METROPOGIS: A SEMI-AUTOMATIC CITY DOCUMENTATION SYSTEM

Andreas Klaus* **Joachim Bauer*** **Konrad Karner*** **Konrad Schindler†**

* VRVis Research Center for Virtual Reality and Visualization
email: {klaus,bauer,karner}@vrvis.at

† Institute for Computer Graphics and Vision, Graz University of Technology
email: {schindl}@icg.tu-graz.ac.at

KEY WORDS: GIS, Image Orientation, Image Sequences, Vanishing Points, Bundle Block Adjustment.

ABSTRACT

In this paper we report on a new system to augment a 3D block model of a real city with geo-referenced terrestrial images of the facades. The terrestrial images are taken by a hand-held digital consumer camera using short baselines. The relative orientation of the photographs is calculated automatically and fitted towards the 3D block model with minimized human input using vanishing points. The introduced city documentation system delivers a fully 3D geographic information data set and is called *MetropoGIS*.

1 INTRODUCTION

Virtual city models attract attention due to their usability for 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. All these applications have different requirements on the data set which have to be fulfilled. Some of these requirements are:

- 3D geometric model
- high geometric accuracy
- high geometric resolution
- high resolution texture of all objects for photo-realistic visualization
- topological description of building features like windows and doors.

In our system we attach importance to all these above mentioned requirements. In this paper we concentrate on the first part of the whole city modeling process, the orientation of the input images. In Section 2 we give an overview of related work. Our work flow starting with line and vanishing point extraction, the relative orientation of image sequences and the mapping of the images to the 3D block model are given in Section 3. Some results which illustrate the robust search for corresponding points to estimate the relative orientation between image pairs and the image to 3D model fitting using the vanishing points are shown in Section 4. Section 5 presents the conclusions and gives some ideas on future work.

Figure 1: Overview of the virtual 3D block model of the inner city of Graz.

2 RELATED WORK

A working system for calibrated, terrestrial image acquisition in urban areas can be found in Teller [11]. The system works with uncalibrated images and provide calibrated, geo-referenced images as output. To enable a fully automatic processing a GPS-based position estimation is required. The sensor provide omni-directional images for better pose estimation. Additional information can be found in Antone [2], [3] and Bosse [8]. Our approach is quite similar but differs in some important aspects. The most relevant one is that we can handle occlusions in a much easier way by evaluating multiple adjacent images simultaneous as described in Bornik [1]. In addition we are much more flexible with our hand-held camera, that can be used in stop and go as well as in dynamic mode.

3 OVERVIEW OF OUR WORK FLOW

In our approach we concentrate on the refinement of the facades of buildings. We assume that a 3D block model with roof lines exists. For our pilot project, the inner city of Graz, we started with a simple three dimensional (3D) block model obtained from converting 2 1/2 dimensional

GIS (geographic information system) data. An overview of a part of our 3D block model can be seen in Figure 1.

The 3D block model is augmented using image sequences captured by a digital consumer camera from arbitrary positions. Currently, we use a Canon D30 with a geometric resolution of 2160 x 1440 and 12bit per pixel radiometric resolution. The images have to be captured using short baselines, thus a digital video camera with a reasonably high resolution will work as well. Our work flow consists of five consecutive steps which will be explained in the following subsections.

3.1 Line Extraction and Vanishing Point Detection

The line extraction stage starts with an edge detection and edge linking process and yields contour chains with sub pixel accuracy. For all contour chains of sufficient size a RANSAC [4] based line detection method is applied. Pairs of contour points are randomly picked from the contour and a potential line segment is formed. For this line segment inlier points, that are points with a small perpendicular distance to the segment are searched. The line segment with most inlier points is considered the best hypothesis for the line. The final line parameters are estimated by a least squares adjustment over the inlier points. A final grouping process merges collinear segments that lie close to each other.

The vanishing point detection is based on the method proposed by Rother [10]. In this approach the previously extracted line segments are used for the detection. Each intersection is treated as potential vanishing point and the weight for the intersection is determined by testing against all other line segments. The smaller the angle difference between a line segment and the vector pointing from the mid point of the segment to the intersection, the higher the contribution to the weight of the accumulator cell for the potential vanishing point. If the angle difference is too large, the weight of the accumulator cell is not increased. The intersection with the maximal weight is then reported as vanishing point.

3.2 Relative Orientation of Image Pairs

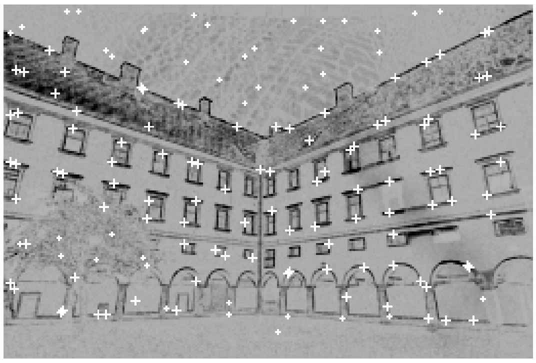
We developed two completely different methods to calculate the relative orientation of image pairs. The first method is based on calculating corresponding points within an image pair. Therefore an area based hierarchical matcher is used. In the second approach vanishing points are used to solve the relative orientation problem without the necessity to perform time consuming point to point correlation in image pairs.

3.2.1 Relative Orientation from corresponding points

In order to determine the relative orientation of an image sequence, we need to find corresponding points in all adjacent image pairs. 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 et al. [9]. For each input picture 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 when having repetitive structures within a facade.

(a) First input image

(b) Second input image



(c) Visualization of our accuracy measure; Crosses indicate corresponding points used for the calculation of the relative orientation.

Figure 2: Input images and visualization of our accuracy measure.

A reliable calculation of the relative orientation can be done using a set of corresponding points which should fulfill some properties. They have to be well distributed over the images with a good location accuracy and a low outlier rate. To achieve this requirement it is necessary to calculate an accuracy measure for all calculated corresponding points. This is done during the matching process and is calculated within a cost function based on the distribution of local minima and maxima. The better the distinction of local minima the higher the accuracy measure. Figure 2(c) shows this accuracy measure where darker areas indicate higher location accuracy. The two input images are shown above.

Bucketing over the image is used to get a well distributed point set out of all corresponding points. Therefore, the images are divided into a number of regions and for each region the point with the highest accuracy measurement is used. These points are displayed in Figure 2(c) as crosses. As shown in the image there are regions where it is not possible to find correct corresponding points. Thus, only a fraction (about 50%) of all regions are used. Those are indicated by a larger cross.

So far our corresponding points still contain a few outliers. To get rid of them we use a robust estimator – the RANSAC (RANdom SAmple Consensus) algorithm of Fischler and Bolles [4]. Afterwards the fundamental matrix is calculated with the accurate Gold Standard algorithm [14]. Using the fundamental matrix the minimum epipolar distance is calculated for each remaining corresponding point and points with great distances are indicated as outliers.

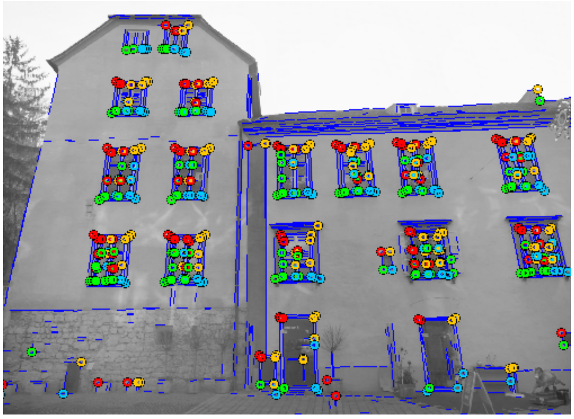
Due to the fact that we work with almost planar facades the observed corresponding points may lie on a plane which is a critical surface for the estimation of the relative orientation (see Horn [6] for a detailed discription of critical surfaces). Thus, we distinguish for the estimation of the relative orientation between two configurations. In the first configuration the corresponding points are well distributed in space. The solution for this case is explained in the following section. In the second case all points lie close to a plane. This critical configuration is solved using homographies between the images.

Estimating the Relative Orientation using at least 5 Corresponding Points The inliers (shown as large crosses in Figure 2(c) are used to determine the relative orientation. We use the algorithm of Horn [6], who proposed to use quaternions to depict the relative orientation. It is possible to determine the relative orientation from the fundamental matrix directly. We do not recommend this widely used procedure due to numerical instabilities like in the case that the optical axes are almost parallel (the epipole lies at infinity).

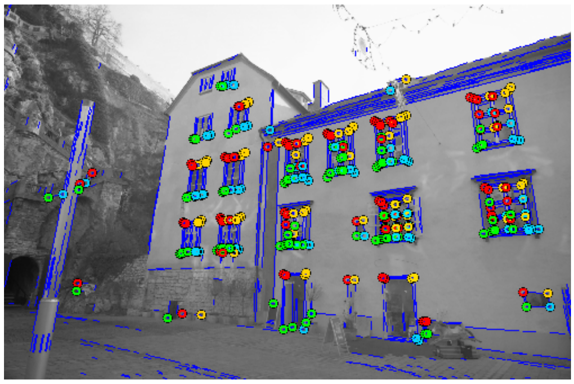
Estimating the Relative Orientation from a Homography A homography describes the projective transformation of a plane in 3D. As shown in Wunderlich [13] and later reformulated by Triggs [12] it is possible to calculate the relative orientation from a homography. Thus, we use a RANSAC algorithm to find a subset of corresponding points to estimate the homography.

3.2.2 Relative Orientation from Vanishing Points In our second approach we do not need point to point correlation in image pairs; instead we are using vanishing points and line intersections.

The known position of vanishing points in the image are used to extract all lines pointing to these vanishing points.



(a) Image of a building at the Schlossbergplatz.



(b) Another view of the same building.

Figure 3: The color of the detected points of interest indicates their category.

The extraction is based on a sweep line approach. In a preprocessing step sub-pixel edgels are extracted using a low threshold. 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. For each densely chained subset of the inliers a line segment is constructed by computing a least squares adjustment over the inlier points. Overlapping parallel segments are grouped after the sweep. In a post processing step intersections for line pairs from different vanishing points are computed. These intersections serve as points of interest (POI) for the computation of the relative orientation.

A set of POIs that belongs to two vanishing points can be subclassified into 8 categories. We distinguish between 4 line formations that lead to a POI. These are left upper, right upper, right lower and left lower corner. In addition we are using the gradient information of each line. The gradient information indicates which side of the line has brighter pixels. E.g. for horizontal lines it indicates if either the upper or the lower pixels are brighter. This information can easily be determined for both lines that form a POI by calculating the average direction of all edgels that belong to the line. Thus we have another 4 reasonable possibilities and therefore in total 8 categories. In Figure 3

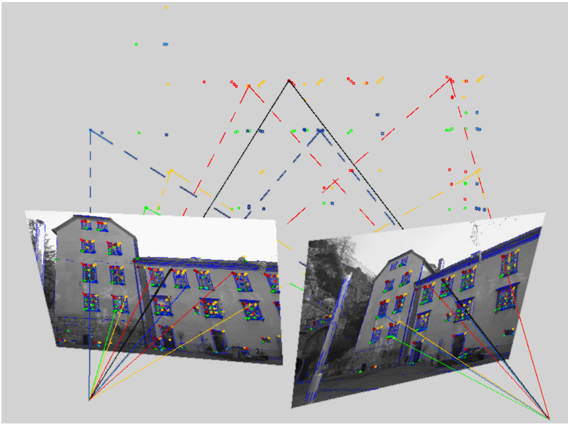


Figure 4: Illustration how to determine relative orientation of an image pair using POIs.

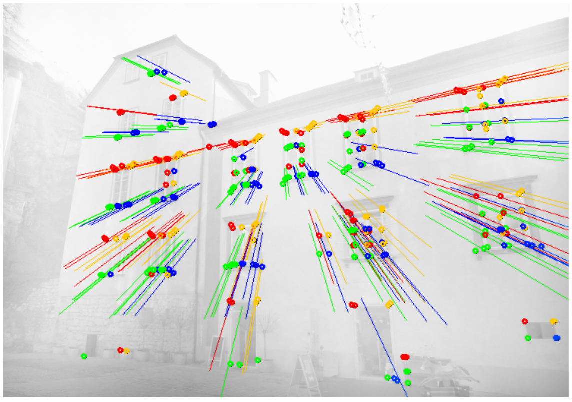


Figure 5: If one corresponding point is given the second camera can only be shifted on the given ray.

POIs of different category are indicated by different color. As can be seen the POIs are very resistant to affine transformation, thus a great fraction of POIs keeps unchanged.

The relative orientation of a camera has 5 degrees of freedom - three for the rotation and two for the pose direction. We assuming that two adjacent images view the same plane to some extent. If we extract two vanishing points belonging to a plane, the rotation of the camera relative to this plane is determined. Furthermore the corresponding vanishing points (derived from the same physical plane) of adjacent images give use the relative rotation between the cameras. The estimation of the rotation angles between the physical plane and both cameras are described in Kraus [7]. In our approach the remaining 2 degrees of freedom for the base direction are found by searching for one corresponding point.

This correspondence is obtained by testing a qualified set of possible POI pairs. For each potential pair the support is measured using an image based method and the pair with the maximum support is selected. The position of the left camera is fixed at an arbitrary position. Furthermore the rotation of the left camera is used to determine a plane, on which the POIs of the left image are projected as it is shown in Figure 4. A correct point pair is found by sampling through all possible combinations. If one potential corresponding 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 4 this ray is indicated by the black continuous line.

If the right image is shifted along the corresponding ray the 3D points on the determined plane are projected on varying positions on the right image plane. If we add the constraint that the distance of a 3D point to the left and right camera centers only differs to a certain extend, each projected 3D point lies on a bounded line which is projected into the right image and is shown in Figure 5. To calculate the support for a POI pair we are using an additional image with reduced resolution where the POIs of the right image are plotted. For each entry a special bit according to its category is set within the region around the POI. Therefore the support for a given correspondence can be determined with low calculation cost. The support is increased for each line that crosses a region of the same category.

Once the best corresponding points are found, the right camera has still one degree of freedom as described above. The final position is found by shifting the camera on the corresponding ray until the support is maximized.

The maximum support has to be greater than a special threshold to assign the images to be adjacent. For all image pairs which are assigned to be adjacent the corresponding points are determined for later use. A POI pair is assigned to be a corresponding point if the corresponding ray crosses a suitable region in the additional image plane.

The average calculation time for the mentioned steps on a Pentium 4 with 1800 MHz is below three seconds. Therefore it is possible to automatically find adjacent images within a large set. Although this step requires $O(n^2)$ computation time if no other information is available. Such additional information is provided either by a GPS or INS based position estimation or by simple using the information of adjacent images from the acquisition step.

3.3 Relative Orientation of Image Sequences

So far we have determined the orientation of image pairs and corresponding points of adjacent images. In order to calculate the orientation of a continuous sequence we perform the following steps:

1. Without loss of generality, we assume a fixed baseline to calculate 3D points from corresponding points and the relative orientation of the first image pair.
2. An adjacent image is added to the sequence. The rotation is obtained from the vanishing points as described before. The position is determined by minimizing a cost function that sum up the reprojection errors of suitable 3D points. Suitable means that there exist correspondences between a POI in the new image and POIs that led to the 3D point.
3. The rotation of the new image is improved by minimizing the same cost function we used above.

4. The corresponding points of the new image are either used to calculate new or to improve old 3D points.

5. As long as adjacent images are left, go to step 2.

3.4 Image to 3D Model Fitting

So far we have only obtained a relative oriented sequence, where the position and orientation in geo-referenced coordinates as well as the scale is not known. The upgrade from the relative orientation to a geo-referenced orientation of all images needs at least three well distributed control points. This upgrade involves a non neglecting manual effort, especially working with large image sequences. In our approach only two images of a continuous sequence have to be fitted in a semi-automatic way, so the manual work is minimized. Due to the fact, that the vertical direction of the images is known from vanishing points only two control points are necessary to transform the image into a geo-referenced coordinate system. This two control points are taken from the aerotriangulation. With the so obtained camera positions in geo-referenced coordinates for at least two images and the corresponding relative positions we calculate a transformation matrix that solves the orientation upgrade for the whole sequence.

3.5 Bundle Block Adjustment

For each building block a bundle adjustment is carried out to optimally fit the images into the existing block model.The block geometry ressembles the classical photogrammetric case of aerotriangulation: most of the image sequence is connected with corresponding points only, while known control points in some images help to fix the transformation to the world coordinate frame and stabilise the block, particularly in long sections of translational camera movement. The control point accuracy (i.e. the point accuracy of the GIS data) is known, so that correct weights can be assigned to the control points in order to account for their limited relative accuracy. Since the vertical direction is known from the detected vertical vanishing points in the terrestrial images, a constraint is added to keep the mean vertical direction of the cameras unchanged.

4 RESULTS

4.1 Accuracy

We distinguish between two different kinds of accuracy; the relative and the absolute accuracy. The absolute accuracy of the reconstruction is limited by the accuracy of the ground truth, in our case of the GIS data acquired from aerial images. The control points are the intersections of the roof lines from aerial photogrammetry, which were measured with an accuracy $< 10cm$.

Our obtained relative accuracy is about 0.25 pixel, far below one pixel reprojection error. The corresponding object point accuracy strongly depends on the recording configuration. It can be obtained from the bundle block adjustment, if the accuracy of the image measurements is known. Thus, further analysis about the accuracy of the corresponding points delivered by our algorithm has to be undertaken. A good starting point will be the work of Förstner [5] where an optimal estimation for uncertain geometric entities is given.

4.2 Geo-Referenced Orientation

Figure 6 shows two sequences of images which were oriented semi-automatically.

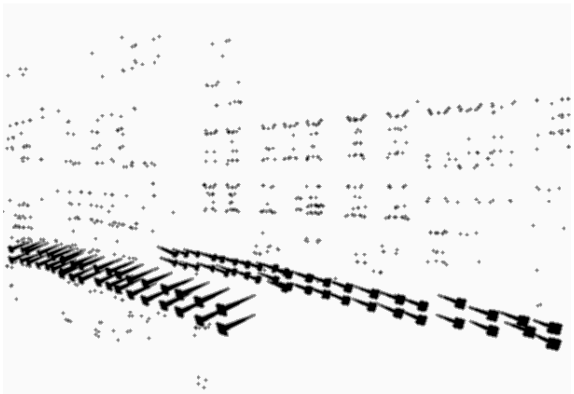


Figure 6: Geo-referenced orientations of two image sequences of the Schlossbergplatz. Each camera is represented by a small plane and a direction arrow. At each location we shot two images with a short vertical baseline. The reconstructed POIs are indicated by dots.

4.3 Image to 3D Model Fitting

The graphical user interface (GUI) used for fitting of some images to the 3D block model can be seen in Figure 7. After an aerial and an image sequence is loaded, it is possible to update the whole sequence from a relative to the geo-referenced coordinate system with only a few mouse clicks. To verify the right geo-referenced orientation the terrestrial images are superimposed by roof lines.

5 CONCLUSIONS AND FUTURE WORK

We have presented on image based documentation system called MetropoGIS in which 2 1/2D GIS data is augmented with terrestrial photographs. We showed that its possible to determine the geo-referenced orientation of a large image sequence with only a few mouse clicks without any position estimation. Because we use a hand-held camera, our image acquisition is straightforward and allows high flexibility. By exploiting improved lines and vanishing points we developed a robust and fast method to determinate relative orientation even a large baseline is present. So far we have a semi-automatic system. An operator is still involved in our work flow. We are planning to avoid this manual interaction by a fully automatic step called sky line fitting. In this step we try to match roof lines extracted from the terrestrial images with 3D lines from the block model.

ACKNOWLEDGMENTS

Parts of this work have been done in the VRVis research center, Graz and Vienna/Austria (http://www.vrvis.at), which is partly funded by the Austrian government research program Kplus.

REFERENCES

[1] J. Bauer F. Leberl A. Bornik, K. Karner and H. Mayer. High-quality texture reconstruction from multiple views. In *to appear in Journal of Visualization and Computer Animation*, pages 13:1–14, 2002.

[2] M. Antone and S. Teller. Automatic recovery of relative camera rotations for urban scenes. In *CVPR*, pages 282–289, 2000.

[3] M. Antone and S. Teller. Scalable, absolute position recovery for omni-directional image networks. In *CVPR*, 2001.

[4] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the Association for Computing Machinery, 24(6):381–395.*, 1981.

[5] W. Förstner. Algebraic projective geometry and direct optimal estimation of geometric entities. In *OEAGM Proceedings*, pages 67–86, 2001.

[6] B. Horn. Relative orientation. *International Journal of Computer Vision, 4:59–78.*, 1990.

[7] K. Kraus. *Photogrammetrie II*, pages 57–60. Duemmler, 1996. ISBN 3-427-78653-6.

[8] D. de Couto M. Bosse and S. Teller. Eyes of argus:georeferenced imagery in urban environments. In *GPS World*, pages 20–30, 1999.

[9] Andre Redert, Emile Hendriks, and Jan Biemond. Correspondence estimation in image pairs. *IEEE Signal Processing Magazine*, pages 29–46, May 1999.

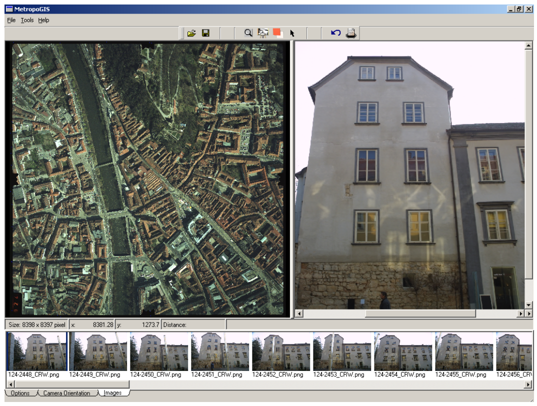
[10] C. Rother. A new approach for vanishing point detection in architectural environments. In *Proceedings of the 11th British Machine Vision Conference*, pages 382–391, 2000.

[11] S. Teller. Calibrated, registered images of an extended urban area. In *IEEE Proceedings of CVPR*, 2001.

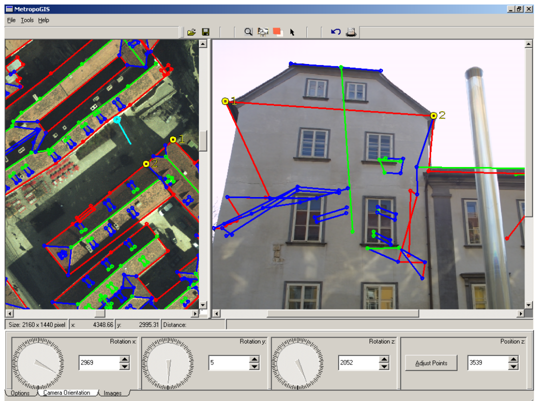
[12] Bill Triggs. Autocalibration from planar scenes. In *Proceedings of the 5th European Conference on Computer Vision (ECCV'98)*, pages 89–105, 1998.

[13] W. Wunderlich. Rechnerische Rekonstruktion eines ebenen Objekts aus zwei Photographien. *In Mitteilungen der geodaetischen Institute, TU Graz, Folge 40 (Festschrift Karl Rinner zum 70. Geburtstag)*, pages 365–377, 1982.

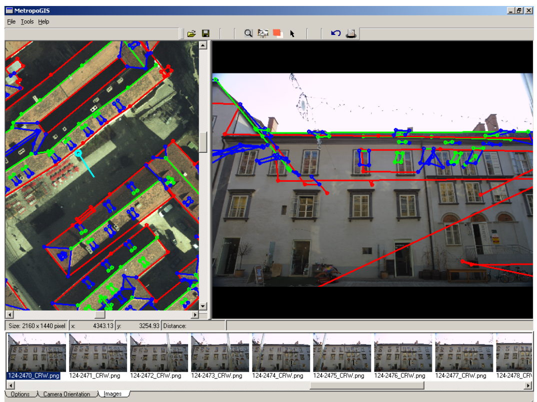
[14] A. Zisserman and R. Hartley. *Multi View Geometry*, chapter 10, pages 262–294. Cambridge University Press, 2000. ISBN 0-521-62304-9.



(a) The left part shows an aerial image of Graz. The lower strip shows a preview of an image sequence. The active image is shown on the right side.



(b) The left part shows a close up of one aerial image with superimposed roof lines and two selected control points. The right part shows the terrestrial image with superimposed roof lines using the vanishing points and the corresponding control points.



(c) The upgrade of a whole image sequence from relative to geo-referenced orientation is done by selecting two control points in two images.

Figure 7: GUI for the semi-automatic fitting of one terrestrial image to the 3D model.