.6.1 Line Modeling

The set of line segments per image together with the known orientation of the image sequence are the input for the line matching algorithm. Our approach closely follows the one described by Schmid and Zisserman. The result of the line matching process is a set of 3D lines in object space.

Basically the algorithm works as follows: For a reference line segment in one image of the sequence potential line matches in the other images are found by taking all lines that are enclosed by the epipolar lines induced by the endpoints of the reference line segment.

Each of these potentially corresponding line pairs gives a 3D line segment (except for those, which are parallel to the epipolar line, since in this case no intersection between the epipolar line and the image line can be computed).

The potential 3D lines are then projected into all remaining images. If image lines are found which are close to the reprojection, the candidate is confirmed, else it is discarded. Finally a correlation based similarity criterion is applied to select the correct line. Figure 6 shows two views of the extracted 3D line set. Obviously, due to the small vertical baseline the geometric accuracy of the horizontal line segments is limited.

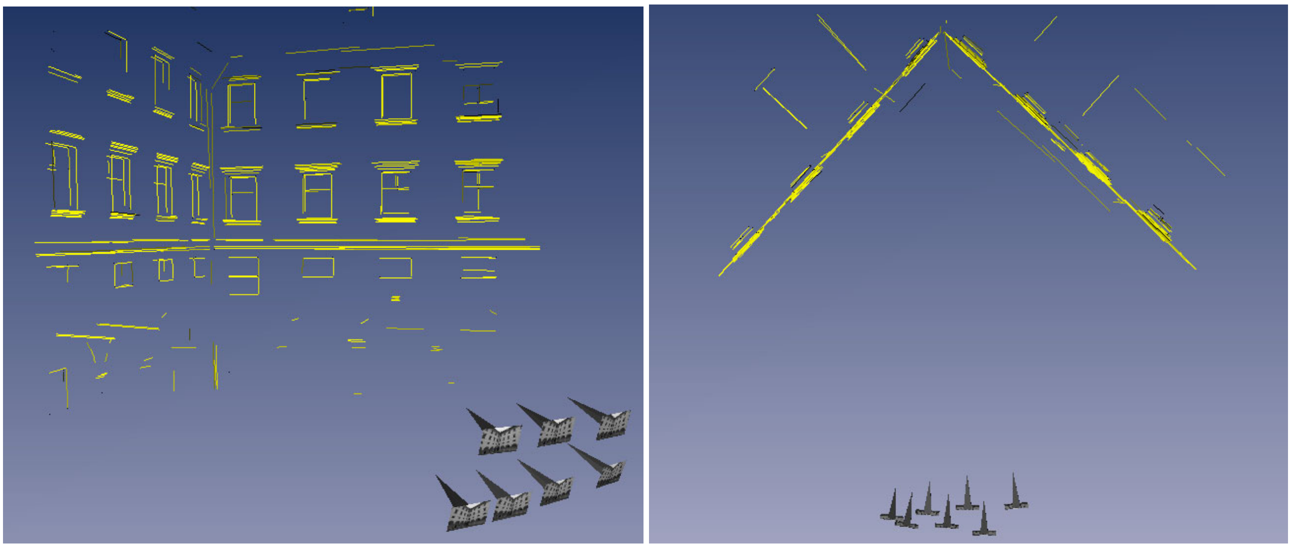

Figure 6: Two views of the extracted 3D line set of the facade in Figure 2. The left image shows a front the right a top view.

4.6.2 Dense Point Matching

In order to calculate 3D points we need to find corresponding points in all adjacent images of our oriented image sequence that contains the same projected 3D point.

In our approach we focus on an iterative and hierarchical method based on homographies to find this corresponding points inspired by a work published by Redert. For each input image an image pyramid is created and the calculation starts at the coarsest level. Corresponding points are searched and upsampled to the next finer level where the calculation proceeds. This procedure continues until the full resolution level is reached.

This hierarchical method convergences fast and avoids local minima solutions especially having repetitive structures within a facade. Figure 7 shows two reconstructed sides of the „Grazer Hauptplatz". If the dense point matching algorithm is applied on the geo-referenced image sequence, the calculated 3D points align to the measured roof lines from the block model. The on site time for capturing the input images was about 5 minutes. The processing of the input images (feature extraction, calculation of the orientation, dense point and line matching) took 43 minutes on a 1.6GHz PC with 512MB RAM. The geo-referenced orientation of the image sequence, the only step where a human operator is needed, requires less than 1 minute for a trained person.

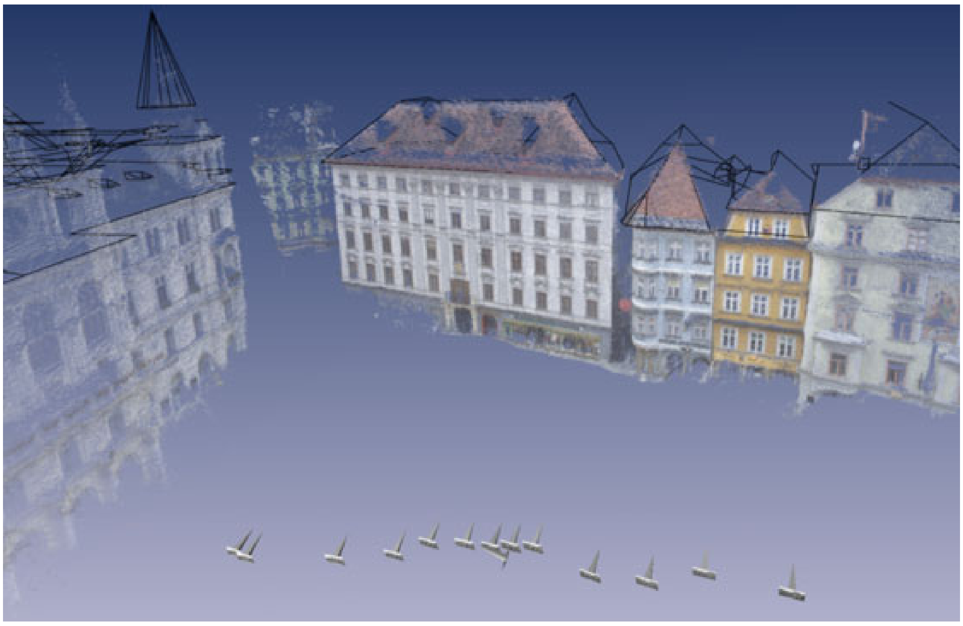

Figure 7: Two compounded geo-referenced sequences of the Grazer Hauptplatz augmented with roof lines.

Figure 8 and 9 show different views of the extracted 3D points of a sequence with detailed facades in Vienna. The reconstruction contains more than 2.4 million points.

Figure 8: Fully automatic result from a sequence of 26 images acquired at ´Am Hof˝ in Vienna.

Figure 9: Two other views of the reconstructed sequence shown in Figure 8.

5 CONCLUSIONS AND FUTURE WORK

We have presented an image based modeling system called MetropoGIS in which a 3D block model of a city is augmented with terrestrial measured data. We have shown that it is possible to determine the geo-referenced orientation of a large image sequence with little human interaction without any position estimation. Because we use a consumer camera, our image acquisition is straightforward and allows high flexibility. By exploiting lines and vanishing points we have developed a robust and fast method to determinate relative orientation for even large baselines.

So far we have a semi-automatic system where an operator is still involved in the work flow. We are planning to avoid this manual interaction by a fully automatic step where we try to match roof lines extracted from the terrestrial images with 3D lines from the block model. Furthermore, we are working on a plane sweeping approach to find object planes connected to 3D lines. Our approach is similar to the one presented in Zisserman with the difference that we use a feature based correlation criterion instead of area based cross correlation.

6 ACKNOWLEDGEMENTS

This work has been done in the VRVis research center, Graz/Austria (http://www.vrvis.at), which is partly funded by the Austrian government research program Kplus. We would also like to thank the surveying office of Graz (Stadtvermessungsamt Graz) for providing us with their GIS data and aerial images.

7 REFERENCES

Antone M. and Teller S.: Automatic recovery of relative camera rotations for urban scenes; *IEEE Proceedings of CVPR*, 2000

Antone, M. and Teller S.: Scalable, absolute position recovery for omni-directional image networks; *IEEE Proceedings of CVPR*, 2001

Bornik A., Karner K., Bauer J., Leberl F. and Mayer H.: High-quality texture reconstruction from multiple views; *Journal of Visualization and Computer Animation*, 2002

Bosse M., de Couto D. and Teller S.: Eyes of argus: georeferenced imagery in urban environments; *GPS World*, 1999, pp. 20–30

Klaus A. and Karner K.: Accurate roof modeling using 3D lines and aerial images; *Proc. of the 26th workshop of the Austrian Association for Pattern Recognition*, 2002, pp. 341-346

Redert A., Hendriks E, and Biemond J.: Correspondence estimation in image pairs. *IEEE Signal Processing Magazine*, pages 29–46, May 1999.

Rother C.: A new approach for vanishing point detection in architectural environments; *Proceedings of the 11th British Machine Vision Conference*, 2000, pp. 382–391

Schmid C. and Zisserman A.: The geometry and matching of lines and curves over multiple views, *IJCV*, vol. 40, no. 3, pp. 199–233, 2000

Schaffalitzky F. and Zisserman A.: Planar grouping for automatic detection of vanishing lines and points; *IVC*, vol. 18, no. 9, pp. 647–658, 2000

Teller S.: Calibrated, registered images of an extended urban area; *IEEE Proceedings of CVPR*, 2001

Zisserman A., Werner T. and Schaffalitzky F.: Towards automated reconstruction of architectural scenes from multiple images; *Proc. of the 25th workshop of the Austrian Association for Pattern Recognition*, 2001, pp. 9–23